

# Expert Evidence

The International Digest of Human Behaviour Science and Law

## Articles: Double issue

*Given the length of some of the papers, in this issue, the Editors decided that a double issue would be preferable to splitting the articles.*

**Eyewitness memory research: Probative v. prejudicial value**  
 Ebbe Ebbesen and Vladimir Konecni ..... 2-28

**Recognising Daubert: What judges should know about falsificationism**  
 G. Edmond and D. Mercer ..... 29-42

**Methodological differences make a difference when evaluating eyewitness confidence**  
 Tara Hollins and Tim Perfect ..... 43-48

**Police interviewing and mentally disordered offenders: changing the role of the legal adviser**  
 John Pearse and Gisli H Gudjonsson ..... 49-53

**The House of Commons Defence Committee Report on the 'Gulf War Syndrome'**  
 R. Goldberg ..... 54-57

**Competence to determine 'protected person' status under the Mental Health (Scotland) Act 1984**  
 Michael T Carlin and David J Cooke ..... 58-60

**Gender differences in assessments of rape victim reports**  
 Aldert Vrij ..... 61-64

## Recent legal developments

Court may appoint a single expert on several issues and expect report to be taken as reliable. No privilege for experts' reports in child care proceedings because they are

investigatory. Experts' evidence admissible on the credibility of children's evidence. Expert appointed as assessor to a coroner may question witnesses but not give expert evidence. Juries should not to apply Bayes' Theorem or assess evidence on probabilities of its parts. Time can dim memory so apt to strike out cases where there has been inordinate delay. Expert's opinion must be based on evidence proved in court or admitted by the other parties. .... 65-67

Compensation for injury to feelings where have abortion rather than lose job. Causation in claims for compensation for 'non-physical' loss. Adults must sue for the consequences of childhood sexual abuse before they become 24. .... 67

Can recall a conditionally discharged patient to hospital where already resident. .... 68

Collecting evidence of a child's doli incapax. For direct contact with children to be ended the other parent's opposition must be such as to impose a serious risk of major emotional harm upon the children. .... 68

Only parents can appeal to the Special Educational Needs Tribunal. Availability of resources irrelevant when assessing needs under 1970 Act. Dismissal because of gender reassignment is unequal treatment. .... 69

Silent phone calls can be an assault causing actual bodily harm. Obscene phone calls to a number of people constitutes a public nuisance. Mental impairment cannot excuse provocation. .... 69-70

Consequences of abolishing mandatory warning of jurors about convicting on the uncorroborated evidence of a complainant of sexual assault. Fixing the tariff period of imprisonment for children who kill. Discharging jurors with incapacities. .... 70

## Recent research

More 'neutral' expert witnessing recommended for England and Wales. .... 71

Clinical or actuarial prediction of risk of violence?..... 71-72

Recovered memory. Suggestibility in young children. Children's testimony. Anatomical dolls, sexual abuse, evaluations, assessments. .... 72-76

Verbal cues to deception in the context of interrogation. .. 76

False confessions. .... 76-77

## Recent books

Sporer, S, Malpass, R.S. and Köhnken, G. 1996, *Psychological Issues in Eyewitness Identification*, New Jersey: Lawrence Erlbaum Associates ..... 78-80

# Eyewitness memory research: Probative v. prejudicial value

---

Ebbe B. Ebbesen

and

Vladimir J. Konečni,

University of California, San Diego, Department of Psychology, La Jolla, CA 92093-0109, USA.

*Abstract: Psychologists often testify in court about eyewitness memory research. A critical review of those research areas most frequently testified about suggests that such testimony has greater prejudicial than probative value and therefore should not be allowed in court. Not only does a generally accepted theory for eyewitness identification not exist, but the evidence in many areas is inconsistent, the procedures and measures used to study various relationships are not well tied to legal procedure, and there is no evidence that the experts who testify would be any better at detecting witness inaccuracy than uninformed jurors. Finally, the nature of what is known about human memory is so complex that an honest presentation of this knowledge to a jury would only serve to confuse rather than improve their decision-making.*

## Introduction

Since the late 1960s, several US higher court decisions (e.g., *People v. Cardenas*, 31 Cal.3d 897, 184 Cal.Rptr. 165 (California State Supreme Court)1982; *People v. Carr*, 88 Daily Journal D. A. R. 12150 (Calif. Court of Appeal, 4th Dist.) 1988; *People v. McDonald*, 37 Cal.3d 351 208 Cal. Rptr. 236 (California Supreme Court) 1984; *People v. Shirley*, 31 Cal.3d 18 181 Cal.Rptr. 243, 641 P.2d 775 (California Supreme Court) 1982; *People v. Wright*, 43 Cal.3d 399 (California Supreme Court) 1987 and 1988; *State v. Chapple*, 135 Ariz. 281, 660 P.2d 1208 (1983); *United States v. Amador-Galvan*, 1414 F.3d 1417 (9th Cir. 1993); *United States v. Amaral*, 488 F.2d 1148 (9th Cir. 1973); *United States v. Binder*, 769 F.2d 595 (9th Cir. 1985); *United States v. Brown*, 557 F.2d 541 (6th Cir. 1977); *United States v. Downing*, 609 F.Supp. 784 (D.C.Pa. 1985); *United States v. Fosher*, 590 F.2d 381 (1st Cir. 1979); *United States v. Green*, 548 F.2d 1261 (6th Cir. 1977); *United States v. Langford*, 802 F.2d 1176 (9th Cir. 1986); *United States v. Poole*, 794 F.2d 462 (9th Cir. 1986); *United States v. Rincon*, 921 F.3d 28 (9th Cir. 1994); *United States v. Russell*, 532 F.2d 1063 (6th Cir. 1976); *United States v. Sebetich*, 776 F.2d 412 (3rd Cir. 1985); *United States v. Smith*, 736 F.2d 1103 (6th Cir. 1984); *United States v. Tyler*, 714 F.2d 664 (6th Cir. 1983); *United States v. Wade*, 388 U.S. 218 (United States Supreme Court 1967)) have discussed the admissibility of expert testimony concerning factors affecting the reliability of eyewitness identification. In struggling with this issue, higher courts have generally felt that the decision to admit the testimony of 'eyewitness memory experts' was within the discretion of the trial judge but that they would consider a number of factors in reviewing trial court decisions on appeal. The underlying logic of these factors was recently outlined in *U.S. v. Rincon* (1994). A lower court had refused to allow Dr. Kathy Pezdek to testify about the reliability of eyewitness identifications and the defense appealed. The U.S. Supreme Court asked the 9th Circuit Appellate Court to review the trial court decision in light of the Supreme

Court's latest thinking about expert testimony and scientific evidence, a view they had expressed in *Daubert v. Merrell Dow Pharmaceuticals*, 509, U.S. 113 S.Ct. 2786, 125 L.Ed.2d 469 (1993).

In *Daubert*, the U.S. Supreme Court noted that Fed.R.Evid. 702 supersedes the *general acceptance* standard established in *Frye v. United States*, 1013, 283 F., D.C. Cir. (1923) - a standard frequently cited by eyewitness memory researchers (e.g., Kassin, Ellsworth and Smith, 1989; Konečni and Ebbesen, 1979; McCloskey and Eggeth, 1983; Wells, 1993) when discussing whether experts should be allowed to testify in court. *Daubert* stated that lower court judges must "ensure that any and all scientific testimony or evidence admitted is not only relevant, but reliable." To establish this, the trial judge is supposed to apply a two part test: 1) Is the expert proposing to testify about scientific knowledge? and 2) Will the expert testimony assist the trier of fact to understand or determine a fact at issue? They further established that evidence that passed both of these tests could still be excluded, "if its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury." (p. 2798) They defined scientific knowledge as an inference or assertion derived from the scientific method and stated that any testimony about such knowledge must be supported by appropriate validation, i.e., good grounds, based on what is known. To determine whether a theory or technique constitutes scientific knowledge, the trial court may consider such things as, 1) whether the theory or technique can be or has been tested, 2) whether it has been subjected to peer review and publication, 3) the known or potential error rate, and 4) the particular degree of acceptance within the scientific community. The court added that these four factors were not meant to be an exhaustive check list.

In *Rincon*, the court concluded that the trial court had not abused its discretion in excluding Dr. Pezdek's testimony, despite its belief that Dr. Pezdek's testimony clearly passed the relevancy test. The main reasons given were that the defense had not presented sufficient evidence to convince the court that the research on which Dr. Pezdek's testimony would be based was related to a scientific subject (only one article, by Kassin, Ellsworth, and Smith, 1989, consisting of the results of a survey of eyewitness expert opinion about the generally high reliability of many eyewitness memory results was presented to support Dr. Pezdek's proposed testimony), and that they were also not convinced that such testimony would be more helpful to the jury than a set of cautionary instructions read by the judge prior to jury deliberation. Finally, the *Rincon* court was very careful to leave the door open about whether such testimony might be allowed in the future were the defense to present a better set of supporting research studies.

Prior to *Rincon*, many courts also expressed concern about whether eyewitness experts should be allowed to testify in court (Loftus and Schneider, 1987), however, the issues that concerned them were somewhat different from those outlined in *Rincon*. For example, *United States v. Amaral* (1973) was concerned about the expertise of the defense expert. *United States v. Binder* (1985) worried that testimony about witness memory might invade the province of the jury. *United States v. Fosher* (1979) and *United States v. Poole* (1986) wondered whether the jury might not already be well aware of the things about which the expert might testify. The extent to which other evidence exists to corroborate the eyewitness' identification was raised as an issue in *People v. McDonald* (1984). The relevance of the factors about which the expert testifies to the particular facts of the case was of importance in *United States v. Downing* (1985). Finally, *United States v. Smith* (1984) questioned whether the expert testimony about the effect of particular facts of the case on eyewitness reliability would add to the general knowledge of the jury. Despite these concerns, many (but by no means all) courts seemed to have concluded that the psychological research on which eyewitness experts base their testimony is, in fact, sufficiently extensive and conclusive that there is some probative value to the testimony and/or that the theory underlying the 'field' is generally accepted. For example, in *People v. McDonald* (1984), after citing a series of texts that described eyewitness memory research, the California Supreme Court concluded that, "The consistency of the results of these studies is impressive, and the courts can no longer remain oblivious to their implications for the administration of justice." The United States Court of Appeals (6th Circuit) argued as follows about the testimony of Dr. Fulero, an expert called by the defense to testify about eyewitness identification research: "The day may have arrived, therefore, when Dr. Fulero's testimony can be said to conform to a generally accepted explanatory theory," (*United States v. Smith*, 1984). Finally, in a recent

extensive review of legal opinion Handberg (1995) argues, "...courts should admit expert testimony on eyewitness identification in much the same way that they allow it on CSAAS (Child Sexual Abuse Accommodation Syndrome) and RTA (Rape Trauma Syndrome)," and "...courts should permit eyewitness expert testimony to correct the misperceptions that many jurors have about the reliability of eyewitness identifications."

The thesis of this paper is that, *Rincon* notwithstanding, the courts have been misled about the validity, consistency, and generalizability of the research in the area, in part because of a lack of understanding by many members of the judiciary about the nature of science, especially social science, and in part because researchers in eyewitness memory have been overconfident in their own expertise. Further we argue that a generally accepted theory of eyewitness identification that is capable of predicting witness accuracy in a particular real world situation does not exist. Although the science of psychology has developed many useful and interesting models of memory, the fact remains that no theory of memory has been proposed that would allow researchers to predict how accurately people will be able to identify a defendant whom they have seen commit a crime. Accurate and exact prediction is prevented in part because the phenomena are complex, in part because we may be unable to measure the appropriate variables, and in part because the theories are not sufficiently developed (they do not tell us how the many potentially relevant variables combine) to allow prediction (Lykken, 1991). Thus, like others (Egeth, 1993; Elliott, 1993; Konecni and Ebbesen, 1979, 1986; McCloskey and Egeth, 1983; Wells, 1993; Yuille, 1989) we believe that substantial evidence supports the claim that research on eyewitness memory continues to lack external validity or generality and, therefore, that testimony by psychologists about factors affecting eyewitness memory should not be allowed in court or if allowed, should be attacked vigorously. Finally, because the conclusions drawn by defense 'experts' (e.g., that factors such as stress, racial dissimilarity, weapon focus, confidence, selective attention, reconstructive memory, short exposure durations, suggestion, and unconscious transference detrimentally affect the accuracy of eyewitness identifications and testimony) are specious when applied to the real world and because their testimony is often limited to a discussion of eyewitness identification in isolation of other evidence heard by the jury, we argue that it is highly questionable whether they can help juries reach more accurate decisions about the probable guilt of defendants despite the frequent opposite claims by a number of researchers (e.g., Bothwell, Brigham and Malpass, 1989; Cutler, Dexter and Penrod, 1989; Cutler, Penrod and Dexter, 1989, 1990; Cutler, Penrod and Stuve, 1988; Kassin, Ellsworth and Smith, 1989; Loftus, 1983, 1986, 1993; Maass, Brigham and West, 1985; Wells, 1984, 1993; Wells, Lindsay and Tousignant, 1980).

## On the validity of procedures used in eyewitness memory research

It is important to make clear at the outset that a substantial majority of the studies conducted in the 'eyewitness' memory area involves *simulation* research (Yuille, 1989). That is, researchers create conditions (often in laboratories at universities, but sometimes in other settings) that are claimed by the researchers to capture the essence of the conditions that real eyewitnesses experience. Before results from studies that claim to deal with eyewitness memory can be applied to real witnesses of real crimes, however, researchers must establish that they have created the same (or at least very similar) memory processes and motivational states in their test subjects as are experienced by witnesses and victims of actual crimes. Unless the research is designed to ensure that the underlying processes have been adequately simulated in the laboratory settings, it is unscientific and unwise to generalize the results to real witnesses.

The requirement that simulation research create the same or nearly the same processes that are assumed to work in the settings to which one hopes to generalize is a common one in medical and other scientific areas. For example, in medical research one often speaks of laboratory simulations of biological systems using animal preparations as 'animal models' of the processes of interest in the human. We should demand no less careful construction of simulation studies of eyewitness memory. Unfortunately, memory research has not, in general, been designed with an eye toward accurate simulation of relevant processes. Partly this is because we do not yet have agreed-upon theories of what the relevant processes are and partly this may be because it is difficult to create the relevant processes in laboratory studies (e.g., we cannot reconstruct in a laboratory the experiences of an actual rape victim).

Several different procedures can be used in an attempt to test whether the conclusions drawn from particular simulation procedures can be generalized to actual crime situations. The weakest method for achieving generality, one that most methodologists (e.g., Crano and Brewer, 1973) argue is entirely insufficient, is 'face validity.' A study is said to have a high degree of face validity if it appears, *on its surface*, to have simulated adequately the processes under study. The terms 'forensic relevance' and 'legal verisimilitude' are sometimes used in a manner synonymous with face validity possibly suggesting that face validity is the only correct method of assessing generality. That is, some researchers express concern about the 'forensic relevance' of studies because an experimental situation does not 'look like' situations actual witnesses experience. Although the forensic relevance of research in eyewitness memory is a crucial, if not central, issue, one *must* establish the generality to the legal system of results from simulation research by conducting additional

empirical research that uses methods and procedures different from those used in the simulations (Crano and Brewer, 1973; Webb, et al., 1981). That is, generality and forensic relevance is determined by empirical research and not by the subjective judgment of (frequently biased) observers.

Much better than face validity is whether the use of a wide array of different procedures and methods, all designed to study the same issue, produce similar results. If a conclusion is general, then empirical results should be consistent with that conclusion regardless of the particular methods and subjects used to test it. However, an even stronger test of external validity than convergent validation by a series of different simulation procedures is to assess the accuracy of a conclusion in the conditions and situations to which one hopes that conclusion will generalize, that is, with witnesses to actual crimes.

## Face validity (legal verisimilitude)

Most researchers would agree that the *particular* procedures used to assess the effect of a variable or variables on eyewitness accuracy represent only a few of many different possibilities. Thus, most would agree that the use of pictures of faces to study the effect of, say, duration of exposure on accuracy is but one of several different procedures. Others might involve showing videotapes of simulated criminal events in which the culprit was visible for different lengths of times. In still others unsuspecting individuals might be asked about their memories for different length interactions they had with confederates in field settings. However, although the number of studies using procedures other than face-memory tasks is on the rise (Cutler and Penrod, 1995; Maass, 1996), the procedures employed in the overwhelming majority of memory studies whose results defense experts have claimed generalize to the real world simply do not 'look like' conditions eyewitnesses frequently experience and therefore seem to have low face validity. Being a clerk in a store who is paid with pennies or who has a brief interaction with a patron (Brigham, Maass, Snyder, and Spaulding, 1982), much less looking at 40 or so pictures of faces and then being asked to pick them out of larger set of faces, simply does not 'look' the same as a being a victim of a rape or an attempted murder and then having to decide whether the defendant really is the culprit. Most studies from which experts seem to draw their conclusions in the eyewitness area appear to lack face validity.

It might be argued that this claim is unreasonable given that so little is currently known about the kinds of experiences of actual eyewitnesses because, with one or two exceptions (Moore, Ebbesen, and Konecni, 1994; Tollerstrup, Turtle, and Yuille, 1994), no research has attempted to establish the kinds and relative frequency of conditions that are typically experienced by eyewitnesses. Nevertheless, one obvious example makes the point.

Virtually all of the studies conducted on eyewitness memory involve witnesses, whereas it is, in fact, the victims who supply evidence in the majority of crimes (with the exception of murder) in which eyewitness identification is part of the evidence against the defendant.

We do not even know the distribution in the real world of the 'size' of most factors in which eyewitness researchers are interested. For example, what is the distribution of actual duration of exposures of victims and witnesses to crimes (and how do these distributions vary with crime type)? We know from an extensive review of the facial memory literature by Shapiro and Penrod (1986) that in face memory studies subjects average a little more than six seconds of study time per face. In the only study (Moore, Ebbesen and Konecni, 1994) that has attempted to collect data about the average amount of time that real witnesses and victims had to study the face of the criminal, the median exposure duration was estimated to be somewhere between 5 and 10 minutes (not seconds). In other words, the few seconds of exposure used in most laboratory studies of face memory may well be considerably shorter than the time that the large majority of witnesses to real crimes have to study a face. Similar conclusions might apply to retention interval, stress level, relative time a weapon is present, and so on.

A similar problem exists on the measurement side. Measures of subject memory may not provide the accuracy information that is most appropriate to the legal system. For example, in studies involving recall of witnessed events (e.g., Clifford and Scott, 1978; Yuille and Cutshall, 1986), researchers typically report percent correct of the total number of possible 'facts' witnessed, where the researchers defined what was and was not a fact. However, the legal system rarely knows what the total number of facts are that witnesses might recall in any given situation and, in any case, are more concerned with knowing whether any of the facts that witnesses report are in error. In other words, the rate of false relative to accurate reports is of interest to the legal system rather than the rate of accurate reports of facts to the total possible facts that could have been remembered. This distinction is an important one because it is quite possible that many variables will cause witnesses to recall fewer total facts, but have no effect on the relative accuracy of the facts that are reported. In addition, because researchers do not analyze fact memory on a fact-by-fact basis, but simply count up the total number of facts recalled, they rarely present results about the rate at which witnesses recall certain types of facts compared to other types. For example, it is not known whether recall of hair color is generally better or worse than recall of the color of a culprit's shirt (although Yuille and Cutshall, 1986, have suggested that eyewitnesses' color memory is not as good as their memory for events). In the legal system, however, such issues may be central to determining the guilt of the defendant. Guilt will often depend not on how much a

witness recalls, but on the accuracy of the witness's memory of one or two specific highly probative facts, e.g., a license plate number, which one of several different people fired a gun, and so on. Thus, even the way we measure eyewitness accuracy may lack legal verisimilitude.

The question of the 'forensic relevance' of the research that has been conducted in the psychology and law area has been of considerable concern to a number of researchers. For example, the Devlin (1976) Report in England, Ebbesen and Konecni (1980), Egeth (1993), Konecni and Ebbesen (1979, 1982, 1986), Loh (1984), Lloyd-Bostock and Clifford (1983), Lindsay and Wells (1983), McCloskey, Egeth, and McKenna (1986), Malpass and Devine (1981), Pachella (1986), Wells (1993), Yuille (1989) and Yuille and Cutshall (1986), to name a few, have all made note of the fact that many of the procedures used by researchers in the area of eyewitness memory appear to lack relevance (*on their surface*) to the legal questions for which the authors typically claim relevance. For example, as Yuille and Cutshall (1986) suggest,

*It is readily apparent, for example, that the use of slide sequences and filmed events in eyewitness research does not qualify as a 'forensically relevant paradigm' and may be of limited value for generalizing to witnessing situations in the real world.*

Yuille and Cutshall went on to report that a computer search of psychology journals for studies dealing with eyewitness testimony found 41 articles published between 1974 and 1982 and fully 92% of those used college students as subjects, exclusively. Although the situation may have improved since the publication of their search, it is the case that the large majority of published studies of adult eyewitness memory still involve college students.

On the other hand, it could be argued that to attack eyewitness research on the grounds that it lacks face validity is a weak attack. It is still possible that research using very different appearing, even if seemingly unrealistic, procedures might, nevertheless, yield consistent results. That is, although the procedures may look unrealistic, they may all be tapping into the same basic memory processes that exist when actual witnesses view a crime. With this in mind, it might be useful to turn to the next method of determining generalizability, namely, the consistency of the results across different methods and procedures.

### **External validity, convergent validation, and consistency of results**

Clifford and Lloyd-Bostock (1983) argued that "before communication with legal personnel [about psychological research on eyewitness memory], any finding should be shown to be impervious to the use of different subjects, different research settings, different experimental materials, and different research designs or

methodologies." Despite many claims by experts to the contrary (Kassin, Ellsworth and Smith, 1989), it is our position that the results for many of those factors that have been fairly extensively studied have yielded a picture that is far from consistent. This is true even for factors with such intuitively obvious effects as the length of the retention interval. Furthermore, for many of the factors, even those that are often testified about by defense experts, some researchers agree that the results are far from consistent (e.g., Penrod, Loftus and Winkler, 1982). If the effects that the factors have on accuracy vary with the type of subjects, tasks, settings, materials, and so on, then conclusions such as, "X interferes with eyewitness accuracy," lack external validity. That is, the conclusion cannot be applied in a general way to particular witnesses of particular crimes because X is only a factor for some subjects, some tasks, some settings, some materials, and/or some measures. This section examines, in detail, the consistency of research dealing with several of the 'factors' that are said to play a role in the accuracy of eyewitness identification.

The question of whether results are or are not consistent is a more complex issue than it might at first appear. To explore systematically this question requires that we agree with regard to a measure or measures of consistency. For example, one might decide that the results of two independent experiments designed to test the effects of, say, exposure duration on accuracy should be considered consistent if both experiments yield statistically significant (e.g.,  $p < .05$ ) results. Elliott (1991, 1993) seems to have used this definition in reaching similar conclusions to ours about the consistency of research in several different areas, including eyewitness memory. Alternatively, one might follow Rosenthal's (1991) lead and define consistency in terms of effect size. For example, if the effect size of duration of exposure manipulations over all studies in which it has been examined is statistically significantly different from zero, then this would constitute evidence for consistency. Still another possibility is to define consistency in terms of the nature of the functional relationship between a given factor and accuracy (e.g., is it a linear function or a power function). Finally, one might demand that consistency be measured in terms of the actual parameter values of fitted functions. For example, if accuracy of face recognition increases linearly with increasing duration of exposure, one might ask how consistent the intercepts and slopes of the linear functions are over studies.

Which of these different conceptions of consistency one chooses depends on what one hopes to achieve after deciding whether the results from a number of experiments are or are not consistent. If the goal is to know whether one factor has an effect (any directional effect) on accuracy, then an examination of the variability in effect sizes seems appropriate. If, on the other hand, the aim is to estimate the odds that a given individual's identification

after a 25 second exposure is correct, then consistency needs to be defined in terms of the parameter values of functions relating exposure duration and accuracy. That is, in order to predict accuracy from information about a witnessing situation, we need to have an understanding of how, exactly, accuracy varies with particular features of the situation. For example, suppose an eyewitness expert is told that a witness saw a culprit for one minute. Assume that the expert knows that a meta-analysis of all duration of exposure studies found the effect size to be significant such that longer durations produced higher accuracy than shorter ones. What can the expert say to a jury that might help the jury decide, more accurately, whether the witness's identification is correct. The expert cannot conclude, "Well, witnesses who observe someone for one minute are correct only 40% of the time." Such a statement is only possible if a precise functional relationship between duration and accuracy has been consistently found such that the expert can 'read off' the expected accuracy given the duration of exposure in the particular case. This form of consistency is a much more stringent criterion, one that has, to our knowledge, never been applied in the eyewitness accuracy area. Instead, consistency is generally measured in terms of whether different studies produce significant effects or in a few areas whether the effect size is significantly different from zero. What would an expert who believes that this less stringent type of consistency has been adequately demonstrated for duration say about the accuracy of the one-minute witness? She could say that people who have seen defendants for more than one minute will be much *more* accurate (how much more can't be known) than someone who has seen a culprit for only a minute. Alternatively, she could say that people who have seen defendants for less than one minute will be much *less* accurate than someone who has seen a culprit for an entire minute. Clearly, both conclusions follow equally well from this weaker type of consistency, however, the former sounds much better for the defense while the latter sounds much better for the prosecution.

Despite our claim that consistency has never been adequately assessed with an eye toward the prediction of eyewitness accuracy, we can still ask whether the research is consistent according to weaker standards. If it is not consistent at the weakest of levels, then courts should understand that nothing experts can tell jurors will improve their ability to make more accurate guilt decisions.

## Exposure duration

Witnesses can make two types of errors when identifying faces: they can fail to identify a face that they had seen before (a miss) and/or they can falsely identify a face they had not seen before (a false alarm). Clearly, although misses are of concern to the prosecution (since they might mean that an otherwise strong case against a defendant is not corroborated by witness identifications), false alarms

are of most concern to the defense (since they represent a witness falsely identifying an innocent person as the perpetrator). Unfortunately, when testifying, most defense experts fail to note the difference between these two types of errors and instead simply speak of 'eyewitness reliability' or 'eyewitness accuracy.'

There is little denying the manifest intuition that longer exposure durations should lead to more reliable identifications. However, despite broad defense claims, the literature suggests that this conclusion may only apply to hits versus misses (that is, the accuracy with which previously seen faces are recognized) and not to false alarms. In an extensive review of the literature on facial memory, Shapiro and Penrod (1986) examined the results of eight experiments that systematically varied the exposure duration of faces and measured miss and false alarm rates. Comparing the error rates for the shorter with the longer exposure duration conditions in all eight studies, they found that length of viewing time had a significant effect on the rate of hits versus misses (as expected, subjects averaged more misses in the shorter duration conditions) but had no effect on the rate of false alarms. Furthermore, in a meta-analysis of over 190 studies, they found that across all studies (consistent with most people's intuition), as the exposure time increased from experiment to experiment, the hit rate also increased, but, quite unexpectedly, the false alarm rate increased, as well. Although the latter result is by no means conclusive, it does represent an important type of inconsistency among findings in the field, namely that some factors may affect one accuracy measure differently than they affect another.

From another point of view, even if both hits and false alarms are affected by exposure duration, we currently do not know what the *functional relationship* is between exposure duration and accuracy. Although it is almost surely the case that longer durations will show diminishing returns, how soon those diminishing returns take effect is unknown. How much extra accuracy can we expect between 30 seconds and 30 minutes of exposure; how much between one minute and two minutes? The answer to such questions is not known at this time, but should be of considerable interest to jurors who are being lectured by defense experts about the terrible effects of short exposure durations on eyewitness accuracy. Reasonable jurors should want to know whether a witness who may have seen the culprit for two minutes produces identifications that are more like those obtained with 1/2 second of exposure or more like those obtained with 20 minutes of exposure. As it now stands, eyewitness experts called by the defense simply testify that shorter durations of exposure reduce the accuracy of witnesses' identifications. After reading the transcripts and listening to the testimony of many eyewitness defense experts in over 50 cases, not one mentioned that he or she was

uncertain about the effect duration might have on the rate of false alarms.

## Retention interval

Most people would agree that memory fades with time and most experts agree that it fades faster immediately after exposure but then tends to level off (Kassin, Ellsworth and Smith, 1989). Even if this verbal description is accurate, when the results of a number of studies are considered together, the picture that emerges is far from consistent. Penrod, Loftus, and Winkler (1982), on pages 135 and 136, and Shepherd (1983) review most of the major studies in this area published prior to the 1980s. Penrod, Loftus and Winkler concluded, after citing several inconsistent findings, that the longer the retention interval, the worse the performance, but they also noted that, "unless one knows a great deal about the specific conditions under which the incident is viewed, it is impossible to predict the precise forgetting curve." We would agree with the latter and extend it to include needing to know how memory is measured, who the subjects are, the motivations of the subjects, and so on. Even then, we would argue that the current state of knowledge is such that although one might be able to say that the form of forgetting is best described as a power function (Wixted and Ebbesen, 1991, 1996), one can not accurately predict the exact shape of the forgetting curve (that is, the parameter values of the power function) in any given instance. A careful examination of eyewitness memory studies that have included retention interval as a factor is quite consistent with Penrod, Loftus and Winkler and our conclusion (Cutler, Penrod and Martens, 1987a,b; Egan, Pittner and Goldstein, 1977; Krafska and Penrod, 1985; Laughery, Fessler, Lenorovitz and Yoblick, 1974; Shapiro and Penrod, 1986; Shepard, 1967; Shepherd, 1983; Shepherd and Ellis, 1973; Yuille and Cutshall, 1986).

There may be good theoretical explanations for the apparent inconsistencies in the results of studies that have examined the effect of retention interval on eyewitness memory. One factor may be the extent to which the procedure for measuring memory provides the subjects with information about 1) the likelihood that previously seen people are present in the test stimulus set and 2) the consequences to the subjects for making false, positive identifications. In particular, there can be little doubt that, in general, people's memories for some things, including what other people look like, fade with time. In a recognition memory test that asks witnesses whether they have seen a face that they saw before, one would expect the odds of saying 'yes' to decrease as the retention interval increases -- because the witnesses would be forgetting what the face looks like. However, what might one expect when people are shown a face that they had not seen before? Surely memory for this never seen face does not become stronger as time goes on. But then, why would people be more likely to positively identify a previously unseen face the longer the retention interval? An answer to

this question might involve the 'pressure' to say 'yes' during the recognition test. The greater the pressure, the more likely subjects might be to pick someone they did not recognize. When the pressure is weak, people could simply say they cannot remember what the face looked like or that no one looks familiar. Pressure to pick someone can come from several sources. In laboratory tasks in which the subjects are shown a large number of faces in a test, the subjects might be told that they have seen half of the test faces before and thus say 'yes' about 50% of the time. In other test procedures, the experimenter might imply that one of an array of choices is a previously seen face (even though none of the faces were actually seen before) thereby increasing the odds that the subject will pick someone. Finally, the greater the costs of falsely picking a previously unseen face, the less likely people will be to say 'yes.'

The above reasoning might explain why some studies find an increase in false alarms with greater retention intervals and others do not. Some memory tasks put pressure on subjects to hold the odds of 'yes' (or positive identification) responses fixed across all retention intervals. For example, if the subjects know that they have seen half of the test faces before, they might try to say "yes" about 50% of the time regardless of the length of the retention interval. Alternatively, when a lineup task is used, experimenters might imply that the culprit is present, even in target absent lineups. If this reasoning is correct, as memory for the faces that were seen fades with time and witnesses become more likely to say "no" they have not seen these faces before (because they cannot remember them), they necessarily must become more likely to say "yes" they have seen other faces that they had not seen in order to keep the odds of saying "yes" fixed at some level. This analysis raises the possibility that accuracy results will depend not only on variables such as duration and retention interval but also on whether witnesses are given the opportunity to say, "I can't remember," or even the opportunity to indicate that they are less than completely confident (Ebbesen and Wixted, 1996).

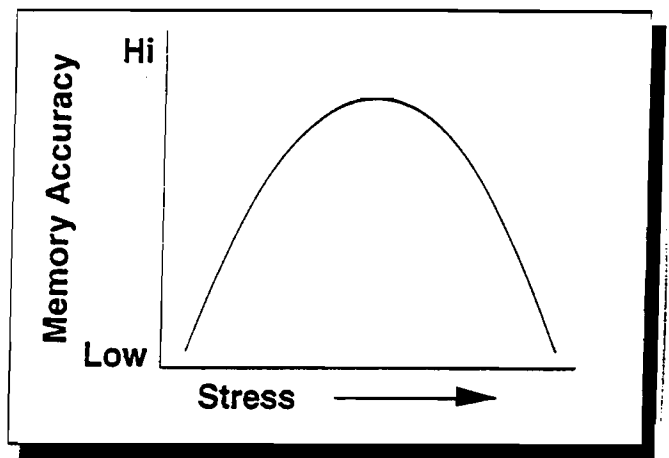
Thus, whether one will find an increase in the rate of false alarms as retention interval increases may depend on the extent to which subjects believe that they have to pick someone, even if they do not really remember having seen that person before. The requirement noted by Clifford, et al. (1983) that minor changes in procedures should not have a substantial effect on the findings has not been met even for a factor whose effects seems so intuitively obvious, namely, the length of the retention interval.

## Stress and memory

Deffenbacher (1983) reviewed some 21 studies that he claimed examined the relationship between arousal and the accuracy of eyewitness memory. He noted that, "Ten of

them have produced results that suggest that higher arousal levels increase eyewitness accuracy – or at least do not decrease it ... The remaining 11 studies have produced just the opposite result - lower accuracy of memory was yielded by experimentally manipulated increases in arousal or higher individually assessed arousal levels." The Deffenbacher review provides a thorough enough description of the procedures and studies that we do not have to provide further description here. Instead, we will comment on Deffenbacher's conclusions and on the conclusions frequently reached by defense experts regarding stress and memory.

Figure 1.  
Inverted U-shaped function that some claim relates stress and memory.



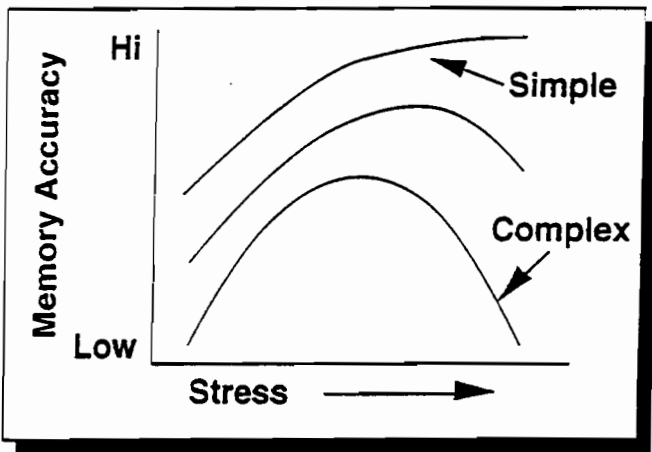
A frequent claim by defense experts is that high stress causes more mistakes, not fewer. Or more strongly, they claim that the consensus of scientific judgment is that emotional arousal is destructive to the perception process and hence to memory (Kassin, Ellsworth, and Smith, 1989). Since defense experts rarely mention the details of Deffenbacher's review directly in their testimony, it is difficult to know how they would deal with the fact that half of the studies found one effect and half found another. On the other hand, in our experience and reading, the usual method of handling this dramatic inconsistency in results follows a portion of Deffenbacher's own conclusions: Namely that the relation between stress and memory is not a simple decreasing function, but an inverted U-shaped function (see Penrod, Loftus and Winkler, 1982). That is, both low levels and high levels of stress are assumed to produce poorer performance than medium stress levels (see Figure 1, above). To draw this conclusion, Deffenbacher argued that those studies which found a positive relationship between stress and memory generally used lower stress levels in all conditions and studies finding the opposite relationships used higher levels of stress. It was further argued, with no independent empirical evidence, that witnesses and victims of crime must all be experiencing very high stress levels. (If they



were experiencing medium levels of stress, then Deffenbacher's ideas would predict superior, in fact, the best, memory for witnesses of crimes.)

Deffenbacher further argued that the shape of the inverted U depends on the complexity of the memory task, such that the peak of the curve, or the stress level at which performance would be maximal, moves to higher and higher stress levels as the memory task gets simpler and simpler (see Figure 2, below). It is of considerable interest to note that none of the testimony by defense eyewitness experts that we have read and heard in court (including that of Robert Bjork, Robert Buckout, Scott Fraser, Solomon Fulero, Elizabeth Loftus, Kathy Pezdek, Steven Penrod, and Robert Shomer) mentions this part of Deffenbacher's explanation.

**Figure 2.**  
Deffenbacher's original ideas about how memory and stress are related to each other.



Nonetheless, although Deffenbacher's full theory may sound like a reasonable explanation for the inconsistent findings, it is only one of several reasonable explanations that fit the results and the details of the studies. For example, one of the studies that Deffenbacher argues belongs on the higher stress side of the inverted U is the Clifford and Scott (1978) study. In that study high stress consisted of watching a 1.2 minute *film* in which four blows were exchanged and low stress consisted of watching a similar film in which angry words were exchanged between confederates. One might be tempted to conclude from this that defense expert reasoning assumes that college students watching a brief film while knowing that they are in an experiment involves about the same stress as being a victim in an armed robbery. This would be incorrect, however, since Deffenbacher placed another study (Sussman and Sugarman, 1972) in which subjects watched one of two films that varied in terms of the violence they portrayed on the lower arousal side of the curve. In this study, the high violence film depicted a victim of an armed robbery being threatened with a gun and then being beaten about the head with the gun in such

a manner that the victim's head was shown bleeding profusely. The low stress film eliminated the gun, the beating, and the blood. Interestingly, possibly because subjects were forewarned about their having to identify the culprit, identification accuracy was equally good across the two films. It is tempting to use such a conclusion, if true, to wonder whether the effects of stress on accuracy (whatever they may be) might be irrelevant when the witness attempts to remember who the perpetrator is. If so, whether a witness was, at the time of the exposure, trying to remember the perpetrator's face should be something a defense expert considers before agreeing to testify about the effect of stress on witness accuracy.

A more recent study by Cutler, Penrod, and Martens (1987a) completed after Deffenbacher's review appeared, also varied the degree of violence that subjects observed in a videotape of a simulated crime. In the violent tape, the robber pushed a store clerk around, fired his gun into the floor, and threw the victim down before leaving. In the non-violent tape, the robber remained calm throughout and neither fired his gun nor manhandled the victim. Like the Sussman and Sugarman (1972) study, Cutler, et al. reported no effect on witness identification accuracy of the violence depicted in the videotaped robbery, despite the fact that subjects rated the high-arousal episode as much more violent and despite the absence of forewarning.

The placement of a number of the remaining studies reviewed by Deffenbacher is equally suspect. For example, one of the studies placed on the lower-stress side of the abscissa was a study by Leippe, Wells, and Ostrom (1978). In this study three levels of stress were produced by having subjects believe that they were witnessing real crimes of different seriousness (the theft of a researcher's calculator, the theft of a pack of cigarettes, or no theft). Accuracy was measured using a six-person photo-ID spread. Accuracy was 56% in the calculator theft group and only 19% correct in the cigarettes group. And in another study (Johnson and Scott, 1976), also claimed to be on the low overall arousal side of the curve, subjects ostensibly waiting for an experiment to begin either saw a person run into the room for about four seconds with grease covered-hands holding a pen and muttering something about broken machinery, or, in the high stress condition, after over hearing an argument and then sounds of glass breaking and chairs crashing, saw a person run into the room, again for about four seconds, with blood-stained hands, holding a bloody letter opener. Memory was measured using free and controlled narrative reports similar to that used by Clifford and Scott and a mug-shot identification task. Although male and female subjects remembered different things, those witnessing the bloody-handed event recalled more correct facts about the target's actions and the crime scene than those witnessing the greased-hand event.

Still another study placed on the lower arousal side of the curve was conducted by Hosch and Cooper (1982). This study compared the identification accuracy (from a six person photospread) of someone who entered a room while the subject was engaged in another task and apparently stole the subject's own watch, another person's calculator, or nothing. Identification accuracy was 71%, 67%, and 33%, respectively. In a second study (conducted after Deffenbacher's review) Hosch, et al. (1984) found similar results. Victims of an apparent theft of their own wrist watches were no more likely to falsely identify an innocent foil in a lineup than were witnesses to the same theft despite the fact that there was a non-significant tendency for the victim's overall memory to be worse than that for witnesses.

Interestingly, studies using almost identical methods for manipulating arousal were placed on different sides of the arousal curve. Three studies (Clifford and Hollin, 1978; Giesbrecht, 1980; Majcher, 1974) varied the amplitude of white noise to which subjects listened while they were exposed to slides of faces. The former two were placed on the higher arousal side of the function, while the last was placed on the left, lower-arousal side of the curve - with no apparent reason other than this placement was consistent with Deffenbacher's inverted-U claims.

Although not a necessary argument, the above point makes it easier to understand that there are other reasonable explanations that Deffenbacher did not present in his review for the inconsistent effects of 'stress.' One is that the function relating stress and memory is not an inverted U as he suggests but is, instead, a U. That is, memory is best at low and at high levels of stress and worst at medium levels. Seeing someone with bloodied hands or stealing someone's calculator or one's own watch seems more stressful than watching a movie. In fact, a study by Yuille and Cutshall (1986) of the accuracy of 13 real witnesses to an actual robbery/killing supports this conclusion. Although there are many interesting findings in this study, one is that witnesses who reported greater arousal while seeing the crime recalled fewer incorrect facts about the events and individuals involved in the crime than witnesses who reported being less aroused. Very high levels of arousal were associated with better memory than medium levels of arousal. Although some defense experts have correctly pointed out that those witnesses who were most stressed had a better view of the crime and it was the better view rather than the extra stress that may have caused their more accurate memory, it is still the case that whatever extra stress those closer to the crime experienced, it was not enough to cause them to have worse memory. More importantly, it is precisely these kinds of correlations among factors, nearness to crime and stress, that make generalizations about the effects of any one factor (in isolation of all others) to real crime scenes virtually impossible.

In still another paper, Tooley, Brigham, Maass, and Bothwell (1987) reported the results of a study in which among other variables, the level of 'stress' that subjects felt was varied by delivering blasts of white noise and threatening the same subjects with electric shock while they looked at faces. When measured in terms of hit rates, recognition memory was better for those subjects who were threatened than for those who were not. The fact that the threat manipulation was presented in such a manner that subjects thought they might be able to avoid the noise and shock by discovering a hidden cue in each face might explain the better memory in the higher arousal group. More importantly, if this explanation is correct, it points out the potentially different effects that stress may have depending on how the subjects are motivated to deal with that stress. This is a conclusion that is never reached by defense experts.

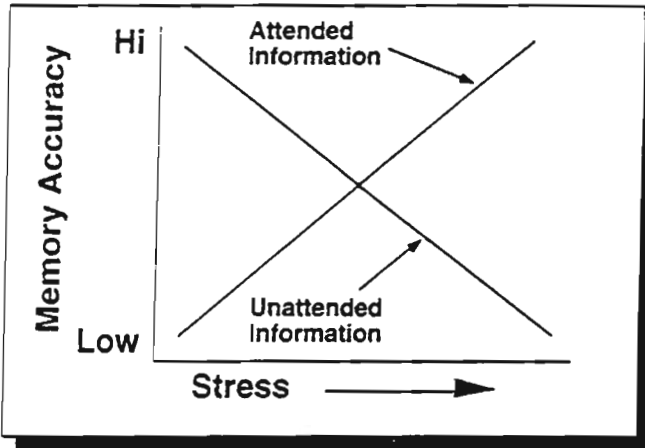
Using a very different paradigm, Brown and Kulik (1977), Pillemer (1984), Winograd and Killinger (1983) studied what they called flashbulb memories. Stated simply, the concept of flashbulb memory is that sudden, dramatic, and very emotional events leave very detailed and very long lasting memories for events surrounding the experience. Common examples are people's very good memories for what they were doing and where they were when they heard about President Kennedy being shot, or for blacks, when they heard about Martin Luther King being killed. Researchers have found that those who expressed the most emotional involvement had the strongest and most detailed memories. For example, only those who were strongly upset by the attempt on Reagan's life (primarily Republicans) had vivid memories of what they were doing when they heard of the event.

Despite what some consider support for the U-shaped-function explanation of the inconsistent findings, there is another reasonable view, namely, that the effect that high stress levels have on memory is conditional on other things, i.e., the effect varies with other as yet to be specified and understood factors (see Figure 3, below.). Such interaction models are very common in psychology. It could be, for example, that stress produced by fear improves memory but stress produced by anger decreases memory. Or it could be that stress enhances memory for some things and decreases memory for other things. Evidence for the latter comes from a study that is often cited as supporting the conclusion that high stress interferes with memory. Loftus and Burns (1982) found that subjects who watched a film of a robbery that included a scene depicting a little boy being shot in the face were less likely to remember the number on the jersey that the boy was wearing than those seeing a similar film that did not include the shooting. What is never mentioned is that for almost every other recalled detail that was coded for accuracy, the two groups were virtually identical and highly accurate (averaging around 85% correct for all but one item, although there was a slight

tendency for subjects in the nonviolent film to be higher than lower, by very small percentages, on more items). In short, out of 17 recalled facts, the only detail not remembered equally well by both groups was the number on the boy's jersey. The robber's clothes, the robber's hair, the note to the teller, the robber's mustache, the alarm button, and so on, were all recalled equally well with and without stress.

Figure 3.

Interaction idea about how stress and memory might be related that is similar to Easterbrook's (1959) cue-utilization hypothesis.



The effect of stress may vary with other factors. Some types of stress might affect males and other types of stress might affect females. Still another possibility is that common types of stress produce one effect and uncommon, or novel, types produce other effects. But the most reasonable possibility is that stress enhances memory for some things and reduces memory for other things. In particular, after reviewing hundreds of studies, Christianson (1992) argued that stress causes people to attend more closely to some things and less closely to others. If researchers measure memory for those things to which people pay more attention when stressed, they will find that memory improves with stress. If they measure those things to which people do not attend, they will find that memory worsens with greater stress. Christianson (1992) believes that this old explanation (Easterbrook, 1959) is to be preferred over the inverted-U explanation. Of course, the list of potential explanations is endless. And the reason that it is endless is that not nearly enough research has been done to eliminate the many plausible rival hypotheses about the relation (or relations) between stress and memory.

Other researchers, not normally cited by defense experts, argue that considerable evidence supports the claim that emotion generally improves memory, both for peripheral and central details. For example, Heuer and Reisberg (1992) describe work in which subjects are more likely to

recall accurately and correctly answer multiple-choice questions about that part of a story that contains emotional content than that same part of a similar story, but without the emotional content. Some (McGaugh, Introini-Collison, Cahill, Castellano, Dlamaz, Parent and Williams, 1993) have even suggested that the amygdala may be responsible for the enhanced memory that emotion produces implying the memory for emotional information is driven by different brain processes from memory for non-emotional information. Thus, according to this view, not only is the inverted-U explanation not true, but memory for emotional events will be more accurate than memory for non-emotional events.

Even if the Deffenbacher model (a range of inverted U-shaped functions that vary with task complexity) is correct, one wonders how knowledge of this could possibly help a juror determine the reliability of a given witness in a given case. Not only would the juror have to know exactly how much stress the witness experienced (in order to 'look up' on the inverted-U chart the expected level of memory for that amount of stress), they would also have to know how complex the memory task was to that witness (in order to know which inverted-U to use). Is recognizing a face a complex or a simple memory task? Which is harder, remembering the words to a song, remembering what a robber said, remembering what someone was wearing, remembering a license plate number, etc.? No one knows the answer to questions such as these because no one agrees how complexity should be defined or measured. Once again, the state of knowledge in the field is such that a complete explanation of what is known about stress and memory would only serve to confuse the jury.

Not only do eyewitness experts not have an agreed upon measure of task complexity, but they also do not know how to estimate the amount of stress that particular witnesses were probably experiencing during the crime. Part of the difficulty arises because the crime is already over, and therefore physiological indicators of stress, such as, heart rate and blood pressure, may have returned to normal by the time identifications and descriptions are given by the witnesses. Psychologists do not even know whether the physiological arousal produced by recalling a mildly stressful crime would be different from that produced when recalling a very stressful crime. No one knows how to reliably measure the amount of stress that was experienced hours or days earlier. Equally important, the criminal investigation system does not measure the amount of stress that different witnesses may have experienced in a standardized manner. Instead, the jury is often left to their own intuitions, possibly 'helped' by some verbal statements by the witness in court, to judge the amount of stress the witness might have experienced from a description of the events taking place as the crime unfolded. We have been unable to find any research that examines the relationship between conclusions that the

witnesses, defense experts, and/or jurors reach about the amount of stress a witness or victim experienced during a crime (real or simulated) and the actual stress experienced by that witness.

## Weapon focus

Briefly, the logic of the 'weapon focus' effect is that people will look at a weapon more than at other things and therefore when a weapon is present in a crime, memory for other things will be less accurate than when no weapon is present. Some might also argue that the presence of a weapon increases stress and thereby causes still further reductions in accuracy. However, once again, a detailed analysis of different studies of the 'same' phenomenon suggest some inconsistencies in results (compare Loftus, Loftus and Messo, 1987; Kramer, Buckout and Eugenio, 1990; Cutler, Penrod and Martens, 1987a,b; Tooley, et al., 1987; and Maass and Köhnken, 1989).

Regardless of what one makes of these studies, it is clear that only a few studies have attempted to systematically explore this issue. And despite claims that increased attention to one aspect of the environment causes decreased attention to and memory for other aspects of the environment (an obvious fact that jurors certainly already know), the crucial issue is whether and how much weapons attract attention and if they do, does the degree of attraction vary with the type of weapon, other aspects of the environment, the motivation of the witnesses (are they trying to remember the face of the perpetrator so they can identify him later?), the length of exposure to the scene, the retention interval, and so on.

For example, while it seems likely that memory for what a 'criminal' looks like depends on the amount of time that the witness looks at the culprit rather than something else (e.g., the weapon), a 'law of diminishing returns' should apply to time spent looking at the culprit. If so, it follows that the effect of a weapon might disappear if the witness has enough time to look long enough at the culprit, even when a weapon is present. A meta-analysis (Stebly, 1992) of 19 different tests of the weapon focus effect (a minority of which found significant effects on identification accuracy) concludes,

*The data support the hypothesized weapon focus effect...The data also show that both dependent measures - lineup identification accuracy and feature accuracy - are sensitive to the weapon focus effect...The presence of a weapon does make a significant difference in eyewitness performance. (p 420)*

However, in another line of the article Stebly notes:

*Thus, it appears that scenarios (and more specifically, lineups) that produce low identification accuracy for subjects in general (i.e., control subjects) accentuate the weapon-focus effect. (p 420)*

In other words, when the procedures allowed the subjects to learn what the target looked like, the presence of a

weapon had much less, if any effect. Unfortunately, we can not tell from Stebly's analysis how long the exposure duration needs to be before attention to a weapon no longer matters, yet defense experts do not tend to qualify their conclusions about the effects that a weapon might have on accuracy according to the time that a particular witness might have had to look at the suspect.

Furthermore, these studies have ignored a crucial aspect of real-world witnessing when a weapon is present, namely, the witnesses' self-reports about the focus of their attention. In our experience, some victims of and witnesses to actual crimes do report looking at the weapon, but others do not. Some say they looked into the eyes of the culprit to judge his intentions. Others say they studied his face in order to identify the person who put them in such a terrible situation. Often prosecutors will use witnesses who say they can only remember the weapon to help identify a weapon found in the defendant's possession, but will not use that same witnesses to identify, directly, the culprit. Before data from weapon focus experiments can be generalized to real-world victims and witnesses, researchers should be required to report accuracy results separately for those who said that they looked at the weapon and therefore felt unable to identify the culprit and those who said they looked at the culprit despite the presence of the weapon. In fact, it is conceivable that witnesses who report looking at the culprit when a weapon is present may actually have better memory for the culprit than those witnesses who saw the same event without a weapon present. The stress-induced narrowing of attention may improve later recall and recognition performance.

Finally, it might be noted that an agreed-upon theory for a weapon focus effect does not exist (Bosworth and Ebbesen, 1996). Some have suggested that the effect is mediated by a narrowing of the focus of attention, in much the same manner that a spot light shining on a stage can be made smaller. Others argue that it is merely the direction of gaze (on the weapon or on the face) that produces the effect. Until we know which of these, if either, is correct, defense experts can not tell a jury whether it is important to listen to witness reports of what they were looking at during crimes.

## Post-event information

This is the one area among those frequently mentioned by the defense experts, for which, until the mid-1980s, one might have justifiably argued that the results had been fairly consistent and had supported the idea that memory for events after exposure to a 'crime' might become integrated with memory for facts about the 'crime.' However, articles by McCloskey and Zaragoza (1985), Bekerian and Bowers (1983), Bowers and Bekerian (1984), Zaragoza (1987) and Loftus, Schooler and Wagenaar (1985) have suggested that the consistency of the earlier findings

was potentially the result of the relatively common use of a fundamentally flawed measurement procedure and sloppy theorizing about the nature of memory. The major points of this discussion were made by McCloskey and Zaragoza.

One of the major unresolved issues in this area is exactly what one means by memory and whether all errors that witnesses make should be classified as mistakes brought on by faulty memory for the source of the remembered information (Zaragoza and Lane; 1994) or by some other process (e.g., response bias or strong desires to help the experimenter). For example, Zaragoza and her associates (Zaragoza and Koshmider, 1989; Zaragoza and Lane; 1994) argue that post-event suggestions do not consistently produce "source misattributions" nor do they consistently result in subjects saying they remember things that they really know they do not remember. Some defense experts might argue that this distinction is irrelevant because in both cases the witnesses will falsely identify something they have not seen before. However, despite over 20 years of work on this problem, the results have been so mixed that we do not have a theory that allows us to predict under what circumstances witnesses are likely to say they can't recall, to knowingly 'lie' about what they saw, or to misattribute the source of the memory.

### Unconscious transfer

Although unconscious transfer or 'photo-biased memory' appears in several different conceptual forms, the logic of the defense position is clear. The defense argues that many identifications made by witnesses may be based on memories of prior events rather than on an independent memory of the criminal obtained during the commission of the crime (*United States v. Wade*, 1967; Sobel, 1987). Four different procedures have been used. In one, people's memory for where a face was seen has been shown to be worse than memory for the face, itself (Brown, Deffenbacher and Sturgill, 1977). Another tests whether presence of a bystander can reduce the accuracy of later identifications of a criminal (e.g., Read, Tollestrup, Hammersley, McFadzen and Christensen, 1990; Ross, Ceci, Dunning and Toglia; 1994). A third has tested whether seeing someone in a mugshot or photo lineup can influence who will be picked from a later lineup (Cutler, Penrod and Martens, 1987a,b; Davies, Shepherd and Ellis, 1979; Deffenbacher, Leu and Brown, 1979). The last examines whether the act of picking someone from an earlier photo lineup commits the witness to choose the same person again even if the first choice was incorrect (Gorenstein and Ellsworth, 1980).

Our reading of the research in these areas suggests that the results have little or no relevance to eyewitness identification in the real world and/or are inconsistent. For example, the fact that we remember faces without being able to remember where we saw those faces is a

problem only if we do not know that we can not recall where the face was seen or come to believe that we saw the face in one location when, in fact, we saw it in another. Unfortunately, researchers have not allowed subjects to indicate their reasons for their lineup choices or if they have, they have not broken down the results by those reasons (Gorenstein and Ellsworth, 1980). Thus, we do not know whether subjects who think a face is familiar, but do not know why, would be willing to testify that this familiar person is the culprit. Stated differently, if there is an effect, it may well be limited to people whose confidence would be so low that they would never be used as witnesses in a real case.

In the 'bystander effect' area, Read, Tollestrup, Hammersley, McFadzen and Christensen, (1990) reported that their results "repeatedly failed to reveal more misidentifications of an innocent bystander by witnesses who had been previously exposed to the bystander than by control eyewitnesses who had not." (p. 3) On the other hand, Ross, Ceci, Dunning and Toglia (1994) reported that subjects were three times more likely to pick a bystander as the culprit when they saw a lineup that contained the bystander and not the culprit. However, this effect went away when the subjects were informed prior to seeing the lineup that the bystander and the culprit were not the same person.

Finally, studies of the effects on lineup choices of seeing or choosing a mugshot of an innocent person have been designed in such a manner that their results are diagnostically useless. In particular, since in the real world the person in the mugshot and the identified defendant are almost always one in the same, it makes no sense to focus exclusively on the detrimental effect on lineup choices of seeing a mugshot of an innocent individual. If it is the case that seeing or choosing a mugshot of someone increases the odds that witnesses will pick that person out of a later lineup, then the defense may have a point but only if the mugshot was of an innocent person. If the mugshot was of the guilty culprit, seeing it should increase, not decrease, the odds that the witness will choose the guilty person from the later lineup. Thus, to use results from this area, a jury must first decide if whether the person in the mugshot is the culprit, an issue about which unconscious transfer research is silent.

In short, the evidence testing the unconscious transfer or 'photo-biased memory' effect seems inconsistent and/or irrelevant, at best.

### Confidence and accuracy

Since Deffenbacher's (1980) review of the literature, research evidence on the relationship between confidence and accuracy has proven inconsistent with the common defense expert claim that there is little or no relationship

between witness confidence and accuracy. To quote Fleet, Brigham, and Bothwell (1987):

*The claims of previous reviewers of the confidence-accuracy literature (Deffenbacher, 1980; Leippe, 1980; Wells and Murray, 1984) that confidence is an unreliable predictor of accuracy are perhaps premature. In addition to the unresolved issues of how to subdivide the research samples, there are the issues concerning ecological validity. For example, several recent field studies have found a significant correlation between confidence and accuracy (Brigham, et al., 1982; Hosch and Platz, 1984; Krafska and Penrod, 1985; Pigott, et al., 1985).*  
(p 183)

Although it is clear that the size of some types of correlations between confidence and accuracy is not large, it is becoming clearer that when witnessing conditions allow subjects to perform at better than near chance levels on identification tasks, the correlations are positive (Brigham, Maass, Snyder and Spaulding, 1982; Bothwell, Deffenbacher and Brigham, 1987; Deffenbacher, 1980; Krafska and Penrod, 1985). In addition, when the relationship is measured only for hits and false alarms (e.g., choosers or 'yes' responses) and the confidence is in those responses rather than predictive of yet to be made identifications, the relationship is even stronger (Sporer, Penrod, Read and Cutler, 1995; Wells and Lindsay, 1985). Finally, recent work by Ebbesen and Wixted (1996) showing that the confidence and accuracy relationship may be understood in terms of signal detection theory suggests that at the level of individual identification responses, more confident identification responses are virtually always much more likely to be correct than less confident responses, despite the fact that certain correlational measures of the association will be small and sometimes not significant.

These empirical facts have two very important implications. The first, suggested recently by Elliott (1993), is that because confidence, like response latency (Sporer, 1993; 1994; Sporer, Penrod and Cutler, 1995) and the reasons that subjects give for their identification responses (Dunning and Stern, 1994), probably reflects the strength of people's memory for the people and faces that they identify as having been seen before, it is possible, and even likely, that such response measures will prove to be much better predictors of the accuracy of identifications than other situational factors, such as, stress, duration of exposure, or weapon presence. In fact, it may even be the case that such memory strength measures will capture a good portion of whatever effects such factors have on identification accuracy (Ebbesen and Wixted, 1996).

The second implication follows from the intuitive fact that the legal system tends to use confidence (and other certainty indicators) to select witnesses and to determine the facts about which witnesses will tend to testify (Wells and Turtle, 1987). As such, juries will tend to hear mostly

witnesses who express high confidence in their memories of the things about which they testify. However, researchers continue to report results of the effects that different factors have on the accuracy of all subjects, including those who confidence would almost surely prevent them from every testifying in court. Until the effects of such factors as racial similarity, stress, duration, etc. are examined separately for confident and non-confident identification responses, the external validity of conclusions about the effects of those factors is highly suspect.

## Lineup Fairness

Some (Wells and Lindsay, 1985) have suggested that the most forensically relevant test of the confidence-accuracy relationship is to compare the ratio of 'suspect' choices in suspect present lineups with 'suspect' choices in blank lineups for each level of confidence. This ratio should be highest for those subjects who have expressed the greatest confidence (that is, the most confident witnesses should be the ones best able to discriminate the culprit from a nearly identical look-alike). At one level, the Wells and Lindsay position suggests that the system's reliance on confidence as an indicator of witness reliability (e.g., *Neil v. Biggers*, 409 U.S. 188 (1972) and *Manson v. Braithwaite*, 432 U.S. 98 (1977)) is premature because except for a very few reported studies, this test is not performed. At another level, if one accepts their argument, it suggests that the frequently made claim by defense experts that accuracy and confidence are unrelated is also premature.

Yet, without yielding any ground to the present defense-expert claims, one should reject the Wells-Lindsay procedure because it requires the use of a blank lineup without defining the procedures that should be used to construct that lineup (e.g. Gonzalez, Davis and Ellsworth, 1995). How similar in appearance to the culprit should the look-alike in the blank lineup be? Is sophisticated similarity scaling to be done on a case-by-case basis? Obviously, the more similar the look-alike is to the culprit, the more likely it is that the witness will pick the look-alike, even though the witness has close-to-perfect memory for the criminal. Imagine creating two lineups, one with the actual criminal and another with his identical-twin brother. Would the fact that a witness picked both with high confidence tell us about the unreliability of the witness's memory or that the first twin was probably not guilty or that the construction of the blank lineup was specifically created to cast aspersions on a perfectly good witness?

Another point is that lineups serve different purposes in different cases. For example, in the large majority of cases the lineup is apparently used merely as a method to establish that a witness's memory is good enough to allow the witness to testify that a defendant who is already known, or very likely to be the perpetrator (because of

various types of corroborating evidence) is the person that the witness saw. It is only in a small proportion of actual cases that the lineup serves as the primary (and sole) method of discovering who the perpetrator was (Moore, Ebbesen, and Konecni, 1994). And, even in the latter cases, the Wells-Lindsay argument only makes sense if the defendant was arrested solely on the basis of his looks.

Consider, for example, a range of actual arrest situations in which defense experts actually testified about eyewitness identification. A defendant was arrested because he was near the scene of a crime and was wearing clothing that matched a victim's description. In this case, the arrest was not based on the facial characteristics of the defendant. If the defendant was not the culprit, then *none* of the individuals in what would have been a blank lineup would look like the actual culprit. In another example, a victim was asked to view a lineup because the 'MO' of someone arrested at the scene of a different crime matched that of the crime perpetrated against the victim. If such a lineup were 'blank,' it is extremely likely that none of the individuals in that lineup would have looked like the culprit. In still another case, an arrest was made on the basis of the culprit's name. In each of these instances, there is a high probability that none of the people in the lineup would have looked like the culprit unless the police had arrested the actual culprit. This is an important issue because the prevailing consensus in the field seems to be that the fairest way to test accuracy is to present witnesses with two lineups, one with the culprit and one without, and to construct the target-absent lineup in such a way that the culprit is replaced with a look-alike. Presumably this belief is based on the assumption that all lineups contain people who look like the culprit and what the system needs to guard against are witnesses whose memory of the culprit's looks is so poor that they will be more than happy to pick someone who only looks somewhat like the culprit. However, as the former examples are designed to show, many real-world lineups are not constructed on the basis of the looks of the defendant and therefore it is impossible to generalize the results of the research that has used blank lineups to many real world identifications because in virtually all simulated blank lineups at least one person looks a lot like the culprit.

## Functional size of lineups

Lineup fairness is often discussed in terms of the lineup's 'effective or functional size' rather than in terms of its actual size. Thus, the odds of picking the suspect in a lineup of six individuals would be much higher than one in six, if, for example, the witness knew that the culprit was black and the six-person lineup contained three African Americans and three Caucasians. In this view of lineup fairness, the fairest lineup is one in which the *a priori* odds of picking the suspect by individuals who know various aspects of the suspect's looks but never saw

the suspect should be close to one in six. This logic argues for lineups in which all of the foils are as similar in appearance to the suspect as possible. While this argument seems to make sense at first thought, there are several reasons to question it. First, at the extreme, the argument cannot be correct (Wells, Seelau, Rydell and Luus, 1994). Assume that all of the foils were virtually identical in appearance to the suspect. A witness with perfect memory would be unable to detect which of the six individuals matched her memory best because all of the individuals would do so equally well. Such a lineup would not tell us anything about the accuracy of the witness's memory. Second, it is unclear how to measure accurately the effective size of a lineup. One method is to tell people who did *not* see the culprit something about the culprit's looks and then measure how these people would distribute their choices over the individuals in the lineup. If most of the subjects chose the culprit, this would imply that the lineup was biased. On the other hand, how much should the subjects be told about the culprit's looks in such a test (Gonzalez, Davis and Ellsworth, 1995)? Again, at the extreme, the test fails. Suppose the subjects are told in great detail exactly what the culprit looks like. Would we not expect the odds of picking the culprit to be a lot higher than one in six, even in a lineup in which the foils looked something like the culprit? In fact, might not we conclude that the person so many people picked on the basis of the detailed description was indeed the actual culprit. Third, in live lineups especially, it is unclear whether guilty individuals display cues of their guilt in their behavior (e.g., eye contact, micro-facial expressions, and body language), cues that non-witnesses might use even if they were to know nothing about the looks of the culprit. This reasoning suggests that tests of lineup bias should include a group of non-witnesses who are told nothing about the culprit's looks. In sum, it is unclear exactly how similar foils should be to the culprit and how much non-witnesses should be told in tests of lineup bias. Not only have appropriate calibration experiments not been done - that is experiments that examine how these variables, strength of witness memory for the culprit and witness confidence, interact - but even if they had been, a fair evaluation of lineup bias would have to include an independent assessment of the odds with which culprits tend to be in lineups. And the latter, importantly and incredibly, is also not known.

Many researchers and defense experts seem to believe that the functional size problem is the worst at one extreme of the lineup size, namely a lineup of size one. In such cases, often called show-ups in the legal system, a witness is shown one picture or one suspect and asked if that one person is the culprit they saw. Most experts and the research community seem to believe that this is the most biased identification procedure because there seems to be so much pressure on the witness to pick the suspect and because there are no opportunities for the witness to pick a non-suspect. Despite these claims and protests, several

recent studies have suggested that show-ups actually result in a lower probability of false alarms than multi-suspect lineups (Ebbesen and Boley, 1994; Gonzalez, Ellsworth, and Pembroke, 1993; Moore and Ebbesen, 1994). These results have been explained by assuming that witnesses use a different judgment strategy in lineups than showups. In particular, it is suggested that in lineups the witness picks the person who looks most like the culprit but in showups, she decides whether the suspect is or is not the culprit. Other explanations are possible, however. One argues that the differences are due to the fact that witnesses who see a showup are simply less willing to identify someone because they realize that there is a greater chance of a mistake being undetected. In addition, in the real-world, showups often occur a very short time after the crime has occurred when all of the cues are still fresh in the victim's mind. Lineups, on the other hand, may occur days and even months after the crime.

### Cross race effects

Loftus (1979) and others (Malpass and Kravitz, 1969; Wells, Lindsay and Ferguson, 1979; Yarmey, 1979) have suggested that cross-race identifications tend to be less accurate than same-race identifications. Despite the apparently intuitive nature of this conclusion, an early review of the cross-race literature (Lindsey and Wells, 1983) suggested a) that research outcomes are far less consistent than defense experts typically imply, b) that even if we accept the defense conclusion that cross-racial identifications tend to be less accurate than within-race identifications, the size of the effect is small, c) that the size of the effect may depend on the experience of the witness with the other racial group, d) that the research methods used to study cross racial identification lack forensic relevance, and e) that even if these threats to the forensic relevance of the research did not exist, it would be clear that no generally accepted theory exists to explain the results (e.g., Ng and Lindsey; 1994).

Several meta-analyses of the literature (e.g., Anthony, Cooper and Mullen, 1992 and Bothwell, Brigham and Malpass, 1989) have concluded that the evidence for a cross-race effect (for Black and White racial groups only) has increased in consistency since Lindsey and Wells completed their review. However, several potential limits on the external validity of these results have not been carefully examined. One is whether the cross-race effect emerges in hits, false alarms, or both types of responses. If the problem is that similar-race faces are easier to learn, then defense experts need not warn jurors that cross-race identifications are likely to be wrong, however, prosecutors might be concerned that cross-race criminals are not getting arrested. Another limitation concerns the fact that cross-race effects have not been examined under a wide range of levels of other factors that might easily moderate the size of the effect. For example, in the Anthony, et. al. review, the longest duration of exposure to

a face was less than 10 seconds. One might expect the cross-race effect would decrease with increasing strength of memory for faces. Would the cross-race effect disappear if the subjects were to have the opportunity to view each face for, say, two minutes?

Even if there is a tendency for people of one race to be better at identifying people from their own race at all durations of exposure, it is unclear how a jury might use this information to help them decide in a particular case whether a witness's identification is or is not correct. We do not know what it is about the 'other race' that makes them less likely to identify correctly. What about light-skinned blacks? Would Caucasians respond to them more like dark-skinned blacks or more like other Caucasians? What about darker-skinned Hispanics. Are they better at identifying darker-skinned blacks than light-skinned Hispanics?

Whatever the answer to these questions, it is important to realize that the existence of a cross-race effect does not mean that cross-race identifications are inaccurate, only that they would be less accurate than within-race identifications.

### Time estimation

One finding that has consistently emerged in our own simulation studies (Ebbesen, Konecni and Moore, 1989) is the tendency of subjects to over-estimate the duration of short exposure durations. Although the only evidence for this result comes from our and other simulation studies, the consistency of the finding increases the odds that witnesses to actual crimes will show the same kind of effect.

Loftus and others (Loftus, 1979; Wells and Loftus, 1984) have argued that this tendency should be made clear to jurors because they might overweigh a witness's report of the exposure duration when judging the credibility of the witness's identification. Thus, a witness who testifies that an event took 1.5 minutes might actually have been exposed for only 0.5 minutes. Although simulation research suggests that this may frequently occur, it does not follow that the jurors are being misled about the identification accuracy if they remain uninformed of the conclusion. Loftus's argument rests on the assumption that a difference in exposure duration between 0.5 minutes and 1.5 minutes will have a large effect on the false alarm rate. We have already pointed out that the empirical evidence is inconsistent on this issue, with a tendency towards diminishing accuracy returns with increasingly longer exposure intervals. Even if witnesses overestimate the duration of exposure, they may well be as accurate in their identifications (in terms of false alarms) as if they had been exposed for the longer time period.



Furthermore, some research suggests that witnesses tend only to overestimate shorter durations. For example, in Cutler, Penrod and Martens, (1987b), the slope of the relationship between actual duration of exposure and estimated duration was less than one, suggesting that at exposure durations above two minutes witness estimates might begin to underestimate actual durations.

### The problem of interactions

As we have implied throughout the previous sections of this article, the relationship that one finds linking particular factors and the accuracy of eyewitness memory seem to depend on the levels of other factors. Thus, the relationship between stress and accuracy may depend on the complexity of the memory task or the type of stress or the method of measuring memory. To the extent that the nature of the relationship between a particular factor and memory is affected by the level of other factors, the kinds of conclusions that can be reached and presented to jurors about the impact of any given factor must be conditioned on the specific range of procedures, tasks, subjects, settings, and measures that were actually used to study the effect of the factor(s) of primary interest. If we know that the effects of some factors, stress for example, are believed to be sensitive to the level of other factors, such as complexity of the memory task, is it not reasonable to suppose that factors yet to be examined empirically, but always present at some level, might also interact with the factors of interest? If interactions are as common as we believe the previous review suggests, then from a scientific point of view, testimony about the effect of a given factor on memory should be admitted only if supported across a wide variety of different methods, procedures, subject types, measures, motivational conditions, etc.. However, such a requirement puts a considerable burden on judicial expertise during pre-trial motions concerning the admissibility of eyewitness expert conclusions.

Similarly, to the extent that interactions among standard eyewitness memory 'factors' exist, should not the admissibility of expert testimony about these factors be conditioned on full disclosure of those interactions to the jury? How do stress, unconscious transference, confidence, retention interval, exposure duration, lineup fairness, racial similarity, and weapon focus combine to affect memory? Are shorter retention intervals sufficient to eliminate the effects of unconscious transference? Will longer and repeated exposures increase the strength of the relationship between confidence and accuracy and will the size of that increase depend on whether a weapon was present? Obviously, the range of combinatorial questions is very large indeed; and this, in part, probably explains why not much is known about how these factors do interact. But if not much is known, one wonders how defense experts can draw the sweeping conclusions that they do? And if more were known about the nature of

these interactions, would the accuracy of jury decisions be improved by such knowledge?

### The problem of measurement

The external validity of conclusions about eyewitness memory depends not only on the consistency of the results but also on measures that are used to define the variables over which researchers are going to generalize. Without agreed-upon measures of both the independent and dependent variables, the possibility arises that defense experts and jurors will 'over-generalize' from the particular measures used in research to what happens in typical crimes. For example, researchers do not even agree how best to measure confidence. Some researchers use 5 point scales labeled "just guessing" at one end and "very confident" at the other. Other researchers use three point scales. Still others use 10 point scales. Sometimes one endpoint is "willing to testify in court" other times it is "absolutely confident." Interestingly enough, witnesses in real cases rarely fill out confidence scales when they make identifications. Instead, observers infer confidence from witness descriptions such as, "I'll never forget those eyes," or "That's him. That's definitely him," or "It looks like him, but I can't say for sure. If I could see him in person, then I'd know." How do we translate these descriptions into a 10 point scale of confidence? Even more interesting are situations in which witnesses provide what appear to be conflicting confidence estimates, such as one we heard in a recent case: "That one looks the most like the person who molested me, but he has gained weight," and then a few minutes later, after looking at the person for some time says, "I'm 100% sure that's him."

This problem of measurement is not limited to confidence. Agreed upon measures do not exist for almost every concept or factor that has been studied in the field. Experts do not even agree how to measure identification accuracy. For example, researchers have not standardized the degree of similarity that should exist between the suspect's picture and the suspect, not to mention the suspect's picture and pictures of foils used in a lineup (Wells, Seelau, Rydell and Luus, 1994). How much should the suspect's picture have to look like the suspect? After all, we can all agree that people can be made to be look quite different with different lighting, different camera angles, and so on. Surely, a witness with a good memory will be more likely to pick the culprit, the more the culprit's picture looks like the culprit. In fact, all of the problems in constructing voice samples described by Hammersley and Read (1996) that arise when trying to compare the results from different earwitness identification studies apply to the selection of people and pictures in eyewitness identification studies. We simply do not have an agreed-upon system for the construction of lineups.

Furthermore, without agreed-upon measures of the factors claimed to affect eyewitness accuracy, experts do not

know how to 'assess' (expect by intuition) the situations in real crimes to know how stressed, how 'cross-raced,' how distracted, how pressured, and so on a given criminal situation is likely to make a witness. Without such information, it is virtually impossible to translate the conclusions from experiments (even if the results were consistent) to predictions about the accuracy of witnesses to a real crime. And if the experts can not do so, how can the court expect jurors to do so?

Courts do not seem to understand that experts in other sciences are able to go from the general to the specific only by measuring attributes in the specific case and using the results of those measures and general theory to draw conclusions about the specific case. For example, a serologist can use specific measures of the presence of DNA markers and generally accepted theory about the distribution of those markers in the population to draw a conclusion about whether a particular blood sample came from the defendant. What is the generally accepted measure of stress that would allow a jury to infer a given witness's accuracy from the type of general theories of memory that we have seen are available to eyewitness experts?

### **Fundamental misunderstandings about psychological research**

One of the main goals of experimental psychologists who study human memory is to discover which factors influence memory, especially detrimentally. To obtain evidence about whether a particular factor might have an effect, researchers use what Pachella (1986) reminds us is the "fixed-effect" model. Two or a few different levels of the factor of interest are constructed and their effects on some measure are examined holding all other things constant, each at their own fixed levels. The choice of levels at which to hold all other factors constant is often arbitrary or based on ease of collecting data. The choice of levels for the factor of interest is also arbitrary with the exception that the researcher attempts to choose levels that are different enough from each other that weak causal effects might be observed even if the chosen levels do not represent those that frequently occur in the real world.

Although this may seem like a reasonable way to do science, it has long been known that the fixed-effect approach does not provide information to the researcher about the robustness of a phenomenon nor about its ubiquity (Campbell and Stanley, 1966; Ebbesen and Konecni, 1980). That is, knowing that there is some set of conditions under which a particular causal result will be found does not tell us how susceptible that causal relationship is to the modifying influence of other factors and phenomena nor does it tell us how often the particular set of conditions in which the phenomenon was observed occurs in the real world.

Because human memory is a function of a large number of factors all of which can have (interactive) effects at the same time, the fact that most research conclusions depend on the fixed-effect model adds to the already discussed concerns about how juries can use our discoveries to help make more accurate decisions about a defendant's guilt. The defense is asking the jury to use the expert's general claims about directional effects of different factors on memory. For example, the 'facts' that more stress, shorter durations, dissimilar-race, longer retention intervals, biased lineups, prior exposure to mugshots, and so on, tend reduce accuracy should help decide the reliability of a particular witness's memory. But how should the jury use these facts when each comes from a small number of studies that have sampled only a very small range of levels not only of each of these factors, but also of all of the other factors that affect memory (e.g., length and nature of rehearsal, context effects, depth of encoding, intelligence, and so on)? Since fixed effect research provides so little information about ubiquity, one wonders whether jurors would be able to make more diagnostic decisions were they simply told our best guesses about the rate of mistaken identifications in past real world cases.

Another problem arising from the extensive use of the fixed-effect approach is that there is nothing in the list of factors that tells jurors how to balance a particular level of one factor against the level of another factor. How long does the duration of exposure have to be before it can overcome the detrimental effects of a longer retention interval?

### **The problem of individual differences**

Because experimental psychologists are looking for general principles, they often ignore differences between people and, unlike the medical sciences, rarely report results in the form of the percent of subjects for whom the crucial factor produced the predicted effect. However, knowledge of the latter statistic is far more important in trying to assess the reliability of a given witness than the fact that a given factor can have a causal effect on some people. For example, let us assume that the defense expert's description of the current state of research in the field is correct and that stress has been consistently shown to decrease memory performance. The fact that experimenters report results suggesting that stress causes memory impairment in no way tells the defense expert or anyone else what percentage of the population showed the effect nor what types of individuals were most susceptible to the effect. Are Type-A people more or less susceptible than Type-Bs? Are people high in achievement motivation more or less susceptible? Are some types more likely to be stressed by crimes than others? What role does IQ play? How about race? And so on. The point is that quite large individual differences are not incompatible with finding consistent causal effects with the fixed-effect research strategy.

It might even be the case that individual differences are far more potent determinants of memory performance than such things as the weapon effect, stress, and so on. After all, even in those conditions which produce low average memory scores, some individuals do very well, or as well as, if not better than, the average of those subjects in the conditions producing higher average memory scores. How is the jury going to know whether the personality, training, and background of the witness is important? Such issues have been infrequently studied by memory researchers and seem to be ignored by most defense experts.

The possibility of individual differences naturally raises concerns about the relative diagnostic role of measures of witness behavior and measures of situations (assuming the field could agree on some). Would a jury be better off knowing information about a particular witness (his confidence in his identification, his memory for other aspects of the case, his willingness to be swayed under cross examination, the latency of his answers, the reasons for his identification), or information about the effects that specific levels of circumstances that were present during and after witnessing have on 'typical' witnesses (the duration of exposure, the retention interval, the number of bystanders, whether a mugshot was seen)? Clearly, this is a complicated issue about which psychologists have had little to say despite several discussions of the *Neil v. Biggers* (1972) criteria (e.g., Wells and Murray, 1983). Nevertheless, the issue is crucial for a full understanding of the position we are taking in this paper. In particular, many defense experts argue that identifications by witnesses to real crimes are not to be trusted because it is known from simulation research that high stress produces lower accuracy. This argument rests on the assumption that the witness was very highly stressed by the crime (as well as that the task of remembering what someone did seems complex). However, no evidence is presented in court, other than a description of the crime, about the stress that a particular witness experienced. No measurements are even made of the average stress that the crime situation in the particular case causes in individuals, in general. The expert merely claims that crimes are stressful and that memories of highly stressed witnesses are less trustworthy. Even assuming that stress measurements could be taken, it is important to ask whether a measure of, say, the witness's confidence in their identification is a better predictor of identification reliability than knowledge of the relationship between stress and memory and measures of the circumstances of the crime scene or self-reports of stress.

### Is juror knowledge about eyewitness memory accurate?

Many defense experts argue (especially in motion hearings to have their testimony admitted) that (a) jurors misunderstand the way eyewitness memory works

(Yarmey and Jones, 1983), (b) if left uncorrected, jurors will draw incorrect conclusions about the accuracy of witness memory (Wells, et al., 1984), and (c) defense testimony about general principles of human memory, i.e., an introductory lecture on memory, will be sufficient to eliminate juror misunderstandings (Cutler, Penrod and Dexter, 1989; Cutler, Penrod and Dexter, 1990; Cutler, Penrod and Stuve, 1988; Kassir, Ellsworth and Smith, 1989; Penrod and Cutler, 1987; Wells, Lindsay and Tousignant, 1980). These arguments were used to convince the California Supreme Court (in *People v. McDonald*, 1984) to allow eyewitness testimony by 'experts' routinely into the court. Some of the evidence to support the first premise of this argument comes from studies that ask potential jurors what they believe about the effect that different factors, e.g., stress, racial differences, etc., have on memory and then compare these beliefs to the known results of 'scientifically' valid experiments.

The argument that jurors' knowledge of factors that affect eyewitness memory can be tested by asking them a few questions about their beliefs in this area and comparing their answers with what some experts say is correct (based on their understanding of the research) is fundamentally flawed. In order to accept the argument that jurors are misinformed about stress, cross-racial factors, and so on, one has to believe that the experts are correctly informed. We have tried to explain why the defense expert view may well be incorrect. Obviously, if we are right, then one has no idea whether jurors are misinformed or actually even better informed than 'experts.'

Even if the former argument is ignored, another feature of the methodology used to assess juror knowledge is defective and misleading, perhaps deliberately so. Stated simply, the results obtained from the questionnaires used in this research may depend as much on the way in which the questions are worded and the set of response alternatives that are offered as they depend on juror knowledge. For example, a look at most of the questions used in this research shows that respondents are never offered the most obvious answer: "It depends." That is, the survey may ask about the memory of two people, one who sees a gun and one who does not. The choices never include an option such as: Whether the man who sees the gun will remember things better than the man who does not see the gun depends on who is more intelligent, what is being remembered, who tries harder, how long they have to look, what racial groups they are from, whether they are sleepy or not, and so on.

In addition, the surveys never ask questions such as: Imagine that one hundred witnesses saw the following event: a man drives up to an all-night gas station, talks with the attendant for a minute or two, pulls out a gun, after a brief but angry verbal exchange takes all of the money from the cash register, and drives away. How many of the witnesses do you think will correctly identify

the defendant from a photo line-up 10 months later? How many would falsely pick an innocent suspect from the lineup with sufficient certainty to be willing to testify in court?

One reason that such questions are not asked might be because 'eyewitness experts' have no idea what the correct answer to these types of questions are.

Another line of research (Cutler, Dexter and Penrod, 1989; Cutler, Penrod and Dexter, 1990; Goodman and Loftus, 1988; Lindsay, Lim, Marando and Cully, 1986; Lindsay, Wells and O'Connor, 1989; Wells and Lindsay, 1983; Wells and Turtle, 1987) attempts to show that simulated jurors make more 'accurate' decisions about witness testimony if they have heard the testimony of a defense expert than if they have not. Several of these studies share a common design. Subjects observe one of several simulated crimes and then testify about what they saw and make identifications. Some subjects see a criminal with a weapon, others do not. Some testify after a long delay, others after a short delay. Their testimony is videotaped. The experimenters then show simulated jurors videotapes of the subject-witnesses and ask the 'jurors' to judge the witness's accuracy. Finally, half of the jurors hear expert testimony about factors that affect eyewitness identification and half do not, before they reach a decision. Although the results from different studies are not entirely consistent, many experts believe that these studies show that jurors make 'better' decisions after hearing the expert testimony because jurors' judgments of witness-accuracy are influenced more by the 'proper' factors after testimony. Thus, defense experts argue that jurors become more accurate in their judgments because they learn to ignore factors that are not associated with witness accuracy (e.g., confidence) and concentrate on those that do (e.g., stress).

Unfortunately, these studies share a common logical error. In particular, they assume that experts know the relative predictive utility of different factors compared to each other as well as to other aspects of a case when in fact they do not. The decision task facing the jurors is not whether the witness is correct but whether the defendant is guilty (Loh, 1981, 1984). To make this decision, jurors need to consider such things as the number of witnesses who identify the defendant, corroborating evidence linking the defendant to the crime, as well as other witness evidence (e.g., alibi witnesses, testimony by police about the behavior of the defendant when arrested, and so on). It might be very reasonable for witnesses to ignore stress, ignore racial dissimilarity, ignore the presence of a weapon, and concentrate on the fact that two witnesses independently picked the same person from the lineup and that the person they picked has no alibi. What is needed are studies that compare the guilt diagnosticity of witness identifications (produced under a wide range of circumstances) with other indicators of defendant guilt, e.g., multiple IDs, nature of arrest, corroborating evidence,

etc. Since there are no studies like these, the conclusion that expert testimony helps jurors may only apply when the jurors only hear one witness who cannot remember much about what she saw.

## The relevancy test

In our experience defense eyewitness experts do not testify about the same set of factors in all cases. That is, they tailor their testimony to some extent to the facts of the case. For example, the 'cross-race effect' has frequently been cited by experts who have been called by the defense to testify about factors affecting eyewitness accuracy but only when the defendant and the witness(es) have been from different racial groups. Although we have tried to show that a general theory of cross-racial identification effects does not exist, the claim is always made that the empirical evidence suggests that cross-race identifications are less accurate than within-race identifications. However, when the defendant and witness are from the same race, the experts (and some courts) seem to believe that data supporting the cross-race effect are irrelevant. For example, in one case with which we are familiar (*U.S. v. Sebetich*, 1985), the witness and the perpetrator were from the same racial group. As such, the cross-race conclusion implies that the reliability of the identification should be higher than were the witness and perpetrator from different racial groups. Nevertheless, both defense experts called to testify in that case never mentioned the absence of the 'cross race effect' despite the fact that one of them had testified in many other cases in which he extensively discussed this 'cross race effect.'

The analysis is relevant to an issue raised by a federal district court (*United States v. Downing*, 1985) in deciding whether expert testimony about eye-witness reliability was to be admitted. In particular, in *Downing*, the court argued that admission of the expert depends on a proffer that the testimony of the expert will focus on particular characteristics of the eyewitness identification at issue and how those factors might affect the reliability of the identification. How does an expert (and the court) decide which factors are relevant to the case? Should the expert focus only on those aspects of the identification process that might decrease reliability even though other factors might exist in the case that would completely negate their effects? An affirmative answer to this question substantially increases the probability that expert testimony will be more prejudicial than probative because almost any given witnessing situation will contain levels of some factors that make reliable identification more likely than it might be in other witnessing situations (assuming that we accept as relevant the research evidence typically relied on by defense experts).

The fixed-effect nature of eyewitness research is important here, as well. For example, in a cross-race study the accuracy of Caucasian witnesses who are identifying

blacks might be compared to the accuracy of Caucasian witnesses who are identifying Caucasians. Assuming that a difference in accuracy between the two groups exists, racial dissimilarity becomes a potential candidate as a relevant factor in real witnessing situations. But, note that every study that finds that cross-race accuracy is lower than within-race accuracy did so by comparing two conditions. Is it reasonable for a defense expert to suggest that racial similarity is a relevant factor only in those real world instances in which a witness's race is different from the defendant's race? After all, the expert must know that accuracy is higher in within-race identifications as well as that accuracy is lower in between-race identifications. If racial similarity affects accuracy, then regardless of the racial similarity between witness and defendant, racial similarity should be considered a relevant factor.

This logic suggests that all factors that are known to affect eyewitness reliability might be considered relevant whether the actual level of the factor seems to suggest reliability will be higher or lower. If this logic conflicts with the *Downing* test it is because the kind of evidence on which defense experts base their conclusions is inconsistent with the *Downing* criterion and not because the *Downing* criterion is unusable. In general, results from experiments do not supply relevance information. They merely tell us whether a change in the level of a factor is associated with a change in eye-witness reliability. They do not tell us how high a given factor needs to be before a reduction or increase in witness reliability is seen. They do not tell us when several factors might negate each others' effects. They do not tell us the odds of witnesses in similar situations being correct or incorrect.

The idea that experts should limit their testimony only to factors or issues that are relevant to the facts of the case makes considerable sense for expert testimony that refers directly to specific facts that are in contention in the case. For example, a ballistic expert might testify about whether the bullet found embedded in a car door was fired from a gun found on the defendant but would be prevented from testifying about theories of how powder burns are formed. Or a serologist might testify about the type of blood found at the scene of a stabbing and whether it matched the defendant's blood but would not talk about how semen is typed. Unfortunately, for two reasons, the eyewitness expert is not required to testify only about relevant case facts. First, US courts have quite clearly resolved that eyewitness experts are not allowed to testify about the capacity of particular witnesses to make correct identifications. This means that an eyewitness expert, "informs the jury of certain factors that may affect such an identification in a typical case; and to the extent that it may refer to the particular circumstances of the identification before the jury, such testimony is limited to explaining the potential effects of those circumstances on the powers of observation and recollection of a typical eyewitness." (People v. McDonald (1984), 371). Second, it seems that in

practice, it is the expert rather than the court who decides what factors he or she will testify about and since the expert is hired by the defense, the factors are almost always selected and discussed in a manner to make witnesses appear inaccurate. In a witnessing situation involving good lighting, a long exposure duration, a very brief retention interval, high stress, and a weapon, most defense experts will testify only about the detrimental effects on accuracy of the latter two factors.

It might be argued that experts called by the defense should not be expected to testify about factors that increase the odds that witnesses will be correct in their identifications. After all, prosecutors can call their own eyewitness experts who could lecture the jury about those case factors that might increase the odds of the witnesses' identifications being correct. On the other hand, if courts continue to use 'relevance' as an admissibility criterion, one would hope the concept would be better defined than it has been in the US appellate courts.

### Hypotheticals and testimony about facts of case

In *United States v. Smith* (1984) the court suggested that eyewitness experts might be able to help jurors understand the facts of a particular case by analyzing the role that specific perceptual and memory factors would play in a factually identical hypothetical situation. For example:

*In the hypothetical, three witnesses were shown a line-up containing the same defendant and four months later they were shown a photospread containing the same defendant. The defendant was the only "common" element in each showing. Dr. Fulero offered that the later line-up was not "independent" of the earlier photospread and that the eyewitnesses "incorrectly transferred" the "familiar" figure from one procedure to the next. What they identified was the picture of the defendant at the earlier photospread, not the figure of him at the bank. (p. 1110)*

That the court accepted the expert's testimony about the hypothetical as if he was testifying about the facts of the specific case is astonishing, especially given the generally accepted idea that the expert is to refrain from testifying about the capacity of particular witnesses in the instant case. How could the expert conclude that such witnesses would 'incorrectly transfer' the 'familiar' figure, without first knowing whether the witnesses were correct in their initial choices? Equally important, one wonders how the expert was able to weigh the combined impact of all of the circumstances and prior experience of the hypothetical witness in reaching his conclusion since virtually no research has examined the combined effects of even three factors at one time.

Now imagine the following hypothetical facts were described to a defense expert:

*The event takes place in a well-lit store. The distance between the victim and the culprit varies from four to twenty feet throughout the course of the robbery. The victim recalls looking up from behind the counter and noticing a man standing in front of the magazine rack. She sees his face, but only for a few seconds. A brief time later, the man approaches the counter and says to the victim, "How's business today?" A brief, but pleasant, exchange about work takes place and then the robber says, I hate to tell you this but this is a robbery." He then opens his coat and shows the victim that he has a gun and says, "I will kill you unless you do as I say." Next the robber closes his coat but leaves his hand inside, presumably on the gun that is now hidden by the coat. The victim takes money out of the cash register and fills a paper bag with the money saying, "I will do as you say. Please don't hurt me." After all of the money is in the bag, the robber takes it and says, "Sorry to have to do this to you, but I needed the money. Don't call the cops until I leave." After the robber leaves, the victim calls the police. They arrive about five minutes later. In the interview, the victim says the robber was about 5 feet 8 inches or about the same height as the officer, the robber was black (she is white), she thinks the robber may have had a mustache but is not sure because she only remembers looking him in the eyes. She thinks his coat was light tan. She remembers his hair was on the short side. He was more muscular than normal. She cannot recall his pants and she does not recall any distinguishing features. She says that she was very upset. She thought she might be killed. She thinks that she would be able to recognize the man again because she looked him right in the eyes to see if he really intended to kill her. She estimates the entire event took about four minutes. The police radio in the description. About ten minutes later, additional police arrive at the scene with a black man who is 5'11", muscular, is wearing a tan coat and has a three day stubble, but no distinct mustache. When the arresting officers attempted to detain this man, about five blocks from the scene, he started to run into an alley where they lost sight of him for about 30 seconds.*

After being told by the police that they had someone in their car who may or may not be the robber and that it is just as important to tell them if the man is not the robber as it is to tell them that he might be the robber, the woman looks at the man for about 30 seconds and says, "He looks like the guy; yes I think that's him; no, I am sure that is him. Yes, that's him. I can tell by his eyes."

Then the expert is asked by the defense attorney, "If 100 randomly selected people experienced exactly the same thing as described in this hypothetical, using everything you know about eyewitness memory research, what percent of them would make an inaccurate identification?" We cannot imagine how a psychologist (or anyone else) who knew all of the research in the field could possibly answer such a question (although we have heard defense experts give estimates to similar questions). The problem

facing the expert becomes more obviously difficult if we now ask, "By how much would your estimate of the percentage of errors change if the victim had confidently described the robber as having a stubble instead of possibly having a mustache?" Alternatively, suppose the victim said that she was not that upset by the gun because he looked so embarrassed about what he was doing, that she felt kind of sorry for him. And so on. If experts were to provide answers to these questions, it would be of interest to ask how they accomplished the task. That is, how did they compute the odds? What principles of eyewitness memory allowed them to compute a probability estimate? Did they use the error rates obtained from witnesses in studies of simulated crimes that were identical or nearly identical to the hypothetical? Unlikely, since so few simulated studies actually involve victims of crimes and, at the time of this writing, no study has examined eyewitness accuracy of subjects who were told they might be killed. What types of errors did they consider, misses and false positives or just the latter? And when thinking about the hypothetical, did they assume that the police had arrested the culprit or the wrong individual and how did they decide this? Would their 'scientifically based' estimates of the identification error rate change if, in the hypothetical, a brown paper bag that matched those used by the store was found on the suspect when he was arrested? If the suspect had a gun in his possession?

Extrapolation from the witnessing conditions and settings used in experiments to the conditions described in the hypothetical can only be based on the defense expert's intuitions because scientific rules for extrapolation do not exist in these research domains. The main reasons that such rules do not exist are a) researchers in the eyewitness area have not developed measures of the variables about which they theorize that can be used to assess the strength of a particular witness's memory and b) studies of memory in isolation of other 'identity' evidence fail to recreate the decision problem facing the jury. Whether a gun is found at the scene of the crime with a finger print that matches the defendant has implications for the accuracy of a witness's positive identification of the defendant. Such factors have never been studied in eyewitness memory research because they do not involve *memory*, despite the fact that they might reasonably be expected to affect the odds that the victim's identification is an accurate one. It is quite possible, therefore, that defense expert testimony (especially in the form of hypotheticals) that focuses on how eyewitnesses have poor memories will cause jurors to underweight other evidence indicating that the witnesses in the particular case have made correct identifications. Clearly, this is exactly what most defense attorneys hope will happen when they seek the help of eyewitness experts.

Interestingly, unlike forensic clinical judgments (Faust and Ziskin, 1988; Dawes, 1994), we know of no research on the accuracy of predictions made by eyewitness experts about

witness reliability. In particular, many studies have evaluated the accuracy of clinical psychologists' and psychiatrists' diagnoses and judgments. These studies provide information to the clinicians about patients and then compare the judgments of different clinicians to each other or to some objective standard. As Faust and Ziskin (1988) reported in an extensive review of this literature, clinical judgment is not particularly accurate. The same research strategy could and should be used to test the ability of eyewitness experts to select accurate from inaccurate witnesses. That is, eyewitness experts could be given descriptions of actual witnessing situations and asked to make judgments about witness accuracy. Those judgments could then be compared to other experts or to actual witness performance. If the research on which experts base their judgments is reliable and valid and if it provides them with a better sense of witness accuracy, then experts should be more accurate in their judgments of witness accuracy than uninformed jurors. Until such research is done, the claim that expert witnesses would be better at estimating the odds of witnesses being correct in hypothetical examples than lay individuals seems unwarranted.

### Is there a generally accepted theory?

The question of whether a generally accepted theory of eyewitness memory exists has frequently been used by the courts as a criterion in deciding whether to admit the expert's testimony. Interestingly enough, this question is difficult to answer because it is not clear which theory of memory underlies the experts' opinions. There are many theories of memory currently being considered and discussed in psychology journals. Some speak about spreading activation, some about parallel distributed processes, some about visual images and auditory traces, some about short and long term storage, some about feature lists, some about schemas, some about reconstructive processes, some about implicit and explicit forms of memory, some about episodic and semantic forms, some about specific areas of the brain (e.g., the hippocampus), some about specific neurotransmitters, and so on. No theory combines these into a single coherent whole. In addition, most of the factors that experts claim affect eyewitness memory are not, in general, related to these theories. Instead, various authors have proposed very narrow ideas to explain particular findings in the eyewitness area, e.g., the effect of a weapon on identification accuracy, or the reason confidence and accuracy are sometimes unrelated, without attempting to relate these specific explanations to a broader, more comprehensive theory of human memory. In short, it can be safely said that there is no generally accepted theory that attempts to link all of the effects of these factors together to explain how eyewitnesses remember (or forget) things.

From a different point of view, inconsistent theories about how various factors affect witness reliability is not limited to psychologists. The courts have not presented a thoroughly consistent view either. For example, the thread of court decisions dealing with due process violations introduced by suggestive procedures in eyewitness identification terminated with the U.S. Supreme Court case, *Neil v. Biggers* (1972), which clearly defines a set of criteria that are to be considered in determining accuracy: "(a) The opportunity of the witness to view the criminal, at the time of the crime, (b) the witness' degree of attention, [c] the accuracy of the witness' prior description of the criminal, (d) the level of certainty demonstrated by the witness at the time of confrontation, (e) the length of time between the crime and the confrontation." (p.199) Yet, these are not the only factors about which experts normally testify, nor does the testimony have to agree completely with the assumptions in *Neil v. Biggers* about the effects that these five factors have. In particular, experts testify about factors that were not discussed in *Neil v. Biggers*, such as, cross-race effects, unconscious transference, stress, the witness's need to be consistent or appear intelligent, line-up fairness, and post-exposure memory effects, just to name a few. In addition, experts frequently testify that the level of certainty demonstrated by the witness at the time of confrontation is *not* diagnostic of higher accuracy, quite contrary to *Neil v. Biggers*. Finally, several courts seem to have accepted the defense expert testimony as scientifically established, despite their disagreement with *Neil v. Biggers*. In short, it seems clear that a generally accepted theory of how witnesses remember things is not available, either among psychologists or among courts.

On the other hand, a paper by Kassin, Ellsworth and Smith (1989) seems to suggest that the opinions we have expressed in the present article regarding the failure of research results to provide consistent support for a number of very specific defense conclusions is a minority opinion among 'eyewitness experts.' In particular, these authors report the results of a survey of 63 'experts,' almost all of whom had a Ph.D. in psychology. Three quarters of the respondents reported authoring at least one publication dealing with eyewitness testimony and 54 percent indicated they had testified in court at least once on eyewitness matters.

It is of considerable interest, however, that the 'experts,' collectively, were asked to testify almost 20 times more often by criminal defense lawyers (a total of 1063 times) than by prosecuting attorneys (56 times). Although Kassin, Ellsworth and Smith (1989) note that the respondents were equally likely to agree to testify when asked by the defense as when asked by the prosecution, the fact that defense lawyers are so much more likely to ask is completely consistent with the type of testimony that most 'experts' are willing to give and have given, namely testimony about factors that supposedly *reduce* (rather than enhance)

the reliability of eyewitness testimony. Kassin, et. al. use these same facts to conclude that "there is no evidence to support the fear that eyewitness experts are inherently biased in favor of criminal defendants," but rather that the disproportionate interest shown by defense lawyers in their testimony is due to the fact that prosecutors introduce eyewitnesses whom the defense lawyers try to counter. What is left unsaid is why defense lawyers are so interested in what the experts have to say. The answer is obvious. The testimony of the experts almost always consists of an attack against the reliability of eyewitness testimony. Whether this reflects 'inherent bias' or merely *de facto* bias is unknown; the fact remains that experts almost always tell juries why eyewitness testimony is not to be trusted.

The reasons for expert willingness to testify about factors that threaten the reliability of eyewitness memory may have little to do with the whether experts generally agree about how memory works and more to do with other things. In particular, the huge majority of conclusions in journal articles about memory are phrased in terms of memory deficits rather than memory enhancements, not because human memory is so bad, but because perfect performance is less informative than patterns of errors when trying to discover how a system (any system, including memory) works. Second, the growth in financial incentives for eyewitness testimony is much greater for the defense than the prosecution. Prosecutors are not likely to call an expert to explain to the jury all of the factors that increase the odds that a confident witness is correct. They will simply rely on the testimony of the witnesses. On the other hand, every case that includes identification testimony by a prosecution witness is an opportunity for experts who are willing to attack witness reliability to earn high fees. For example, we have heard some defense experts testify that they charge as much as \$2,500 per day for their time. Third, as Pachella (1986) noted, most academics may hold personal values consistent with the belief that the legal system leans too far to the side of the prosecution and therefore assume that many innocent defendants are being found guilty by a corrupt system, even if they lack direct evidence about the actual rate of innocent people being found guilty (Konecni and Ebbesen, 1986). In short, the mere fact that most experts have testified for the defense rather than the prosecution offers little in the way of evidence that their testimony represents the truth about how eyewitness memory works.

On the other hand, the results of the Kassin, Ellsworth, and Smith (1989) survey suggest that the majority of respondents agreed with many "specific" claims about eyewitness memory. For example, one claim with which over 80 percent of the respondents agreed was that the accuracy of memory drops off rapidly after the event and then tends to level off. Although this is not a theory about memory, it is a claim that is frequently made by defense experts in an attempt to warn jurors that they should have

less faith in identifications made just several weeks after a crime than those made within hours. Unfortunately, although the general conclusion about forgetting may be correct, the implication that jurors should have less faith in positive identifications after a delay does not follow from that conclusion. The former says that average memory for things gets weaker with time. The latter is concerned with the probability that an identification is a false alarm. As we noted earlier, there is both sufficient theory and data to suggest that these are not the same. In short, it is possible to agree with many of the general claims about memory made in the Kassin, Ellsworth, and Smith (1989) study, but still feel that the theories have not been developed enough to allow us to generalize to the performance of actual witnesses.

The problem is that the causal hypotheses that Kassin and colleagues tested are *not* theories. They do not provide us with nearly enough information to predict how accurate witnesses will be in particular situations. Even assuming that every single causal hypothesis in Kassin, et al., is empirically correct, they are not stated in ways that allow either experts or jurors to predict the odds that eyewitness identifications will be correct. They cannot even narrow the range of odds that might be expected. All these hypotheses do is describe the sets of conditions under which eyewitnesses tend to be more or less accurate; how accurate or inaccurate is not specified. Therefore, even though it might be argued that some agreement exists among eyewitness 'experts,' the things about which they agree is far from what most scientists would call a theory. There is no way to use this 'knowledge' to predict, with even a moderate degree of precision, how accurate witnesses will be after being involved in any given crime setting. Given these problems, it seems impossible to argue that a generally accepted theory of eyewitness identification exists.

## Conclusions

Research in the area of eyewitness memory, although extensive and interesting to psychologists, lacks the external validity necessary to be useful to jurors when deciding whether a particular defendant is guilty. Despite some court decisions and surveys of some psychologists, there is no direct evidence that the kinds of testimony offered by defense experts about factors that might affect eyewitness memory can improve the accuracy of jury decisions. Because the evidence is either inconsistent or insufficient in almost every area in which eyewitness experts testify and because there is no research that provides the experts, much less the jurors, with rules to use in translating the evidence to particular decisions in particular cases, we believe that eyewitness expert testimony is more prejudicial than probative and should not be allowed in court.



## Authors' note

This document was initially prepared in 1989 in response to a request by the California District Attorney's Association for a position paper about the scientific value of testimony by eyewitness memory 'experts.' It has been updated and revised a number of times as new research has emerged and in the light of the reviewers' comments. Reprint requests may be obtained by writing to Ebbe B. Ebbesen, or Vladimir J. Konecni, University of California, San Diego, Department of Psychology, 9500 Gilman Drive, La Jolla, CA 92093-0109.

## References

- Anthony, T, Copper, C. and Mullen, B, 1992, Cross-racial facial identification: A social cognitive integration, *Personality & Social Psychology Bulletin*, 18: 296-301.
- Bekerian, D. and Bowers, J, 1983, Eyewitness testimony: Were we misled? *Journal of Experimental Psychology: Learning, Memory and Cognition*, 9: 139-145.
- Bosworth, C. and Ebbesen, E.B, 1996, Weapon focus: The effect of weapon locality and reported attentional gaze on accuracy and confidence, Unpublished paper, University of California, San Diego.
- Bothwell, R.K, Brigham, J.C. and Malpass, R.S, 1989, Cross-race identification, *Personality and Social Psychology Bulletin*, 15: 19-25.
- Bothwell, R.K., Deffenbacher, K.A. and Brigham, J.C, 1987, Correlation of eyewitness accuracy and confidence: Optimality hypothesis revisited, *Journal of Applied Psychology*, 72: 691-695.
- Bowers, J. and Bekerian, D, 1984, When will postevent information distort eyewitness testimony? *Journal of Applied Psychology*, 69: 466-472.
- Brigham, L.C, Maass, A., Snyder, L.D. and Spaulding, K, 1982, Accuracy of eyewitness identification in a field setting, *Journal of Personality and Social Psychology*, 42: 673-681.
- Brown, E, Deffenbacher, K. and Sturgill, W, 1977, Memory for faces and the circumstances of encounter, *Journal of Applied Psychology*, 62: 311-318.
- Brown, R. and Kulik, J, 1977, Flashbulb memories, *Cognition*, 5: 73-99.
- Campbell, D.T. and Stanley, J.C, 1966, *Experimental and quasi-experimental designs for research*, Chicago: Rand McNally.
- Christianson, S.A, 1992, Emotional stress and eyewitness memory: A critical review, *Psychological Bulletin*, 112: 284-309.
- Clifford, B.R. and Lloyd-Bostock, S.M.A, 1983, Witness evidence: Conclusions and prospect, (in) Lloyd-Bostock, S.M.A. and Clifford, B.R, (eds.), *Evaluating witness evidence: Recent psychological research and new perspectives*, New York: Wiley.
- Clifford, B.R. and Hollin, C.R, 1978, *Experimentally manipulated arousal and eyewitness testimony*, Unpublished manuscript, North East London Polytechnic.
- Clifford, B.R. and Scott, J, 1978, Individual and situational factors in eyewitness testimony. *Journal of Applied Psychology*, 63: 352-359.
- Crano, W.D. and Brewer, M.B, 1973, *Principles of research in social psychology*, New York: McGraw-Hill, Inc.
- Cutler, B.L, Dexter, H.R. and Penrod, S.D, 1989, Expert testimony and jury decision making: An empirical analysis, *Behavioral Sciences and the Law*, 7: 215-225.
- Cutler, B.L, Penrod, S.D, 1995, *Mistaken Identification: The Eyewitness, Psychology, and the Law*, New York: Cambridge University Press.
- Cutler, B.L, Penrod, S.D. and Dexter, H.R, 1989, The eyewitness, the expert, and the jury. *Law and Human Behavior*, 13: 311-332.
- Cutler, B.L, Penrod, S.D. and Dexter, H.R, 1990, Juror sensitivity to eyewitness identification evidence, *Law and Human Behavior*, 14, 185-191.
- Cutler, B.L, Penrod, S.D. and Martens, T.K, 1987a, The reliability of eyewitness identification: The role of system and estimator variables, *Law and Human Behavior*, 11: 233-258.
- Cutler, B.L, Penrod, S. and Martens, T.K, 1987b, Improving the reliability of eyewitness identification: Putting context into context, *Journal of Applied Psychology*, 72: 629-637.
- Cutler, B.L, Penrod, S.D. and Stuve, T.E, 1988, Juror decision making in eyewitness identification cases, *Law and Human Behavior*, 12: 41-55.
- Dawes, R.M, 1994, *House of cards: Psychology and psychotherapy built on myth*, New York: The Free Press.
- Deffenbacher, K.A, 1980, Eyewitness and confidence: Can we infer anything about their relationship? *Law and Human Behavior*, 4: 243-260.
- Deffenbacher, K.A., 1983, The influence of arousal on reliability of testimony, (in) Lloyd-Bostock, S.M.A. and Clifford, R.B, (eds.), *Evaluating witness evidence: Recent psychological research and new perspectives*, Chichester: John Wiley & Sons.
- Deffenbacher, K.A, Leu, J.R. and Brown, E.L, 1979, *Remembering faces and their immediate context*, Paper presented at the annual meeting of the Psychonomic Society, Phoenix.
- Devlin, Rt. Hon. Lord Patrick, Chair, 1976, *Report to the Secretary of State for the Home Department of the Departmental Committee on Evidence of Identification in Criminal cases*, London: H.M.S.O..
- Dunning, D. and Stern, L.B, 1994, Distinguishing accurate from inaccurate eyewitness identifications via inquires about decision process, *Journal of Personality and Social Psychology*, 67: 818-835.
- Easterbrook, J.A, 1959, The effect of emotion on cue utilization and the organization of behavior, *Psychological Review*, 66: 183-201.
- Ebbesen, E.B. and Boley, S, 1994, A comparison of simultaneous, sequential, and showup lineup procedures, Unpublished technical report, University of California, San Diego.
- Ebbesen, E.B. and Konecni, V.J, 1980, On the external validity of decision-making research: What do we know about decisions in the real world? (in) Wallsten, T.S, (ed.), *Cognitive processes in choice and decision behavior*, Hillsdale, N.J.: Lawrence Erlbaum.
- Ebbesen, E.B, Konecni, V.J. and Boucher, R, 1990, Effects of exposure duration and retention interval on face memory: The mediating role of confidence, Unpublished technical report, University of California, San Diego.
- Ebbesen, E.B. and Wixted, J, 1996, A signal detection analysis of the relationship between confidence and accuracy in face recognition memory, Submitted to *Journal of Experimental Psychology, General*.
- Egan, D, Pittner, M. and Goldstein, A.G, 1977, Eyewitness identification: Photographs vs. live models, *Law and Human Behavior*, 1: 199-206.

- Egeth, H.E, 1993, What do we not know about eyewitness identification? *American Psychologist*, 48: 577-580.
- Elliott, R, 1990, On going public with psychological research: When should we just say no? Paper presented at the annual meeting of the American Psychology Association Meetings, August, Boston, MA.
- Elliott, R, 1991, Social science data and the APA: The Lockhart brief as a case in point. *Law & Human Behavior*, 15: 59-76.
- Elliott, R, 1993, Expert testimony about eyewitness identification: A critique, *Law and Human Behavior*, 17: 423-437.
- Ellis, H.D, Davies, G.M. and Shepherd, J.W, 1977, Experimental studies of facial identification, *National Journal of Criminal Defense*, 3: 219-234.
- Faust, D. and Ziskin, J, 1988, The expert witness in psychology and psychiatry, *Science*, 241: 31-35.
- Fleet, M.L, Brigham, J.C. and Bothwell, R.K, 1987, The confidence-accuracy relationship: The effects of confidence assessment and choosing, *Journal of Applied Social Psychology*, 17: 171-187.
- Giesbrecht, L.W, 1980, The effects of arousal and depth of processing on facial recognition, Dissertation Abstracts International, 40: 4561.
- Goodman, J. and Loftus, E.F, 1988, The relevance of expert testimony on eyewitness memory, *Journal of Interpersonal Violence*, 3: 115-121.
- Gonzalez, R, Davis, J. and Ellsworth, P, 1995, Who should stand next to the suspect? Problems in the assessment of lineup fairness, *Journal of Applied Psychology*, 80: 525-531.
- Gonzalez, R, Ellsworth, P. and Pembroke, M, 1993, Response biases in lineups and showups, *Journal of Personality and Social Psychology*, 64: 525-537.
- Gorenstein, G.W. and Ellsworth, P.C, 1980, Effect of choosing an incorrect photograph on a later identification by an eyewitness *Journal of Applied Psychology*, 65: 616-622.
- Hammersley, R. and Read, J.D, 1996, Voice identification by humans and computers, (in) Sporer, S.L, Malpass, R.S. and Köhnken, G, (eds.), *Psychological issues in eyewitness identification*, Mahwah, NJ: Lawrence Erlbaum Associates, Inc., pp. 117-152.
- Handberg, R.B, 1995, Expert testimony of eyewitness identification: A new pair of glasses for the jury, *American Criminal Law Review*, 32: 1013-1064.
- Heuer, F, and Reisberg, D, 1992, Emotion, arousal, and memory for detail (in) Christianson, S.-A, (ed.), *The Handbook of Emotion and Memory: Research and Theory*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc..
- Hosch, H.M. and Cooper, S.D, 1982, Victimization as a determinant of eyewitness accuracy, *Journal of Applied Psychology*, 67: 648-652.
- Hosch, H.M, Leippe, M.R, Marchioni, P.M. and Cooper, D.S, 1984, Victimization, self-monitoring, and eyewitness identification, *Journal of Applied Psychology*, 69: 280-288.
- Hosch, H.M. and Platz, J.J, 1984, Self-monitoring and eyewitness accuracy, *Personality and Social Psychology Bulletin*, 10: 289-292.
- Johnson, C. and Scott, B, 1976, Eyewitness testimony and suspect identification as a function of arousal, sex of witness, and scheduling of interrogation, paper presented at the annual meeting of the American Psychological Association, September, Washington, D. C.
- Kassin, S.M, Ellsworth, P.C. and Smith, V.L, 1989, The 'general acceptance' of psychological research on eyewitness testimony: A survey of experts, *American Psychologist*, 44: 1089-1098.
- Konecni, V.J. and Ebbesen, E.B, 1979, External validity of research in legal psychology, *Law and Human Behavior*, 3: 39-70.
- Konecni, V.J. and Ebbesen, E.B, 1982, Social psychology and the law: The choice of research problems, settings, and methodology, (in) Konecni, V.J. and Ebbesen, E.B, (eds.), *The criminal justice system: A social-psychological analysis*, San Francisco: W. H. Freeman.
- Konecni, V.J. and Ebbesen, E.B, 1984, The mythology of legal decision making, *International Journal of Law and Psychiatry*, 7: 5-16.
- Konecni, V.J. and Ebbesen, E.B, 1986, Courtroom testimony by psychologists on eyewitness identification issues: Critical notes and reflections *Law and Human Behavior*, 10: 117-126.
- Krafka, C. and Penrod, S, 1985, Reinstatement of context in a field experiment on eyewitness identification, *Journal of Personality and Social Psychology*, 49: 58-69.
- Kramer, D, Buckout, R. and Eugenio, P, 1990, Weapon focus, arousal, and eyewitness memory, *Law and Human Behavior*, 14(2): 167-184.
- Lane, M.J, 1984, Eyewitness identification: Should psychologists be permitted to address the jury? *Journal of Criminal Law and Criminology*, 75: 1321-1365.
- Laughery, K.R, Fessler, P.K, Lenorovitz, D.R. and Yoblick, D.A, 1974, Time delay and similarity effects in face recognition, *Journal of Applied Psychology*, 59: 490-496.
- Leippe, M.R, 1980, Effects of integrative and memorial and cognitive processes on the correspondence of eyewitness accuracy and confidence, *Law and Human Behavior*, 4: 261-274.
- Leippe, M.R, Wells, G.L. and Ostrom, T.M, 1978, Crime seriousness as a determinant of accuracy in eyewitness identification, *Journal of Applied Psychology*, 63: 345-351.
- Lindsay, R.C, Lim, R, Marando, L. and Cully, D, 1986, Mock-juror evaluations of eyewitness testimony: A test of metamemory hypotheses, *Journal of Applied Social Psychology*, 16: 447-459.
- Lindsey, R.C. and Wells, G.L, 1983, What do we really know about cross-race identification? (in) Lloyd-Bostock, S.M.A. and Clifford, B.R, (eds.), *Evaluating witness evidence: Recent psychological research and new perspectives*, Chichester: John Wiley & Sons.
- Lindsay, R.C, Wells, G.L. and O'Connor, F.J, 1989, Mock-juror belief of accurate and inaccurate eyewitnesses: A replication and extension, *Law and Human Behavior*, 13: 333-339.
- Loftus, E.F, 1979, *Eyewitness testimony*, Cambridge, MA: Harvard University Press.
- Loftus, E.F, 1983, Silence is not golden, *American Psychologist*, 38: 564-572.
- Loftus, E.F, 1986, Ten years in the life of an expert witness, *Law and Human Behavior*, 10: 241-263.
- Loftus, E.F, 1993, Psychologists in the eyewitness world, *American Psychologist*, 48: 550-552.
- Loftus, E.F. and Burns, D, 1982, Mental shock can produce retrograde amnesia, *Memory and Cognition*, 10: 318-323.
- Loftus, E.F, Loftus, G. and Messo, J, 1987, Some facts about 'weapon focus,' *Law and Human Behavior*, 11: 55-62.
- Loftus, E.F. and Schneider, N.G, 1987, "Behold with strange surprise": Judicial reactions to expert testimony concerning eyewitness reliability, *University of Missouri-Kansas City Law Review*, 56: 1-45.

- Loftus, E.F., Schooler, J. and Wagenaar, W., 1985, The fate of memory: Comment on McCloskey and Zaragoza, *Journal of Experimental Psychology: General*, 114: 375-380.
- Loh, W., 1981, Psycholegal research: Past and present, *Michigan Law Review*, 79: 659-707.
- Loh, W., 1984, *Social Research in the Judicial Process: Cases, Readings, and Text*, New York: Russell Sage.
- Lloyd-Bostock, S.M.A., Clifford, B.R., 1983, *Evaluating Witness Evidence: Recent Psychological Research and New Perspectives*, New York: Wiley.
- Lykken, D.T., 1991, What's wrong with Psychology anyway? (in) Cicchetti, D., Grove, D.M., (eds.) *Thinking Clearly about Psychology V. 1: Matters of Public Interest*, Minneapolis: University of Minnesota Press.
- Maass, A., 1996, Logic and methodology of experimental research in eyewitness psychology, (in) Sporer, S.L., Malpass, R.S. and Köhnken, G., (eds.) *Psychological issues in eyewitness identification*, Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Maass, A. and Köhnken, G., 1989, Eyewitness identification: Simulating the 'weapon effect,' *Law and Human Behavior*, 13: 397-409.
- Maass, A., Brigham, J.C. and West, S.G., 1985, Testifying on eyewitness reliability: Expert advice is not always persuasive, *Journal of Applied Social Psychology*, 15: 207-229.
- Macmillan, N.A. and Creelman, C.D., 1991, *Detection Theory: A User's Guide*, New York: Cambridge University Press.
- Majcher, L.L., 1974, *Facial recognition as a function of arousal level, exposure duration and delay interval*. Unpublished thesis, University of Missouri.
- Malpass, R.S. and Devine, P.G., 1981, Eyewitness identification: Lineup instructions and the absence of the offender, *Journal of Applied Psychology*, 66: 482-489.
- Malpass, R.S. and Kravitz, J., 1969, Recognition of faces of own and other race, *Journal of Personality and Social Psychology*, 13: 330-334.
- McCloskey, M. and Egeth, H.E., 1983, Eyewitness identification: What can a psychologist tell a jury? *American Psychologist*, 38: 550-563.
- McCloskey, M., Egeth, H. and McKenna, J., 1986, The experimental psychologist in court: The ethics of expert testimony, *Law and Human Behavior*, 10: 1-13.
- McCloskey, M. and Zaragoza, M., 1985, Misleading postevent information and memory for events: Arguments and evidence against memory impairment hypothesis. *Journal of Experimental Psychology: General*, 114: 1-16.
- McGaugh, J.L., Introini-Collison, I.B., Cahill, L.F., Castellano, C., Dalmaz, C., Parent, M.B. and Williams, C.L., 1993, Neuromodulatory systems and memory storage: Role of the amygdala, (Special issue: Emotion and memory, *Behavioural Brain Research*, 58: 81-90.
- Moore, P.J. and Ebbesen, E.B., 1994, Lineups versus showups another look at relative versus absolute judgments, Unpublished technical report. University of California, San Diego.
- Moore, P.J., Ebbesen, E.B. and Konecni, V.J., 1994, What does real eyewitness testimony look like? An archival analysis of witnesses to adult felony crimes, Technical report: University of California, San Diego, Law and Psychology Program.
- Ng, W.-J., and Lindsay, R.C.L., 1994, Cross-race facial recognition: Failure of the contact hypothesis, *Journal of Cross-Cultural Psychology*, 25: 217-232.
- Pachella, R.G., 1986, Personal values and the values of expert testimony, *Law and Human Behavior*, 10: 145-150.
- Penrod, S.D. and Cutler, B.L., 1987, Assessing the competency of juries, (in) Weiner, I. and Hess, A. (eds.), *The Handbook of Forensic Psychology*. New York: John Wiley & Sons.
- Penrod, S., Loftus, E.F. and Winkler, J., 1982, The reliability of eyewitness testimony, (in) Kerr, N. and Bray, R. (eds.), *The Psychology of the Courtroom*, New York: Academic Press.
- Perfect, T.J., Watson, E.L. and Wagstaff, G.F., 1993, Accuracy of confidence ratings associated with general knowledge and eyewitness memory, *Journal of Applied Psychology*, 78: 144-147.
- Pillemer, D.B., 1984, Flashbulb memories of the assassination attempt on President Reagan, *Cognition*, 14: 63-80.
- Pigott, M.A., Brigham, J.C. and Bothwell, R.K., 1985, *A field study on the relationship between quality of eyewitnesses' descriptions and identification accuracy*, Florida State University.
- Read, J.D., 1979, Rehearsal and recognition of human faces, *American Journal of Psychology*, 92: 71-85.
- Read, J.D., Tollestrup, P., Hammersley, R., McFadzen, E. and Christensen, A., 1990, The unconscious transference effect: Are innocent bystanders ever misidentified? *Applied Cognitive Psychology*, 4: 3-31.
- Rosenthal, R., 1991, Cumulating psychology: An appreciation of Donald T. Campbell. *Psychological Science*, 2: 213-221.
- Ross, D.R., Ceci, S.J., Dunning, D., and Toglia, M.P., 1994, Unconscious transference and mistaken identity: When a witness misidentifies a familiar with innocent person, *Journal of Applied Psychology*, 79: 918-930.
- Shapiro, P.N. and Penrod, S., 1986, Meta-analysis of facial identification studies, *Psychological Bulletin*, 100: 139-156.
- Shepard, R.N., 1967, Recognition for words, sentences, and pictures, *Journal of Verbal Learning and Verbal Behavior*, 6: 156-163.
- Shepherd, J.W., 1983, Identification after long delays, (in) Lloyd-Bostock, S.M.A. and Clifford, B.R. (eds.), *Evaluating Eyewitness Evidence*, New York: Wiley.
- Shepherd, J.W. and Ellis, H.D., 1973, The effect of attractiveness on recognition memory for faces, *American Journal of Psychology*, 86: 627-633.
- Shepherd, J.W., Ellis, H.D. and Davies, G.M., 1982, *Identification Evidence*, Aberdeen, Scotland: Aberdeen University Press.
- Sporer, S.L., 1993, Eyewitness identification accuracy, confidence, and decision times in simultaneous and sequential lineups, *Journal of Applied Psychology*, 78: 22-33.
- Sporer, S.L., 1994, Decision Times and Eyewitness Identification Accuracy in Simultaneous and Sequential Lineups, (in) Ross, D.F., Read, J.D. and Toglia, M.P. (eds.) *Adult Eyewitness Testimony: Current Trends and Developments*, Cambridge University Press, New York, NY., pp. 300 - 327.
- Sporer, S.L., Penrod, S., Read, D. and Cutler, B., 1995, Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies, *Psychological Bulletin*, 118: 315-327.
- Sobel, N.R., 1987, *Eyewitness Identification: Legal and Practical Problems* (end ed.) New York: Clark Boardman Comp. Ltd.
- Stebly, N.M., 1992, A meta-analytic review of the weapon focus effect, *Law and Human Behavior*, 16:413-424.
- Sussman, E.D. and Sugarman, R.C., 1972, The effect of certain distractions on identification by witnesses, (in) Zavala, A. and

- Paley, J.J. (eds.), *Personal Appearance Identification*, Springfield, Ill.: Charles C. Thomas.
- Tollerstrup, P.A., Turtle, J.W. and Yuille, J.C., 1994, Actual victims and witnesses to robbery and fraud: An archival analysis, (in) Ross, D.F., Read, J.D. and Toglia, M.P. (eds.) *Adult Eyewitness Testimony: Current Trends and Developments*, New York: Cambridge University Press.
- Tooley, V., Brigham, J.C., Maass, A. and Bothwell, R.K., 1987, Facial recognition: Weapon effect and attentional focus, *Journal of Applied Social Psychology*, 17: 845-859.
- Webb, E.J., Campbell, D.T., Schwartz, R.F., Sechrest, L. and Grove, J.B., 1981, *Nonreactive Measures in the Social Sciences*, (2nd. ed.). Boston: Houghton Mifflin.
- Wells, G.L., 1984, Do the eyes have it? More on expert eyewitness testimony, *American Psychologist*, 39: 1064-1065.
- Wells, G.L., 1993, What do we know about eyewitness identification? *American Psychologist*, 48: 553-571.
- Wells, G.L. and Lindsay, R.C.L., 1983, How do people judge the accuracy of eyewitness identifications? Studies of performance and a metamemory analysis, (in) Lloyd-Bostock, S.M.A. and Clifford, R.B. (eds.), *Evaluating Witness Evidence: Recent Psychological Research and New Perspectives*, Chichester: John Wiley & Sons.
- Wells, G.L. and Lindsay, R.C.L., 1985, Methodological notes on the accuracy-confidence relation in eyewitness identifications, *Journal of Applied Psychology*, 70: 413-419.
- Wells, G.L., Lindsay, R.C.L. and Ferguson, T.J., 1979, Confidence, accuracy, and juror perceptions in eyewitness identification, *Journal of Applied Psychology*, 64: 440-448.
- Wells, G.L., Lindsay, R.C. and Tounignant, J.P., 1980, Effects of psychological advice on human performance in judging the validity of eyewitness testimony, *Law and Human Behavior*, 4: 275-285.
- Wells, G.L. and Loftus, E.F., 1984, *Eyewitness Testimony: Psychological Perspectives*. Cambridge: Cambridge University Press.
- Wells, G. and Murray, D., 1983, What can psychology say about the *Neil v. Biggers* criteria for judging eyewitness accuracy? *Journal of Applied Psychology*, 68: 347-362.
- Wells, G. and Murray, D., 1984, Eyewitness confidence, (in) Wells, G. and Loftus, E.F. (eds.), *Eyewitness Testimony: Psychological Perspectives*, Cambridge: Cambridge University Press.
- Wells, G.L., Seelau, E.P., Rydell, S.M. and Luus, C.A.E., 1994, Recommendations for properly conducted lineup identification tasks, (in) Ross, D.F., Read, J.D. and Toglia, M.P. (eds.), *Adult Eyewitness Testimony: Current Trends and Developments*, New York: Cambridge University Press., pp. 223-244.
- Wells, G.L. and Turtle, J.W., 1987, Eyewitness testimony research: Current knowledge and emergent controversies, (Special Issue: Forensic psychology), *Canadian Journal of Behavioural Science*, 19: 363-388.
- Winograd, E. and Killinger, Jr., W.A., 1983, Relating age at encoding in early childhood to adult recall: Development of flashbulb memories, *Journal of Experimental Psychology: General*, 112: 413-422.
- Wixted, J. and Ebbesen, E.B., 1991, The mathematics of forgetting functions, *Psychological Science*, 2: 409-415.
- Wixted, J. and Ebbesen, E.B., 1996, Genuine power curves in forgetting, *Memory & Cognition*, (in press).
- Yarmey, A.D., 1979, *The Psychology of Eyewitness Testimony*, New York: The Free Press.
- Yarmey, A. D. and Jones, H., 1983, Is the study of eyewitness identification a matter of common sense? (in) Lloyd-Bostock, S.M.A. and Clifford, B.R. (eds.), *Evaluating Eyewitness Evidence*, New York: Wiley.
- Yuille, J.C., 1989, Expert evidence by psychologists: Sometimes problematic and often premature, *Behavioral Sciences & the Law*, 7: 181-196.
- Yuille, J.C. and Cutshall, J.L., 1986, A case study of eyewitness memory of a crime, *Journal of Applied Psychology*, 71: 291-301.
- Zaragoza, M.S., 1987, Memory, suggestibility, and eye-witness testimony in children and adults, (in) Ceci, S.J., Toglia, M.P., Ross, D.F. (eds.), *Children's Eyewitness Memory*, New York: Springer-Verlag.
- Zaragoza, M.S. and Lane, S.M., 1994, Source misattributions and the suggestibility of eyewitness memory, *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 20: 934-945.
- Zaragoza, M.S. and Koshmider III, J.W., 1989, Misled subjects may know more than their performance implies, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15: 246-255.