EYEWITNESS' CHARACTERISTICS AND MEMORY: AN IN SITU ANALYSIS

By

Judith Lynne Cutshall

B.A., The University of North Carolina, 1982

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF ARTS

in

THE FACULTY OF GRADUATE STUDIES

Department of Psychology

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

© October 1985
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

_________________________
Judith L. Cutshall

Department of Psychology
The University of British Columbia
1956 Main Mall
Vancouver, Canada
V6T 1Y3

Date October 10, 1985
ABSTRACT

An extensive review of eyewitness testimony research reveals that (1) research findings are inconsistent in their assessments of the fallibility of eyewitness memory and in their attempts to delineate the factors which affect eyewitness' memory; (2) the criminal justice system questions the validity of generalizing these research findings to real—world eyewitnessing situations. It is suggested in this thesis that researchers' over—reliance on experimental methodology has led to a confused image of eyewitness' memory which is of questionable relevance to actual eyewitness situations.

A non—experimental examination of real—world eyewitnesses is therefore proposed as a first step toward collecting meaningful data relating to eyewitness memory. Two types of non—experimental research are presented in this thesis: archival research and a case study of eyewitnesses to a violent crime. The archival research, which was drawn from R.C.M. Police files, was designed as a means of gathering baseline data concerning actual eyewitnesses, for example, their gender and age and how often they are also victims of crime. The case study involved interviewing thirteen witnesses, all of whom had witnessed the same gunshooting incident. Witnesses were questioned by police at the time of the incident and interviewed by our research team four to five months later. Both statements were analyzed for the number and type of details reported and the accuracy of those details. Higher accuracy rates than those reported in the experimental literature were found in the witnesses' statements both at the time of the incident and several months later.

Although the data base established in these studies is insufficient to judge the validity or invalidity of prior eyewitness research, it does appear that experimental research has not appropriately assessed eyewitnessing ability in regard to witnessing a striking, violent real—world event. The generalizability
of experimental research to this type of event is therefore highly suspect. It is concluded that the combination of archival and field research will yield the data base needed for developing an understanding of the behavior of actual eyewitnesses.
# TABLE OF CONTENTS

List of Tables .............................................. v
List of Figures ........................................... vi
Acknowledgements ....................................... vii
Introduction ............................................. 1
Historical Review ....................................... 5
An Alternative Approach to Eyewitness Memory Research ........... 62
A Profile of Eyewitnesses to Violent Crimes ....................... 66
A Case Study of Eyewitness Memory of a Violent Crime .......... 79
Conclusion .............................................. 115
Bibliography ............................................. 121
LIST OF TABLES

Table I. Crime Statistics for 1983. ........................................... 70
Table II. Violent Crimes Reported in Burnaby. ......................... 71
Table III. Ages of Witnesses by Crime Category. ...................... 74
Table IV. Total Reported Classifiable Details. ......................... 90
Table V. Reported Classifiable Details for
Two Groups of Witnesses. .................................................. 91
Table VI. Accuracy of Classifiable Details. ............................. 94
Table VII. Distribution of Errors. .......................................... 96
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Distribution of Errors in the Police and Research Interviews.</td>
<td>93</td>
</tr>
</tbody>
</table>
Evidence supplied by witnesses or victims of crime is of crucial importance to the criminal justice system. According to a Rand Corporation study of the investigation process, the single most important determinant of whether a case will be solved is the information supplied by the witness or victim of crime (Rand Corp., 1975). Given its vital role in the investigative process such evidence is afforded considerable weight by the court, being generally accepted as direct evidence rather than circumstantial evidence (Tapp, 1976). Because of the weight given eyewitness testimony, such testimony has led to convictions which would not otherwise have been possible. Unfortunately, however, not only the guilty are convicted on the basis of witness testimony. Wrongful convictions due to mistaken identification, for example, have been documented by several authors (i.e., Frankfurter, 1927; Borchard, 1932; Buckhout, 1974). Sobel (1972) suggests that mistaken identification by witnesses or victims has been responsible for more miscarriages of justice than any other factor, "more so, perhaps than all other factors combined" (p.vi). Another lawyer, Woocher (1977), suggests that such miscarriages of justice are largely due to the fact that "most juries and even some judges, are unaware of the sources of error in eyewitness testimony and subsequently place undue faith in its veracity" (p.970). A vitally important question is thus posed by the judicial system: what are the sources of error in eyewitness testimony and how might these errors be detected and/or reduced? With so much at stake, it is little wonder that psychologists have become involved in trying to unravel this extremely complex problem. The first research efforts were made at the turn of the century. Investigators tried to uncover the psychological factors which influence a witness's initial perception of an event and his or her subsequent reporting of the event, but these early efforts disappeared as behaviourism shifted the attention of psychologists to other problems. Eyewitness testimony
research re-emerged with renewed vigor in the early 1970's, however, focusing especially on eyewitness identification. Psychologists interested in perception, attention, memory, language, and thought pursued their research interests within the domain of eyewitness testimony. In less than a decade, the field had grown to such an extent that three separate volumes appeared summarizing research findings (Clifford & Bull, 1978; Loftus, 1979; Yarmey, 1979). Except for a legal textbook written in 1931, these texts represented the first full-length treatment of eyewitness testimony by a psychologist since Hugo Munsterberg's *On the Witness Stand* appeared in 1908. Since the publication of these texts, growth in the field has continued at an accelerated pace. In 1980, the first conference devoted solely to "The Psychology of Eyewitness Testimony" was held at the University of Alberta. The degree of activity in the field is further evidenced by a review of eyewitness testimony by Penrod, Loftus, and Winkler (1982) which reports on more than 60 research articles and papers appearing in 1979-1980 which the authors consider "highly relevant to the problems of witness reliability that are encountered every day in legal settings" (p.121). This veritable explosion in the psychological literature concerning eyewitness testimony has produced three more volumes during 1982-1984 devoted to the psychological aspects of eyewitness testimony (Lloyd-Bostock & Clifford, 1983; Trankell, 1982; Wells & Loftus, 1984). Such rapid, extensive growth is quite impressive, but it raises several questions. First, one is led to ask why an explosion in the literature has occurred. Secondly, what is the image of the eyewitness that has emerged from this literature? And finally, considering the interest in the psychology-law interface, of what relevance is the psychologist's view of the eyewitness to the criminal justice system? An historical review of the literature will be presented, focusing on three historical "eras" within the development of eyewitness research: (1) Early Aussage research (1900-1920)
which provided the prototype for current studies; (2) The "silent" behaviorist era (1920–1960) during which very little eyewitness research was conducted; and (3) The contemporary period (1960–present) which encompasses the so-called "cognitive revolution". In answering the above questions, this review of the literature provides the foundation upon which this thesis is built. It will be shown that eyewitness research, thus far, has presented an inconsistent, confused image of the eyewitness. The only consistent conclusion which can be drawn from this research is that eyewitness memory is fallible. In fact, the following review indicates that during the past eighty years, research has focused almost exclusively on painting an extremely negative view of the unreliable eyewitness. The reasons for this negative focus will be explored along with questions concerning the validity of this view of the eyewitness. Furthermore, it will be argued that no conclusions concerning eyewitness performance in real life situations can be drawn on the basis of hitherto conducted research which has exclusively employed experimental paradigms. It is the contention of this author that the generalizability of subject—witness' memory for filmed or contrived staged events is highly questionable. The purpose of the present thesis, therefore, is to offer an alternative approach to eyewitness research which will provide information about real—life eyewitnesses to actual crimes. Two studies illustrating this approach will be presented. One study consists of archival research conducted with police files and a second one focuses on a case study of thirteen witnesses to a gunshooting incident. These studies not only provide an initial view of eyewitness performance in situ, but they also serve as a touchstone for assessing the generalizability of previous eyewitness research. Before discussing these alternatives to past and present eyewitness research, however, a review of the literature and the context in
which the field developed will be presented in order to demonstrate the necessity for gathering data from real world eyewitnesses.
HISTORICAL REVIEW

The Aussage Era

An historical review of eyewitness testimony begins at the turn of the century with Binet's *La Suggestibilite* (1900) in which he called attention to "the advantage that would accrue from the creation of a practical science of testimony." The following year, in a lecture before the Berlin Psychological Society, Wilhelm Stern proposed the practical use of expert tests to ascertain testimonial certitude (Wigmore, 1909, p.410).

In Germany psychologists worked in earnest to make Stern's proposal a reality. "A new special science" was created which dealt "exclusively with the reliability of memory" (Munsterberg, 1908, p.46). The German word *Aussage*, translated as testimony or report (or as Stern defined it: "the verbal expression of a recollection"), became the name of this new subfield within psychology (Stern, 1910, p.271). In Germany the new science was sufficiently established by 1903 to have founded its own journal; and by 1911 European courts found *Aussage* convincing enough to allow expert testimony in a murder trial in Belgium. This trial is the first instance on record of a psychologist giving expert testimony concerning the reliability of other evidence submitted to the court (Whipple, 1912).

This case was heralded by Whipple as a landmark for the psychology of testimony. By admitting the psychologist's testimony as evidence, this case represented formal acceptance of *Aussage* research by the court. In addition, because the psychologist's presentation of *Aussage* findings served to discredit the testimony of witnesses against the accused, *Aussage* psychology contributed to a "not guilty" verdict. This was interpreted by Whipple as demonstrating
the value of Aussage research; the psychology of testimony had, in effect, "saved a man's life" (Whipple, 1912, p.268). Whipple reported two additional cases in 1913 in which psychologists served as expert witnesses in Germany, attesting further to the acceptance of Aussage research by European courts. Wilhelm Stern also noted that Aussage research concerned with the testimony of children was beginning to directly affect German jurisprudence; the testimony of children was "less highly valued (by the court) than formerly" (Stern, 1910, p.274). From these reports, it appears that Aussage psychology met with a certain degree of success in Europe. Also, we see that an essentially negative attitude is being taken toward the eyewitness. The emerging view of eyewitness memory is that it is highly unreliable. In North America, however, the impact of Aussage was more problematic.

By 1908 the call for a "practical science of testimony" was being heard and popularized in North America. Munsterberg's On the Witness Stand and Whipple's reviews in the Psychological Bulletin (1909, 1910, 1911, 1912, 1913, 1914, 1915, 1917) along with the English writings of Wilhelm Stern served to import Aussage from Germany. Also, the move of Munsterberg from Germany in 1906 brought a major supporter of the Aussage movement to the United States.

A summary of Aussage research as it was presented to North American psychologists is presented below. Before looking at specific research findings, however, it will be instructive to first examine the methodology employed in Aussage research. This methodology, as outlined by Whipple (1909), involved two basic choices: an experimental method and an historical method. Whipple immediately dismissed the latter as "less promising" and embraced the experimental method because of its "obvious advantages". The materials to be used within the experimental framework consisted primarily of picture-
tests and event—tests, but Whipple also suggested the future use of the "moving-picture show". As the name suggests, the picture—test consisted of presenting the subject with a series of pictures or photographs depicting an event. The event—test, on the other hand, consisted of a staged, live event which was often enacted before unsuspecting subjects. An event—test or "reality experiment" will be described in more detail in the following section concerned with the fallibility of eyewitness memory.

Exposure time of these events varied from five seconds to seven minutes, with 45–60 seconds being most usual. The time interval between exposure to a test and the witness' report varied from zero to nine and one half weeks and the report itself generally consisted of both a narrative, or free report, and an interrogatory report. The report was used to assess both the accuracy and completeness of subject—witnesses' memory of the previously witnessed event.

Using these techniques, Aussage psychologists focussed on several aspects of eyewitness testimony. In the summary of Aussage research presented below, research findings have been broken into several topics which represent the prevalent research interests. Demonstrating the fallibility of witness' memory was the most overarching concern of this era, therefore, the research summary begins with a discussion of this topic. In addition, more specific aspects of the witness himself/herself were explored by Aussage researchers. These findings, which include such things as sex and age differences and differences due to intelligence, will be presented below under the topic of individual differences.

Aussage research also focussed on aspects of the reporting procedure, dealing with such questions as the effect of the time interval between the event and the report; the effect of the form of the report; and the effect of suggestive questioning. Each of these topics will be dealt with separately in
the following research summary. In addition, *Aussage* psychologists conducted research which dealt with the relationship between a witness's confidence in his/her recollection and the accuracy of that recollection, and the impact of stress upon witness' memory. Presentation of these two topics will conclude the summary of *Aussage* research.

**Fallibility of eyewitness memory**

In 1902 the first recorded "reality experiment" was conducted by Professor von Liszt in Berlin. In *On the Witness Stand* (1908), Munsterberg noted that these dramatic experiments "have recently become the fashion and almost a sport...which will have to be continued with a great variety of conditions if the psychological laws involved are really to be cleared up" (p.53—54). Thus, by 1908 this type of event—test had become the preferred experimental design among *Aussage* psychologists.

Liszt's original "reality experiment" involved a university seminar which was interrupted by an argument between two men. As the argument escalated, one man pulled a revolver; the professor then stepped between the two men, grasped the man's arm, raising it toward the ceiling as the gun discharged. This was followed by a general uproar in the classroom which the professor calmed by explaining that the event had been staged and everyone was to write a report of what had transpired. The resulting reports were evaluated according to the number of mistakes within each report. Mistakes were calculated based on the total number of omissions, additions, and alterations. Using this procedure it was found that the written reports contained from 26% to 80% mistakes. Besides omissions, errors ranged from putting words into the mouths of silent spectators to attributing nonexistent actions to participants (Munsterberg, 1908, p.49—51).
Wigmore (1909) reported the following experimental findings based only on testimonial correctness, excluding omissions: (1) Research conducted by Guenther found that individuals range from 35.8% correct to 100% with an average of 80.6% correct (2) Lipmann's research indicated that individuals range from 71% to 100% correct with an average of 89% correct (22 subjects) (3) Research by Jaffa found that individuals' errors ranged from 26% to 114% (sic), with 6 of 28 subjects making less than 45% errors and 8 making over 80% errors.

It is important to note here both the accuracy rates and the ranges of accuracy reported in this research. The variability among witnesses is extensive in both Guenther's and Jaffa's research, but not in Lipmann's research. There is clearly no consensus among these researchers as to the accuracy of reports or the range of errors. It is also interesting to note that although some witnesses' reports were highly accurate, researchers focused on reporting the fallibility of eyewitness testimony.

Research which was aimed toward demonstrating the fallibility of eyewitness' testimony also indicated specific sources of errors in witnesses' reports. It was generally found that witnesses could report the main outline of an event quite accurately (Whipple, 1909; Stern, 1910). Witnesses' reports of secondary features such as quantities and colors, however, were often quite poor. Poor performance on the recollection of colors was consistently reported by several investigators (Munsterberg, 1908; Whipple, 1910; Stern, 1910; Whipple, 1913). In fact, Nacke reported in 1911 that testimony "descriptive of colors is practically worthless" (reported in Whipple, 1913, p.266). In addition, it was found that witnesses' reports of an individual's height and weight were often erroneous (Munsterberg, 1908; Whipple, 1909). These findings led to the
general consensus among Aussage psychologists that "testimony descriptive of an individual's appearance has only a limited reliability" (Whipple, 1913, p.266). Although research differed in reported accuracy rates, the findings were consistently interpreted as demonstrations of the fallibility of "honest" eyewitness' memory. In fact, in his 1939 review of Aussage research, Stern suggested that demonstration of the fallibility of witness memory was the single most important finding of Aussage psychology. According to Stern, the common-sense notion (of jurors, if not jurists) holds that testimony is correct so long as the witness is mentally competent and has the intention of telling the truth. In other words, deviation from objectivity is due to pathology or intentional lying. Aussage psychology, however, convincingly demonstrated that this notion was patently false: the memory of honest, unbiased witnesses often deviates to a considerable extent from "objective" reality (Stern, 1939).

Individual differences

Concerning the effect of sex differences on witness' testimony, Aussage psychologists found no systematic pattern. Munsterberg (1908) reported, for example, that some experiments found essentially no differences in memory of the two sexes, while other experiments showed considerable differences (p. 54). Similarly, Whipple (1909) reported an experiment by Stern which showed men superior in accuracy; while, in the same review, Whipple also reported research conducted by Wreschner (1905) and Borst (1904) which showed women superior in accuracy and number of correct statements. The following year, Whipple (1910) referred to research by Breukink (1909) in which men reported more items than women, but with less accuracy. And, in 1912 Whipple reported an experiment conducted by Schramm (1911) which demonstrated a slight superiority in women's recall ability. Whipple, however, pointed out that the
differences do not "exceed the probable error of the results" (Whipple, 1912, p.267).

Aussage psychologists also failed to establish a relationship between accuracy of testimony and intelligence (excluding mental defects or retardation). Bruekink (1909) reported that "persons of culture" gave more extended and accurate reports (reported in Whipple, 1910). However, most other researchers agreed with Whipple's observation that there were no conclusive experiments on the relationship between accuracy and intelligence (Whipple, 1909).

Concerning age differences, Whipple (1909) spoke for most Aussage psychologists when he claimed that "the reports of children are in every way inferior to those of adults" (Whipple, 1909, p.162; see also Stern, 1910; Munsterberg, 1908; Whipple, 1911). Gross (1910), however, strongly disagreed, claiming that children "make different errors but no worse ones that do adults" (reported in Whipple, 1911, p.308). In fact, Gross went so far as to assert that "a healthy half—grown boy is the best possible witness for simple events" (p.308).

Age differences were also reported concerning witness' susceptibility to suggestion. These findings will be presented below in the section concerned with the effects of suggestive questioning.

On balance, the findings concerning individual differences were inconclusive. No consistent patterns were found either for sex differences or intelligence. There appears to be a greater degree of consensus among Aussage psychologists concerning age differences; however, there was dissension on this issue as well.

Effect of time—interval between event and report

Once again, Aussage psychologists reported conflicting findings. For example, Munsterberg (1908) reported that "(i)n some cases it was shown that
the mistakes made after a week were hardly more frequent than those made after a day. Other experiments seemed to indicate that the number of mistakes steadily increases with the length of time which has elapsed" (p.54).

Experiments conducted by Stern (1907) and Borst (1904), for example, found a fairly constant decrease of accuracy over time. Stern reported a loss of 0.33% per day over 3 weeks and Borst reported a loss of 0.27% per day over 6 days (reported in Whipple, 1909). This reported loss was exceedingly small; in Borst's experiment, an accuracy rate of 89.5% after 3 days dropped to 87.9% after 6 days (reported in Whipple, 1909, p.163).

Like Munsterberg, Whipple (1909) also found no clear pattern among research findings. He reported that lengthening the time-interval between an experience and the reporting of that experience generally exerted an "unfavorable influence", but he asserted that "there is nothing like the loss in efficiency shown in curves of memory for nonsense syllables, as in the familiar Ebbinghaus test: indeed, for some reporters the report seems to be somewhat improved after several days have elapsed" (p.163). Whipple's conclusion summarizes the research findings of his day: the effect of the time interval between event and report varies among reporters and across events.

**Effect of form of report: Narrative v. Interrogatory report**

Experiments reported by Munsterberg, Stern, and Whipple agreed that less information is given in a narrative report, but the information is more accurate than that of an interrogatory report. For example, Whipple (1909) reported an experiment in which the range of interrogatory reports was 50% greater and inaccuracy was as much as 550% greater. One tenth of the narrative was incorrect; one fourth of the interrogatory was incorrect.

Stern (1910) reported a picture-test in which the narrative contained 5–10% errors and the interrogatory contained 25–30% errors. An event-test in
which reports were written one week following the event produced 25% errors in the narrative and 50% errors in the interrogatory.

Given the consistency of these findings, Aussage psychologists were in agreement that witnesses should first be permitted to give a free report, followed by an interrogatory report if the initial report was inadequate in scope or obscure in details (Whipple, 1918).

Suggestibility of questions

Aussage psychologists were in general agreement that as the suggestibility of questions increased, accuracy decreased. For example, Whipple (1909, 1918) reported Binet's findings which showed a 26% error rate for indifferent questions, a 38% error rate for moderately suggestive questions and a 61% error rate for strongly suggestive questions.

Whipple (1918) pointed out that Binet's figures pertained to results with children, "who are well known to be more open to suggestion than are adults, but the relative influence of the different types of questions may be deemed about the same for children and adults" (p.245).

Stern (1910) demonstrated that the degree of susceptibility to suggestion was largely a function of age. Stern found that seven year olds were incorrect 50% of the time when given a suggestive question, whereas eighteen year olds were incorrect only 20% of the time.

Muscio (1915) extended this research to adults and found a similar pattern to that reported by Binet: increasing the suggestibility of questions led to an increase in the number of errors. Muscio also classified different types of misleading questions. He noted, for instance, that greater suggestibility resulted when sentences contained more than one negative. He also found that questions posed with the definite article the elicited more incorrect responses than the same questions posed using the indefinite article a. Muscio also
pointed out that if misleading questions were blatantly obvious, their negative effect would be nullified.

In summary, we see that Aussage researchers agreed that suggestive questioning led to increased errors and that this increase was especially severe with children. So certain were Aussage psychologists about the deleterious effects of this practice that it was recommended in a 1911 survey concerning the value of testimony that "any question that carries a suggestion should be avoided; in especial, children have a very slight resistance to suggestion" (Nacke, 1911, reported in Whipple, 1913, p.266).

Relationship between confidence and accuracy

Whipple's (1909) remark that "generally speaking, attestation does not guarantee accuracy" was supported by the findings of Aussage psychologists (p.161). Research consistently demonstrated that "sworn" testimony was not error free. Although sworn testimony did not guarantee accuracy, such testimony was found to contain fewer errors than unsworn testimony. Stern (1907) and Borst (1904), for example, showed that sworn testimony contained approximately 10% errors, whereas unsworn testimony contained from 20% to 25% errors. Boring (1916) found a similar pattern with sworn testimony being more accurate than unsworn testimony by 8%–9% for adults and 14% for children (reported in Gardner, 1933, p.408).

Research was also extended to include an assessment of varying degrees of confidence. Aussage research reported by Burtt (1931) showed 92% accuracy under oath; 86% accuracy when witnesses felt fairly certain; and 56% accuracy when witnesses believed their testimony to be true but were unwilling to commit themselves (p.153). Similarly, Dallenbach (1913) reported an accuracy rate of 87% for items marked "would swear under oath"; 71% for "fairly certain" items; and 41% for items marked "little better than a guess."
The research findings reported above demonstrated that sworn testimony was not free from error. Aussage research also found a consistent pattern in the relationship of confidence and accuracy; the higher the degree of certainty exhibited by a witness, the more accurate his/her testimony was likely to be.

Impact of Stress

Very little Aussage research dealt directly with the effects of stress. However, results from the Liszt "reality experiment", recounted above, showed that the second part of the written reports which pertained to the more strongly emotional aspects of the event, contained 15% more errors than the first part of the report.

Whipple (1915) reported that research conducted by Kobler (1914) suggested that excitement improved observation and memory of witnesses up to a given point (which varied for different persons) and impaired performance beyond that point. From such scanty research findings, Aussage psychologists were unable to make any conclusive statements about the impact of stress on testimony. Stress, however, was generally considered by Aussage researchers to have a negative rather than a positive effect on witness' testimony.

This brief review of the findings of Aussage psychology affords us at least a glimpse at the kind of research that was being conducted during the first two decades of this century. Perhaps more important than the particular findings of Aussage psychology, however, was the excitement which was generated by the idea of an Applied Psychology. Munsterberg, Whipple, and Stern all wrote about a new psychology whose "time had come." Their writings are replete with the enthusiasm, optimism, and over-generalizations which characterized many of the first generation experimental psychologists. Stern wrote of an "Applied Psychology...now springing up...(which) must provide a
scientific basis for practical knowledge of, and judgments upon, human mental acts and qualities...and must give assistance in the practical manipulation of human minds" (Stern, 1910, p.270). Stern considered this to be the task of Applied Psychology in the fields of education, law, and medicine (p.270).

From the scanty research findings which we have reviewed, it seems incredible that Stern would make such far-reaching claims for the application of psychology to the law. We must keep in mind, however, the Zeitgeist within which psychologists of the early 1900's were operating. Psychology was still a very young science, bent on proving itself amidst the scepticism of the established physical sciences. Successful practical applications of psychological research would perhaps serve to legitimize the new science of psychology by bringing public recognition and acceptance to the field.

Indeed, with the spectacular success of Binet's pioneering work in mental testing, published in 1905, the idea of a successful applied psychology began to take hold. The testing movement quickly made strides toward developing testing methods applicable in education, medicine, and the military. In fact, acceptance of psychological techniques by both education and medicine led Munsterberg to proclaim that "(t)he lawyer alone is obdurate" in his refusal to accept psychological techniques (Munsterberg, 1908, p.10). With this in mind Munsterberg wrote On the Witness Stand (1908); the expressed purpose of the text being to garner public support for the application of psychological techniques in the court system which would, in turn, exert pressure on the courts to accept psychological evidence. To underscore this intention, Munsterberg closed his introduction to the text by saying, "my only purpose is to turn the attention of serious men to an absurdly neglected field which demands the full attention of the social community" (p.12).
Thus, the primary interest in the psychology of testimony, at least in North America, was in its value as an applied science rather than a theoretical one. This emphasis in turn greatly influenced the methodology employed by the "new special science". Both Stern and Munsterberg, in advocating the new science of Applied Psychology, warned that experimental psychology cannot "simply pick up some bits of theoretical psychology and throw them down before the public" (Munsterberg, 1908, p.8). "The attempt to carry over into it (Applied Psychology) unchanged, the methods of pure psychology, (would be)...a cardinal error." (Stern, 1910, p.270).

Stern suggested that because "practical life does not deal with elements, but with very complex mental processes; the special methods of applied psychology must therefore take a middle position; they must combine the necessary nearness to life with that degree of exactness which is indispensable for the drawing of reliable inferences" (Stern, 1910, p.270). "The necessary nearness to life" criterion led to increased criticism of picture-tests (Whipple, 1910; Stern, 1910; Whipple, 1911) on the basis that in practice, reports are of dynamic events, not static pictures. There was an expressed preference among these researchers for "reality experiments" or event-tests which followed the format of the previously discussed Liszt experiment during which unsuspecting subjects would view a carefully enacted scene and then write a report of the scene which they had witnessed. The "exactness" criterion would be met by the experimenter's control over the scene and by quantitative and qualitative analyses of the written report.

We see, therefore, that due to their interest in the applications of their research, these early Ausseage psychologists were facing the problem of "ecological validity". Whipple stated the problem succinctly in 1909:

The experimental method possesses the obvious advantages, standardizations of conditions, control of the determining factors,
possibility of repetition and of quantitative and statistical evaluation of the results. But the method has been criticized, particularly by jurists, on the ground of oversimplification, general artificiality and consequent nonapplicability to the problems of real life—criticism which is, I think, entirely intelligible, however incorrect. (p.155)

Stern, however, was less certain concerning the applicability of experimental methodology and he suggested that experimentation be supplemented with an historical method. As its name implies, the historical method involved the more interpretative, idiographic approach employed by the historian. Using the historical method, an investigator would reconstruct a particular unique event by taking into account the context in which the event occurred and all relevant material which could be gathered concerning the event. Such an approach was necessary because, according to Stern, "even reality experiments offer an artificial simplification compared to the seriousness and complications of actual events....Each trial and each individual present unique features and call for particular procedures. Sometimes it is not so much an experimental as an historical method that must be applied in order to reconstruct the history of a given body of testimony" (Stern, 1939, p.14).

The issue of the appropriate methodology for the study of eyewitness testimony was largely settled by fiat, however. During the Aussage era most mainstream North American psychologists such as Whipple and Munsterberg accepted experimental methodology without question. Thus, while Whipple (1909) mentioned Stern's historical methodology, he did not pursue it because of experimental methodology's "obvious advantages". Advocacy of a non-experimental methodology was thus summarily dismissed. Stern's findings were widely cited, but his call for a supplementary historical methodology was largely ignored. This was in keeping with the general goal within psychology at this time—the development of an experimental science patterned after 19th
century physics. Viewed retrospectively, the choice of experimental methodology may be seen as serving political ends. That is, psychology adopted the experimental method in order to establish itself as a discipline separate from philosophy, rather than basing the choice of methodology on the specific needs of the subject matter of psychology (Koch, 1961).

One means of assessing the validity of experimental methodology is to consider the extent of its success. As we shall see, the new experimental science of testimony met with only limited success. The aspirations of Munsterberg for the creation of a "new special science of testimony" applicable to the criminal justice system in North America were not realized. Munsterberg attributed the lack of successful application of Aussage psychology in North America to the intransigence of the court system. Others, such as Whipple held out hope for future accomplishments within experimental psychology although current accomplishments were meager. The "time will tell" maxim operated; history would prove the success of experimental methodology. After all psychology was a young science at this time and if Aussage failed it could be due to the relatively unsophisticated research designs and the rather primitive statistical techniques in use during the early stages of experimental psychology. The survey of current research, provided below, involving more sophisticated research methods places the "time will tell" maxim in serious doubt.

The preceding discussion presents the psychology of testimony as it stood during the first two decades of this century. Before leaving the Aussage era, however, we need to look more closely at the failure to which I have alluded of Aussage psychology in North America. Aussage psychology was considered a failure in North America because of the lack of acceptance of its findings by the American criminal justice system.
Munsterberg's presentation in *On the Witness Stand* of the value of psychological research to the court system was considered, at best, to be premature and in the eyes of many it was irresponsible "yellow Psychology". Wigmore's 1909 insightful, scathing assessment of Munsterberg's claims, presented in the Illinois Law Review in the form of a mock libel trial, exposed many of the problems encountered or perhaps ignored by *Aussage* psychologists in North America. In order to assess the failure of *Aussage* as an applied discipline, a closer look at Wigmore's critique of *Aussage* psychology is presented below.

A major criticism raised by Wigmore concerned the methodology employed by *Aussage* psychologists and the claims offered by Munsterberg for this methodology. After examining the "new and exact methods" lauded by Munsterberg, Wigmore concluded that *Aussage* methodology was "still so little exact and so incapable of forensic use that even their well—wishers confess that 1000's of experiments and years of research will be required before they will be practicable, if ever" (Wigmore, 1909, p.415). In addition, Wigmore suggested that many of the research findings were neither new nor practicable. He claimed that the research revealed little that was not common sense knowledge.

The most serious problems discussed by Wigmore, however, dealt with the lack of consensus among researchers and their undue haste in generalizing their findings to actual witness situations. It was readily demonstrated by Wigmore that *Aussage* researchers were not in agreement concerning many fundamental issues. Our survey of *Aussage* research has also shown this to be the case. In addition, considering the sketchy research findings and the lack of consensus among researchers, Wigmore questioned the haste of *Aussage* psychologists to generalize their research findings: "The conclusions drawn from
this and similar researches seem to me in general to be over—hasty....Of course these researches must be much more extended before the jurist can obtain any practical results for use in trials; for that purpose a long and systematic study will be required" (Wigmore, 1909, p.412).

Wigmore's appraisal of the applicability of Aussage research to the North American criminal justice system served to thwart Munsterberg's aim to promote a law—psychology interface. As Hutchins and Slesinger observed in 1929, "Munsterberg was too confident and overshot his mark. The lawyer snapped back and put the psychologist in his place. The psychologist then retired from the courtroom and left the law to muddle on its own way" (Hutchins & Slesinger, 1929, p.14). If the criterion for success of Aussage psychology was its application to the criminal justice system, then Aussage was clearly a failure in North America. Aussage research was not accepted by the courts in North America; instead, it remained solely an academic enterprise.

The Behaviorist Era

With the advent of World War I, Aussage psychology effectively ceased to exist in Europe as did most all forms of psychology. By the end of the war, with the shift of emphasis to North American psychology, that is, behaviorism, the psychology of testimony became irrelevant to mainstream psychological research. With Watson's declaration that "the behaviorist never uses the term memory....it has no place in an objective psychology," (Watson, 1929, p.177) the study of witness' memory was outside of the province of mainstream psychology. Thus, within the mainstream, psychologists remained virtually silent concerning witness testimony from 1920—1960.

A limited amount of research pertaining specifically to eyewitness testimony was conducted during this era (Cady, 1924; Marston, 1924; Whitely
and McGeoch, 1927; Snee and Lush, 1941). In addition, there were several reviews of the literature on witness testimony and commentaries concerning the psychology—law interface (Hutchins and Slesinger, 1929; Slesinger and Pilpel, 1929; Gardner, 1933; Stern, 1939). Also, the first text by a psychologist since Munsterberg's text on the psychology—law interface appeared in 1931, H.E. Burtt's *Legal Psychology*. Although no striking, new findings are reported during this period, there was a perceptible shift in tone. Burtt's 1931 text presented a much more conciliatory attitude toward lawyers than Munsterberg's *On the Witness Stand* as evidenced by his introduction to *Legal Psychology*. Discussing his reason for writing the text (which was primarily designed to serve as a textbook for a legal psychology university course), Burtt said that, "It is felt that the science of psychology has developed a very considerable number of principles bearing on legal problems that are worth presenting" (Burtt, 1931, p. 211). Whereas Burtt may have overreached the data in his claims, the rhetoric has toned down considerably from the days of Munsterberg.

The research findings of the behaviorist era are presented below in a format similar to that used previously to report *Aussage* research. In this case, however, it should be noted that all of the research conducted during the behaviorist era dealt primarily with the effects of various types of reports. The advantages and disadvantages of narrative and interrogatory reports and combinations of the two types of reports were a major focus of this research. These researchers also addressed questions concerned with the incomplete nature of testimony, the effect of fore—warning, and the effect of various time intervals on recall. These questions, however, were asked in the context of the effect of various types of report and therefore all of the findings will be presented below under this single topic.
Effect of form of report: Narrative v. Interrogatory Report

Cady's (1924) research confirmed findings of earlier Aussage psychologists; a greater number of errors are committed with specific questioning than with the narrative. In contrast to earlier reports, however, Cady reported a greater range of details with the narrative (88) than with specific questions (50). The specific questions, however, elicited information which was not volunteered in the narrative and which was deemed important to the completeness of the report. Thus, in agreement with the earlier Aussage psychologists, Cady concluded that the most effective reporting procedure was a combination of narrative and interrogatory reports.

Marston (1924) added some additional variables to his event-test research. He presented an event to one group unannounced and he forewarned another group concerning the event. Marston found that under the "expected" condition, there was a slight gain in completeness for all types of reports (ranging from 4.5% to 8.8%), but there was also a slight loss in accuracy (averaging 7.9%). Generally, Marston found that the narrative was less complete than the interrogatory (both direct and cross—examination) and the direct interrogatory was both more complete and more accurate than cross—examination. In all cases, however, Marston found the reports to be basically very incomplete. Narrative reports were 94% accurate and covered 23% of the details; direct interrogatory reports were 83% accurate and 31% complete; and cross examination was only 76% accurate, covering 29% of the details.

Whitely and McGeoch (1927) examined the effect of the report type and combinations of reports on recall at various time intervals. They found that retention improved for 30, 60, 90, 120 days if recall was initially tested with
a combination of narrative report and interrogatory as opposed to either report alone.

Snee and Lush (1941) tested the relative effectiveness of free narrative, controlled narrative, and questionnaire. In keeping with previous Aussage research, they reported that the questionnaire elicited a greater number of details and a higher rate of errors than the narrative. When the free narrative was compared with the controlled narrative, Snee and Lush found that the controlled narrative yielded more correct response items (28) than the free narrative (22) with no increase in errors. When various combinations of reports were compared, Snee and Lush found that administering the questionnaire after the free narrative produced no change in the frequency of incorrect responses on the questionnaire, but there were increases in the number of correct responses and decreases in the number of "don't know" responses. When the order was reversed, however, the number of incorrect responses on the narrative increased along with an increase in the number of correct responses. Snee and Lush (1941) concluded that the narrative should precede the questionnaire and that the controlled narrative was more effective than the free narrative.

The above reported findings essentially represent all of the witness testimony research conducted during the forty-year period from 1920–1960. Other discussions and reviews focused primarily on the earlier Aussage research and general psychological principles pertaining to memory. Gardner's 1933 review which was directed primarily to lawyers was representative of the general approach and attitude during this period and therefore it will be examined more closely in order to give us a flavor of the psychology-law interface at this time.
Gardner expressed the frustration of lawyers attempting to learn about the nature of memory when there was no consensus among psychologists. According to Gardner, "the objectivists (behaviorists) use 'memory' in a strained, technical sense, the imagists consider it almost solely a question of imagery, the gestaltists are fairly losing themselves in metaphysics" (p.391). Gardner thus concluded that it was "little wonder that lawyers and jurists have been unable to learn anything of the true nature of memory", (p.391) re-echoing the sentiment expressed by Wigmore concerning the lack of agreement among psychologists.

Gardner then proceeded to discuss memory according to the basic principles of the behaviorist era, noting that the basic principle of memory was the law of association (p.392). With this as the basic framework, Gardner discussed witness memory in terms of frequency, recency, and intensity, illustrating these "laws" with *Ausssage* research. This S–R approach to memory was followed by a discussion which hinted at an entirely different theoretical framework for memory research. In discussing the role which expectation and imagination play in one's memory and the fact that one fills in the gaps or missing details in order to have a completed picture of an event, Gardner alluded to the creative aspects of memory, the focal point of current research.

In fact, at the time of Gardner's review the groundwork for a "revolutionary" view of memory was being laid by F.C. Bartlett at Cambridge. In 1932 Bartlett published his seminal text, *Remembering*, in which memory was conceived as an active, reconstructive process rather than a passive, associative process. The hold which behaviorism had over mainstream psychology at this time, however, prevented wide acceptance of Bartlett's approach. It was not until the late 1960's that the reconstructive view of
memory took hold within psychology and eyewitness research was undertaken with renewed vigor.

As we have seen, very little witness research was conducted during the behaviorist era due largely to the fact that the study of memory itself was severely circumscribed. But as the reviews written by various lawyers suggest, interest in the research continued to be expressed. In 1957 an issue of the *Journal of Social Issues* was devoted entirely to witnesses and testimony research, presenting the topic as "A Social Problem in Need of Social Research." The literature review presented in this journal consisted primarily of research conducted during the *Aussage* era and hence the thrust of the issue was a call for renewed research efforts concerning witness testimony. This call, however, met with little response until the early—1970's at which time interest in the field began to be expressed by experimental psychologists interested in human cognitive processes. Within five years of its re—emergence as a viable area of research, the study of eyewitness testimony had generated a considerable amount of data. This data—generation has continued at an accelerated pace as interest in the applicability of the research to the court system continues to be expressed. It is to this resurgence of interest in eyewitness research to which we now turn.

**Current Research**

As previously noted, renewed interest in eyewitness research in the early 1970's was primarily due to the theoretical shift in modern experimental psychology brought about by the so—called "cognitive revolution". Although behaviorist methodology continued to be employed by most cognitive psychologists, the focus within psychology shifted from concentration on the study of behaviour to interest in the cognitive processes which mediate that
behaviour. Rather than viewing the human being as a passive recipient of environmental input, psychologists assigned an active role to the individual in construing his or her environment. In memory research this "new" conception of human nature led to an emphasis on the constructive role of perceptual processes and the reconstructive nature of memory. The importance of meaning as the basis for organizing memories also began to be recognized among psychologists studying memory.

Along with these theoretical changes within psychology, a movement developed, calling for a shift away from an over-reliance on laboratory techniques toward an emphasis on "ecologically valid" research. (e.g. Neisser, 1976). This movement generated a great deal of discussion and debate which resulted in the term "ecological validity" becoming the psychological buzz phrase of the late 1970's. Although the laboratory continued as the primary focal point in memory research, innovative research outside of the laboratory also began to appear. In his text, *Memory Observed* (1982), Ulric Neisser chronicled studies focussing on such things as memory for the Talmud, John Dean's memory, memory for oral poetry, childhood memories, memories for critical events, and eyewitness memory.

These theoretical and methodological shifts within psychology, along with the faint, but insistent call for a renewal of research applicable to the criminal justice system, precipitated the re-emergence of eyewitness research. The "new" field of eyewitness memory research began primarily with two different approaches, that of Elizabeth Loftus and her colleagues at the University of Washington and that of Robert Buckhout and his colleagues at the Center for Responsive Psychology in New York.

The impetus for Loftus's eyewitness research was clearly theoretical. The purpose of her research was (and, is) to demonstrate the reconstructive
nature of memory, and eyewitness research provided an ideal means for such a demonstration. As Loftus explained in the introduction to one of her studies (Loftus, 1975), we rarely report our memories in any precise way, but there are occasions when we are called upon to do so, such as when witnessing a crime. In an eyewitness situation, the witness is called upon to relate his/her memory of an incident precisely and in great detail. By providing a subject—witness with a simulated or filmed crime, a psychologist could compare the original incident with the subject's memory or "reconstruction" of that incident. Eyewitnesses would thus provide an excellent vehicle for the study of reconstructive memory.

The paradigm employed by Loftus included the presentation of a slide sequence or a film which depicted either a purse snatching or an automobile accident, followed by a series of questions, some of which contained misleading information or presuppositions. Loftus and her colleagues found that this misleading information was often incorporated into the witnesses' memory and these findings were interpreted as demonstrations of the malleability of human memory and its reconstructive nature (Loftus & Palmer, 1974; Loftus, 1975; Loftus & Zanni, 1975; Loftus, Miller, & Burns, 1978). While the ramifications of these findings for the criminal justice system were discussed by Loftus and her colleagues (e.g. Loftus, 1975; Loftus, Miller, & Burns, 1978), it is clear that the research was primarily designed to address theoretical issues concerning the nature of memory.

Robert Buckhout's research, on the other hand, was designed solely as a means of demonstrating the unreliability of eyewitness testimony. The focus of his research was clearly on the applicability of his findings to the criminal justice system, and not on the theoretical aspects of memory. His concern for real—world applicability led also to an emphasis on designing ecological valid
research. The paradigms employed by Buckhout and his colleagues, therefore, included a staged live event, sometimes followed by a statement from the witness, but always including an identification procedure during which the "witness" attempted to identify a target person from a either a photospread or a filmed or live line-up. Using this procedure Buckhout found consistently low accuracy rates in both the witnesses' statements and identification procedures (less than 50% accurate). These findings led Buckhout to conclude that eyewitness testimony is highly unreliable (Buckhout, 1974; Buckhout, 1975; Buckhout, Figueroa, & Hoff, 1975).

The renewed eyewitness research of the 1970's began with these two varying approaches: Loftus et al's focus on theory and laboratory research and Buckhout's focus on application and live, staged events. The focus on theory, however, has faded to the point that today eyewitness research is generally considered in terms of its applicability to the criminal justice system. As stated by Malpass & Devine (1981):

The purpose of eyewitness identification research is to contribute to the solution of the practical problems of obtaining accurate criminal identification, to assist legal fact finders in evaluating eyewitness testimony, and to assist lawmakers in formulating procedures for developing valid eyewitness evidence.

This focus on eyewitness research as applied psychology has also led to an emphasis on "real world" events rather than the use of pictures, slides, or films as events. The current situation is thus remarkably similar to that of the earlier Aussage psychology with its applied orientation and emphasis on the use of "event tests" rather than static pictures. In order to assess other similarities and differences between modern eyewitness research and Aussage psychology, current research findings are presented below in a format similar to that used previously in the review of Aussage psychology. The following presentation differs from the earlier review, however, in that it offers a more
extensive analysis of research efforts. Methodological problems and criticisms will be raised throughout this review and current research will be compared with previous Aussage psychology.

**Fallibility of eyewitness memory**

In keeping with previous Aussage psychology, current research has largely focussed on demonstration of the fallibility of eyewitness memory (Clifford & Bull, 1978; Loftus, 1979; Yarmey, 1979). A major shift in the focus of the literature has occurred, however, in that fallibility is assessed more in terms of identification errors than in terms of report errors.

Staged crime research has demonstrated a high degree of fallibility among witnesses (e.g., Buckhout, 1974; Buckhout, Figueroa, & Hoff, 1975; Leippe, Wells, & Ostrom, 1978; Lindsay, Wells, & Rumpel, 1981; Malpass & Devine, 1981; Wells, Lindsay, & Ferguson, 1979). Buckhout (1974), for example, found that for 141 witnesses of a staged event, their sworn statements were only 25% correct and only 40% of their identifications (made seven weeks after the incident) were accurate. Similarly, Leippe, Wells, & Ostrom (1978) reported only 31% accuracy for 43 witnesses across conditions using a photospread with a staged theft paradigm. Gorenstein & Ellsworth (1980), likewise, found very low identification rates. Four to six days after a confederate interrupted a class, students were correct in their identifications of the confederate only 39% of the time.

Higher degrees of accuracy have been found by some researchers using the live event paradigm, however. Wells, Lindsay, & Ferguson (1979) staged a calculator theft individually for 127 individuals: 58% made correct identifications; 20% made false identifications; 21% made no identification. Similar results were obtained by Wells & Leippe (1981) using the same paradigm for 107 witnesses:
53% made correct identifications; 29% made false identifications; 16% made no choice.

The fallibility of eyewitness memory has also been demonstrated by showing "witnesses" videotapes or films of incidents, followed by written statements and/or identification procedures (e.g., Buckhout, 1975; Clifford & Hollin, 1981; Clifford & Scott, 1978; Sanders & Warnick, 1981; Yuille, 1984). For example, Buckhout (1975) arranged for a showing of a twelve second mugging on a New York television news program. Viewers were then shown a six-person line-up and asked to telephone in their choice of the perpetrator. Out of 2100 call-ins, only 14.1% picked the correct person.

Similarly, Yuille (1984) showed six different versions of a videotaped bank robbery to small groups of subject-witnesses followed by an identification procedure. Witnesses' performance on the identification task was only slightly greater than chance; only 13.2% of 212 witnesses made a correct identification. Clifford & Hollin (1981) also found very low accuracy on a photospread identification task following a videotaped crime; out of 60 subjects, only 27% choose the correct photo.

Such high rates of identification errors led Yarmey to conclude in his 1979 text summarizing eyewitness research that identification errors are the rule rather than the exception. This essentially negativistic view of the eyewitness is in keeping with the image of the witness presented by Aussage psychology and leads one to conclude that research thus far has been "concerned with showing the witness's fallibility and his inability to recall accurately physical actions, person descriptions or verbalizations, and his inability to identify unequivocally a once seen person when that person is presented in a lineup, photo spread or mugshot book" (Clifford & Lloyd-Bostock, 1983, p.286).
The range of errors and the varying degrees of accuracy, however, have been quite extreme. The above cited findings range in accuracy from 14% to 58%, but the overall range is even greater. Malpass & Devine (1981) staged a vandalism, for example, in which in one condition, 83% of the witnesses correctly identified the vandal from a live line-up and no witnesses made a false identification. Similarly, in one condition of a staged calculator theft, Lindsay, Wells, & Rumpel (1981) found witnesses to be 74% accurate in choosing the thief from a photospread. Why is there such a range among findings? And, given that some studies show high rates of accuracy, why is an over-all negative view of the eyewitness presented in the literature?

Wells (1978) suggested that the negative assessment of the eyewitness may be preferred among researchers as being "infinitely more interesting, more publishable, and more socially important" (p.1551). Such a preference may lead to a demonstration of low accuracy rates through the adoption of particular experimental designs which are conducive to that purpose. Let us consider some of these design possibilities.

First, there is the question of the presentation medium. As we have seen, this has been a matter of controversy throughout the history of eyewitness testimony research. There seems to be little disagreement among researchers that pictures and slides are inappropriate for recreating dynamic events. The use of filmed or videotaped events, on the other hand, has been more widely accepted within the field. Yuille & Cutshall (1984) have demonstrated, however, that when the same incident is viewed either via videotape or live, the videotaped presentation consistently yields fewer recall items and lower identification accuracy. Thus, videotape or film presentations may produce artificially low accuracy rates.
There also are problems with staged events which may serve to artificially inflate inaccuracy rates. For example, Wells (1978) suggested that by choosing a confederate perpetrator who has "no outstanding features" identification may be made much more difficult than most real-life situations warrant. Also, the experimenter is in control of how long the perpetrator will appear before the witnesses and the degree to which he/she will be visible. The range for manipulation here was demonstrated in a study by Lindsay, Wells, & Rumpel (1981) in which they successfully produced three distinctly different accuracy conditions by altering the appearance of the confederate perpetrator and the length of time during which he was visible. In one condition, the "thief" wore a hat pulled down low and he was present for 12 seconds. In a second condition, the cap was worn higher while the time present remained at 12 seconds. In a third, optimal condition, the "thief" wore no hat and he was present for 20 seconds. These three conditions elicited the following accurate identification rates: Condition One, 33%; Condition Two, 50%; Condition Three, 74%. This suggests that by arranging the situation properly, it is possible for an experimenter to produce a particular range of accuracy rates.

In addition, accuracy rates will vary according to the identification procedures used. For example, Wells, Ferguson, & Lindsay (1981) staged individual calculator thefts for 80 witnesses and then presented a photospread identification task. When the "thief" was in the line-up, accurate identifications were made by 57% of the witnesses and there were 25% false identifications and 18% no choice decisions. When the "thief" was not in the line-up, however, false identifications were reported 80% of the time, with correct no-choice decisions being made only 20% of the time. Thus, accuracy rates differ as a function of perpetrator—present or perpetrator—absent lineups.
This situation is further complicated by whether or not subjects receive biased identification instructions, that is, whether or not they are led to believe the perpetrator is actually in the lineup. Malpass & Devine (1981a) tested the impact of biased instructions after presenting a staged vandalism to 350 student observers, 100 of which attended a subsequent lineup procedure. When the "vandal" was present in the lineup, biased instructions (implying he was present) elicited a 100% rate of choosing with a 75% rate of accuracy. Unbiased instructions in the "vandal—present" condition yielded a 83% choosing rate and a 83% rate of accuracy (there were no false identifications). When the "vandal" was not present in the lineup, biased instructions led to a 78% choosing rate and therefore a 78% error rate. Unbiased instructions in the "vandal—absent" condition led to a significantly lower error rate of 33%. Note that this 33% error rate is markedly lower than the 80% error rate reported by Wells, Ferguson, & Lindsay (1981) in their perpetrator—absent lineup.

Another aspect of staged events which may serve to decrease accuracy rates relates to the fact that subject—witnesses are usually informed that they are participating in an experiment before they are given the recall task or identification procedure. Awareness of the staged nature of the incident removes the consequences of the witnesses' actions both for the witness himself/herself and for the suspect (Wells, 1978; Murray & Wells, 1982). The various emotional and social factors which impact the real witness are thus not present. Experiments with increased realism therefore have been designed in an attempt to test this possibility.

Malpass and Devine (1981b) staged a vandalism incident and kept participants blind to the deception through the identification procedure. During identification, subjects were led to believe that the suspect would either suffer trivial consequences ("a good talking to from the Dean") or severe
consequences ("prosecution could mean a felony conviction and possibly some time in jail"). In addition, an equal number of subjects in each condition viewed either a perpetrator—present or a perpetrator—absent lineup. Results showed no difference in overall error rate between the severe consequences condition (47% errors) and the trivial consequences condition (48% errors). However, there was a striking interaction between these two conditions and the composition of the lineup. In the severe punishment condition 83% of the subjects made a lineup choice. When the "vandal" was present in the lineup, this high choosing rate led to a high accuracy rate (75%). When the "vandal" was absent, however, the high choosing rate led to correct rejection of the lineup only 27% of the time. The reverse trend occurred with the subjects in the trivial consequences condition. The overall choosing rate of these subjects was only 26%. When viewing the "vandal—present" lineup, this low choosing rate led to a low accuracy rate (30%). When the "vandal" was absent, however, the low choosing rate led to correct rejection of the lineup 78% of the time. These results illustrate that a complex of factors influence eyewitness behaviour which have perhaps been obscured by the use of simple experimental designs. It also appears that describing eyewitnesses in broad terms such as "the fallibility of eyewitnesses" does not capture the complexity of the phenomenon as it operates in more realistic situations.

One study reported in the literature deals with identification of someone seen previously in a real—world setting. A field experiment was conducted by Brigham, Maass, Snyder, & Spaulding (1982) in which 73 clerks at various convenience stores were asked to identify from photograph lineups a customer who had been in the store approximately 2 hours previously. Excluding no— guesses, identifications were correct less than half of the time (46.8%). This identification rate is similar to those found in simulated crime events; however,
as Brigham et al. pointed out, their paradigm was "more directly analogous to an alibi situation (Was the person x at a given place at a given time?) than to a potentially violent robbery attempt. It could be expected that the clerks' degree of attention to our customers would have been considerably less than in a robbery, depressing the recognition rate" (p.679).

The low accuracy rates reported in the eyewitness literature might also be questioned in terms of "volunteerism". As Wells (1978) has pointed out, the accurate/inaccurate ratio might be substantially higher if only volunteer witnesses are used. In most criminal cases, witnesses are those who indicate overtly that they can make an identification; whereas most research includes identification attempts by all witnesses present. Thus, the accuracy rates in experimental research may be artificially lowered. For example, in the field experiment reported above, Brigham et al.(1982) found an overall accuracy rate of 46.8%, but among clerks who agreed to swear in court to their identification, the accuracy rate was 85%. In a real—life situation these are the only witnesses who would actually appear in court; therefore, caution should be exercised when an assessment of the fallibility of real—world witnesses is based on laboratory findings.

Another feature of the experimental design employed in eyewitness research which could serve to influence accuracy rates concerns the type of questionnaire used to assess accuracy. Forced—choice questionnaires, i.e. multiple choice questionnaires, are often preferred because of the degree of control they afford the experimenter and because of their ease in scoring. Most of Loftus' research, for example, has employed this type of questionnaire to assess accuracy. A study by Lipton (1977), which is discussed more completely in a subsequent chapter, indicated that accuracy rates were 83% with the use of open—ended questions, but only 56% with the use of multiple
choice questions. This difference may be explained, in part, by the respondents' increased tendency to attempt to answer multiple choice questions (75% complete), perhaps often guessing; whereas, respondents declined to do so when answering open-ended questions (32% complete). The task demands engendered by multiple choice questions could therefore serve to inflate error rates.

In summary, we see that current research has focussed upon the fallibility of eyewitnesses in much the same fashion as Aussage psychologists did seventy-five years ago. The suggestion of sensationalism as a motive also appears today as it did earlier in reference to Munsterberg's claims. It also appears that certain features of experimental design can contribute to deflated accuracy rates. In addition, we see present-day psychologists, like the Aussage psychologists before them, faulting research findings on the basis that they fail to meet certain reality criteria.

Individual differences

The effect of male—female differences in reporting an incident or identifying a suspect varies considerably according to current research findings. Females are superior in facial identification according to Goldstein & Chance (1970), Ellis, Shepherd, & Bruce (1973), Going & Read (1974), Lipton (1977), and Yarmey (1974). Witryol & Kaess (1957) reported that males are superior for male faces and females are superior for female faces. McKelvie (1978) reported no reliable differences between males' and females' facial recognition abilities.

Powers, Andriks, & Loftus (1979) reported that females are more accurate in recall and more resistant to suggestion concerning female-oriented details, such as women's clothing; whereas, males are more accurate in recall
and more resistant to suggestion concerning male-oriented details, such as a male thief.

Research focussing on the impact of violence and stress on witnesses has reported the following differential effects according to the sex of the witness. Clifford & Scott (1978) showed subjects a one minute videotape of either a violent or a non-violent incident. Males and females performed equally well when reporting details of the non-violent incident; however, males showed superior recall of details for the violent incident. In an archival study of violent crimes, Kuehn (1974) reported that female victims are less complete in their verbal descriptions of assailants than male victims. Contrary to the above findings, Yarmey & Jones (1983) presented a slide sequence of a criminal assault-implied rape and found no difference between males and females in recall or identification of suspect or victim. More women, however, reported that the suspect was not in the lineup and males who reported strong anti-rape attitudes were more confident in their ability to identify the suspect even though they were not more accurate.

Once again these findings are similar to those reported by Aussage psychologists in that there is no consensus as to the effect of sex differences on eyewitness behavior. There is the suggestion, however, that females respond differently to arousal than males. The tenuous nature of this suggestion will be explored more fully in our discussion of the overall effects of stress. Concerning the above finding, however, I would suggest that differences in performance vacillate more as a function of experimental design than as a function of sex differences.

Current research has generally not pursued the question of intelligence differences among witnesses. This perhaps is due to the fact that research such
as that reported by Goetze (1980) has found no significant relationship between IQ ratings and witness' ability.

Effects of age differences have been explored by current researchers, however. Studies concerned with children as witnesses have been few in number, but they have agreed in their conclusion that children perform more poorly than adults in unstructured recall situations such as free narration (Marin et al., 1979; Goetze, 1980; Cohen & Harrick, 1980; King & Yuille, 1985). Researchers have disagreed, however, concerning children's performance on more structured tasks. Marin, Guth, Holmes, & Kovac (1979), for example, have suggested that children perform more poorly than adults on questionnaire tasks; whereas King & Yuille (1985) have demonstrated a high level of performance with children who were asked specific questions about a previously witnessed event.

Research on age differences has been recently expanded from comparisons of children and adults to comparisons of the witnessing abilities of young adults and the elderly (Yarmey & Kent, 1980; Yarmey, 1983; Yarmey, Jones, & Rashid, 1983). Yarmey et al. (1980; 1983) reported that the elderly performed poorer than the young adults on verbal recall. Yarmey & Kent (1980), for example, found an error rate of 19% for young adults as opposed to a 29% error rate for elderly subjects. There are conflicting findings, however, concerning the identification performance of the elderly. Yarmey & Kent (1980) found the elderly to be as accurate as the younger subjects; whereas Yarmey (1983) reported that the elderly, in contrast to young adults, are more likely to misidentify an innocent bystander to a crime if he looks like a stereotypical criminal. Similarly Yarmey et al. (1983) reported that the elderly had substantially higher false alarm rates. These findings led Yarmey et al. (1983)
to conclude that "the elderly, on the average, are less reliable witnesses to a crime than are young adults."

Yarmey's conclusion, however, may be unwarranted. All of the research conducted by Yarmey and his colleagues employed slide presentations, making the generalizability of the results highly suspect. In addition, since verbal recall was measured by written questionnaires, the young adults' better performance may merely reflect their exam-taking expertise as students rather than their greater recall ability. Furthermore, the Yarmey & Kent (1980) study tested 40 elderly subjects after rejecting 39 other elderly adults because "of their apparent inability to follow instructions, or for their lack of cooperation" (p.361). These experimental design problems make this research highly questionable.

In summary, we see that current research finds no systematic patterns in individual differences among witnesses due to age, sex, or intelligence. The inconclusive nature of these findings is remarkably similar to that of the Aussage research. In addition, we continue to encounter problems with experimental procedures and questionable generalizations.

Effect of Time Interval Between Event and Identification

It is generally accepted that information fades with the passage of time, but the question of interest to eyewitness researchers is "how long is it likely to be until it fades to a level of complete unreliability?" (Shepard, 1983). This question is of particular interest because it is included as one of the five critical factors discussed by the U.S. Supreme Court in the case of Neil v. Biggers and offered by the Court as guidelines for determining the accuracy of a witness's testimony (Neil v. Biggers, 1972). There is no unequivocal answer to this question, but the following research findings bear upon it.
Rather rapid memory loss has been demonstrated in several studies. Lipton (1977) found that after one week, subjects generated 18% fewer recall items with 4% more errors than those questioned immediately. Sanders & Warnick (1981) found a significant reduction in recall of a target person's physical features after only a four minute delay compared to immediate recall. In a field study, Brigham et al. (1982) found that clerks could correctly identify a confederate—customer slightly less than half of the time after a two hour delay. After a twenty four hour delay, however, the rate of correct identifications dropped to chance level.

Research findings such as these led Penrod, Loftus & Winkler (1982) to conclude in their summary of research findings that "research on the length of the retention interval supports Ebbinghaus's (1885/1964) original finding that the longer the interval, the worse the performance" (p. 136). Other research findings, reported below, paint a much more complex picture, however.

Egan, Pittner, & Goldstein (1977) presented two live targets to subjects for 15 seconds and asked them to imagine that the targets had just robbed a bank. Subjects were recalled after two days, 21 days or 56 days and were asked to view either a live lineup or photospread in order to identify the previously seen target. Although identification rates were higher for the live lineup than for the photospread, the rate of correct identifications did not vary over the three time intervals. The rate of selecting a non—target, however, increased from 48% after two days to 62% after 21 days and 93% after 56 days, suggesting that the rate of false identifications increases over time.

A study by Malpass & Devine (1981b), however, found both a decrease in accurate identifications and an increase in incorrect identifications. An act of vandalism was staged before unsuspecting subjects, some of whom were recalled for identification parades on the next three days. Eighty three percent of these
witnesses correctly identified the vandal, none made a false identification. Five months later, a separate group of witnesses from the initial incident were shown a photospread; 36% correctly identified the suspect, 35% made false identifications. These findings represent a substantial drop in identification accuracy both in the decreasing rate of correct identifications and the increasing rate of false identifications. However, if there is a differential effect due to viewing a live lineup as opposed to a photospread (see for example, Dent, 1977; Egan et al., 1977; Yuille & Cutshall 1984), the results of this study are confounded by the use of live lineups for the immediate recall and photos for the delayed recall.

Shepard (1983) reports a series of experiments designed specifically to show the impact of delay on identification. Interruption of a pencil and paper test was staged for a non-university population. The intervention lasted approximately 45 seconds and subjects remained uninformed concerning the purpose of the experiment. Groups of subjects were recalled either after one week, one month, three months or eleven months and asked to identify from a videotaped lineup, the man who had previously interrupted testing. The percentage of correct identifications over three months showed no significant decline (65%, 55%, 50%); nor did the percentage of incorrect identifications increase significantly (15%, 20%, 20%). However, there was a sharp decline in correct identifications after eleven months (down to 10%). Incorrect identifications did not increase even after eleven months; instead 75% of the subjects made no choice. Two other similar experiments were conducted with identification tasks spread over a four month period. No significant effects were found for either decreasing correct identifications or increasing false indentifications.
The impact of delay, however, may not be so much time per se as the activities which have occurred during the intervening time. The work of Loftus and her colleagues, for example, indicates that erroneous and misleading information received subsequent to an event may become incorporated into the witness's memory, thereby distorting the original memory (Loftus, 1975; Loftus & Palmer, 1974; Loftus, Miller, & Burns, 1978). The accuracy of identification has also been shown to be adversely affected by viewing mugshots after the incident but before the lineup (e.g., Davies, Shepherd, & Ellis, 1979; Gorenstein & Ellsworth, 1980). Deffenbacher, Carr, & Leu (1981), however, have shown that faces are less susceptible to the effects of mugshot interference if testing is delayed for two weeks. Shepherd (1983) found that the interpolated tasks of viewing 40 mugshots, or providing a description of the target, or answering a general questionnaire about memory had no effect on identification performance measures at either a one month or four month interval.

In summary, we see no clear consensus concerning the impact of delay or interpolated activity during the delay period; instead, it appears to differ as a function of differing experimental designs. It does appear, however, that Aussage psychology was correct in its assessment that eyewitness memory does not follow the simple pattern of Ebbinghaus's forgetting curve. This stands in marked contrast to Penrod et al.'s (1982) conclusion, which appears to be based on a restricted range of research. In assessing the degree of progress of current research, in this instance there appears to be digression from the days of Aussage psychology rather than progression.

Effect of form of report: Narrative v. Interrogatory report

Although the emphasis in current research has been upon identification rather than report, some research has been directed toward the issue of the
effect of the form of report. Lipton (1977) showed a filmed murder sequence to subjects and assessed their recall in terms of accuracy and completeness by means of various types of report. Lipton found that narrative or free recall reports were 91% accurate and 21% complete. Open-ended questions were 83% accurate and 32% complete; whereas, multiple-choice questions were 56% accurate and 75% complete. Marquis, Marshall, & Oskamp (1972) conducted a similar study using a film of an automobile accident and in agreement with Lipton they found that the narrative was more accurate, but less complete than the interrogative.

Clifford & Scott (1978), on the other hand, failed to obtain a difference in accuracy for free narrative versus structured questions. This finding, however, is the only reported research which has failed to confirm the above-mentioned conclusion. Yuille (1984), for example, has shown the robustness of this phenomenon by demonstrating that the effect carries across medium of presentation. Subjects who completed a narrative and structured recall after viewing a videotape of a robbery were 90% accurate in the narrative and 72% accurate in completion of the structured recall. Likewise, subjects who viewed a live, staged event were 96% accurate in their narrative reports and 79% accurate in the questionnaire.

Other research has looked more closely at the particular aspects of recall. Tickner & Poulton (1975) found that subjects who viewed a videotaped street scene detected more criminal actions than target persons. Similarly, Clifford & Scott (1978) reported that subjects provided more complete information about actions in a one minute film than about persons in the film. Yuille & Cutshall (1984) found similar results for subjects who viewed a live-staged event; however, they found a reversed trend for subjects who viewed a videotaped bank robbery. These subjects reported more descriptive details
point out that these findings raise the question of task demands, suggesting that the nature of the event will elicit a particular pattern of responding.

The fact that recall of the sequence of an event varies independently of recall of descriptive details, however, suggests that "eyewitness recall is mediated by complex processes, and not by some isomorphic representation of the event" (Yuille, 1984). Furthermore, it has been demonstrated that there is no relationship between accuracy or congruence of a person's description of a suspect and his or her identification of the suspect (Yuille, 1984). This finding calls into question another one of the factors specified by the U. S. Supreme Court in Neil v. Biggers as representing a guideline for determining the accuracy of a witness's testimony. In Neil v. Biggers the Court suggests that identification evidence can be evaluated by considering "the accuracy of the witness' prior description of the criminal" (Neil v. Biggers, 1972).

On balance, we see that research has been consistent from Aussage psychology to the present in noting the higher degree of accuracy with narrative reports and the greater degree of completeness with structured recall. Present research, however, has expanded on earlier research to include a comparison of recall and identification. Such a comparison shows that present Court guidelines for determining the accuracy of identification evidence may be in error.

Suggestibility of questions and post-event information

As previously mentioned, the work of Elizabeth Loftus and her colleagues has consistently demonstrated witnesses' susceptibility to suggestive questions (Loftus & Palmer, 1974; Loftus, 1975; Loftus & Zanni, 1975; Loftus, Miller, & Burns, 1978). For example, Loftus & Palmer (1974) demonstrated the
effect of question wording by showing that use of the word *smashed* in reference to a filmed car accident elicited significantly higher speed estimates (avg. of 40.8 mph) than use of the word *hit* in the same context (avg. of 34 mph). In a similar vein, using a slide sequence as the witnessed event, Loftus & Zanni (1975) showed that use of the definite article *the* elicited higher false answer rates (15%—20%) than use of the indefinite article *a* (6%—7%).

Loftus, Miller & Burns (1978) also employed a slide sequence to demonstrate that when inconsistent post-event information is given to a "witness", the rate of correct responses to a questionnaire decreases significantly. Proportions of correct responses dropped from 80% when only consistent information was available to 46% when misleading information was given immediately after viewing the slide sequence, and when misleading information was delayed by one day to one week, correct responses fell to 20%. These results led Loftus, Miller, & Burns (1978) to conclude that new information "writes over" old information so that the old memory no longer exists.

The work of Loftus and her colleagues has been interpreted as a demonstration of the malleability of eyewitness memory; however, other research suggests there may be severe limitations to the generalizability of these findings. First, there is the problem of replication; Zanni & Offermann (1978) failed to obtain a significant effect for definite articles versus indefinite articles using the Loftus paradigm. Secondly, there is the crucial question of the generalizability of the slides/film paradigm employed by Loftus et al. to real world situations. It has been argued, for example, that slides cannot capture the dynamics of a real-world situation (Whipple, 1909; Stern, 1910). In addition, as previously discussed, Yuille (1984) has shown that filmed presentations are of dubious validity in terms of assessing eyewitness performance in real-world situations. Witnesses of filmed sequences, for
instance, may be more susceptible to misleading information due to fewer available cues and the consequent need to "fill in the gaps".

Research by Christiaansen, Sweeney, & Ochalek (1983) demonstrated the impact of postevent information on a live event by informing subjects that a man they had seen was either a truck driver or a dancer. Information that the man was a truck driver led to consistently larger weight estimations. Also, referring to the man as a young man as opposed to as a man influenced age estimates. It seems obvious, however, that these findings are a blatant example of the impact of task demands rather than a demonstration of the malleability of memory.

In addition, Dodd & Bradshaw (1980) have argued that these findings are largely due to the pragmatics of language in experimental situations where it is not clear who the speaker is and what his/her intentions are. Dodd & Bradshaw (1980) demonstrated that resistance to misleading information increases dramatically when the credibility of the source is taken into account. They found that the misleading effect was cancelled when the information was attributed to a biased source.

Bregman & McAllister (1982) offer a further qualification concerning the effects of misleading information. These researches found that subjects who made a previous committment to their "testimony" (by signing their statement) were less susceptible to subsequent misleading information.

In summary, once again we find conflicting opinions among psychologists concerning appropriate experimental designs for the study of eyewitness behavior. Laboratory experiments have demonstrated a drastic reduction in accuracy due to misleading information; however, the ecological validity of these experiments has been severely criticized, calling into question the applicability of these findings to witnesses in the real world. The research of
Dodd & Bradshaw (1980) and Bregman & McAllister (1982) suggest that the dynamics of a real-life eyewitness situation provide a context which greatly reduces susceptibility to misleading information compared to laboratory research.

Comparing current research with Aussage psychology, we find general consensus, although current research has refined the original question and expanded research efforts. Whipple's (1909) report of increasing inaccuracy with increasing degrees of suggestibility is in keeping with much of current research efforts. Muscio's (1915) findings concerning the suggestibility of the definite article as opposed to the indefinite article have been replicated and expanded by Loftus, although subsequent research by Zanni & Offerman (1978) failed to replicate these findings. I have not addressed the question of the suggestibility of children, but current research efforts generally agree with Stern's observation that young children are more susceptible to misleading questions than adults (King & Yuille, 1985). The contrast between current research and Aussage psychology lies largely in the qualifications which have been put forward by recent researchers, calling into question the extent to which actual eyewitnesses' memory is malleable.

**Relationship between confidence and accuracy**

A great deal of attention has been focussed on this issue among current researchers due primarily to the emphasis given this relationship by the court. Referring once again to the guidelines set by the U. S. Supreme Court in the case of Neil v. Biggers, we find that one of the criteria for determining the accuracy of a witness's testimony is "the level of certainty demonstrated by the witness at the time of confrontation" (Neil v Biggers, p. l99).
In his review of research findings concerning the relationship between confidence and accuracy, Deffenbacher (1980) cites eleven current, "forensically relevant" studies which show a positive correlation between confidence and accuracy (e.g., Lipton, 1977; Malpass & Devine, 1980; Lindsay, Wells, & Rumpel, 1981); thirteen studies which show no correlation or a slightly negative one (e.g., Buckhout et al, 1975; Leippe et al, 1978); and six studies which show both positive and non-significant correlations in different conditions (e.g., Wells, Ferguson, Lindsay, 1981; Loftus et al, 1978; Clifford & Hollin, 1981).

It is patently obvious that there is no consensus among researchers on this issue, and due to the different experimental designs employed, it is extremely difficult to uncover the extent of the confidence-accuracy relationship in these various studies. Identifications are made from photographs in some of the studies, from live lineups in others. Sometimes subjects are informed that the event has been staged before identification; sometimes they are not. In addition, confidence is sometimes assessed by Likert-type scale and sometimes by open-ended statements, and in others, by whether or not the witness would be willing to swear in court (Wells & Murray, 1983). Amidst this confusion, let us consider some of the variables which some researchers have suggested affect the confidence-accuracy relationship.

Malpass & Devine (1981a) demonstrated that confidence may be related more to merely making a choice than to the accuracy of the choice. They found that when witnesses of a staged vandalism viewed a lineup in which the perpetrator was absent, those making a choice expressed more certainty than those who made no choice, even though the former were actually incorrect and the latter were actually correct.

Leippe (1980) suggested that confidence can be influenced without affecting accuracy and vice versa, indicating that there is not necessarily a
direct correspondence between confidence and accuracy. In a previously mentioned study by Lindsay, Wells, & Rumpel (1981) conditions were varied in order to effect accuracy rates. The manipulations were successful in producing three distinctly different accuracy conditions of 33%, 50%, and 74%. Confidence ratings varied as a function of these conditions, from 62% to 66% to 77%. The rate of change in certainty, however, was not sufficient to match the changes in witnesses' degree of accuracy. This finding suggests that one's confidence may not accurately reflect the degree of accuracy.

A study conducted by Murray & Wells (1982) suggests that the confidence—accuracy relationship is affected by the witness's belief in the reality of the crime. Following a staged theft, subjects who were informed that the theft was staged before making their identification showed a strong confidence—accuracy correlation. This correlation was not found however for the subjects who were not informed and therefore believed the theft to be real. This research suggests that the confidence—accuracy relationship is even more tenuous in more realistic identification situations than has been demonstrated in most experimental studies.

In contradiction to the above, Brigham et al (1982) suggest that increasing realism strengthens the relationship between confidence and accuracy. Whereas the staged—incident paradigms discussed above found no relationship between confidence and accuracy, the field study conducted by Brigham et al (1982) found confidence to be strongly related to accuracy ($r=.50$). As mentioned previously, clerks who felt certain enough to swear in court were accurate 85.4% of the time. Clerks who felt "pretty sure" were accurate 55.3% of the time; whereas, clerks who only "thought" they chose the correct person were right only 28.2% of the time. Clerks who admitted that they were guessing were correct only 17.2% of time, i.e., they performed at chance level.
While these results contrast sharply with the above reported research of Murray & Wells (1982), this field study, like Murray & Wells' study, maintained realism throughout the identification procedure. It should be noted, however, that the clerks in this study viewed only target—present photospreads and therefore the issue raised by Malpass & Devine (1981a) was not addressed.

It is extremely difficult to draw any conclusions from the above research findings concerning the confidence—accuracy relationship. With the shift in focus to identification and the attendant intervening variables, this issue has become decidedly more complex since the days of Aussage psychology. The initial observation of Whipple (1909) still holds, however: "attestation does not guarantee accuracy". In fact, it appears that even after scores of experimental studies, this remains the only firm conclusion which can be drawn concerning the relationship between confidence and accuracy.

Impact of stress

The literature on the effects of arousal on eyewitness memory is also extremely difficult to interpret due to differing experimental designs and different definitions and measures of arousal. For example, in twenty one studies reviewed by Deffenbacher (1983), ten studies suggest that arousal increases eyewitness accuracy or has no effect; eleven studies conclude that arousal decreases eyewitness accuracy. It is interesting to note that of the former ten studies, the majority are live—event presentations whereas of the eleven studies which conclude that arousal decreases eyewitness accuracy, only one study is a live—event study.

In the studies cited by Deffenbacher (1983), manipulation of arousal ranged from white noise to application of shock to varying the level of violence in films and videotapes. Measurements of arousal are generally self—report
scales, due to the unreliability found with various physiological measures (Brigham, Maas, Martinez, & Whittenberger, 1981). Some representative research findings are reviewed below.

Hosch & Cooper (1980) staged individual live events in which a subject witnessed the theft of a calculator, or the theft on his/her own watch, or the target person in a no-theft condition. Identifications were then made from a photospread; accuracy rates were 67% in the calculator-theft condition, 71% in the personal watch theft condition, and 33% in the no-theft condition. These results were interrupted to suggest that low levels of arousal (the arousal associated with personal theft) increase identification accuracy.

Clifford & Scott (1978) showed two versions of a film, one of which was considered arousing (a policeman assaults a third person as opposed to weakly restraining him). They found that recall accuracy of the violent film was significantly poorer (35% overall, 32% for descriptions, 38% for action) than recall of the nonviolent version (46% overall, 41% for descriptions, 51% for action).

Greenberg, Wilson, Ruback, & Mills (1979) led subjects to believe they had been victims of several thefts. It was found that males who expressed high levels of anger were more likely to accurately identify the target person in a lineup (71% accurate) than those who expressed moderate anger (50% accurate) or low anger (17% accurate). Accuracy was unrelated, however, for females, who were 58%, 57%, & 55% accurate across the three levels from high anger to low anger.

In a correlational study, Siegel & Loftus (1978) found that after viewing a slide sequence, performance on a questionnaire designed to test eyewitness ability was negatively correlated with anxiety as measured by the Multiple
Affect Adjective Checklist \( (r=-.20) \). There was not a significant correlation, however, with stress as measured by the Life Experiences Survey.

An archival study conducted by Kuehn (1974) investigated police reports completed by victims of violent crimes. Although accuracy could not be assessed, completeness of reports was assessed. Kuehn's findings do not relate directly to stress, but they are perhaps suggestive and since these findings represent the only non-experimental data, they have been included here. Kuehn found that the threat of weapon was not related to completeness of report; suggesting that arousal engendered by a weapon does not effect the amount of details remembered.

In order to assess the degree of confusion among eyewitness researchers concerning the impact of stress, consider the following widely-cited, unpublished study conducted by Johnson & Scott (1976). Johnson & Scott (1976) exposed subjects to either a target person who was overheard having an innocuous conversation concerning mechanical failure, followed by his appearance before the subject holding a pen in greasy hands or a target person who was overheard having a heated argument, followed by crashing sounds and his appearance before the subject holding a bloodied letter opener. Subjects then completed a free narrative report, a questionnaire, and a mugshot identification task. Indicating their level of arousal, for the most part witnesses rated themselves on the lower end of the arousal scale. This unpublished study is reported by Deffenbacher (1983) as showing both higher recall and recognition under relatively higher arousal conditions for males. Females, likewise, exhibited the same pattern when tested on their memory for the setting and target's actions; however, they performed worse when describing the target. Deffenbacher, thus, included this study among those which show "witness memory facilitated by greater arousal." Leippe, Wells, & Ostrom
(1978), on the other hand, cite Johnson & Scott (1976) as demonstrating that eyewitness accuracy is reduced by highly arousing crime. And finally, Penrod, Loftus, & Winkler (1982) present Johnson & Scott's (1976) results showing that "overall performance on recognition (picking the target person out of 50 photographs) was poorer under high arousal (33% correct) than low arousal conditions (49% correct)" adding, however, that "in general the results are inconclusive with respect to arousal." (p.127) Thus, we see that Johnson & Scott (1976) can be selectively interpreted to show that stress increases accuracy or that stress decreases accuracy or that the findings are inconclusive.

These varying interpretations illustrate a recurring theme in the present review of the experimental literature—There appear to be fundamental problems in our attempts to employ experimental methodology in the study of eyewitnesses. Not only do researchers disagree in their interpretations among various studies, but they also disagree in their interpretations of a particular study. In fact, different aspects of one study can be cited to illustrate opposing viewpoints. It is little wonder, therefore, that jurists and laypersons outside of psychology exhibit skepticism in their assessments of the validity of our experimental findings.

One final study relating to stress will be presented in order to illustrate some of the problems found in experimental research and to point to a recurring problem specific to eyewitness research. In assessing "The Effect of Arousal on Facial Recognition", Brigham, Maass, Martinez, & Whittenberger (1981) manipulated arousal level by applying either "a single electric shock" (moderate arousal) to 10 males and 10 females, or by applying "safe but uncomfortable shocks" (high arousal) to 10 males and 10 females. Arousal was monitored both by self reports and physiological measures (heart rate, GSR, &

54
finger pulse volume response). None of the physiological measures correlated with arousal manipulations or self reports, however. Physiological measures were, therefore, discarded. In addition, it was discovered that self-reported levels of anxiety correlated with the arousal manipulations only for the females, therefore, the data from the males was discarded, leaving only 10 female subjects in each cell.

A recognition task was employed whereby subjects viewed a set of slides of various faces and then viewed another set of slides which were to be identified as "new" or "old" (having been seen in the previous set). The results showed that "recognition performance deteriorated as arousal increased....a previously—seen face was correctly identified 36.5% of the time in the high arousal condition and 59.2% of the time in the moderate arousal condition. By the same token, the frequency of false alarms (faces wrongly labeled as previously seen) was 42.1% in the high arousal condition and 34.2% in the moderate arousal condition" (p. 287).

Keeping in mind that there were only 10 subjects in each condition and all of these subjects were female introductory psychology students, consider the following conclusion drawn in the final discussion section:

In summary, the present study has a number of implications for the legal system, the most important of which is the finding that arousal does impede facial recognition. Considering the generally stressful experience of crimes, eyewitnesses can be expected to provide relatively unreliable evidence. One might expect that eyewitness identification will be particularly unreliable in those crime situations that evoke a more than average degree of stress. Such situations might include violent crimes, unpredictable or unexpected crimes, and cases in which the witness is also the victim (page 291).

Considering the small number of subjects, the fact that all are females and all are university students, the artificial nature of the recognition tasks (slides), the tenuous nature of the arousal manipulation and the unknown
correlation with this measure and stress as it operates in real-life witnessing situations, the above conclusion is, frankly, outrageous. It brings to mind criticisms launched against Munsterberg which accused him of making unwarranted generalizations in "pathetic haste." Also, we once again see researchers focusing on a negative view of the eyewitness, bent on demonstrating his/her unreliability. This attitude persists throughout the majority of research findings. Two reviews of the literature on stress (Penrod, Loftus, & Winkler, 1982; Deffenbacher, 1983) present a host of conflicting results and yet both conclude that high levels of arousal, such as that encountered by witnesses of violent crimes will necessarily adversely affect witness testimony. One wonders to what degree this conclusion is based on evidence as opposed to bias.

The above criticisms serve as a summary of the research which has been reported on stress. As we have seen, research findings conflict, and even though interpretations vary both among and within particular studies, the general conclusion seems to be that arousal has a detrimental effect on witnesses. This conclusion, however, has been drawn from very tenuous experimental findings; the most crucial problem being the lack of an established relationship between the types of arousal produced in contrived experimental settings and the kinds of arousal engendered by witnesses of violent crimes. In addition, individual responses to arousal and individual coping strategies for stress have virtually been ignored.

Thus, the plethora of research which has been conducted since the days of Aus sage psychology appears to have confused the issue rather than clarified it. The Yerkes-Dodson principle which was expressed in reference to eyewitnesses in Whipple's 1915 review continues to be the theoretical backbone
of current research efforts; however, its applicability to a particular eyewitness in a particular situation remains unknown.

LITERATURE REVIEW SUMMARY

With this survey of eyewitness research complete, we can now step back and view the field as it has developed over the past eighty years and as it exists today. One recurring theme clearly emerges from this overview—eyewitness memory is fallible. When we look closer, however, in an attempt to assess the degree of fallibility and the factors which influence eyewitness memory, the picture becomes increasingly hazy. No clear image of the eyewitness, beyond the descriptor "fallible", emerges from the psychological literature. The research findings of Aussage psychology were scanty, inexact, and conflicting. Seventy five years later, eyewitness research fills volumes, sophisticated statistical techniques generate reams of "exact" data, and yet research findings present a confused, conflicting picture of the eyewitness. What has gone wrong? Either eyewitness memory cannot be assessed in a coherent, useful way or psychology has, until now, employed an inappropriate methodology in its attempts to understand eyewitness testimony.

For the present, I will dismiss the first of these options and concentrate on the second possibility. As we have seen, the question of the proper methodology for eyewitness research has continually arisen throughout the history of the discipline. It was obvious to Aussage psychologists, Stern, Whipple, and Munsterberg, that standard laboratory techniques were inappropriate and therefore the "event test" was employed in order to yield "forensically relevant" information. Standard experimental methodology, however, continued to be used; the event test was merely interjected into the already existing experimental framework.
When eyewitness research re-emerged in the early 1970's, there was a "new" theoretical approach which emphasized the reconstructive nature of memory and the importance of context and social and personality variables. This theoretical shift, however, continued to be expressed in terms of turn of the century methodology. The standard experimental framework allowed contextual, social, and personality variables to play only a minor, circumscribed role in actual research. Eyewitness research has moved to the edges of this framework with the use of the event test outside of the laboratory, however research continues to be restricted by experimental methodology.

The review of witness research has indicated some of the restrictions and problems engendered by adherence to experimental methodology. First, there is the recurring problem of ecological validity. This concern is particularly acute in the case of eyewitness research since the overriding purpose of the research is its applicability to real-life situations. Eyewitness researchers have tried to circumvent this restriction by developing "forensically relevant paradigms." Lindsay & Wells (1983), for example, suggest that: "Research employing forensic paradigms is a necessary step before any strong conclusions can be drawn on a forensic issue. Thus employing a forensic paradigm is not merely useful, it is imperative!" (p. 225). As our survey of research has shown, however, there are difficulties with "forensically relevant" paradigms too.

There is the problem, for example, of the experimental population. A computer search of psychology journals produced a list of forty one research articles pertaining specifically to eyewitness testimony which appeared from 1974-1982; of these, 38 or 92% had tested college students exclusively. If we follow the suggestion of Clifford and Lloyd-Bostock (1983) that "before communication with legal personnel, any finding should be shown to be
impervious to the use of different subject populations, different research settings, different experimental materials, and different research designs or methodologies", we will be communicating very few findings except among ourselves (p.289).

Besides the problem of homogeneous subject groups, we have also seen that many of the reported findings are extremely sensitive to "different experimental materials, and different research designs or methodologies". In fact, the assessment of a witness's fallibility seems largely dependent upon the use of particular experimental procedures. By varying aspects of the experimental design, an experimenter can produce a wide range of eyewitness performance, thus casting doubt on the validity of the entire enterprise.

The validity of eyewitness research is also called into question due to the lack of consensus among psychologists concerning research findings. This problem was evident in Aussage research and it persists among current researchers. As we have seen, disagreement exists not only within a particular area of research such as the effect of stress, but disagreement extends to the interpretation of a specific research finding.

One of the consequences of the lack of consensus among psychologists and the questionable generalizability of experimental research has been the existence of a credibility gap between psychology and the law. We have seen the extent of this gap during the Aussage era. The Devlin report (1976) reflects a similar state of affairs today:

It has been represented to us that a gap exists between academic research into the powers of the human mind and the practical requirements of courts of law and the stage seems not yet to have been reached at which the conclusions of psychological research are sufficiently widely accepted or tailored to the needs of the judicial process to become the basis for procedural change" (p. 73).
Malpass & Devine (1981) concur, suggesting that "(w)e have a deserved credibility problem" which can only be rectified by showing that laboratory research is generalizable or by designing studies which deal with more realistic events (p. 348). However, Malpass & Devine conclude that "(n)o matter how well executed or elegant our studies are, they will be of questionable relevance at best without a knowledge of the differences between eyewitnessing in real situations compared with research situations" (p.398).

This last statement hits at the crux of the matter. In order for eyewitness research to become a viable area of inquiry, we must come to understand the dynamics of real—world eyewitness situations. Without this knowledge, we are severly limited in making generalizations from experimental research to real—world situations.

The preceding review of the literature has shown a surprising lack of knowledge concerning real witnesses' performance. There has been only one field study (Brigham et al, 1982) and one archival study (Kuehn, 1974) which address the issue of real eyewitnesses and each of these studies provide only tangentially relevant information about witnessing a crime. Brigham et al's study, as previously noted, was analogous to an alibi situation. Kuehn's 1974 study used data from police reports to assess the relationship between the degree of violence of a crime and the completeness of witness's description of the perpetrator. In Kuehn's study, data were not collected directly from witnesses and the accuracy of witness' reports was not assessed.

Thus, after eighty years of experimental research designed to study eyewitnesses, we emerge with a confused picture of how subject—witnesses perform in simulated crime situations and we have virtually no understanding of how actual eyewitnesses perform in real—world situations. Considering the lack of information about eyewitnesses' memory for actual crimes and the need
for such information, the present research was designed to study eyewitness' behaviour *in situ*.
AN ALTERNATIVE APPROACH TO EYEWITNESS MEMORY RESEARCH

Due to the fact that there has been no eyewitness research involving real-life witnesses, designing and executing such research presented an extraordinary challenge. In an unexplored area of research, one is initially overwhelmed by the plethora of unanswered questions. The first step, therefore, in meeting the present challenge involved selecting a manageable number of the most fundamental and pressing questions concerning real-life eyewitnesses.

The questions deemed most appropriate for an initial exploration of actual eyewitness performance were selected to guide the research reported in this thesis. These questions are encapsulated by the following: (1) How can actual eyewitnesses to violent crimes be characterized? (2) How well or how poorly do actual eyewitnesses perform when reporting the details of a violent crime? The former question provides the basis for the first study to be presented here, entitled "A Profile of Eyewitnesses to Violent Crimes", and the latter question forms the basis for the second study which is entitled "A Case Study of Eyewitness Memory of a Violent Crime."

After formulating these questions the next challenge was the choice of an appropriate methodology. As indicated in the preceding discussion, it is my contention that experimental techniques have proved inappropriate for exploring eyewitness' performance. Since virtually all previous eyewitness research efforts have employed experimental techniques, there was little guidance in the eyewitness testimony literature for a methodology appropriate for the study of real-world eyewitnesses. Stern's call for an historical methodology is an exception to the singular focus on experimental research within the earlier literature, but since there was no elaboration of this technique, it was of little use in formulating a viable methodology for the study of eyewitness'
performance. Thus, a methodology had to be found outside of the confines of mainstream psychological research concerning eyewitnesses.

In the case of the first study, the choice of a methodology appropriate to compiling a profile of actual eyewitnesses was relatively unproblematic. Archival research was employed to secure these data. For the second study, however, the methodology employed in assessing the accuracy of actual eyewitness reports required the development of an interpretative technique peculiar to this type of research, and therefore exploratory in nature. Being non-experimental, both research methodologies employed in these studies differ radically from most previous research efforts in eyewitness testimony. Neither of these studies has as its goal prediction and control; instead, the more modest goal of increased understanding of eyewitnesses in real-world situations guided the present research.

"Beginning at the beginning" was the priority of the first study presented in this thesis. With no information concerning real-world witnesses, the primary task of this study was to begin to build a data base profiling actual eyewitnesses. In order to make this task manageable, the profile was limited to violent crimes (homicides, assaults, and robbery) and focused on providing basic information about eyewitnessing situations such as the number of witnesses involved in various types of violent crimes and the gender and age of these witnesses. How often witnesses are also victims of crime and how often witnesses are acquainted with either or both the perpetrator and the victim were additional questions guiding this research. The profile of eyewitnesses to violent crimes which constitutes the data presented in the first study of this thesis was drawn from actual police files. As archival research the methodology employed in this study was straightforward: police files concerning homicides, robberies, and assaults were examined and information
was extracted relating to eyewitnesses as outlined above. The profile of eyewitnesses to violent crimes which emerges from this study provides essential information regarding actual eyewitnesses; it is the first step toward establishing a data base which reflects some of the characteristics of real-world eyewitness situations. This study also allows an assessment of the research needs of the criminal justice system based on actual eyewitness situations and thereby suggests avenues for future "forensically relevant" research.

From the overall profile provided by the first study, the second study, "A Case Study of Eyewitness Memory of a Violent Crime", narrows its focus to an in-depth analysis of eyewitness' memory for one particular criminal episode. A case study of thirteen witnesses, all of whom witnessed the same gunshooting incident, was conducted. Witnesses were questioned by police at the time of the incident and they were interviewed by our research team four to five months following the incident. Both statements given by each witness were parsed into action and description details. The accuracy of these details was assessed based on a rigorous reconstruction of the event drawn from forensic evidence, additional interviews, and logical constraints on the event. Questions previously explored by eyewitness researchers such as the fallibility of eyewitness memory, the loss of memory over time, and the effect of stress were examined in this study, but the context of this exploration (the witnessing of an actual violent crime with life and death consequences) differed dramatically from previous research efforts. This study thus demonstrates the memorial effects of witnessing a striking real-world violent event as opposed to witnessing more mundane, contrived, or filmed events. Whereas differences in the findings of this study and those reported in the experimental literature will be noted; the major purpose of this case study is to provide an analysis of actual eyewitness performance.
All of the research presented here was conducted with the co-operation of the Royal Canadian Mounted Police (RCMP), specifically the detachment which serves as police force for the community of Burnaby, British Columbia. Burnaby, with a population of 136,465 (1981 census), is a part of the metropolitan Vancouver area. The groundwork for this research was laid with the establishment of a working relationship between the RCMP detachment and our research team which consisted of Dr. John Yuille, the co-ordinator of the project; Ms. Doreen Kum, a research assistant; and myself.

The first step in this research involved our receiving nationwide security clearance from the RCMP which allowed us access to all police files and pertinent forensic materials. Then we each accompanied different police officers and constables on patrol for one week in order to observe the procedures used when making calls and interviewing witnesses. In addition to accompanying police officers on patrol, Ms. Kum and I each spent several weeks at RCMP headquarters in Burnaby, examining police files and interviewing police officers. Throughout the year during which the research was conducted, we maintained close contact with members of the detachment and this allowed us to establish a good working relationship with the detachment. Throughout the project, RCMP members exhibited a great deal of interest in our research and they were very co-operative in supplying information and assistance.
A PROFILE OF EYEWITNESSES TO VIOLENT CRIMES

A primary objective of the present research project was to provide forensically relevant information about eyewitness memory as it operates in real-world crime situations. In order to achieve this objective, baseline data are needed outlining basic information about a variety of witnessing situations. For example, we need to know what types of crimes most often involve witnesses and whether there are typically multiple witnesses involved in certain crimes. It would also be useful to know the ages and gender of witnesses and how this varies with type of crime. Also, how often are witnesses innocent bystanders; how often are they victims? In how many cases are the witnesses acquainted with either or both the victim and the perpetrator of the crime? Answers to these questions would provide a perspective on eyewitness testimony based on the types of situations which actually occur in real-world criminal cases. Eyewitness research could then be conducted in a manner relevant to the needs of the criminal justice system.

The present study was designed as a first step toward building this sorely needed data base. This initial research effort focuses on violent crimes which include murder, assault, and armed robbery. Police files of violent crimes committed during one calendar year were examined and information was extracted from the files which addressed each of the questions indicated above. These particular issues were chosen for specification because of their potential effects on eyewitness testimony. The memorial account of an innocent bystander, for example, may differ from that of a victim. It therefore becomes important to know how often witnesses are also victims. In addition, knowing the inter-relatedness between the type of crime witnessed and the status of the witness (bystander or victim), the age of the witness (child or adult), and the number of other witnesses present would be valuable information for an
eyewitness researcher interested in conducting research based on the needs of
the criminal justice system.

Whereas gender and age of witnesses are included in this study, it
should be noted that detailed demographics were not included in the
information extracted from the police files. No attempt was made to describe a
"typical eyewitness" in terms of either demographics or personality
characteristics. The nature of eyewitness situations is such that any one can
witness a crime at any time and thereby be suddenly thrust into the role of
an eyewitness. Eyewitnesses, therefore, do not comprise a distinct subgroup
within the human community which can be characterized in terms of
personality characteristics, economic status, education, etc. There is no "typical
eyewitness". There may, however, be typical and atypical patterns within crime
categories. For example, one might expect that sexual assaults typically
involve only victim witnesses. If such patterns exist, the present study will
delineate them.

Finally, it should be noted that the present study relies solely on
archival data. This is strictly an information gathering enterprise; no
hypotheses are being postulated or tested. The purpose of this study is to
provide an initial data base, characterizing particular types of real–world
eyewitnessing situations on a number of selected dimensions.

METHOD

The present research was conducted at the RCMP headquarters in
Burnaby, British Columbia. Crime statistics were obtained from Statistics
Canada, indicating the number and types of crimes committed in Canada
during 1983. Along with the national statistics, crime statistics were also
obtained for the province of British Columbia and the municipality of Burnaby. This study focused on violent crimes, which comprise approximately 8.0% of all crimes committed in Canada. For the purpose of this analysis, violent crimes were broken into three primary categories: homicides, robberies, and assaults. Homicides include both first and second degree murder, attempted murder, and manslaughter; robberies include all violent robbery attempts which involve weapons and/or bodily harm or the threat of bodily harm; assaults involve bodily harm or attempted bodily harm and include both sexual and non-sexual assault.

After establishing the total number of violent crimes committed in Burnaby during 1983, a list of the individual case file numbers was obtained for each crime in the three categories. Using this list, police files were manually accessed for a randomly selected group of cases in each category. Seventy five percent of the homicide cases reported in Burnaby during 1983 were examined; 45% of the robbery cases were examined; and 29% of the assault cases were examined.

Each accessed case was entered on a log sheet identified only by its file number. No names or confidential information was recorded from the police files. The type of crime was identified, along with specifics concerning each of the eyewitnesses such as age and gender and his or her relation to the crime (victim or non-victim). This information was compiled for each case examined in the three categories of violent crimes.

It is important to note that this research focused exclusively on eyewitnesses, that is, witnesses who actually witnessed some aspect of the crime during its commission. Material witnesses who supply information concerning actions of the victim or perpetrator which occurred prior to or subsequent to the actual crime have not been included in this study.
RESULTS

Table 1 reveals the breakdown of violent crimes across Canada and in British Columbia and in Burnaby, B.C. When the total number of violent crimes are considered in terms of population (1981 census), we find that Canadawide there were approximately 686 violent crimes per 100,000 persons. In British Columbia there were approximately 1,052 violent crimes per 100,000 persons and in Burnaby there were 815 violent crimes per 100,000 persons.

Whereas Burnaby's violent crime rate is higher than that of Canada as a whole and lower than that of British Columbia, the pattern of violent crime is very similar across the three populations. Across Canada and British Columbia and in Burnaby, homicides comprise less than 1% of violent crimes (.93% Canadawide, .77% in B.C., .72% in Burnaby). Assaults are by far the largest category of violent crimes, accounting for approximately 85% of violent crimes (84.5% Canadawide, 87% in B.C., 84% in Burnaby). Robberies comprise from 12% to 15% of reported violent crimes (14.5% Canadawide, 12% in B.C., 15% in Burnaby).

A total of three hundred and fifty five individual cases of violent crimes were examined from Burnaby police files. Table 2 shows a breakdown of these cases according to the type of crime witnessed and the number of victim and non—victim witnesses in each crime category. The assault cases, which include both non—sexual and sexual assault, are presented both as the total number of assaults and with a breakdown according to the two categories of assaults. As Table 2 indicates, the pattern differs among homicides, robberies, and assaults in regard to the proportion of victim and non—victim witnesses. In the case of homicides, since the victims rarely live long enough to tell their stories, it is not surprising that all of the witnesses
<table>
<thead>
<tr>
<th></th>
<th>Canada</th>
<th>B.C.</th>
<th>Burnaby</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Crimes of Violence</td>
<td>167,102</td>
<td>28,861</td>
<td>1,109</td>
</tr>
<tr>
<td>Homicides</td>
<td>1,562</td>
<td>223</td>
<td>8</td>
</tr>
<tr>
<td>Robberies</td>
<td>24,274</td>
<td>3,445</td>
<td>164</td>
</tr>
<tr>
<td>Assaults</td>
<td>141,266</td>
<td>25,193</td>
<td>937</td>
</tr>
</tbody>
</table>

TABLE I
CRIME STATISTICS
FOR 1983
TABLE II

VIOLENT CRIMES REPORTED IN BURNABY FOR 1983

<table>
<thead>
<tr>
<th>Number of Cases</th>
<th>Victim</th>
<th>Non-Victim</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homicides</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>Robberies</td>
<td>75</td>
<td>85</td>
</tr>
<tr>
<td>Assaults</td>
<td>274</td>
<td>288</td>
</tr>
<tr>
<td>Non-sexual</td>
<td>229</td>
<td>243</td>
</tr>
<tr>
<td>Sexual</td>
<td>45</td>
<td>45</td>
</tr>
</tbody>
</table>
are non—victim witnesses. These witnesses were not unbiased bystanders, however. Twenty of these witnesses were acquainted with the victim and 17 of them also knew the perpetrator of the crime, indicating that homicides are often committed among family and friends.

The robbery cases present a different picture with victim witnesses playing an important role. Of the total 155 witnesses, slightly more than one half (55%) were victims. The non—victim witnesses were often relatives, friends, or acquaintances of the victim; 61% of these witnesses were acquainted with the victim. None of the 155 witnesses, however, reported knowing the perpetrator.

Turning to the breakdown of assault cases, there is a difference in the proportion of victim witnesses in sexual assault cases compared to non—sexual assaults. In non—sexual assault cases, 68% of the witnesses were victims; in sexual assault cases, 45 of 48 witnesses or 94% were victims. Different patterns also emerge between these two groups of victim witnesses in regard to their acquaintance with the perpetrator of the crime. Most of the victims of non—sexual assault (70%) reported that they were previously acquainted with their attacker. The victims of sexual assault, however, often did not know their attacker. Only 44% of these victims reported knowing their attacker. Regarding the non—victim witnesses of assaults, 73% of these witnesses who viewed a non—sexual assault knew the victim and 58% knew the attacker. There were only three non—victim witnesses to a sexual assault and all of them knew the victim, but not the attacker.

Besides the distinction between victim and non—victim witnesses, witnesses were distinguished in terms of gender and age. Gender was most clearly a factor in sexual assaults. Since sexual assault is primarily a crime against females, it is not surprising that 41 of the 45 victims of sexual assault
were females. The remaining three non-victim witnesses were also females. The only other crime category with a gender bias was that of homicides. In the homicide cases, all of the witnesses (21) were males. In both the robbery cases and non-sexual assault cases, the witnesses were fairly evenly divided between males and females. One might expect that victim witnesses of non-sexual assault would be predominantly male, but of the 243 victims examined in these cases, only 52% were males.

Table 3 indicates the ages of witnesses according to crime category. Except for sexual assaults, most witnesses (including victims and non-victims) of violent crimes are young adults ranging in age from 16 to 35 years. Approximately one half of the witnesses to robberies and non-sexual assaults, for example, were between 16 and 25 years of age. Witnesses to sexual assaults (94% of whom were victims) comprised a younger group. Seventy three percent of these witnesses were under 26 years of age, 40% of whom were children under the age of 15.

A final question posed by this study concerns the number of witnesses involved in various types of crimes. The number of witnesses noted for each crime category in Table 2 were not evenly distributed among the cases examined. The range in the number of witnesses per case was the most extreme in the homicide cases. Three of the cases had no witnesses, whereas one case had 16 witnesses. In the remaining two cases, 4 witnesses were involved in one case and 1 witness in the other. All of the other crime categories had at least one witness per case. Even when there were no bystander witnesses, in these cases there was always a victim witness able to report what had occurred. In fact, in the majority of the cases there were only victim witnesses available. In 53% of the robbery cases, 69.5% of nonsexual assaults, and 96% of the sexual assault cases, the victim was the only witness.
### TABLE III

**AGES OF WITNESSES AS A FUNCTION OF CRIME CATEGORY**

<table>
<thead>
<tr>
<th>AGE</th>
<th>Homicides</th>
<th>Robberies</th>
<th>Non-sexual</th>
<th>Sexual</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;15</td>
<td>0</td>
<td>4</td>
<td>33</td>
<td>18</td>
</tr>
<tr>
<td>16–25</td>
<td>4</td>
<td>72</td>
<td>123</td>
<td>15</td>
</tr>
<tr>
<td>26–35</td>
<td>11</td>
<td>19</td>
<td>96</td>
<td>9</td>
</tr>
<tr>
<td>36–45</td>
<td>4</td>
<td>22</td>
<td>49</td>
<td>2</td>
</tr>
<tr>
<td>46–55</td>
<td>1</td>
<td>21</td>
<td>27</td>
<td>1</td>
</tr>
<tr>
<td>56–65</td>
<td>0</td>
<td>6</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>&gt;65</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>N/A</td>
<td>1</td>
<td>7</td>
<td>16</td>
<td>0</td>
</tr>
</tbody>
</table>
Robberies involved more multiple witnessing situations than any of the other crime categories. There were 2–5 witnesses in 42% of the robbery cases and 6 or more witnesses in 5% of the cases. Nonsexual assaults were viewed by the victim and one other witness in 15% of the cases; by 3–5 witnesses in 13.5% of the cases; and by 6–10 witnesses in 2% of the cases.

DISCUSSION

The findings from this study offer an initial profile of eyewitnesses to violent crimes. Drawn from actual police files, this profile delineates some of the parameters of real-world eyewitnessing situations. Before discussing these findings, however, it is important to note that the results of this study were drawn from a relatively small sample and from perhaps an atypical community in terms of the amount of violent crime. It would be a mistake, therefore, to over-generalize from these findings. This is particularly true in regard to the findings relating to homicide which deal with only six cases and twenty-one witnesses. I would not suggest, for example, on the basis of these findings that all witnesses to homicides are males or that it is typical for sixteen witnesses to view a homicide. Such sweeping statements would be wholly inappropriate in this instance. There are, however, some overall patterns in these findings which deserve attention and are discussed below.

The first recurring pattern of interest concerns the role of the victim witness. In the homicide cases there were no victim witnesses—victims rarely survive long enough to serve as witnesses. In all other violent crime cases surveyed in this study, however, the majority of the witnesses were themselves victims. This is, of course, not surprising since each of these crimes necessarily entails at least one victim whereas other witnesses may or may not be present. In fact, in over half of the robbery case and two thirds of the
non-sexual assaults, and virtually all of the sexual assaults, the victim was the only witness available. The primacy of the victim witness in real-world violent crimes (excepting homicide) appears obvious from the findings of this study. As victims, the testimony of these witnesses is liable to influences not present in experimental paradigms, such as high levels of stress, complex motivations, extreme fear, and active participation in the event. Research conducted with "unbiased" witnesses who are less susceptible to these influences is therefore probably not generalizable to victim witnesses. Eyewitness research thus presents an image of the eyewitness (the "unbiased" witness) which does not apply in the majority of violent crime situations. Both the relevance and generalizability of "unbiased" eyewitness research is therefore suspect.

Another important factor to emerge from this study concerns the relationship between the witness and the perpetrator of the crime. Murders and assaults, for example, appear to be committed primarily among friends and family. In these cases both victim witnesses and non-victim witnesses were often acquainted with the perpetrator of the crime. This was true less often in sexual assaults where victim witnesses reported knowing their attackers in less than half of the cases. Robberies, on the other hand, present a very different picture. Out of one hundred and fifty five witnesses, none of the witnesses reported recognizing the perpetrator. Robberies are not crimes of passion where friends or family are the target victims. In addition, the perpetrator of a robbery may disguise himself/herself to make recognition difficult. Regardless of the reasons for this difference between robberies and other crimes of violence, the dynamics created by these differences have important implications for eyewitness research. Identification evidence is of greater significance, for example, in robbery cases where the perpetrator is unknown than in homicide or assault cases in which the perpetrator's identity is known. In these cases,
the witness' account of what happened would be the primary focus. Personal acquaintance with the perpetrator could also significantly bias the witness' account of the crime, making the determination of the truthfulness of his/her testimony especially problematic. The dynamics of these different situations thus call for research strategies which are sensitive to these issues.

Turning to the patterns found among the cases for age and gender of witnesses, across all violent crime cases young adults constitute the largest group of witnesses. It may be the case that this age group constitutes the largest group in society as a whole and therefore also is the most highly represented group of witnesses. Or, as the most mobile and visible age group, this group may be involved more often in violent crime situations. There were no gender differences among witnesses in robbery cases or non-sexual assault cases. The male bias in the homicide cases is most likely artifactual. In sexual assault cases, however, both age and gender produce distinct patterns which differ from the other crime categories. Since sexual assault is primarily a crime against women, it is not surprising that the witnesses of this crime, who are most often also its victims, are females. It is also not surprising, in light of recent widespread publicity, to see the high incidence of children victims to sexual assault revealed in these data. These data suggest therefore that eyewitness research should investigate children's eyewitnessing abilities.

In summary, it appears that the innocent bystander eyewitness is a relatively rare phenomenon in actual criminal cases. Such a witness may happen to witness a robbery, but most often as a witness, one is also a victim. Witnesses also are rarely able to be "objective"—they often are acquainted with either the perpetrator or the victim. Most importantly, these data indicate that witness characteristics differ in systematic ways with type of crime witnessed. Age and gender patterns, for example, differ for sexual assault
crimes and personal acquaintance with the perpetrator is less of a factor in robbery cases. These differences among cases call for specific research strategies which take into account the dynamics engendered by particular types of crime. Eyewitnesses do not constitute a generic category; the context in which the eyewitness is functioning must be considered.
A CASE STUDY OF EYEWITNESS MEMORY OF A VIOLENT CRIME

The preceding study focused on establishing some of the most salient characteristics across eyewitnessing situations and within particular types of crimes. The present study moves from the expansive survey methods of the preceding study to focus on eyewitnesses within one particular case. Rather than reporting eyewitness characteristics, this study focuses on eyewitness performance. The purpose of this study is to determine how accurately eyewitnesses in a real-world situation report the occurrence of a violent crime.

For this study, a specific case was chosen for in-depth analysis from police files of violent crimes committed in Burnaby during 1983. This case involved a gun shooting incident which occurred on a Spring afternoon in Burnaby, outside of a local gunshop in full view of several witnesses. In a face-to-face encounter on the street, a gunshop owner exchanged gunshots with a man who, minutes earlier, had robbed him. Witnesses viewed the incident from various vantage points along the street, from adjacent buildings, or from passing automobiles; and they witnessed various aspects of the incident, either prior to and including the actual shooting or after the shots were fired. A statement was taken from each witness at the time of the incident by police officers. Four to five months later our research team interviewed thirteen of these witnesses. The witness statements taken at these two points in time have been analysed to yield an image of eyewitness performance at the time of the incident and several months following the incident.

This particular case was chosen for analysis for several reasons: (1) There was a sufficient number of witnesses to allow an assessment of accuracy across witnesses. (2) Because the perpetrator was killed and the weapons and
stolen money were confiscated, a great deal of forensic evidence was available by which to assess the validity of witnesses' testimony. (3) The death of the perpetrator closed the police file, thereby allowing our research to proceed without interference in the judicial process. (4) The variety of items visible during the incident (i.e., weapons, perpetrator's automobile, gun boxes, a blanket) provided materials for assessing witness' memory in several ways. For example, the perpetrator's automobile was visible to all witnesses and yet peripheral to the action of the incident, making it an appropriate subject for misleading questions. In addition, witnesses could be questioned in the research interview about items not germane to the police investigation and therefore not mentioned in the original police interview (e.g., the color of the blanket covering the perpetrator's body; the clothing worn by the perpetrator). This allowed an assessment of memory across time unaided by a previous recounting to police.

Whereas the primary purpose of this research was to collect data concerning the performance of actual eyewitnesses to real crimes; both the collection of data and their analyses have been organized to examine the issues emphasized in previous eyewitness research and discussed in the preceding literature review. The nature of the case chosen for analysis allows examination of several of these issues including the accuracy of witness' verbal reports, the effect of question wording, and the impact of stress, but it also precludes the exploration of some issues such as identification accuracy. The case being examined in this study is also comparable to the type of eyewitness situation explored in the experimental research in that this case involves multiple witnesses and these witnesses were neither victims nor were they acquainted with the perpetrator of the crime. These similarities with the
experimental literature allow a comparison between the findings of this study and those reported in the literature.

**METHOD**

**Participants.** Twenty-one witnesses to a gun shooting incident were interviewed by police officers within two days of the incident. Twenty of the twenty-one witnesses were located and asked to participate in the research; thirteen of the witnesses agreed to a research interview. Two refusals resulted because the witnesses had moved from the area and five others did not wish to participate. One of these was the victim, who did not wish to re-live the trauma of the incident in a retelling of it. The 13 witnesses who agreed to an interview included all of the major witnesses except for the victim and one other witness who was in prison at the time of our interview. The 13 witnesses included in the study ranged in age from 15 to 32, with a mean age of 23.5. This group included 3 females and 10 males. Three of the witnesses were high school students; nine witnesses were high school graduates, three of whom had had some post-secondary school training.

**Interview Procedure.** Police officers interviewed 16 witnesses immediately after the incident at the scene of the crime. Four other witnesses were interviewed the following day at either their place of employment or residence. One witness was interviewed two days after the incident. The research team was not present at these interviews, but our observations of police officers' interviewing techniques when accompanying them on patrol had revealed consistency among RCMP members when interviewing witnesses. Witnesses are first asked to recount in their own words what has happened. This is followed by specific questions aimed at clarifying earlier points or
soliciting specific details. Both the questions asked by police officers and the responses given by witnesses are recorded in a verbatim statement written by the police officer. After the statement is completed, its accuracy is verified by the witness, who reads the statement, makes any necessary corrections, and then signs it.

All of the witnesses who had been interviewed by the police were contacted by means of a letter which explained the research project and asked for their consent to be interviewed by our research team. This was followed by a telephone call and the arrangement of an interview with the 13 witnesses who consented to participate in the research. Interviews were conducted four to five months after the incident at a time and place chosen by the witness, usually in his or her home or place of business. Interviews were from 45 to 90 minutes in length and each interview was recorded on audiotape and transcribed. Interviews were conducted in a format which followed police procedure; a free recall of what happened was first solicited followed by specific questions. In a departure from police procedure, two misleading questions were incorporated into the research interviews. One question concerned a headlight on the thief's car. Following the procedure employed by Loftus (1974), half of the witnesses were asked if they had seen the busted headlight on the perpetrator's automobile. The remaining witnesses were asked the same question except that the definite article the was changed to the indefinite article, a. The second question concerned the color of a quarterpanel on the car. Half of the witnesses were asked if they had seen the yellow quarterpanel, while the remainder heard the same question with the indefinite article, a. These two questions were chosen because the car was in a prominent position during the incident and all of the witnesses had noticed its presence. These suggestions were reasonable since the car was a 1967 model
with a large dent on the passenger door. It could easily have had a busted
headlight and it did, in fact, have an offcolor quarterpanel although it was blue
rather than yellow.

In the research interviews, witnesses were also questioned about the
degree of stress which they had experienced at the time of the incident.
Witnesses were asked to rate their stress on a 7-point scale with 1
representing "perfectly calm" and 7 representing "extreme anxiety or stress". In
addition witnesses were asked about their emotional state prior to the incident
and about any negative effects engendered by the incident such as sleeplessness
or nightmares. The degree of contact between a witness and the perpetrator
or the victim or any of the weapons was also ascertained and considered as a
stress factor.

Scoring Procedure. The scoring procedure employed in the present
study evolved from that employed by John Yuille and his associates (e.g.,
Yuille, 1984; King & Yuille, 1985; Yuille & McEwan, 1985; Yuille & Cutshall,
1984) in their experimental eyewitness research. In essence, this procedure
involves partitioning witness statements into single bits of information based
primarily on verb and adverb phrases (action details) and noun and adjective
phrases (descriptive details). After tallying the total number of both action
and descriptive details, each detail is then scored as correct or incorrect.

The nature of the present research, however, necessitated a refinement of
this procedure along with the employment of non-experimental interpretative
techniques. The most significant departure from experimental procedure
occurred in the development of an instrument for determining the accuracy of
witness' statements. This task required the use of a reconstructive methodology
more similar to that of the historian (Stern, 1939) than the experimentalist.
Because the witnessed event had not been prearranged and staged, it was
necessary to first determine what had transpired at the scene of the crime in order to be able to judge the accuracy of witness' statements.

As previously noted, a major reason for the choice of this particular case was the fact that there was a great deal of forensically relevant material available to aid in our assessment of the case. The first step in determining what had transpired during the gunshooting incident involved the assembly of this material, including the following: (1) police reports of forensic evidence gathered at the scene of the crime including photos of the perpetrator's body and his automobile, descriptions of confiscated weapons and stolen articles, physical descriptions of the perpetrator and the victim, the location of blood stains, and the number of expended bullet casings; (2) reports taken from ambulance attendants and public safety personnel at the scene of the crime; (3) verbatim statements taken by police from 21 witnesses who viewed various aspects of the incident; (4) autopsy and medical reports. In addition, the scene of the incident was inspected and photographed by our research team. The photographs included photos of a partial re-enactment of the shooting incident which was staged in order to determine the effect of the viewing angle of some of the witnesses.

Using these materials, the incident was reconstructed. In keeping with Yuille et al's procedure and that employed by other eyewitness researchers (e.g., Clifford & Scott, 1978), the reconstruction focused on two major types of details: (1) action details which relate what actually happened; (2) descriptive details which provide details concerning the appearance and location of people and objects at the scene. Determination of the descriptive details was straightforward. From police photographs and descriptions, a list of descriptive details was compiled which included physical descriptions of the perpetrator, victim, and central witnesses, and descriptions of clothing, weapons, stolen
property, and the perpetrator's automobile. Following Yuille et al's procedure the descriptive details were subdivided into two categories: (1) details concerning people such as descriptions of hair color and style, clothing color and style, and age, height, and weight estimations; (2) details concerning objects, including descriptions and locations of such items as an automobile, weapons, and stolen property.

Determination of the action details and their veracity, however, was more problematic and required a more interpretative technique. The action component encompassed the events which transpired in front of the gunshop from the moment the perpetrator exited the gunshop until the ambulance arrived at the scene approximately ten minutes later, including the actions of all those who came in contact with the weapons and either the perpetrator as he lay dead in the street or the gunshop owner as he lay wounded in the gunshop. Since none of the actions had been preserved on film as had some of the objects at the scene and because of the fleeting nature of an action sequence, verification of the action details was more difficult than that of concrete, relatively long lasting objects. However, with the combination of forensic evidence and reports from all the witnesses and support personnel who viewed the incident from a variety of perspectives, along with the constraints imposed by a logical analysis of the events, there was enough information upon which to base a reconstruction of the event. Using these sources, an outline of the action sequence was independently constructed by each of the members of the research team. These outlines were compared and a sequence was abstracted which reflected agreement among the three raters.

The resulting outline represented an interpretation of the event rather than an objective construction of the event. The latter was not possible. However, it should be noted that the reconstruction was based on forensic
evidence and a logical analysis of the sequence of events. Consensus among witnesses played a role in the reconstruction, but only when the information could logically be pieced together from reports given by witnesses viewing the incident from different perspectives. Agreement among witnesses about a particular detail, without corroborating evidence or without the benefit of a variety of viewing angles, was not considered sufficient for inclusion in the reconstruction. Following these guidelines the accuracy of all but a few details could be determined. In reporting the results, the details whose accuracy could not be determined on the basis of the reconstruction of the event were tallied separately as unclassified details. The unclassified details constituted 3.65% of the total number of details in the police interviews and 6.02% of the total number of details in the research interviews. These details have not been included in the subsequent analyses unless otherwise noted.

After reconstructing a list of descriptive details and an outline of the action details by which to judge the accuracy of the witnesses' statements, both the witnesses' statements taken by the police and those obtained in the research interviews were analyzed in terms of the number of details given and the accuracy of those details. The first step in this analysis consisted of the transposition of each witness' interview transcript from a narrative format into the two components upon which the reconstruction had been based: (1) action details; (2) descriptive details, subdivided into people and object descriptions. A list of action details and a list of both people and object descriptive details were thus compiled for each witness for both the police interview and the research interview. The two aspects of each statement, the free account and the responses to specific questions, were initially analyzed separately (e.g., Cady, 1924; Lipton, 1977; Marquis, Marshall, & Oskamp, 1972; Yuille & McEwan, 1985). However, because no systematic differences were found between
the free accounts and answers to questions, this distinction was collapsed. Repetitions of details which might occur between the free recall portion of the statement and the responses to questions were thus eliminated from the analysis. Unresolved contradictions, however, were included. For example, if a witness described the automobile as a Valiant at one point during the interview and as a Falcon at another point in the same interview, both descriptions would be entered into the analysis unless it was made clear by the witness in subsequent questioning that one response was correct and the other incorrect.

The reliability of the transposition procedure was ascertained by explaining the procedure to a volunteer undergraduate student who was not familiar with the research. The student then transposed several interview transcripts and a comparison between these transcriptions and my own yielded a variance among raters of less than 5% (.044).

After transposing witness' interviews into action details and two categories of descriptive details, each component was tallied separately in terms of the number of details reported. The details were scored by allotting one point for each specific, unique bit of information. Scoring the descriptive details was accomplished primarily by partitioning the details in relation to noun and adjective phrases. For example, the statement "She was 5'2" and wore a yellow sweater" contains three descriptive details. Action details, on the other hand, are revealed primarily in verb and adverb phrases and were scored accordingly. For example, the statement "He turned around and shot the guy in the shoulder" contains three details. Information lacking specificity, e.g., "He was tall", was assigned a half-point. The total number of person and object descriptive details and the total number of action details were tallied separately.
for each witness' statement from both the police interview and the research interview.

Each component of the report was then evaluated in terms of the proportion of details which were correct. Accuracy was judged on the basis of our reconstruction of the event as outlined above. A range was established for judging the accuracy of some numerical estimations; for example, height and age estimations were judged correct when they were plus or minus two (inches or years) of the actual height or age. Weight estimations were judged correct when they were plus or minus five (pounds) of the actual weight. The time of the event was judged correct when given within an established fifteen minute range. There was no leeway given however on such items as the number of gunshots and the date of the incident.

Two other features of the accuracy criteria deserve attention. Qualifiers indicating degree of certainty were not weighted. For example, the statements "She might have been wearing a red shirt" or "Maybe the car was blue, I'm not sure" were scored without considering the qualifiers. Also, allowances were not made for the possibility that the witness' memory was correct in terms of how they had witnessed the event. For example, when a witness reported hearing "five or six gunshots", this was scored as incorrect although it may have been true for that person. Similarly, one witness reported that "the thief looked like he was in his early 20s." In fact, the thief did look like he was in his 20s, but this was scored as incorrect because his actual age was 35. These points indicate the conservative nature of the scoring criteria. The accuracy of each witness' statement was scored independently by three judges and the results compared. There was a variance among scorers of less than 2%.
RESULTS

Number of Details.

A quantitative summary of the total details provided in the eyewitness accounts is found in Table 4. The total number of details the 13 witnesses reported in each of two interviews have been separated into action details, person description details, and object description details (the proportion of the total details which each category constitutes is given in brackets in Table 4). Clearly, considerably more details were elicited in the research interview than in the police interview. This reflects the fact that many questions were asked in the research interview which were solely of memorial and not forensic interest. For example, the witnesses were asked to describe the blanket which covered the thief's body, a fact of no interest to the police. Witnesses were also probed for additional details concerning points they had only briefly noted in their earlier police interviews.

The data in Table 4 demonstrate that the police elicited a higher proportion of action details (60%) than descriptive details. However, the research interview obtained about an equal proportion of action (52%) and descriptive details. The major reason for this difference is that the police requested fewer object descriptions (12% of total details as compared to 22% in the research interview). Since the police had recovered most of the relevant objects (guns, money, etc.) they did not need detailed descriptions.

There was a considerable amount of inter-witness variability in the number of details provided. As reported in Table 5, witnesses reported between 17 and 95.5 details to police, and the comparable range in the research interview was 38 to 123.5 details. This variability primarily reflected the amount of the event that each witness had seen. It turned out that the witnesses could be organized into two distinct
TABLE IV

TOTAL REPORTED CLASSIFIABLE DETAILS

<table>
<thead>
<tr>
<th>Type of Details</th>
<th>Police Interview</th>
<th>Research Interview</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Percent</td>
</tr>
<tr>
<td>Action Details</td>
<td>392</td>
<td>60.35%</td>
</tr>
<tr>
<td>Person Descriptions</td>
<td>180</td>
<td>27.71%</td>
</tr>
<tr>
<td>Object Descriptions</td>
<td>77.5</td>
<td>11.93%</td>
</tr>
</tbody>
</table>

Total Details

|                |        | 649.5 | 1,056.5 |
### TABLE V

**CLASSIFIABLE DETAILS FOR TWO GROUPS OF WITNESSES**

#### Police Interview

<table>
<thead>
<tr>
<th></th>
<th>Peripheral Witnesses</th>
<th>Central Witnesses</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number</strong></td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td><strong>Range</strong></td>
<td>10–20</td>
<td>28–54</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>13.50</td>
<td>44.43</td>
</tr>
<tr>
<td><strong>Details</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Action</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Descriptive</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Research Interview

<table>
<thead>
<tr>
<th></th>
<th>Peripheral Witnesses</th>
<th>Central Witnesses</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number</strong></td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td><strong>Range</strong></td>
<td>19–4</td>
<td>46–63</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>27.67</td>
<td>55.07</td>
</tr>
<tr>
<td><strong>Details</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Action</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Descriptive</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

91
groups (one of 7 witnesses, and the other of the remaining 6). As illustrated in Table 5, these two groups differed substantially in the number of details they recalled. The larger group (n=7) constituted the central witnesses, those who viewed most of the incident from a central vantage point. They averaged 71.93 details in the police interview and 103.71 details in the research interview. The comparable values for the other group were 24.33 and 55.25 details. The central group provided police with more than twice as many details as the peripheral group, and they gave almost twice as many details in the research interview.

Although the witnesses could be distinguished as central or peripheral in terms of the number of details they recalled, it is important to note that these two groups did not differ in the accuracy of their accounts. In the police interview, the overall accuracy of the central witnesses was 84.56% compared with 79.31% in the peripheral group. This difference is not statistically reliable. The numerical difference was even smaller in the research interview: 81.75% accuracy for the central witnesses and 79.39% for the peripheral witnesses. This makes it clear that the accuracy of reported details was not related to the number of those details. The correlation between number and accuracy of details was .23 for the police interview and .05 for the research interview.

Accuracy of Recall.

Each detail was scored as correct, incorrect, or unclassifiable. Unclassified details are not considered in this treatment of the results. Table 6 provides a summary of the accuracy scores for the classifiable details. Both mean and median scores are provided for the two interviews, and the three types of detail. Considering the police interviews, the reports of the witnesses were generally very accurate. Descriptions of people produced the lowest accuracy,
FIGURE 1

POLICE INTERVIEW
TOTAL OF 107.5 ERRORS

RESEARCH INTERVIEW
TOTAL OF 198.5 ERRORS
### TABLE VI

**ACCURACY OF CLASSIFIABLE DETAILS**

<table>
<thead>
<tr>
<th>Type of Detail</th>
<th>Police Interview</th>
<th>Research Interview</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Range</td>
<td>Mean</td>
</tr>
<tr>
<td>Action</td>
<td>40-98</td>
<td>81.90</td>
</tr>
<tr>
<td>People Desc.</td>
<td>33-100</td>
<td>75.57</td>
</tr>
<tr>
<td>Object Desc.</td>
<td>50-100</td>
<td>88.53</td>
</tr>
<tr>
<td>Desc. Subtotal</td>
<td>62-100</td>
<td>82.03</td>
</tr>
<tr>
<td>Total Details</td>
<td>59-96</td>
<td>82.14</td>
</tr>
</tbody>
</table>
whereas the descriptions of objects averaged almost 89% in accuracy. This pattern was the same in the research interviews. Person descriptions were almost 73% accurate, while object descriptions were over 85% correct. Note that there was little change in accuracy over the 4–5 month delay between the two interviews.

Changes in accuracy with time were examined in more detail. The overall accuracy rate for 10 of the 13 witnesses changed by plus or minus one to six percent. Thus there was virtually no change after 4–5 months in most of the witnesses. For one witness accuracy increased 7.65%; for another it decreased 8.72%; and for the third, accuracy decreased 20.13%. The errors committed by the latter witness are discussed below.

Analysis of Errors.

Although errors were relatively rare in the eyewitness accounts, an examination of their nature is useful. The data in Table 7 provide a general picture of the distribution of errors in the two interviews for the three types of details. A comparison of these data with those provided in Table 4 confirms the differential error rate for the three different types of detail. Person descriptive errors are overrepresented and object descriptive errors are underrepresented relative to the frequency of those types of descriptive details recalled. Figure 1 provides a graphic representation of the categories of the errors. Action errors have been separated into Time, Gunshots, and Multiple Action categories. Person descriptions include Statistics, Hair & Clothing, and an Other category. Object descriptions are divided into an Auto category and an Other category.

Considering the action errors, the largest category (Multiple Action) concerned the actions of: (1) the thief or gun shop owner (15 errors in
<table>
<thead>
<tr>
<th>Type of Error</th>
<th>Police Interview</th>
<th>Research Interview</th>
</tr>
</thead>
<tbody>
<tr>
<td>Action</td>
<td>Number</td>
<td>Percentage</td>
</tr>
<tr>
<td></td>
<td>57</td>
<td>53.02%</td>
</tr>
<tr>
<td>People Desc.</td>
<td>44.5</td>
<td>41.40%</td>
</tr>
<tr>
<td>Object Desc.</td>
<td>6</td>
<td>5.58%</td>
</tr>
</tbody>
</table>

Total Errors  
107.5  
198.50
the police interview; 35 in the research interview); (2) other people at the scene (18 errors; and 22 errors); (3) the witness him/herself (2 police errors; 7 research errors). In both the police and research interviews, two-thirds of the errors concerning the thief and gunshop owner were committed by two witnesses. Similarly, two-thirds of the errors relating to the actions of others were committed by only three of the witnesses. The two witnesses common to this error pattern were both teenage males. They reported that physical contact had occurred between the thief and gunshop owner. All of the other evidence made it clear that such contact did not occur. Our research team re-enacted the scene, photographing it from various locations including the perspective of the two teenagers. The sharp angle from which they viewed the shooting and the interposition of a telephone pole limited the view of these witnesses. The distance between the thief and the store owner was collapsed for these witnesses. Thus, in part, their action errors may have originated in a perceptual error. However, they seem to have embellished the error. One of them reported to the police that the two men were "grabbing one another", and in the research interview he was explicit about the points of contact.

Only one witness correctly reported that the pattern of shots was two followed by a slight pause and then six rapid shots. This accounts for the high proportion of errors which concerned the gunshots. It should be noted, however, that although the proportion of gunshot errors dropped between the police and research interviews, their absolute number decreased only from 17 to 12.5. Several of the witness reported that the noises they heard were much softer than they expected, sounding more like firecrackers than gunshots. Five witness reported hearing only five or six shots in rapid succession. Thus, for them the first two shots went unattended, although they may have oriented
the witnesses for the second volley. The one correct witness was familiar with guns, and because of this familiarity he may have recognized immediately the sounds of gunshots and oriented accordingly. In addition, he had handled the gunshop owner's empty gun after the shooting, and therefore may have known how many rounds had been discharged. It would have been very difficult to determine the number of shots on the basis of earwitness evidence alone.

Errors concerning time and date of the incident were, not surprisingly, more common in the research interview. Ten of the witnesses could not remember the exact month of the incident, although three of these witnesses were correct about the day of the week. Only one witness, the wife of the gunshop owner, remembered both the time and the correct date, and this was a date which had clear significance for her.

Turning to errors in the description of people, 23 incorrect statistics (estimates of height, age, and weight) were given to the police (out of a total of 46 statistics provided), and 25 were given in the research interview (out of a total of 49 statistics). It is apparent that there was about a 50—50 chance of a statistic being correct. When asked for weight, height, and age estimates most people lack the training and experience to provide an accurate estimate. One witness, for example, appeared to offer a standard set of statistics which changed little across individuals or over time. This witness described three individuals to the police with the following height and weight estimates: (1) 5'10", 160 lbs. (2) 5'10", 160 lbs. (3) 5'10", 130 lbs. In the research interview he reported similar statistics: (1) 5'11", 165 lbs. (2) 6", 180 lbs. (3) 5'8"—5'10", 130 lbs. Approximately 50% of these statistics are incorrect, including both weight estimations given for individual #2. Since this type of error constituted 52% of the person description errors in the police reports, and 37% of such errors in the research interviews, it appears that much of the lower accuracy
associated with person descriptions is due to the inaccuracy of these statistics. If accuracy of person descriptions is recalculated after removing height, weight, and age estimates, the accuracy in the police interview was 82.47%, and 79.97% in the research interview.

Errors reporting the style and color of hair, and the style and color of clothing constituted the bulk of the remaining person description errors. In the police interview, hair color was reported correctly 77% of the time, and hair style was correct 72% of the time. The comparable values for the research interview were 80% and 71%. In the case of clothing, colors were correctly provided 66% in the police interviews and 59% in the research interviews, while style was correct 88% and 80% respectively. Thus, the color of clothing seems to be the most difficult feature to retain (or notice). For example, in the research interview one witness graphically described the wounds on the body of the thief and she provided detailed descriptions of the body's position and its exact location in the street. Whereas this information was highly accurate, she erroneously described the thief as wearing a T-shirt and a red and black plaid jacket. He actually wore a dark blue sweater and a blue jean jacket. This witness reported that the body was her "main focus of attention" but apparently this did not include his clothing.

The number of object descriptive errors was few (6 to the police, 34 to the researchers), and five of the six given the police and one-half of those in the research interview concerned the car, primarily its make and color. In the police interviews the witnesses were correct 83% of the time concerning the make and color. In the research interview this fell to 56% correct about the make and 57% concerning the color. The remaining object description errors concerned the guns (10 errors) and the stolen property (6.5 errors). These
errors were distributed across several witnesses and fell into no discernible pattern.

Consistency between the Two Interviews.

For all of the witnesses, 60.05% of the information provided during the research interviews (including unclassified details) was new, i.e., it had not been included in the police report. The median accuracy of the classifiable new details was 81.07%, which compared favorably to the 84.14% accuracy of their police accounts. The witnesses, therefore, were not simply repeating their police interview during the research interview. Furthermore, it is unlikely that all of the new information solicited in the research interview had previously been included in typical renditions of this event. For example, witnesses were asked about the color of the blanket used to cover the thief's body, a detail not mentioned by one witness to the police, not reported in the media, and probably not a component of spontaneous retellings. Of the six witnesses who reported noticing the blanket, five correctly reported the color. Detailed descriptions of the dead thief were also not generally solicited by the police. One witness, for example, was asked by the police to describe the dead man's clothing. The witness reported that the dead man wore a brown sweater, light blue jacket, and faded denims. No other descriptors were given. In the research interview, however, this witness was asked for a complete description of the thief and he supplied the following information. He reported that the thief was wearing a light, faded, three-quarter length denim jacket, brown sweater, leather belt, denim pants, and cowboy boots; he was in his 30s, with green eyes and blond hair which was fine and straight, almost to his ears, and he hadn't shaved for a day or two. This witness had apparently retained a vivid memory of the dead man which had most likely not been verbalized in
previous retellings of the event and most of the details were correct. The color of the sweater and the length of the thief's hair were incorrect.

Just under 40% of the details provided in the research interview were also given to the police. Of this total, 81.56% were consistent (whether right, wrong, or unclassified) in the two interviews. The classified repeated details provided in the research interview ranged in accuracy from 41.46% to 96.55%, with an average accuracy of 79.37%. Thus, the repeated and new details of the research interviews were equivalent in accuracy. The following excerpt from one witness' statement describing a hysterical female who was present during the incident illustrates the high degree of consistency in his description over a five month period of time: (1) Police Interview: "She was a white female; 5'4" to 5'5" tall; 17 to 19 years old; skinny; dark brown or black shoulder length hair; cute. She was wearing blue jeans and a yellow, short sleeved pullover, lower—cut square front with frills on the neck. I seem to recall she had on some kind of heels." (2) Research Interview: "She was 5'5". She had on pretty big shoes though or boots so that would make her 5'2"; 100 pounds; she was waisted in (clarified as being 'slender'); 18 to 20, 22, no older than 22; black shoulder length hair; cute face. She had on a yellow top, short sleeves, came down square here (indicating neck), ruffled, but the material pulled up, kind of like this (indicating) without her stomach showing; blue jeans, black leather jacket over top. I couldn't tell whether she was wearing a belt or not. She had on black or really dark brown boots."

The primary difference between these two statements is the increased elaboration in the research interview. The height clarification corresponds to the female's actual height of 5'2" and the weight estimation is also correct. It is also interesting to note that when the police officer asked the witness shortly after the incident if he would recognize the female if he saw her again, he
responded "No, I don't think so." In the research interview, however, the same witness said "I could recognize her again. I can remember details really well. She was extraordinarily upset and I made a mental note of that and (thought) I had better remember what she looked like." This increase in confidence is probably not related to a sharpening of the witness' memory, but instead reflects the differing social dynamics in the two interviewing situations.

There were 54 inconsistent action details between the two interviews, and 28 inconsistent descriptive details. The inconsistent action details included the time and date errors noted above, and the changes in the number of reported gunshots. One witness was inconsistent in reporting action details as follows. In the police interview this witness reported that the thief exited the gunshop carrying a box, and "about five seconds later another guy came out of the gunshop. He walked around the front of the car behind the first guy. From the distance we were at, I couldn't hear if they were saying anything." In the research interview the witness related the same scenario with the following highlighted changes: "About two minutes later, not very long...the owner came out of the store after him, right behind him and went around the front of the car. The owner told him to stop; I think he said stop. I'm not too sure what he said...I heard him say something, but I can't remember what he said now. It was a while ago." The inconsistency in relating the time lapse suggests that the specific time period given was not meant to be taken literally, but rather it was meant to denote a very short period of time. Another interesting example of inconsistency involved two male witnesses (18 yr. and 26 yr.) who reported 6 and 9 inconsistent details, respectively. In each case, half of the inconsistencies were due to egocentric shifts. Each changed his role in the event to one of greater centrality. It is impossible to determine if
this reflected a change in their memory or male bravado (or both). Whatever
the reason, this dropped the accuracy of the older of these two witnesses by
20.13% from the police to the research interview.

The changes which occurred in the descriptions of people and colors
were also interesting. One witness reported to the police that the thief's
automobile was red, but he reported in the research interview that it was
either red or blue. Another witness correctly described the automobile as being
a Falcon in the police interview, but he reported in the research interview that
the car was a gold Chevrolet. The descriptions of the previously mentioned
hysterical woman at the scene were also found to change. One witness told the
police that she wore tan boots, but in the research interview he said that she
wore black shoes. Another witness initially described the woman as wearing
"newer jeans", but reported in the research interview that she wore "faded
blue jeans that were not new." Another witness told the police that the
woman had rotting teeth and wore a yellow sweater and faded denims. In the
research interview, the same witness reported that the woman wore a red or
yellow T-shirt and new denims, and had very white teeth.

Misleading Information.

The wording of the questions had no effect in this study. Ten of the
witnesses replied either in the negative (i.e., that there was no broken
headlight or yellow quarterpanel) or indicated that they hadn't noticed the
detail. One witness was not asked the misleading questions because she
indicated that she had noticed nothing about the automobile. The two
remaining witnesses acquiesced to a misleading suggestion, but one did so in
response to a control question. When asked if he had seen a broken headlight,
this witness replied, "No, I didn't see a broken headlight, but I believe there
was one; the panel of the car was a different color, blue or something." He
then explained that replacing the panel might break the headlight, and added, "I'm just guessing now, it sounds very vague. I don't think I was ever around the car, oh, until afterwards, I was around the front of the car." The other witness was asked if he had seen the broken headlight and he replied, "On the left side? I think there was masking tape. I can't remember which headlight it was, but there was masking tape."

Although unaffected by the wording of the questions, three witnesses reported non-existent events without prompting. In the police reports, 11.5 action details (2.93% of the total actions reported) never occurred. There were 18 such details in the research interviews (3.23% of the total). The three witnesses who reported these erroneous details included the two teenage males noted earlier, and a woman who was driving her automobile past the scene of the incident. Seeing a man lying in the street (the thief), she assumed that a witness who had picked up a gun from the street had done the shooting (by this time the store owner had fallen back inside the gun store). She then believed that she heard the man holding the gun say to another man standing next to him (another witness as it turned out), "Did you see me shoot that guy?" She reported that this was uttered in a menacing tone. No such event took place. However, this witness retained this erroneous version of events from the police interview to the research interview, despite newspaper and television accounts to the contrary. When pressed for more details, none of these three witnesses faltered in their descriptions. The two teenagers (independently) actually embellished their erroneous description of a fight between the gunshop owner and the thief.

**Stress Effects.**

All of the witnesses reported that their emotional state prior to the event was "normal" or "relaxed." The five witnesses who had contact with
either the thief, the gunshop owner, or a weapon, reported the greatest amount of stress. Thus, stress and direct involvement in the incident were confounded in this case. On the 7–pt. scale, one of these witnesses reported a stress level of 5, three rated their stress at 7 (extreme anxiety or stress), and one reported a level of "8." All of these witnesses also reported sleeping difficulties during several nights following the incident. The overall mean accuracy for this group was 93.36% in the police interview and 88.24% in the research interview. Seven of the remaining eight witnesses reported stress levels ranging from 1 to 5, and none indicated any after-effects. The eighth witness indicated no stress at all. The overall accuracy for these witnesses was 75.13% in the police interview and 75.88% in the research interview. The difference in accuracy between the stressed and non-stressed witnesses is reliable in the police interview (t(10)=5.02, p<.05), and marginally so in the research interview (t(10)=2.17, p>.05).

DISCUSSION

This study constitutes the first in situ investigation of eyewitness memory. Given the novelty of the research, the striking character of the event, and the fact that this case involved only thirteen witnesses, any generalizaitons must be made with caution. There are, however, a number of findings in the present study which are intriguing and which paint an image of the eyewitness which differs in many respects from the eyewitness presented in experimental research. Both the similarities and differences between this study's findings and other research are discussed below, along with issues raised by this study which have not been addressed in experimental research.

In keeping with the findings of experimental research, the present study indicates that eyewitness memory is fallible. This, of course, comes as no surprise; the fallibility of human memory is a well established phenomenon and
eyewitness memory is no exception. I would not dispute, therefore, Parker's (1980) claim that "The fact that conscientious and honest people will differ in the reporting of their observations of a crime is one of those immutable phenomenon that will exist as long as man" (p.33). There is no doubt that witnesses do differ in their accounts; some examples were found in the present research. Eyewitnesses are fallible. This description of eyewitnesses, however, can be misleading because despite their fallibility, eyewitnesses can be highly accurate.

Most of the witnesses in the present study were highly accurate in their accounts, both immediately after the event and several months following the event. These results thus call into question the essentially negative view of the eyewitness which has emerged from eyewitness research. As noted in the literature review, researchers have generally been "concerned with showing the witness's fallibility and his inability to recall accurately physical actions, person descriptions or verbalizations" (Clifford & Lloyd—Bostock, 1983, p.286). In the present in situ research, however, a different picture emerges. The witnesses examined in this study are notable for their ability to accurately recall details, rather than their inability to recall details.

Although high accuracy rates were found in the present study, this does not suggest that such a high degree of accuracy is representative of all real-world eyewitnessing situations. The high accuracy rates found in this study are most likely situation specific. The gunshooting incident was striking, arousing, and highly unusual. It is rare for anyone to witness a "shoot out" in the middle of a busy street in a Canadian city. Several of the witnesses commented on their surprise at seeing such an event unfold before them; five months later they were still struck by their initial sense of disbelief. In this respect, the memories of these witnesses are similar to the "flashbulb
memories" reported by Brown and Kulik (1977) which were triggered by surprising, tragic events and which often persisted for years. Thus, the salience and uniqueness of this event probably played a major role in producing vivid memories. It is, of course, a practical and ethical impossibility to stage such an event for laboratory research even though such events are a concern for the criminal justice system. Perhaps the emphasis on negative aspects of eyewitness performance in experimental research has been aided by ignoring the effect of unique and striking events.

Another factor which differs between this case and most experimental research and which might influence accuracy rates was the degree to which witnesses were actively involved in the event. The finding of significantly higher accuracy rates among the five witnesses directly involved in the event suggests that details may be retained more vividly by those who actively participate in an event. However, because witnesses directly involved in the event were necessarily in closer proximity to the incident, no firm conclusions concerning the effect of direct involvement on witness memory can be made. The higher accuracy rates may have been due, in part, to these witnesses having better or closer viewing positions. Nevertheless, it my contention that the degree of witness' involvement in an event is an important factor in evaluating eyewitness testimony. Memory research has suggested, for example, that events which are processed more elaborately are more vividly retained (Craik & Lockhart, 1972) and personal involvement in an event could engender additional processing. Having the self as a referent for the memory also might enhance recall. The passivity demanded of witnesses in laboratory research has precluded examining the effect of active involvement on subsequent recall; however, the possibility of such an effect in real—life witnessing situations should not be ignored.
I am suggesting therefore that the relatively high accuracy rates found in this study may have been due, in part, to factors usually absent in experimental research: a particularly salient event with obvious life and death consequences, and the opportunity for active involvement by some witnesses. However, it should also be noted that although accuracy rates were high in this study, the scoring procedure employed probably underestimated accuracy levels. As noted earlier, conservative rules were incorporated in the scoring criterion. For example, expressions of uncertainty were ignored when scoring a witness' account. It has been suggested that uncertainty is correlated with inaccuracy in eyewitness accounts (e.g., Lipton, 1977; Malpass & Devine, 1980). As a consequence, the scoring procedure used in this study may have inflated the level of inaccuracy by ignoring qualifiers.

One of the more striking results of the present study was the lack of memory loss over time. Accuracy rates remained virtually unchanged five months after the incident. This result calls into question the general conclusion that eyewitness memory deteriorates rapidly with time. Penrod, Loftus, and Winkler's (1982) suggestion that "research on the length of the retention interval supports Ebbinghaus's (1885/1964) original finding that the longer the interval, the worse the performance" (p.136) does not apply to the present study. Once again I would suggest that the striking nature of the gunshooting event was a primary factor in producing accurate long-term memories. Thus, to a large extent, the mundane nature of the witnessed events reported in the Penrod et al. review may account for the discrepancy between their conclusion and the present findings.

Given the striking nature of the shooting incident, it is possible that repeated renditions formed the basis of the excellent retention after several months. Some aspects of the results are inconsistent with this possibility,
however. Many of the details solicited in the research interview had not been given in the police interviews, and some of these were unlikely candidates for inclusion in any spontaneous retelling of the event. Yet these details, such as the detailed descriptions of the thief’s body and descriptions of the blanket covering the thief’s body, were as accurately recalled as repeated details. For related reasons, accurate retention does not seem to have been aided by media accounts. There were a number of inaccuracies in the press and television reports of the event. None of these were found in the eyewitness accounts. Also, witnesses repeated idiosyncratic details which they could have corrected if they had been influenced by the media accounts. It appears that the memory persistence results from the nature of the event, and that an Ebbinghaus decay curve simply does not apply for this type of striking event.

Whereas the general trend was for memory persistence, some specific aspects of the event were forgotten. Some colors, particularly referring to clothing, were not remembered. Alternatively, hair color and the color of the blanket were accurately retained. Nacke’s (1911) conclusion that "testimony descriptive of colors is practically worthless" is too sweeping, and does not indicate the selective nature of color memory. However, it is appropriate to indicate that color memory deserves special attention.

The present results suggest that witnesses should not be asked to provide descriptive statistics of people. This finding is in keeping with long standing experimental research (i.e., Munsterberg, 1908; Whipple, 1909). It is clear that attempts to guess at height, weight, and age are pointless. Unless the witness has special expertise or training in this regard, police should be content with relative estimates of these characteristics. They should ask each witness to make height judgments relative to some environmental fixture, or
another individual. The basis for the judgment should also be elicited, and used to evaluate the value of the estimate provided.

As demonstrated in the preceding literature review, the findings relating memory to stress are confusing. Deffenbacher (1983) summarized ten studies which concluded that arousal either had no effect on subsequent recall or increased accuracy; in contrast, eleven other studies summarized by Deffenbacher (1983) demonstrated a negative effect of arousal on memory. In the present case, self-reports of event related stress were unrelated to memorial ability. The confounding of stress with the involvement of the witnesses in the present instance, however, makes clear attribution of stress effects impossible. Even so, it does appear that in this case stress had no negative effect on memory. Perhaps, because the experimental studies reviewed by Deffenbacher which indicated a negative effect employed primarily recorded events, the negative effect of stress is restricted to memory for recorded events or to the kind of stress that recorded events evoke.

In any event, it is apparent that the definition of stress is problematic. In laboratory research, stress is often equated with arousal which in turn is defined by the presence of white noise, muscle tension, etc. In the context of a crime the pattern of emotional response can be complex. A number of witnesses in the present case reported that they were active during the event, either participating directly or indirectly by calling police or an ambulance. They reported that they felt "adrenalin effects" during the incident, but that they were not aware of feeling stress at the time. The stress response only appeared about a half hour after the incident, with an accompanying feeling of strain. If we wish to study the effects of stress on memory in a forensically relevant manner, we need to devote much more effort to understanding the
stages of the stress response and the relative contributions of the different stages to any memorial consequences of stress.

The attempts to mislead eyewitnesses through the use of biased questions were unsuccessful in the present study. Loftus (1979) has indicated that the effect of such misleading questions should be maximal shortly after the incident. Thus, the fact that the misleading questions were asked four to five months after the event may have reduced the effectiveness of the manipulation. However, as indicated in the preceding literature review, the effect of misleading questions has proved difficult to demonstrate in a number of laboratories (e.g., Dodd & Bradshaw, 1980; Yuille, 1984; Zanni & Offerman, 1978). In any event, this first attempt to mislead real witnesses has demonstrated their resistance to such influences. More research is required to determine the generality of this result.

Other major issues raised in the experimental literature were addressed less directly in this study. A preliminary analysis of the witnesses' statements revealed, for example, no significant difference in accuracy between the narrative and interrogatory forms of report. This finding is contrary to experimental research which has consistently demonstrated lower accuracy rates in the interrogatory portion of witnesses' reports. Perhaps in this instance accuracy did not decrease in the interrogatory report because the witnesses were not pressured to supply additional information. Since the perpetrator was dead at the scene and all stolen items were recovered, there was no need to push witnesses to the point where they might be supplying false information. Thus, before concluding that interrogatory reports will be less accurate than narrative reports, the type of case being investigated and the aim of the interrogator should be considered.
Confidence and accuracy were not systematically assessed in this study. As noted previously, qualifiers such as "I'm not sure" were not considered in scoring the accuracy of witness statements. Hesitations, long pauses, and intonations indicating either certainty or uncertainty were also not systematically examined. The major difficulty in distinguishing degrees of confidence was that witnesses were more likely to explicitly express lack of confidence with hedge phrases such as "maybe" or "I'm not sure"; whereas confidence was more implicit, often being conveyed in the tone of voice. Listening to the audiotaped version of a witness's statement thus presented a different picture than reading the transcribed version. In reading the statements, however, it was obvious that "attestation does not guarantee accuracy". In the case of the previously discussed teenage males who misperceived the action between the perpetrator and victim, both males were confident in their version of the event. On the other hand, witnesses sometimes showed a marked lack of confidence when relating a detail which was accurate. One witness, for example, accurately reported the color of the automobile by saying, "I don't know why, but the color maroon stands out in my mind." From these casual observations, the relationship between confidence and accuracy appears to be extremely complex. Not only is confidence difficult to operationalize (written statements convey a different degree of information than audio versions), but belief in one's own interpretation of an event may have no bearing on the accuracy of that interpretation.

Finally, a methodological issue arose from this study which must be addressed. Nomological summaries of the present data proved useful in describing some patterns in eyewitness recall and these summaries provided a basis by which to compare these data to experimental research. Ideographic analyses, however, played an essential role in understanding the performance of
these witnesses. Consider for example the accounts of the two teenage males who erroneously reported physical contact between the thief and the store owner and the woman who "heard" a man holding a gun acknowledge shooting the dead man and also threaten another man. Each of these witnesses constructed the event from their individual viewing perspectives in line with their own expectations as to what should happen. A qualitative analysis of these witnesses' statements, thus, reveals the central importance of individual interpretation. Witnesses do not simply report an objective memory of an event; witnesses interpret events that transpire in accordance with their viewing perspectives and expectations. Additional findings which emerged from ideographic analysis of the eyewitnesses in this case indicate the importance of personality variables and personal experience. For example, two males, perhaps exhibiting self-presentation motives, changed their initial reports so that they became more central to the action when reporting the event in their research interviews. And, the one witness who correctly reported the number of gunshots was a firearm afficionado. It is most likely that his experience allowed him to immediately recognize the sound of gunshots and thereby correctly note the number of shots. These examples demonstrate that ideographic analysis can reveal important aspects of eyewitness' performance. Most importantly, these qualitative analyses make evident that memory cannot be discussed in a meaningful way as a separate entity, apart from the personality and experience of the individual. Memory does not exist in the abstract; memory is embedded in the context of a particular individual as he/she experiences a particular situation. Qualitative evaluations of individual eyewitness accounts to particular types of incidents, therefore, need to be incorporated into eyewitness research efforts if we are to understand memory as it operates in real-world situations.
In summary, this study represents an important step toward building a data base relating to real-world eyewitnesses. This is only a beginning, however. We need a great deal more research involving different types of incidents, witnesses of varying degree of relation to the crime (i.e., victims, relatives), and so forth, before anything definitive can be said concerning eyewitness performance. This study does indicate, however, that eyewitness memory for an especially dramatic event is generally very accurate and persists in accuracy for months. In addition, stress appears to have no negative effect on memory for this type of incident.

The data base developed thus far is too sparse to judge fairly the generalizability of experimental research; however, this study does suggest some aspects of the present eyewitnessing situation which appear to have been either ignored, exaggerated, or misinterpreted by experimental research. Furthermore, these differences appear to be due to experimental constraints such as the lack of real-world consequences for subject-witnesses, the presentation of less surprising and arousing events, and the passivity of subject-witnesses.
CONCLUSION

A considerable amount of space has been devoted in this thesis to a review of eyewitness testimony research, and in this process a number of questions have been raised concerning the validity, relevance, and generalizability of past and present research findings. These questions have focused primarily on a critique of experimental methodology and its appropriateness for the investigation of eyewitness memory. Critiques such as the one found in this paper are not uncommon in academic psychology; the history of modern psychology is rife with arguments concerning methodological issues. These discussions, however, are generally confined to academic circles and they rarely have consequences for those outside academia. The possible ramifications of eyewitness research, however, extend beyond academia to the decision-making processes within the criminal justice system. Therefore, questions concerning the validity, relevance, and generalizability of eyewitness research should not be dismissed lightly. Not only may long-standing principles of law be modified on the basis of the court's attention to research results, but people's (i.e., defendants, victims, their families) futures may hinge on inferences drawn from eyewitness testimony research (Konecni & Ebbesen, 1979). With so much at stake, it is incumbent upon us as eyewitness researchers to look closely at the validity of our methodologies and the relevance and applicability of our research findings. As noted by Clifford & Lloyd-Bostock (1983):

Perhaps the greatest challenge the future holds for witness researchers is that of the validity of their data and hence its applicability to legal issues and policy. At base all researchers have a deep and genuine concern and consideration for their work and its impact on the judicial system in all its multifarious forms. We do not wish to see changes in policy predicated upon shaky or invalid data. The difficult question, however, is how to achieve firm and valid findings? (p.287)
In regard to this question, the preceding review of eyewitness research suggests that experimental methodology has yielded neither firm nor valid findings. It has been shown, for example, that eighty years of experimentation have led to a confused and conflicting view of eyewitness performance, with research results being highly sensitive to changes in experimental design. In many instances validity has been elusive due to the lack of ecological validity in the event presentation medium (slides), but even when events are presented via live simulations, validity remains an issue. The lack of real-world consequences for informed subject—witnesses and the mundane, non-stressful nature of many simulations, for example, may serve to invalidate simulation research. These issues, however, remain a moot point as long as the validity of these simulations is assessed intuitively, rather than empirically. Only real-world studies of real-world eyewitnesses can validate the results of simulations. The data from the case study presented in this thesis represent a first step in this direction, but at the present time these data are insufficient to judge either the validity or invalidity of eyewitness research in general. It does appear, however, that experimental research has not appropriately assessed eyewitnessing ability in regard to witnessing a striking, violent real-world event and therefore the generalizability of experimental research to this type of event is certainly suspect. The validity of experimental eyewitness research in general, however, remains an empirical question.

Although the data presented in the case study can not conclusively validate or invalidate experimental findings, the two studies presented in this thesis do indicate a productive avenue for eyewitness research—one which may be able to provide the elusive "firm and valid findings" sought within the field. If we wish to examine eyewitness testimony in a forensically relevant way, we must truly "begin at the beginning" and gather data from real—world
eyewitnessing situations. Two means of generating this data have been explored in this thesis—archival research and case studies. The success of both of these methods in the present research encourages further exploration in these directions.

Archival research is proposed as a means of gathering baseline data concerning eyewitnesses. There is a dearth of information pertaining to such issues as the kinds of cases in which witnesses are involved, the number of witnesses per case, the relationship of the witness to others in the case, and the degree of violence witnessed and/or suffered by the witness. This information can be gathered from police files and be used to establish profiles of the types of witnessing situations most common in the criminal justice system. These profiles need to be developed across a wide variety of crimes and across several localities in order to determine what patterns exist both within and between various crimes. From this information we can determine which types of witnessing situations are most relevant to the criminal justice system and devote research time and energy accordingly.

Case studies also can serve as a means of gathering data about real-world witnesses. Eyewitness' statements can be obtained from police files, additional interviews, and court records. Statements from police files will be in a static format, but additional interviews can be structured to examine a number of issues, including memory loss over time and the effect of misleading questions. As the case study presented here has shown, these statements can then be reliably scored in terms of the amount of information contained in the statements and the accuracy of that information. Determining what actually transpired can be problematic and will require choosing cases which have an abundance of forensic evidence, but as the present study demonstrates such
cases do exist. In addition, this problem can be eased by examining crimes such as bank robberies which have been captured on videotape.

It is my belief that the combination of archival and field research will yield the data base needed for developing an understanding of the behaviour of actual eyewitnesses. However, this data collection will be messy and less precise than experimental research. It will be costly in terms of both time and finances. In addition, this research depends upon the co-operation of police, government, and the public. It requires that some unfortunate circumstance must occur, and it must occur in a fashion that is amenable to an interview format similar to the one employed here. Once a suitable event has been found, the messy nature of the data immediately becomes clear. Witnesses have different vantage points, different interpretations of the event, different reasons for cooperating, and different ways of telling their version of the event. The present studies demonstrate, however, that in spite of the problems inherent in field research, it can be done, and useful data can be obtained. Konecni & Ebbesen (1979) reached a similar conclusion in regard to research they conducted on legal decision making. These researchers investigated judges' sentencing procedures using a variety of research methods ranging from archival research to experimental simulations and they concluded that archival analysis was the "most powerful and accurate tool" available for the study of judges' sentencing.

That it would be considered radical to suggest that psychologists interested in eyewitness testimony examine real eyewitnesses reflects the current state of our discipline. Unequivocally, experimentation is afforded center stage and valued above all other methods. Premier statisticians, Campbell & Stanley (1966) have noted, however, that "All problems are not better handled by experimental design and statistical inference." I am suggesting in this thesis
that eyewitness testimony is one such "problem". By clinging to experimental methodology, eyewitness researchers have amassed huge quantities of data which are of questionable relevance to understanding real-world eyewitness performance. The emphasis on "unbiased" subject–witnesses, for example, bears little resemblance to the type of witness found in real world situations. Most witnesses to violent crimes are victims and as victims their testimony is liable to influences not amenable to experimental techniques. The generalizability of research findings gathered from unbiased subject–witnesses to real-world cases involving victim–witnesses is therefore highly suspect.

What I am suggesting is that eyewitness research be structured so that it examines those issues of most relevance to the criminal justice system. Archival research has been proposed as a means for determining what these issues are. This research could be supplemented with interviews with judges, police officers, and lawyers, relating their perceptions of the outstanding problems in eyewitness testimony. As Wallace Loh (1981) points out:

The starting point for the application of psychology are problems posed by and central to the law....The initial question is not "What is known in psychology that can be carried over to the legal process?" Instead, after analyzing the legal dimensions, the question becomes "What kinds of new research can be done by psychologists to illuminate the factual aspects of the legal problem?" In this view, rapprochement between the disciplines does not come about by coordinating existing psychological findings to legal issues, but by conducting afresh investigation tailored to these legal concerns (p.320).

In the more specific domain of eyewitness research and its applicability to the criminal justice system, a similar point needs to be made: The starting point for eyewitness research should be the problems posed by and central to the criminal justice system in regard to actual eyewitness' testimony. Once these problems are identified, issues relating to these problems can be systematically explored in case studies involving actual eyewitnesses. Not only can verbatim witness accounts be examined, but identification accuracy can
also be assessed in carefully selected cases in which witnesses initially viewed line ups or photospreads and the accused later confesses to the crime.

No doubt, some issues pertinent to eyewitnessing can not be studied by the methods I am advocating. In these cases, there are two alternatives. There is the possibility which Clifford and Lloyd-Bostock (1983) suggest of working "backwards from the naturally occurring situation to, if needs must, laboratory simulation, experimenting as early in this backward extrapolation as possible" (p. 288). It would be necessary in this situation, however, to be well acquainted with what actually happens in the real-world circumstances of the to-be-simulated-event. For those more skeptical of the validity of experimental research the second alternative is to accept the limitations of field research and go no further because "erroneous information obtained by scientific methods (and therefore having an aura of truth) is more harmful than no information at all, especially when issues as sensitive as legal ones are being dealt with, and people's futures are quite literally at stake" (Konecni & Ebbesen, 1979, p. 68).
BIBLIOGRAPHY


