

Experimental Psychology and the Art of Judging Evidence

Gary L. Wells, Professor of Psychology
University of Alberta

This is a paper written for the audience to read prior to Professor Wells' presentation on August 23 at the CIAJ seminar "The Art of Judging." Some sections under the heading "How do people judge trial evidence?" have been extracted from Chapter 11 of Wells and Loftus's "Eyewitness Testimony: Psychological Perspectives" (Cambridge University press, 1984).

EXPERIMENTAL PSYCHOLOGY AND THE ART OF JUDGING EVIDENCE

A colleague of mine claims that experimental psychology has as much in common with psychiatry as astronomy has in common with astrology. Although somewhat exaggerated, he makes an important point. Most people understand the important differences between astronomy and astrology, few understand the differences between experimental psychology and its clinical counterpart. In order to appreciate both the value and limitations of what I develop later in this paper, a brief description of experimental psychology is needed.

Experimental psychology is slightly over 100 years old -- very young compared to most sciences. The methods and logic of experimental psychology were borrowed from physics whereas the statistical roots of experimental psychology are from biology. The general idea of experimental psychology is to create controlled conditions, vary those conditions systematically and measure some aspect of human behavior as precisely as possible. Because individual people tend to vary in their responses to controlled conditions, experimental psychologists tend to focus on the average response of people in general.

Traditionally, experimental psychologists have restricted their experiments to issues of theoretical value as defined internally among researchers in the discipline. Often these

issues are highly theoretical and of little immediate practical utility; for example, are pictures coded two ways in the mental system - semantically and pictorially - whereas words are coded only semantically? In recent years, however, experimental psychologists have been asking questions and conducting experiments that may be of greater practical value for current real-world problems. We might refer to this new experimental psychology as applied experimental psychology.

Applied experimental psychology has one common interest with clinical psychology and psychiatry -- an interest in current real-world problems. But it differs radically from the clinical research in that it maintains a close relationship with the experimental, scientific method. An experimental psychologist in court, therefore, would never speak to the case of an individual person; instead the experimental psychologist would speak to the average, modal or median response of people.

Experimental psychologists also differ from the clinical researchers in the kinds of topics with which they have typically concerned themselves. The clinician is typically concerned with matters of mental health in one form or another (e.g., competence to stand trial) whereas the experimental psychologist is concerned more with matters that lend themselves to scientific, objective observation in an experiment (e.g.,

conditions fostering good or poor memory, how people make judgments and decisions and so on).

Experimental psychology as it relates to matters of potential concern to the courts has flourished in the last decade. This has been especially true in two areas: Eyewitness testimony (see Wells & Loftus, Eyewitness testimony: Psychological perspectives, New York, Cambridge University Press, 1984) and the way that people evaluate courtroom evidence (see Kassin & Wrightsman, The psychology of evidence and courtroom procedure, Beverly Hills, Sage Pub., 1985). These two areas are addressed in turn.

Eyewitness Testimony

The question of how reliable eyewitnesses are as sources of evidence is a natural question for experimental psychologists to address. The very first experiments in psychology, over 100 years ago, were concerned with human memory and human memory remains the cornerstone of scientific psychology in the 1980s. The methods for the scientific study of human memory have been well established and there now exist accepted general principles of memory.

Modern psychological theory describes memory as a process that involves the acquisition, storage and retrieval of information. The human memory process operates differently from

a videotape system in several respects. A video system can re-play the recorded event in a relatively faithful manner. Unlike a video system, the human eye does not attend to all that appears in its view but instead is selective. This selectivity in human perception is based on the interests and biases of the human observer and is influenced by the dynamic features of the event. Gaps in memory can result from attention failures or common forgetting and, unlike a video system, these gaps may be filled in later by inferences, guesses or externally-provided information.

Modern experimental evidence regarding the accuracy of eyewitness testimony involves presenting people with carefully controlled events using media such as slide sequences, videotapes or live staged events for which the actual event characteristics are known in detail. In many cases these are simulated crimes, accidents, or other rich and complex events. Following exposure to such events, people are tested for their memory. Witnessing and testing conditions are varied systematically so as to determine the extent to which testimony accuracy is dependent on characteristics of the event, characteristics of the interval between the event and later testing, and characteristics of the testing. Research experiments show clearly that the accuracy of eyewitness

testimony cannot be described independently of the particular witnessing and testing conditions. Within this framework of considering the witnessing and testing conditions, some general conclusions can be reached.

In terms of witnessing factors, witnesses are more accurate under the following circumstances:

1. exposure time is longer rather than shorter.
2. events are less rather than more violent.
3. witnesses were not undergoing extreme stress or fright.
4. witnesses are generally free from biased expectations.
5. witnesses are asked to report on salient aspects of an event rather than peripheral aspects.

In terms of testing conditions, witnesses are more accurate under the following circumstances:

1. witnesses are tested after a short time has passed rather than a long time.
2. witnesses have not been exposed to any biasing information after the event is over.
3. witnesses are questioned in a way that does not suggest what the answer is supposed to be.

In the special domain of identification of previously seen people, there are a number of phenomena that have been discovered by psychologists in this area. An important one is

called "cross-racial identification". This term refers to a situation in which a member of one race tries to identify a member of a different race. Witnesses have more difficulty recognizing individual members of a race different from their own.

Whether a witness is trying to recognize a previously seen person, or simply testify about the color of the traffic light or some other detail, witnesses will give their testimony with some degree of confidence. They say "I am positive" or "I am pretty sure." Some psychological studies have shown that there can be little or no relationship between how confident a witness is and how likely they are to be accurate. This means that inaccurate testimony is sometimes given with a high degree of confidence or certainty by an eyewitness. Thus, one cannot assume that simply because a witness is confident, that witness is probably accurate. People's testimony about an event that they saw can be altered by information acquired after the event (usually called postevent information). Postevent information can be incorporated into the witness's testimony regardless of whether the information is true (i.e., reflects actual aspects of the event) or false. In one widely-replicated experiment, for example, people viewed a slide sequence of an auto-pedestrian accident. In one version of the slides the auto

passed a stop sign and in the other version the auto passed a yield sign. After viewing the slides, some people were asked a leading question about whether or not another car passed the auto while it was at the stop sign while others were asked the same question with the word "yield" substituted for "stop". Later, when tested as to whether they remember the sign as being a stop sign or a yield sign, people tended to report an answer that was consistent with the leading question asked of them earlier, regardless of whether the sign in the original event was a stop or yield sign. Similar phenomena have been observed in experiments involving a wide variety of stimuli, including faces. For example, after briefly observing a clean-shaven face, people who are asked a question like "What color was the man's moustache?" are later likely to describe the person they saw as having a moustache and are likely to falsely identify a moustached person from a set of photographs as being the person that they saw. The precise psychological processes by which postevent information alters testimony remains in doubt but it is clear that postevent information has the greatest influence on testimony when the person's memory for the details of the original event is weak.

How Do People Judge Evidence?

For purposes of continuity, I will begin by addressing the relatively narrow question of how people judge eyewitness evidence and then progress to the broader question of how people judge trial evidence in general.

Psychologists have used four methods to examine the lay person's understanding of eyewitness testimony: (a) the questionnaire method; (b) the prediction method; (c) the written or videotaped trial method; and (d) the method of cross-examining eyewitnesses to staged crimes. Each method has strengths and weaknesses, some shared and some not.

Questionnaire studies

One of the approaches to estimating the knowledge that people have of eyewitness matters has been to use multiple-choice questionnaires. The following example is extracted from a study by Yarmey and Jones (1983):

There are two eyewitnesses to a criminal assault that was committed under poor lighting conditions. When giving evidence some time later, one witness is very positive about his ability to identify the criminal. The other witness is not absolutely positive about his ability to identify the criminal. Which statement best reflects your belief in their testimony?

- (a) The positive person is more likely to be accurate than the less positive person.
- (b) The less positive person is more likely to be accurate than the more positive person.
- (c) The persons are equally likely to be

accurate.

(d) If the less positive person's testimony does not agree essentially with the more positive person's, then the less positive person's testimony will be accurate.

The questionnaire approach has yielded some interesting results. Loftus (1979, chap. 9) for example, asked registered voters in the state of Washington several questions about their knowledge of factors affecting eyewitness accuracy. Her results revealed that some people might not appreciate the effects of stress on perception and recall. Approximately one-third of the respondents indicated a belief that extreme stress would produce an increased ability on the part of the witness to recall details of the witnessed event. Also, approximately two-thirds of the respondents indicated a belief that eyewitnesses would remember details of a violent crime better than they would remember details of a nonviolent crime.

Yarmey and Jones (1983) extended Loftus's (1979) questionnaire to include additional items on such factors as time perception, police versus civilians as eyewitnesses, decay of memory for faces over time, the influence of mug-shot identifications on later identifications from a lineup, the relationship between eyewitness confidence and eyewitness accuracy, and the effects of age of a witness on eyewitness testimony. In addition, Yarmey and Jones tested various

populations, such as legal professionals, law students, other students, and citizen jurors. There was no consistent superiority of any one of these populations over any other population. This is important because studies using university students could be criticized for using respondents who are unrepresentative of actual jurors. The following results represent the findings for citizen jurors: Approximately two-thirds failed to indicate awareness that eyewitnesses are prone to overestimate the length of time involved in a witnessed crime (however, see, e.g., Shiffman & Bobko, 1974); approximately two-thirds believed that police would be superior to civilians as eyewitnesses (however, see, e.g., Tickner & Poulton, 1975); 15% indicated that they believed eyewitnesses' memory for faces would be 90%-95% accurate several months after first seeing the face (however, see, e.g., Egan, Pittner, & Goldstein, 1977); over one-half failed to show awareness that an eyewitness's identification of someone from a set of photographs is likely later to produce an identification of the same person from a lineup regardless of whether the identified person is guilty or not (however, see, e.g., Gorenstein & Ellsworth, 1980); two-thirds indicated that they thought there to be some kind of relationship between eyewitness accuracy and eyewitness confidence (however, see Chapter 8 in this book); and over half

failed to indicate awareness that a physically healthy 70-year-old woman would be as likely accurately to identify a thief as would a 20-year-old woman (however, see Chapter 7 in this book).

Thus, there seem to be many areas where the lay person's intuitions disagree with the results of experimental studies on eyewitnessing. Recent questionnaire studies by Deffenbacher and Loftus (1982) and by Rahaim and Brodsky (1981) seem to corroborate this general conclusion.

A possible problem with the questionnaire approach is that it fails to preserve the ecology of the real-world situation in which these issues are encountered. Actual jurors, for example, have the benefit of discussion and "group knowledge" of these eyewitness factors. Another problem with the questionnaire approach is that people often cannot report accurately on the factors that would influence their decisions in concrete cases (e.g., see Nisbett & Wilson, 1977). Thus, people may report that they would be less likely to believe an eyewitness who made a cross-racial identification yet not behave in a manner consistent with that belief when actually serving as a juror. Care must be taken, therefore, to corroborate self-report questionnaire results with the use of other methods.

Prediction studies

An alternative to the questionnaire approach is one that could be described as a prediction approach. The prediction approach to assessing intuitive knowledge of eyewitness testimony involves the describing of an actual eyewitness experiment to people and asking them to predict the results. These predictions can then be compared with the actual results of the experiment.

Saul Kassin conducted a study exemplifying this approach. Kassin (1979) presented students with a summary description of an eyewitness identification study conducted by Leippe, Wells, and Ostrom (1978). The study involved a staged theft in which the stolen object made it a serious theft (a calculator) or a trivial theft (a pack of cigarettes), and the witnesses either knew the identity of the object before the theft occurred or found out about the relative seriousness of the theft only after the thief had vanished. The students' predictions showed no appreciation for the effects of the seriousness variable or the timing variable. Also, the students tended to overestimate the percentage of witnesses who would be accurate under these conditions. These results have been replicated in a prediction study by Brigham & Bothwell (1982).

I used the same experiment used by Kassin (i.e., Leippe, Wells, & Ostrom, 1978) to do a prediction study involving the

confidence of eyewitnesses. In this study students read the procedure section of the experiment by Leippe and colleagues and were given a target case to predict. The target case was an eyewitness who made an identification and was "completely certain" that he had identified the true thief. Other students were given a target case who was described as "somewhat uncertain" that he had identified the thief. The students were then asked to estimate the probability that the eyewitness had identified the thief. The actual experiment conducted by Leippe and colleagues showed no relationship between certainty and accuracy. Thus, the students' predictions should have been equal in these two conditions. The students, however, predicted a .83 probability that the "completely certain" witness had made an accurate identification and predicted a .28 probability of accuracy for the "somewhat uncertain" witness.

I also used the prediction method to investigate the agreement between scientific evidence and people's intuitions regarding the validity of eyewitness reports when the eyewitness reports were obtained from hypnotized witnesses. Recent evidence indicates that hypnosis has no beneficial effect on eyewitness memory (e.g., Orne, 1979) and that the hypnotized subject is actually more susceptible to the distorting properties of leading questions (e.g., see Chapter 6, by Hall,

Loftus, & Tousignant, in this volume) than is a no-hypnosis control subject. To see if this is congruent with the intuitions of the lay person, I extracted relevant information from the procedure section of a study on this topic conducted by Putnam (1979). Students ($n = 120$) read this procedure and were asked to predict the average percentage of items that hypnotized and nonhypnotized eyewitnesses got correct under conditions of leading versus nonleading questions. The students predicted results that were the precise opposite of the outcome of the data. These students showed a general, indiscriminating tendency to predict superior memory performance under conditions of hypnosis.

I conducted another prediction study based on an experiment by Malpass and Devine (1981a). Malpass and Devine staged an act of vandalism as an eyewitness event and used a lineup that either included the vandal or not. In addition the lineup conductor either told the eyewitnesses that the vandal was probably present (biased instructions) or told witnesses that he did not know whether or not the vandal was present in the lineup (unbiased instruction). I gave the description of Malpass and Devine's procedure section to 80 students (20 in each of the four conditions) and asked them to predict the outcome of the study. (Malpass and Devine's procedure section

was followed closely except that use of the terms biased and unbiased were deleted as these terms suggest the nature of the experimenter's hypothesis.) The results indicated a considerable discrepancy between students' estimates and the experimental data. Those who read a version in which the witnesses received biased instructions were seemingly oblivious to the fact that such instructions greatly increase the likelihood of the witness's choosing an innocent person (especially in the vandal-absent conditions).

The prediction method seems to capture much of what we mean when we question the extent to which eyewitness research findings are intuitive. A particular advantage of the prediction method over the questionnaire method is the fact that the prediction method preserves the original setting. That is, the prediction method describes for the predictor, as well as possible, the same eyewitness setting that yielded the results with which the predictors' estimates will be compared. This contrasts somewhat with typical questionnaire items, which are often cast in a manner that fails to represent the original setting. For example, a questionnaire item dealing with hypnosis might begin, "Suppose an armed robbery took place in a grocery store," when in fact the relevant research has not involved that type of setting. Thus, the questionnaire method

generally involves a greater leap in one's faith regarding the "correct" answer than does the prediction method.

The prediction method is not a perfect means of finding out how people judge eyewitness evidence. Perhaps the major limitation on prediction studies is that such studies do not necessarily tell us how the person would judge a given eyewitness in an actual courtroom setting. There is considerable evidence that people's individual judgments can be quite different from their perceptions of prior probability (e.g., Kahneman & Tversky, 1973). It would not be surprising, therefore, to find a person who believes that 85% of all eyewitnesses would have accurate memories under certain witnessing conditions yet refuses to believe the testimony of individual eyewitnesses under these conditions. Similarly, it would not be surprising to find that a person who believes that only 10% of all eyewitnesses would be accurate under certain conditions is prone to believe most of these eyewitnesses when they are evaluated on the basis of their actual testimony.

Finally, the prediction method fails to account for any beneficial effects that might accrue from cross-examination. For example, would people overlook the damaging effects of biased lineup instructions if the opposing attorney argued that such instructions were unfairly suggestive?

Studies using written or videotaped trials

Another approach to investigating the intuitions of the trier of fact regarding eyewitness matters is to present a written or videotaped trial. Within the fictitious trial one can manipulate eyewitness variables (e.g., high- versus low-confidence eyewitnesses) to see what impact those variables have on the perceived credibility of the eyewitness(es), on the verdicts rendered by mock jurors, or on both. This is a method of assessing people's intuitions regarding eyewitness matters that can preserve most of the crucial elements of an actual trial. In addition, this method requires the trier of fact to distinguish the relevant information from the irrelevant information.

Most eyewitness research using written or videotaped trial information has been devoted to examining the impact of expert testimony. Loftus (1980b), for example, presented subject-jurors with a brief written summary of an actual assault case, including the testimony of an eyewitness. The principal focus of the study was to see if expert psychological testimony on eyewitness identification matters would affect the way in which the subject-jurors processed the trial information. The results showed that the subject-jurors who received expert testimony spent significantly more time in deliberation and were

significantly less likely to render a guilty verdict than were their no-expert-testimony counterparts.

Studies such as Loftus's (1980) that examine the effects of expert testimony do not provide direct information regarding the intuitions of triers of fact. Such studies do provide, however, some indirect evidence on the question. Specifically, if the expert testimony serves to alter the process or outcome of deliberation, or both, then the content of that testimony was probably not redundant with the prior intuitions of the triers of fact. A number of studies indicate that expert testimony on eyewitness matters has a strong impact on the processing strategies and verdicts of subject-jurors (e.g., Hosch, Beck, & McIntyre, 1980; Weinberg & Baron, 1982; Wells, Lindsay, & Tousignant, 1980). Thus, it may be fair to conclude that the way in which expert testimony has been operationalized in the aforementioned studies is not redundant with the intuitive theories of the triers of fact. Alternatively, one could argue that the expert testimony served to give the factors mentioned by the expert more weight than would be given normally.

The expert testimony studies, however, still do not tell us much about how the trier of fact arrives at some decision regarding the credibility of an eyewitness. One approach to this could be to present a case to subject-jurors involving

eyewitness testimony, record the deliberations, and analyze the verbal information for themes or hypotheses regarding the eyewitness testimony. Hastie (1980) has conducted such a study. In this study six-person mock juries were convened and watched a half-hour reenactment of an armed robbery trial. At the conclusion of the trial each jury was set to the task of deliberating to reach a verdict. The deliberations were videotaped and subsequently subjected to an analysis in which all remarks relevant to eyewitness reliability were identified, summarized, and coded into categories. The results of this analysis showed that the jurors were not blindly accepting of the eyewitness testimony. For example, most juries made reference to the characteristics of the crime event (e.g., duration), characteristics of the perpetrator (e.g., race), characteristics of the witness (e.g., witness stress), retrieval conditions (e.g., format of interrogation), the retention interval (e.g., passage of time or interference), and the certainty or confidence of the witness. In general, then, it seems that juries may be capable of identifying pertinent factors. Hastie notes that the typical deliberation included reference to at least 71% of the issues that might be cited by an expert on eyewitness matters.

The Hastie (1980) study, however, also found that many of the issues identified during deliberation seem to conflict with eyewitness findings. Stress, for example, was mentioned in 10 of the 11 juries in connection with the idea that it would benefit the eyewitness (i.e., high stress produces better accuracy), an idea that would seem to conflict with expert opinion (see Loftus, 1979). In addition, the confidence or certainty of the eyewitness was thought to be positively related to eyewitness accuracy, which is also in conflict with empirical evidence (see Wells and Murray, 1985). Also, there was occasional reference to the idea that interpolated descriptions and photographs can increase accuracy in a subsequent face-to-face identification and that memory accuracy might actually increase with the passage of time.

The principal advantage of the written or videotaped trial method is that the relevant information regarding witnessing conditions can be presented in a more forensically valid manner than is true in either questionnaire or prediction study methods. Specifically, the trier of fact receives the relevant information by reading or listening to the testimony of a witness under direct examination and cross-examination in conjunction with other trial information. Because this is the manner in which triers of fact in an actual trial receive their

information about eyewitnessing conditions, results from this method would seem to have greater face validity than would questionnaire or prediction studies.

Existing studies have focused almost entirely in the effects of expert testimony. A more direct approach to studying the intuitions people use in judging eyewitness testimony would be to manipulate eyewitness variables in the trial script. For example, one could manipulate (for the triers of fact) whether the eyewitness's identification of the accused was obtained from a low-similarity lineup or a high-similarity lineup or whether or not the eyewitness had viewed the accused in a set of mug shots prior to viewing the lineup. Both of these factors should make a difference for the trier of fact's assignments of credibility to an eyewitness's lineup identification (see Lindsay & Wells, 1980).

One of the decisions that must be made in the use of written or videotaped trials is how to characterize the eyewitnesses' testimony. Consider, for example, a study by Hatvany and Strack (1980). Hatvany and Strack presented videotaped cases to subjects that varied in whether or not a key eyewitness was discredited. Among other things, discrediting information included the eyewitness's admission that his or her testimony should be disregarded and even the witness apologizing

for testifying. Whether the eyewitness was discredited or not made a considerable difference in the jurors' deliberations. What is unknown, however, is whether or not an eyewitness is likely to withdraw his or her testimony in this manner when exposed to strong cross-examination. In this regard it is important to note that there is evidence that eyewitnesses who are inaccurate in their testimony are not more likely to "give in" under cross-examination than are eyewitnesses who are accurate (e.g., see Wells, Lindsay, & Ferguson, 1979; Wells & Leippe, 1981). Thus, jurors may well believe that they should ignore an eyewitness's testimony if the eyewitness retracts that testimony, but is this a realistic portrayal of the testimony behavior of eyewitnesses? The general question here is how to present triers of fact with eyewitness testimony that is representative of actual eyewitness testimony behaviors.

Studies using testimony of eyewitnesses to staged events

The issue of presenting triers of fact with eyewitness testimony behaviors that are representative of actual eyewitnesses can be partly resolved by using actual eyewitnesses. This was the logic behind several recent studies that used cross-examinations of eyewitnesses to staged events and presented these cross-examinations to subject jurors (e.g., Wells, Lindsay, & Ferguson, 1979).

There are two general advantages to presenting triers of fact with eyewitness testimony given by eyewitnesses who observed staged events. First, the testimony behaviors across witnesses can reflect the variances, means, and relationships that would be expected of witnesses to actual events. The study by Hatvany and Strack (1980), although excellent in many respects, is illustrative of the problem that exists when the eyewitness testimony is actually a script written by an experimenter. As I noted previously, the Hatvany and Strack study portrayed the testimony of an eyewitness who gave identification testimony and then, in some conditions, retracted that testimony under cross-examination. These retraction conditions were referred to as "discrediting information," which Hatvany and Strack argued was generally treated logically by subject-jurors. That is, subject-jurors discounted the eyewitness testimony in those conditions. Although I agree that subject-jurors treated this discrediting information in a logical manner, I believe that the eyewitnesses' behavior was unrealistic. That is, I believe that the testimony retraction behavior of the eyewitness is simply an experimenter's scenario and is not representative of actual eyewitness testimony.

A second general advantage of using the testimony of eyewitnesses to staged crimes is that it is possible to compare

directly subject-jurors' judgments (e.g., belief or disbelief of the eyewitness) to a documented reality (i.e., the eyewitness's testimony was accurate or inaccurate). That is, of course, not generally possible when using eyewitness testimony from actual trials and it certainly is not possible when using experimenter-written scripts of testimony. Therefore, using testimony from eyewitnesses to staged crimes allows for a measure of subject-jurors' judgments that is somewhat closer to what we mean when we ask how adequate human intuition is for judging eyewitness testimony.

At the time of this writing, five studies using this method have been published. The prototype study, conducted by Wells, Lindsay, & Ferguson (1979), involved staging of a theft 127 times for as many eyewitnesses, who attempted an identification of the thief from a set of six photographs. A random sample of 24 eyewitnesses who made an accurate identification and 18 eyewitnesses who made a false identification were then cross-examined individually in the presence of subject-jurors. The subject-jurors' task was to determine whether the eyewitness had accurately identified the thief (belief of witness) or had identified an innocent person (disbelief of witness). The results of this study indicated two important findings. First, there was no evidence overall to

indicate that the subject-jurors could detect an accurate witness from one who had made a false identification. Also, the subject-jurors were extremely responsive to the perceived confidence of the witness. Specifically, although the witnesses' confidence accounted for less than 9% of the variance in witness accuracy overall (and zero percent in the random sample of 42), jurors perceptions of the eyewitnesses' confidence accounted for 50% of the variance in jurors' belief decisions. This result, wherein variation in eyewitness confidence affects subject-jurors' judgments more strongly than it should, has been replicated several times (e.g., Lindsay, Wells, & Rumpel, 1981; Wells, Ferguson, & Lindsay, 1981; Wells & Leippe, 1981; Wells et al., 1980). In addition, juror decisions regarding who is and is not an accurate witness do not correlate with the actual accuracy of the witnesses.

These studies suggest that people cannot distinguish, on the basis of viewing cross-examinations of the eyewitnesses, between an eyewitness who has made an accurate identification and an eyewitness who has made a false identification. A study by Lindsay et al. (1981), however, suggests that under some circumstances people can discriminate between witnessing conditions even though they cannot discriminate between accurate and inaccurate eyewitnesses within conditions. In other words,

jurors show no signs of an ability to discriminate between accurate and inaccurate witnesses, but jurors can discriminate between situations that are likely to produce accurate witnesses and situations that are likely to produce less accurate witnesses. In the study by Lindsay and colleagues, three different thefts were staged and 48 eyewitnesses who identified someone from a photo lineup were cross-examined. One of these thefts produced a high rate of false identifications (67%), one produced a moderate rate of false identifications (50%), and one produced a low rate of false identifications (26%). The subject-jurors who viewed cross-examinations of these witnesses were as likely to believe a witness who had falsely identified someone as they were to believe a witness who had accurately identified the thief. Across theft conditions, however, subject-jurors were less likely to believe a given witness in the high-rate-of-false-identification conditions than in the low-rate-of-false-identification conditions.

It is especially interesting to note that the subject-jurors' performance varied as a function of the certainty or confidence of the eyewitness. Specifically, it was found that subject-jurors did not take account of witnessing conditions when the eyewitness was confident of his or her testimony (i.e., above the median in confidence). Instead, the

subject-jurors seem to have scrutinized the witnessing conditions only when the eyewitness was relatively uncertain of his or her testimony (Lindsay et al., 1981).

As indicated earlier, the cross-examination of eyewitnesses to staged events has two principal advantages over the other methods described in this chapter. First, it provides the triers of fact with eyewitness testimony behaviors that have frequencies, means, variances, and relationships that presumably mimic what one would encounter in actual court cases. Second, and perhaps more important, this method allows for quantitative comparisons between a subject-juror's judgments (e.g., belief or disbelief of a given eyewitness's testimony) and a documented reality (i.e., the accuracy or inaccuracy of an eyewitness's testimony).

Unfortunately, research using cross-examinations of eyewitnesses to staged events has not been entirely satisfactory on several counts. First, this research has been entirely contained in one set of laboratories (University of Alberta). Independent corroboration from other laboratories regarding the general conclusions of these studies is needed. Also, the research has dealt almost exclusively with students, as both witnesses and jurors. Although there is evidence that university students do not differ from jurors in how they

evaluate eyewitness matters (e.g., Hosch, Beck, & McIntyre, 1980; Yarmey & Jones, 1982), such differences have not been explored with the staged-event-cross-examination method. Also, the staged-event-cross-examination method typically has not provided subject-jurors with full trial information (such as other forms of physical or testimonial evidence), and the subject-jurors' judgments have typically been without deliberation (i.e., juror decisions rather than jury verdicts). (See Wells et al., 1981, however, for an exception to these last two shortcomings.)

It should also be noted that the staged-event-cross-examination method is an effortful and costly procedure to implement. It requires repeated stagings of events followed by photograph or lineup identifications and then cross-examinations of the eyewitnesses that are either videotaped or presented live to subject-jurors. In some ways it could be argued that the difficulty of implementing a method should not be considered a shortcoming of that method. In terms of pragmatic considerations, however, the difficulty of implementation is undoubtedly a shortcoming because it reduces the number of researchers who can or will use the method and it slows down the rate of data collection.

Conclusions

As indicated in the previous sections, studies assessing human intuition regarding eyewitness matters have used a variety of approaches. Such variety is beneficial in the long run as it will provide us with evidence of conceptual replication. That is, if two different methods produce similar results we can make better judgments about the generality of the findings. For the same reason, if two methods produce different results, then we must take extra caution in interpreting the results. What, however, do we have in the short run? Do we yet have definitive evidence regarding the adequacy of human intuition for judging eyewitness testimony?

At this point in time I am willing to argue that there is at least one important aspect of eyewitness testimony that is misunderstood by the trier of fact, namely eyewitness confidence. All four methods of assessing people's intuitions converge on the conclusion that confidence and accuracy are perceived to be strongly related. The questionnaire studies by Deffenbacher and Loftus (1982), Rahaim and Brodsky (1981), Yarmey and Jones (1983), and myself (reported in this chapter) show that the majority of respondents believe that there is a meaningful relationship between certainty and accuracy. The prediction study showed that subjects expected accuracy from highly confident witnesses and inaccuracy from witnesses whose

confidence was low. The taped deliberations in the study by Hastie (1980) showed that subject-jurors spontaneously mentioned eyewitness confidence as a cue to the witness's accuracy. Finally, the series of studies using cross-examinations of eyewitnesses to staged events shows that subject-jurors rely heavily on an eyewitness's confidence to infer the eyewitness's accuracy. Thus, when we compare human intuition (via the methods described in this paper) with scientific data on the confidence-accuracy issue, we must conclude that intuition is inadequate on this matter.

Intuitions of the trier of fact regarding eyewitness matters might also be inadequate in other ways. Questionnaire studies suggest that most people do not appreciate the way eyewitnesses overestimate short temporal durations, the interfering effects of interpolated mug shots, and the nonsuperiority of police as eyewitnesses, all of which have been reasonably well documented through experimentation. Prediction studies suggest that people do not appreciate the effects of crime seriousness, the ineffectiveness and distortion properties of hypnosis, and the effects of biased lineup instructions. Studies using videotaped and written trials suggest that people are reasonably good at identifying pertinent factors such as stress but misinterpret their effects on eyewitness memory.

Until these results are corroborated by each of the four methods outlined in this paper we should be cautious in identifying them as shortcomings of the intuitive theorizer. Nevertheless, the evidence might be sufficient already to suggest that the lay person, as trier of fact, be counseled on these matters.

Trial Evidence in General

It is somewhat unfortunate that experimental psychologists have focussed narrowly on the question of how people judge trial evidence. Many of the experiments are, I believe, silly and misleading. For example, a large number of experiments have been conducted in which trial information is presented to mock jurors, either in transcript or videotape or live re-enactment form, and conditions are set up so that the defendant is, for example, attractive or unattractive. Most of these studies report that, with identical trial information, the attractive defendant is judged more leniently than is the unattractive defendant.

There are three main problems with these studies and I and others have been able to prove the irrelevance of these experiments. First, the mock jurors don't take seriously the task before them. David Wilson was the first to show experimentally that if you swear in the jurors and convince them that there are real consequences for their verdict, that there must be reasonable doubt, and so on, then the "attractiveness effect" disappears. Secondly, these experiments created highly unusual sets of materials that were largely devoid of evidence. When evidence is presented at a level of the typical trial, the "attractiveness effect" also disappears.

Finally, these studies tell us little or nothing about how evidence is evaluated -- attractiveness and other such variables are extra-evidentiary. I contend that, for real jurors and judges in real trials, these extra-evidentiary factors have little or no influence on verdicts.

One of the more interesting and informative studies to surface in recent years was conducted by Anthony Doob of the University of Toronto. One of the most common reactions of the public to judicial decisions, when we get out and rub elbows with the person-on-the-street, goes something like this: "Where the hell do they get these judges (jurors)?" or "There is no way I would have let that guy off the hook" or "He only got 7 years ... can you believe it? He should've gotten 30" or "They awarded \$1.2 million ... Jesus ... what's going on with these jurors (judges)?" Indeed, one need not take to the streets for such commentary; just read the editorial pages of North American newspapers. It is argued by commentators that there is a gap between what the public would view as an appropriate verdict or judgment and the verdicts and judgments that the courts reach a la the defendant. Although there is the occasional reaction in which the public thinks the courts were too severe, it almost always runs in the opposite direction in criminal cases, namely that the courts are too lenient. If this is true, there might

be cause for concern. The courts are meant to be an extension of the people and, if jurors are making judgments not in line with the public then we probably would want to know why. Are jurors unrepresentative of the general community? Is this good or bad?

The answer is, I think, contained at least in part, if not whole, in the research by Doob. The answer also is, I think, reassuring. Doob found that there indeed is a gap between jury decisions and the public's decisions on cases when the information on the case is given to the public via a typical newspaper story. Interestingly, however, this is not due to newspaper stories being biased. Instead, the difference seems to be due to the jurors having received judicial instructions while the general public, of course, has not. Doob's research showed that when instructions regarding reasonable doubt and other matters appropriate to the case were given to those who read the newspaper accounts, people reached surprisingly similar judgments to those reached by the jurors in those cases. Experimental social psychologists have repeatedly documented the fact that people believe they will react to a hypothetical situation one way but in fact react differently when actually placed in that situation. The judging of cases from a hypothetical position via newspaper accounts is not the same as

judging the case in the actual courtroom setting.

This is not to say that hypothetical case materials cannot tell us something important about how people judge trial evidence. Doob's study is a case in point; having people judge a trial from a hypothetical, distant newspaper account, he was able to observe something important about the presence versus absence of judicial instructions. In a similar vein, I have been using hypothetical civil case materials for the last 12 months to examine some basic psychological processes that might underly apparent anomalies in how people reach judgments of liability. This work is described in the next section.

Liability and Probability

It should be no surprise to attendants at this seminar that courts are increasingly encountering or are likely to encounter cases that are based in part or whole on statistical evidence. Much of the writings regarding this issue have focussed on concerns about the fact that statistical reasoning presented to courts has sometimes been proven wrong (as when probabilities are treated as independent when they are highly dependent), that the statistical matters are too complex for the juror or judge to evaluate or that the statistical evidence is too readily believed (thereby having more impact than is warranted).

My recent work on the issue of statistical evidence presents a new twist - one that I think is perhaps more interesting and informative. I will state my conclusion from the outset and then discuss the research on which it is based. My conclusion is that pure statistical evidence per se has no impact on civil case judgments of liability. [This might also apply to judgments of guilt in criminal cases, but my research has not yet been extended to these cases.]

In order to make clear what my conclusion entails, consider the following hypothetical case: A suit is launched against the Blue Bus Company for causing an accident resulting in financial costs to Mrs. Prob. Mrs. Prob is able to prove that a bus swerved into her lane of a road, causing her to veer off the road whereupon she hit a tree. Being color blind, Mrs. Prob cannot swear as to whether the bus was blue or grey. Testimony by officers of the Traffic and Licensing Commission show that there are only two bus companies, the Blue Bus Company and the Grey Bus Company and that the Blue Bus Company owns and operates 80% of all the busses on that road. Would you rule in favor of Mrs. Prob on the basis of this evidence? If not, you are in agreement with 97% of the people I've presented this case to. Only 3% of the population would rule against the Blue Bus Company given this "naked statistical" evidence.

If you are not willing to rule in favor of Mrs. Prob, then why not? All of you are well aware of the fact that a civil case is to be decided in effect on "the balance of probabilities." It takes no statistical expert to know that the balance of probabilities favor Mrs. Prob (80%) rather than the Blue Bus Company (20%). Indeed, virtually all of the people who said they were unwilling to rule in Mrs. Prob's favor also freely reported that the balance of probabilities were on her side and against the defendant.

Legal scholars have discussed this hypothetical case and have invoked three general explanations. First, some have suggested that, although the statistical, objective probability favoring the plaintiff is 80%, jurors and judges do not necessarily believe that the probability is 80%. This appears not to be true. I have presented this case to hundreds of people and, when asked what they believe the probability to be, they overwhelmingly endorse the figure of 80%.

A second explanation is based on the general notion of "fairness." Specifically, it is argued that, in order to be consistent, one must rule all (100%) of such cases against the Blue Bus Company when in fact they are responsible for only 80% of the accidents. This is not the basis of people's reluctance either, but the evidence ruling out this explanation is

discussed later.

A third explanation is based on Nesson's (1985, Harvard Law Review) notion of the "message" of verdicts. Nesson argues that the acceptability of a verdict depends critically on the message that the verdict sends to the defendant and the public. He argues that a finding for the plaintiff in the Blue Bus Case sends a message about the volume of business of the Blue Bus Company, an irrelevant message, rather than a message about the act -- i.e., your company is at fault.

Finally, there is a psychological explanation for why people are not willing to rule for the plaintiff. Specifically, it is argued that people fail to see a causal connection between the volume of business and the liability of the company. According to this explanation, a presentation of the statistics in terms of percentage of accidents (rather than the volume of business or the percentage of busses) would result in different judgments. For example, if the evidence were stated as "The Blue Bus Company is responsible for 80% of all the accidents caused by busses on this route" then people presumably would be more willing to rule in favor of the plaintiff.

At the seminar on August 23rd, I will introduce several versions of the blue bus case and solicit audience opinion on whether or not the different reactions arising to these versions

are explainable in legal terms. I will then present data on how people from the general population react to these versions and I will provide a psychological explanation for their reactions.