METHODOLOGICAL ISSUES IN THE STUDY OF EYEWITNESS MEMORY AND AROUSAL

BRIAN H. BORNSTEINT
TIMOTHY R. ROBICHEAUX††

Crimes, especially those of a violent nature, can be upsetting to see. Thus, it makes sense that a great deal of eyewitness memory research has addressed the effect of negative emotional states such as stress or arousal on eyewitness performance. Yet, not all witnesses experience negative emotional states. Moreover, witnesses to crimes might experience any one of a number of emotional states when viewing the crime—fear, anger, anxiety, disgust, etc.—and eyewitness memory research measures those emotional states in a number of different ways.

This Article does not provide a comprehensive review of empirical findings on the question of eyewitness memory and arousal; such reviews exist elsewhere.1 For present purposes, it suffices to say that the expert consensus on eyewitness memory and arousal is that in most respects, arousal exerts a negative effect on eyewitness performance.2 However, this negative effect on eyewitness performance is not

† Brian H. Bornstein, Ph.D., M.L.S., Professor of Psychology and Courtesy Professor of Law, University of Nebraska-Lincoln.
†† Timothy R. Robicheaux, M.A., M.L.S., University of Nebraska-Lincoln.


2. See, e.g., Saul M. Kassin et al., The “General Acceptance” of Psychological Research on Eyewitness Testimony, 44 AM. PSYCHOLOGIST 1089, 1091-94 (1989) (finding 40% of eyewitness memory experts believed that evidence supporting the proposition that very high levels of stress impair the accuracy of eyewitness testimony was generally or very reliable, an additional 40% believed that evidence tended to favor the proposition and 70.5% believed the phenomenon was reliable enough for courtroom testimony). Results of a follow-up survey more than a decade later were comparable,
uniform; negative emotions can facilitate remembering for certain information under certain conditions.3

Rather than attempting to review the empirical research on eyewitness memory and arousal, this Article intends to highlight some of the methodological issues involved in studying negative emotional states in eyewitnesses. This Article proceeds in five sections. First, this Article provides a working definition of the negative emotional states of stress and arousal, and situates the topic within the constellation of other variables known to influence eyewitness performance. Second, this Article reviews people's assumptions about the relationship between negative emotions and memory. Third, this Article describes the various research paradigms available for studying the relationship between negative emotions and memory, and it offers illustrative examples of each research approach. Fourth, this Article summarizes the different techniques commonly employed for measuring negative emotions. Finally, this Article concludes by making recommendations for future eyewitness memory and arousal research.

I. WHAT WE MEAN WHEN WE TALK ABOUT STRESS AND AROUSAL

The terms "stress" and "arousal" are often used interchangeably by the general public and, to a lesser extent, by researchers as well; this Article follows that convention. Both stress and arousal refer to negative emotionality, both terms have physiological and psychological components, and both terms can be applied to either events (e.g., a stressful or arousing situation) or people (e.g., a stressed or aroused witness). Although it is beyond the scope of this Article to describe in detail exactly how stress and arousal differ, it is worth noting that they are not exactly the same.4 For example, stress is invariably perceived as negative, whereas arousal can be positive, such as sexual, thrill-seeking, or exercise-induced arousal. Arousal, especially at relat-

3. Reisberg & Heuer, supra note 1 at 81.
tively modest levels, is generally associated with an orienting response characterized by alertness and attention to informative stimuli; whereas stress, especially at relatively high levels, is generally associated with a defensive response. These separate responses have different implications for which aspects of a situation an individual attends to, and they also have different implications for how best to measure negative emotionality.

A variety of different taxonomies categorize the many variables that can affect eyewitness performance. The most common taxonomies distinguish either between characteristics of the parties and situation, or between system and estimator variables. For example, Wells and colleagues identify three categories of variables: 1) characteristics of the witness, such as the witness’s personality traits and age, 2) characteristics of the witnessing situation and stimulus, such as event duration and the witness’s viewing opportunity, and 3) post-event factors, such as testing procedures and post-event information. Wells and colleagues list arousal under situation/stimulus characteristics, inasmuch as some events elicit more arousal than others. Although this is true, it is also the case that arousal is a subjective response, and therefore more akin to witness characteristics, as opposed to an objective feature like event duration. Moreover, there are substantial individual differences among witnesses in stress response. Thus, stress and arousal occupy a bit of a middle ground, combining characteristics of both the witness and the situation.

Regardless if one classifies stress as a witness or situation characteristic, experts commonly refer to stress as an estimator variable. “Estimator variables are factors over which the criminal justice system exerts no control,” but which nonetheless enable one to “estimate,” or predict, eyewitness performance. “System variables, on the other hand, refer to eyewitness factors that are directly under the

---

5. See Deffenbacher I, supra note 6; Deffenbacher II, supra note 6; Deffenbacher et al., supra note 1, at 687-91.
6. See Deffenbacher I, supra note 6; Deffenbacher II, supra note 6; Deffenbacher et al., supra note 1, at 687-91; see infra notes 86-116 and accompanying text.
9. Id. at 64-65.
10. See infra notes 86-119 and accompanying text.
11. Cutler et al., supra note 7, at 234.
control of the criminal justice system.\textsuperscript{12} Most prominently, system variables include aspects of lineup construction and administration as well as witness questioning techniques.\textsuperscript{13} As system variables, unlike estimator variables, are under the criminal justice system’s control, they lend themselves more readily to policy adoption, and it is therefore tempting to view them as more important and more worthy of research efforts.\textsuperscript{14} However, a better understanding of both system and estimator variables can help criminal justice fact-finders, such as judges and juries, and police investigators weigh the reliability of eyewitness testimony more appropriately.\textsuperscript{15} Proper weighing of eyewitness testimony reliability, in turn, leads to fewer false identifications and false convictions. In addition, legal scholars need to study both system and estimator variables because they often interact; that is, a lineup variation might affect eyewitness performance under some conditions but not others. For example, the effect of biased lineup instructions (a system variable) is greater at moderate retention intervals (an estimator variable) than at both shorter and longer retention intervals.\textsuperscript{16}

Many of the variables that influence eyewitness performance are relevant only to a particular aspect of eyewitness memory, such as recall of event details or face recognition. For example, the number of individuals in a lineup is relevant to the task of identification (i.e., recognition), but is meaningless with respect to the task of reporting details of the event (i.e., recall). Other variables, particularly estimator variables such as eyewitness arousal, might have an impact on both aspects of memory. Therefore, in describing some of the research findings on arousal and eyewitness memory, this Article draws upon studies that have assessed both recall and recognition memory.\textsuperscript{17}

\textsuperscript{12} Id. at 235.
\textsuperscript{13} Wells et al. I, supra note 7, at 68-75; Gary L. Wells et al., Eyewitness Identification Procedures: Recommendations for Lineups and Photospreads, 22 LAW & HUM. BEHAV. 603 (1998).
\textsuperscript{14} Cutler et al., supra note 7, at 235-36; Wells et al. I, supra note 7, at 55.
\textsuperscript{15} Deffenbacher I, supra note 4, at 817.
\textsuperscript{16} See Nancy Mehrkens Steblay, Social Influence in Eyewitness Recall: A Meta-Analytic Review of Lineup Instruction Effects, 21 LAW & HUM. BEHAV. 283, 290 (1997). Biased instructions are those that lead witnesses to believe the suspect is in the lineup or that discourage witnesses from rejecting the lineup. Retention interval refers to the amount of time between when information is encoded in memory (e.g., witnessing an event) and when a retrieval attempt is made (e.g., attempting a lineup identification). For purposes of the study, moderate retention intervals were defined as two to three days. For a more recent review and meta-analysis of lineup instruction effects, see Steven E. Clark, A Re-examination of the Effects of Biased Lineup Instructions in Eyewitness Identification, 29 LAW & HUM. BEHAV. 395 (2005).
\textsuperscript{17} See infra notes 32-87 and accompanying text. It is worth noting that eyewitness recall (i.e., event or person descriptions) and recognition (i.e., identification) are not themselves highly correlated. See Christian A. Meissner et al., Person Descriptions as
II. ASSUMPTIONS ABOUT STRESS AND MEMORY

Emotion has pervasive effects on cognitive processes. Therefore, it is not surprising that people have strong intuitions about the relationship between emotion and memory. Perhaps the best-known example of such intuitions is the case of "flashbulb memory," which refers to "a subjectively compelling recollection of an occasion when we heard an important piece of news." To illustrate, the majority of the general public can report with great confidence and clarity how they learned of the John F. Kennedy assassination, the Challenger space shuttle disaster, the death of Princess Diana, the September 11, 2001 terrorist attacks, or any number of personally meaningful events. People generally assume that flashbulb memories are highly accurate, precisely because of their emotional or consequential nature. However, not only are flashbulb memories much less accurate than commonly believed, but there is no statistically significant correlation between memory and subjective affect. In other words, although the common expectation is that strong emotions will facili-

Eyewitness Evidence, in 2 THE HANDBOOK OF EYEWITNESS PSYCHOLOGY: MEMORY FOR PEOPLE 1, 19-21 (R.C.L. Lindsay et al. eds., 2007).

18. See generally, Handbook of Affective Sciences (Richard J. Davidson et al. eds., 2003); Affect in Social Thinking and Behavior (Joseph P. Forgas ed., 2006).


24. The list of events that can, and do, form flashbulb memories is a long one. In addition to the events cited, researchers have also studied, among other noteworthy public and private incidents, individuals' flashbulb memory for: the assassination of Swedish prime minister Olof Palme in 1986, Steen F. Larsen, Potential Flashbulbs: Memories of Ordinary News as the Baseline, in AFFECT AND ACCURACY IN RECALL: STUDIES OF "FLASHBULB" MEMORIES 32 (Eugene Winograd & Ulric Neisser eds., 1992); the assassination attempt on President Reagan in 1981, David B. Pillemer, Flashbulb Memories of the Assassination Attempt on President Reagan, 16 COGNITION 63 (1984); and the resignation of British prime minister Margaret Thatcher in 1990, Martin A. Conway et al., The Formation of Flashbulb Memories, 22 MEMORY & COGNITION 326 (1994).

25. Neisser, supra note 19 at 68-69; Brown & Kulik, supra note 20, at 95-99.

26. See Neisser & Harsch, supra note 21, at 18-25; Ulric Neisser et al., Remembering the Earthquake: Direct Experience vs. Hearing the News, 4 MEMORY 337 (1996) (finding non-significant correlations between individuals' memory for the 1989 Loma Prieta earthquake and their ratings of how terrified, worried, upset, or emotional they felt).
tate an individual's memory performance, they fail to have this effect on memory.

Surveys of the general public's beliefs about eyewitness memory demonstrate this expectation that strong emotional impact will facilitate memory. Psychological researchers have conducted a number of these surveys over the past thirty years. These surveys assess lay knowledge of factors affecting eyewitness performance, including the impact of stressful and/or violent events that can cause a negative emotional reaction. These surveys have found that the majority of the general public believe that stress has adverse effects on eyewitness memory. Eyewitness memory experts generally agree with this lay consensus that stress adversely affects eyewitness memory.

III. RESEARCH PARADIGMS FOR STUDYING STRESS AND EYEWITNESS MEMORY

There are three principal methods for conducting research on stress and eyewitness memory: 1) laboratory experiments, 2) naturalistic studies, and 3) meta-analyses. Each of these methods has significant pros and cons.


A. LABORATORY EXPERIMENTS

1. Pros and Cons of Laboratory Experiments

Laboratory experiments' major advantage is that they allow researchers to manipulate only those event characteristics, such as violence, that are presumed to influence witnesses' arousal levels. Other aspects of the situation are held constant, providing researchers with reasonable confidence that differences in eyewitness performance are due to the arousal manipulation. In other words, laboratory experiments allow for causal inferences about the effects of arousal on memory; notably, they are the only research method that conclusively allow for such inferences.

However, experimental manipulations of arousal suffer from a number of limitations. First, experimental witnesses are almost always bystanders, not victims or active event participants. Moreover, these eyewitnesses lack the sense of active engagement and threat that even real bystanders might experience because experimental witnesses typically view the event presented on video. Several eyewitness studies have engaged witnesses as active participants in some sort of encounter, but these studies have not, to our knowledge, included significant degrees of witness arousal.

30. Some studies manipulate arousal through something extraneous to the witnessed event, as by having witnesses exercise or by administering a stimulant. See Sven-Åke Christianson, Emotional Stress and Eyewitness Memory: A Critical Review, 112 PSYCHOL. BULL. 284, 285 (1992); Terry M. Libkuman et al., Source of Arousal and Memory for Detail, 27 MEMORY & COGNITION 166 (1999). Because these manipulations tend to have different effects and are less comparable to witnessing a negative emotional event, most studies induce arousal through some change in the witnessed event.

31. A number of older studies presented arousing stimuli using photographs or slides. See, Sven-Åke Christianson & Elizabeth F. Loftus, Memory for Traumatic Events, 1 APPLIED COGNITIVE PSYCHOL. 225 (1987) [hereinafter Christianson & Loftus I]; Sven-Åke Christianson & Elizabeth F. Loftus, Remembering Emotional Events: The Fate of Detailed Information, 5 COGNITION & EMOTION 81 (1991) [hereinafter Christianson & Loftus II]; T.H. Kramer et al., Effects of Stress on Recall, 5 APPLIED COGNITIVE PSYCHOL. 483 (1991); Alafair Burke et al., Remembering Emotional Events, 20 MEMORY & COGNITION 277 (1992). But the current "state of the art" is to use video. This is not to say that the participants do not experience an emotional response to the stimulus—they can, and do, in the same way that moviegoers respond emotionally. Nonetheless, at some level, the moviegoer knows that he is safely ensconced in a theater and is not at risk from the events portrayed on screen.

32. See R.C.L. Lindsay et al., Can People Detect Eyewitness Identification Accuracy Within and Across Situations?, 66 J. APPLIED PSYCHOL. 79 (1981) (describing experiment where participants witnessed the staged theft of a calculator); R.C.L. Lindsay et al., Biased Lineups: Sequential Presentation Reduces the Problem, 76 J. APPLIED PSYCHOL. 796 (1991) (describing experiment where participants witnessed the staged theft of a small item, such as a radio, purse, or calculator); Carol Krafsa & Steven Penrod, Reinstatement of Context in a Field Experiment on Eyewitness Identification, 49 J. PERS. & SOC. PSYCHOL. 58 (1985) (testing convenience store clerks on their memory of a customer encounter); Stephanie J. Platz & Harmon M. Hosch, Cross-Racial/Ethnic Eyewitness Identification: A Field Study, 18 J. APPLIED SOC. PSYCHOL. 972 (1988) (testing
Second, laboratory experiments vary widely in how those experiments measure witness arousal. Some laboratory experiments have not even measured witness arousal, simply assuming that certain event characteristics, such as violence, would naturally induce negative emotions. As this Article discusses in Part IV below, various indices of witness arousal are not necessarily correlated with one another, and different indices can be affected by different event characteristics, such as whether the witness feels personally threatened.

The third and most significant limitation of laboratory experiments is that researchers can ethically only induce relatively modest levels of arousal in research participants. However, this was not always the case. For example, in his 1908 work On the Witness Stand, Hugo Münsterberg describes an experiment by Professor von Liszt, in which a classroom full of students unwittingly became witnesses to an arousing event. Münsterberg describes the incident:

The Professor had spoken about a book. One of the older students suddenly shouts, "I wanted to throw light on the matter from the standpoint of Christian morality!" Another student throws in, "I cannot stand that!" The first starts up, exclaiming, "You have insulted me!" The second clenches his fist and cries, "If you say another word—" The first draws a revolver. The second rushes madly upon him. The Professor steps between them and, as he grasps the man's arm, the revolver goes off. General uproar. In that moment Professor Liszt secures order and asks a part of the students to write an exact account of all that happened. The whole had been a comedy, carefully planned and rehearsed . . . Those who did not write the report at once were, part of them, asked to write it the next day or a week later; and others had to depose their observations under cross-examination.

Modern ethical standards for conducting psychological research are higher than they were in the past. Under modern standards, it is highly unlikely that contemporary Institutional Review Boards would allow a study similar to the study conducted by von Liszt. Thus, researchers are limited to inducing fairly mild negative emotions in eye-convenience store clerks on their memory of a customer encounter); A. Daniel Yarmey et al., Accuracy of Eyewitness Identifications in Showups and Lineups, 20 LAW & HUM. BEHAV. 459 (1996) (testing shoppers and pedestrians on their memory regarding someone who asked them for directions or assistance in locating lost jewelry).

34. HUGO MÜNSTERBERG, ON THE WITNESS STAND 49-51 (1908).
35. Id. at 49-50. The students' overall memory performance was poor, with numerous errors and omissions. Notably, "[t]he reports with reference to the second half of the performance, which was more strongly emotional, gave an average of fifteen per cent. more mistakes than those of the first half." Id. at 50.
witnesses, merely showing gory pictures, R-rated violence, and similar moderate stimuli. Moreover, current informed consent standards dictate that researchers forewarn participants of the possibility that participants will be exposed to stressful material. Ultimately, experimental witnesses differ from actual witnesses in that experimental witnesses know they might observe something upsetting, and as a result, when that observation occurs it is less stressful than the events that real witnesses might observe, especially in the case of severe crimes involving physical threat, bodily injury, and/or trauma.

Despite these limitations, laboratory experiments afford a powerful test of arousal effects on eyewitnesses because, as noted, they allow for the isolation of arousal apart from other variables. The following section provides some illustrative examples of laboratory experiments.

2. Examples of Laboratory Experiments

In one of the earliest experimental studies of arousal effects on eyewitness memory, Elizabeth Loftus and Terrence Burns presented participants with one of two versions of a brief (approximately two and one-quarter minutes) bank robbery video. In both versions of the bank robbery video, after the initial holdup, the robber walked quickly out of the bank, chased by two employees into a parking lot where two boys played. In the “violent version,” the film ended with the robber firing a shot at the bank employees, hitting one of the boys in the face. In the “nonviolent version,” the film ended instead by flashing back to the inside of the bank. In three separate experiments, Loftus and Burns found that the violent, arousing event impaired witnesses' memories, especially for information that occurred just before the violent incident.

In a similar experiment, Bornstein and colleagues showed participants either a violent and arousing or a non-violent, control version of a film. Bornstein and colleagues divided the film into three segments for scoring purposes: the initial segment (identical for both groups), the middle segment (different for the two groups, as it contained a murder scene for one but not the other), and the concluding

37. Id. at 318-19.
38. Id. at 319.
39. Id.
40. Id. at 319-21. The methodology of the three experiments differed only slightly (e.g., using a recognition instead of a recall test, or changing the nonviolent ending).
segment (identical for both groups). Participants in the control condition, that is participants viewing the non-violent version of the film, recalled more details correctly for both the initial (Means = 28.4% vs. 22.9%) and concluding segments (Means = 24.1% vs. 18.7%). However, for the middle segment, which contained the arousing violent event itself, participants in the violent condition did better (Means = 25.5% vs. 10.1%).

These studies, which are representative of other experiments on eyewitness memory and arousal, illustrate two major findings: First, arousal tends to have an inhibitory effect on memory, both for information that comes prior to the arousing event and for information that comes after the event; second, arousal can have beneficial effects on memory for certain types of information, specifically the arousing event itself.

For example, witnesses to crimes almost always remember that they have, indeed, witnessed a crime—the crime itself stands out in the witnesses' memories. Consequently, arousal makes the most central details more salient, while simultaneously drawing attention away from more peripheral details. A problematic result of this effect is that those same peripheral details are often highly forensically relevant, such as providing evidence regarding the identity of an accomplice standing in the background or the license plate number of the getaway car.

B. NATURALISTIC STUDIES

In naturalistic studies, eyewitness memory is studied in situ, that is outside the laboratory. In these studies, researchers assess the memory of real witnesses after they have witnessed a crime. In a very real sense, researchers conducting naturalistic studies are opportunistic. In most cases, these researchers do not create the event or even set out to study it beforehand; rather, they take advantage of naturally-occurring phenomena. That is, these researchers locate eye-

42. Id. at 126-27. All of these comparisons were statistically significant. Participants who viewed the violent film also made significantly fewer errors in recall, although the overall error rate was quite low. Id. at 127.


44. See, Christianson, supra note 1, at 291-93; Christianson & Loftus I, supra note 31, at 237-38; Christianson & Loftus II, supra note 31, at 104-07. Various cognitive and physiological mechanisms have been invoked for this attentional narrowing effect. See Christianson, supra note 30, at 295-302; Christianson & Loftus II, supra note 31, at 104-07. The relationship breaks down somewhat at the extremes very high levels of stress can have negative effects even for central information, such as the appearance of a perpetrator's face. See infra notes 45-63 and accompanying text.

45. In this sense, they are similar to flashbulb memory researchers, who likewise avail themselves of remembering in natural contexts (the phrase "remembering in natu-
witnesses to real crimes and question them about what they experienced. For the most part, the advantages of naturalistic studies make up for the drawbacks of laboratory experiments, and vice versa.

1. Pros and Cons of Naturalistic Studies

Naturalistic studies have diminished ethical concerns compared to laboratory experiments because researchers do not deliberately induce arousal in eyewitnesses. Arousal levels can span the full range of what witnesses' experience, and therefore, can be much higher than what laboratory studies can produce; the events themselves can also be much richer than what researchers can create in the laboratory. To say that ethical concerns are diminished does not render these concerns completely absent. Real witnesses, especially if those witnesses were victimized, might be psychologically vulnerable; questioning them for research purposes could interfere with the police investigation; and discussing their experience with researchers might upset them anew by forcing them to relive the experience. Nonetheless, researchers can generally deal with these ethical concerns, and the enormous advantage of gathering data from real witnesses to actual crimes outweighs the concerns.

On the other hand, the lack of experimental control in naturalistic studies raises a number of significant drawbacks to this type of study. First, each witness has a unique experience, even for the same event, making it difficult for researchers to aggregate witnesses for purposes of comparison (e.g., "high arousal" vs. "low arousal"). Second, as the crimes in naturalistic studies are not planned events in the manner of laboratory experiments, researchers face the impossible task of ascertaining true event facts for purposes of assessing witness accuracy. Some researchers address this obstacle by reconstructing the incident through police reports and forensic evidence; while others do so by observing naturally occurring, arousing events that are not criminal in nature. The former approach to addressing the obstacle of ascertaining true event facts is necessarily imperfect and still precludes testing some sorts of memory, such as face identification. The latter approach is better in most respects, as it affords a degree of experimental control, but the richness and complexity of the situation none-
theless introduces variability. Third, arousal measures necessarily occur long after the event occurred, when the arousal itself has dissipated. This precludes obtaining physiological measures of arousal from the witness, and witnesses’ retrospective reports are not necessarily reliable.\textsuperscript{49}

2. Examples of Naturalistic Studies

As mentioned above, a few especially clever naturalistic studies have exerted a degree of experimental control over naturally occurring, arousing situations. For example, Charles Morgan and colleagues studied the memory of over five hundred active-duty military personnel as they participated in military survival school training.\textsuperscript{50} Part of the military personnel’s training involved a simulated prisoner-of-war (POW) exercise. The POW exercise was realistic and modeled on the actual experiences of actual POWs.\textsuperscript{51} The trainees were deprived of food and sleep for forty-eight hours and detained in harsh conditions. During their captivity in the mock POW camp, the trainees each underwent two separate forty-minute interrogations: a high-stress interrogation with real physical confrontation and a low-stress interrogation without physical confrontation.\textsuperscript{52}

Twenty-four hours after the trainees’ release from the camp, they attempted to identify their interrogators in either a live or a photo lineup.\textsuperscript{53} The study showed a clear effect of stress on the trainees’ abilities to identify their interrogators. Combining percentages across lineup type, approximately 70\% of trainees in the low-stress condition correctly identified the interrogator when the interrogator was actually in the lineup.\textsuperscript{54} Comparatively, approximately 40\% of trainees in the high-stress condition correctly identified the interrogator when the interrogator was actually in the lineup.\textsuperscript{55} Additionally, trainees in the high-stress condition made roughly twice as many false identifications when the interrogator was not in the lineup.

\textsuperscript{49} Subjective reports in experimental studies suffer from this problem as well; but the reports are likely to be closer in time to the arousing event than for naturalistic studies. For example, witnesses in the study by Yuille and Cutshall rated their stress on a 7-point scale 4-5 months after the incident. Yuille & Cutshall, supra note 46, at 293. We discuss these issues more in Part IV.

\textsuperscript{50} Charles A. Morgan, III et al., Accuracy of Eyewitness Memory for Persons Encountered During Exposure to Highly Intense Stress, 27 Int’l J.L. & PSYCHIATRY 265 (2004).

\textsuperscript{51} Id. at 268-69.

\textsuperscript{52} Id. at 268. Each interrogation involved one or two interrogators.

\textsuperscript{53} The live lineup contained 15 persons, whereas the photo lineup contained 16 persons. The photo lineup was presented either sequentially or as two 8-person photospreads. Id. at 269-70.

\textsuperscript{54} Id. at 272.

\textsuperscript{55} Id.
Valentine and Mesout took a different, though equally novel, approach to studying stress in a naturalistic setting. Valentine and Mesout studied the memory of visitors to the Horror Labyrinth (the "labyrinth") exhibit of the London Dungeon, a popular tourist destination in London, England. Designed to induce arousal, specifically fear, Valentine and Mesout described the labyrinth:

[it is a maze of floor-to-ceiling mirrored walls set amongst Gothic vaults. ... It is dark, crowded and there is a sound track playing the rhythm of a heartbeat and various scary noises and screams. ... There are a number of 'scares' in the labyrinth. A screaming skeleton is triggered by an infrared beam when approached. An elderly woman sitting in a rocking chair appears behind a mirror. There was an actor in the labyrinth, dressed in a dark robe and wearing theatrical make up to create a very pale facial skin colour with wounds or scars. This actor ... would step out in front of the participant, and then block their path to prevent them passing.57

After participants finished touring the labyrinth, they attempted to recall what the scarer looked like, and they also attempted to identify the scarer from a nine-person lineup.58

One of the more noteworthy features of Valentine and Mesout's study is that it measured participants' arousal in both physiological (i.e., heart rate) and psychological (i.e., self-report) respects. The study found a very high, positive correlation between the change in participants' heart rates while going through the labyrinth and their state of anxiety,60 suggesting that self-report can sometimes be a reasonable proxy for physiological arousal.61 Most relevant to this Article, participants who experienced greater stress did worse on both recall of details and person identification. For example, low-state anxiety participants made correct identifications 75% of the time, whereas high-state anxiety participants made correct identifications only 18% of the time.62

One could argue that at some level, the military trainees in the Morgan study and the tourists in the Valentine and Mesout studies...
knew that their personal safety was not really in jeopardy and that they were merely participating in a simulated exercise. In this respect, both the Morgan study trainees and the Valentine and Mesout study tourist participants differed from real witnesses to violent crimes. Nonetheless, these witnesses' experiences—both subjective and physiological—suggest that the research subjects responded very much as real witnesses would under similar circumstances. Furthermore, in these naturalistic studies, as in the laboratory studies, stress had a decremental effect on eyewitness performance.

C. META-ANALYSES

Meta-analysis is a statistical procedure used to aggregate research findings from multiple studies on a specific topic. It has become increasingly prevalent in the sciences over the last twenty to thirty years, where meta-analytic findings are often taken as the definitive statement of some phenomenon. Furthermore, meta-analytic findings are increasingly finding their way into court. Psychological and legal researchers have conducted numerous meta-analyses on various aspects of eyewitness performance.

1. Pros and Cons of Meta-analyses

Other articles have detailed, at length, the advantages of meta-analysis. Penrod and Bornstein list four major advantages of meta-analyses. Meta-analyses: 1) "provide a succinct summary of the status of scientific research in a particular domain;" 2) "permit identification of variables where there is substantial variability in findings;" 3) "permit[] tests of interactions or moderating effects across studies that have not been examined within studies;" and 4) "provide guidance about avenues of further research that are likely to be more or

63. E.g. participants in both studies provided informed consent.
65. E.g. meta-analysis typically fares well at satisfying the reliability concerns for expert testimony that prevail in most U.S. courts. See Daubert, 509 U.S. 579 (1993); See also Deffenbacher et al., supra note 64, at 692.
66. See generally Steven Penrod & Brian H. Bornstein, Generalizing Eyewitness Reliability Research, in 2 THE HANDBOOK OF EYEWITNESS PSYCHOLOGY: MEMORY FOR PEOPLE (R.C.L. Lindsay et al. eds., 2007).
68. Penrod & Bornstein, supra note 66, at 535.
69. Id. at 536.
70. Id.
METHODOLOGICAL ISSUES

These advantages are particularly beneficial in forensic settings because courts prefer to rely on an aggregated body of research rather than isolated findings.

Despite the many advantages of meta-analysis, it is no panacea. Meta-analysis is only as good as the data that produce it. Some meta-analyses include only published studies, whereas others also include unpublished studies. As scientific journals rarely publish null (i.e., non-statistically significant) results, meta-analyses of published studies might distort the true state of affairs; on the other hand, unpublished studies might be methodologically flawed. Coding limitations also affect meta-analyses; although meta-analyses can and do code for a number of methodological characteristics and other moderating variables, they cannot code for all potentially meaningful variations across studies. Thus, different results across studies could be explained by any of a number of “silent” factors. Despite these limitations, meta-analysis provides the best tool currently available for aggregating findings across a diverse sample of studies.

2. Meta-analysis of Eyewitness Arousal: An Example

Kenneth Deffenbacher and colleagues recently conducted a meta-analysis of the effect of stress on eyewitness performance. They examined two separate samples: sixteen published papers (with twenty-seven independent estimates and 1727 participants) that tested the effect of stress on a witness's ability to identify a perpetrator; and eighteen published papers (with thirty-six independent estimates and 1946 participants) that tested the effect of stress on eyewitness recall. Deffenbacher and colleagues limited both samples to studies that had induced relatively high levels of stress, which would elicit a defensive rather than an orienting response.

Stress had a negative effect on both face identification and recall. For example, participants made correct identifications 42% of the time under high-stress conditions, compared to 54% of the time under low-stress conditions. Importantly, the researchers found that the negative effect of stress was greater for more naturalistic research para-

71. Id.
72. See Dawn McQuiston-Surrett et al., Sequential vs. Simultaneous Lineups: A Review of Methods, Data, and Theory, 12 PSYCH., PUB. POL’Y & LAW 137 (2006) (discussing these issues at length).
73. Id. at 139.
74. Coding refers to the identification and categorization of relevant variables.
75. Deffenbacher et al., supra note 64.
76. Id. at 692. There was some overlap between the two samples; that is, some papers measured both face identification and recall.
77. Id. at 691.
78. Id. at 694 (indicating a statistically significant comparison).
digms, such as eyewitness (as opposed to face recognition) studies and staged crime (vs. non-staged crime) studies.\textsuperscript{79} This finding is noteworthy because it addresses the question of how confidently researchers can generalize from eyewitness research to more naturalistic settings.\textsuperscript{80} Specifically, this finding suggests that "[c]oncerns that this [eyewitness] research lacks generalizability are unfounded; rather, the more ecologically valid the research becomes, the larger—and potentially more problematic from a forensic perspective—the effects tend to be."\textsuperscript{81}

D. SUMMARY OF EMPIRICAL FINDINGS

Any research domain benefits from the use of multiple methodologies because "one can be more confident in the ‘truth’ of some finding if it has been demonstrated across multiple contexts or exemplars using a variety of techniques."\textsuperscript{82} The value of multiple approaches is especially keen in domains like psychology and law, where research has the potential to influence public policy.\textsuperscript{83} Research on eyewitness memory is a case in point.

With respect to how arousal affects eyewitness performance, findings from laboratory, naturalistic, and meta-analytic studies converge on a fairly consistent, but not quite uniform, pattern of findings. Specifically, the effect of arousal on eyewitness performance is not invariably negative. Arousal can produce an orienting response toward the stressful stimulus, which can be associated with improved memory, especially for remembering that the event itself occurred at all and other central details. Otherwise, however, arousal tends to impair memory. It does so for recall of details of a stressful event, especially peripheral details. High levels of stress also impair memory for what is ostensibly the most central detail of an eyewitness event: the perpe-

\textsuperscript{79} Id. at 696-99. In a face recognition task, participants typically view a relatively large number of faces, followed by a recognition task in which those faces are mixed in with other (new, or distractor) faces. In an eyewitness task, participants typically view a small number of targets, followed by a lineup (or showup) task in which the target individuals may or may not be present. Usually there is only a single target, and participants view the person’s entire body while performing some action (e.g., committing a crime), and not just the face. Studies that did not manipulate stress by exposing participants to a staged crime did so by some other means, such as the threat of an injection.

\textsuperscript{80} Penrod & Bornstein, supra note 66.

\textsuperscript{81} Id. at 551.

\textsuperscript{82} Brian H. Bornstein, Signs for the Future of Civil Justice Research, in Civil Juries and Civil Justice: Psychological & Legal Perspectives 273, 275 (Brian H. Bornstein et al. eds., 2008).

trator's identity. Thus, on balance the research supports the expert consensus that stress has a negative effect on eyewitness performance. Where research findings are inconsistent, one possible explanation has to do with differences in how negative emotional states are measured.

IV. MEASURING NEGATIVE EMOTIONAL STATES

Most studies of stress, arousal, and memory utilize one or more measurements of participants' negative emotional state. Generally, researchers will measure stress levels before and after exposure to a potential stressor, and sometimes researchers measure stress levels during exposure to the stressor. By comparing stress levels before and after exposure to the expected stressor, researchers demonstrate that the stimulus induced a stress response in the participants. Explicitly measuring stress also allows researchers to examine the relationship between different degrees of stress responding (i.e., stress reactivity) and memory. For example, two individuals could both perceive a crime event as stressful, but both the magnitude of that stress response and memory accuracy may differ among individuals. Thus, researchers must employ appropriate measures of negative emotions in eyewitness studies.

Researchers employ multiple methods for measuring negative emotional states. In studies involving stress and eyewitness memory, researchers typically either ask participants to report their levels of negative emotions (i.e., self-report measures), or they measure physical changes associated with stressors (i.e., physiological measures).

---

84. We use the term “ostensibly” advisedly. Although perpetrator identity is often the most important piece of information from a forensic perspective, it is not always so. In many situations, the perpetrator's identity is known and undisputed, but there is disagreement about details of the event (e.g., what was done or said). Moreover, even when identity is the central aspect forensically, it may not be central from the witness's perspective. This is especially true in situations where the witness's attention is drawn to a weapon, where the witness's overriding concern is with his or her own safety, where injured persons need assistance, etc. See Kerri Pickel, Remembering the Identifying Menacing Perpetrators: Exposure to Violence and the Weapon Focus Effect, in 2 THE HANDBOOK OF EYEWITNESS PSYCHOLOGY: MEMORY FOR PEOPLE (R.C.L. Lindsay et al. eds., 2007) (discussing the “weapon focus” effect).

85. See supra notes 30-84 and accompanying text.

86. See, e.g., Jessica D. Payne et al., The Impact of Stress on Neutral and Emotional Aspects of Episodic Memory, 14 MEMORY 1 (2006); Tali Sharot & Elizabeth A. Phelps, How Arousal Modulates Memory: Disentangling the Effects of Attention and Retention, 4 COGNITIVE, AFFECTIVE, AND BEHAVIORAL NEUROSCIENCE 294 (2004).

87. If the experimental stress manipulation is indeed stressful, then levels of the stress measure should increase from before to after (or during) the manipulation.

88. See, e.g., Tony W. Buchanan et al., Impaired Memory Retrieval Correlates with Individual Differences in Cortisol Response but not Autonomic Response, 13 LEARNING & MEMORY 382 (2006).
The pros and cons of both measurement methods will be discussed in detail in the current section.

A. SELF-REPORT MEASURES

Many studies of stress and eyewitness memory rely exclusively on witness self-report measures of stress or arousal. Self-report measures of stress require participants to evaluate their own stress levels; such measures implicitly assume that the participants are aware of and can accurately assess those stress levels. An eyewitness interview conducted at the scene of a crime or during legal proceedings could easily employ a self-report measure of stress. A researcher could simply ask the witness, "Did you find [the eyewitness situation] to be a stressful experience?" In the empirical literature, when researchers employ a single question or item, they generally require participants to rate their perceived level of stress or arousal on a numerical scale with corresponding verbal labels for each point or the two most extreme values. The exact wording of the questions varies, with participants across studies typically asked to rate their degree of stress, anxiety, arousal, fear, upset, or worry. Needless to say, although such emotional states are related, they are not exactly the same thing. Questionnaires with multiple questions are useful when researchers want to investigate multiple attributes of the same concept.

Oftentimes researchers write their own self-report measures of stress or arousal specifically for a given study, while other researchers rely on previously published measures of stress, arousal, or similar constructs. In general, published self-report measures of stress that have already undergone peer-review scrutiny are typically validated for research use. One such self-report measure of stress is the Perceived Stress Scale.


90. See, e.g., Jeffrey M. Brown, Eyewitness Memory for Arousing Events: Putting Things into Context, 17 APPL. COGNIT. PSYCHOL. 93, 96 (2003) (researchers had participants rate the amount of arousal associated with pictures in a slideshow); Kramer et al., supra note 89 at 170 (researchers had participants rate their feelings while viewing stimuli from bored (1) to anxious (7)).

91. See, e.g., the studies described in Part III, supra notes 44-84.

92. See, e.g., Brown, supra note 90; Bornstein et al., supra note 41.

93. Sheldon Cohen et al., A Global Measure of Perceived Stress, 24 J. HEALTH & SOCIAL BEHAV., 385 (1983). The full version of the scale includes 14 items asking participants how often they have experienced certain symptoms in a specified time-span.
1. Pros of Self-Report Measures

The primary benefits of measuring stress and arousal through self-report measures are the simplicity and affordability of data collection, both for research purposes and by investigators in the field. Self-report measures generally require nothing more than paper or a computer. They are efficient tools when researchers merely want to verify that differences exist between those who undergo a stress manipulation and those who do not. Multiple studies also demonstrate the predictive value of self-report measures of stress on memory accuracy.94

2. Cons of Self-Report Measures

Self-report measures of stress also have limitations. Participants must not only have cognitive awareness of their stress levels, but they must also be able to quantify that stress. Quantification of stress is not always easily done, as some aspects of stress occur below cognitive awareness.95 In addition, some commentators criticize self-report measures because of the arbitrary and interpretive nature of the numerical values.96 Although most self-report scales have anchors (i.e., qualitative descriptions of what certain values mean), these measures still are inherently subjective. Self-report measures are useful to compare an individual's stress levels across two or more points in time or across two or more different events, but a single measure by itself tells researchers relatively little information, and it is hard to compare self-reports across individuals who might be interpreting the scale differently.97

Finally, researchers almost always obtain self-report measures after the stressful event has occurred. Such retrospective memory reports by witnesses are notoriously unreliable, and it is likely hard for witnesses to distinguish their current level of arousal from their ear-

---

94. See Part III, supra notes 30-84.
95. For example, some work days may be busier than others, but quantifying the differences in stress levels across those days may be difficult for many. See, e.g., Shlomo Breznitz, Theory-Based Stress Measurement? Not Yet, 1 PSYCHOLOGICAL INQUIRY 17 (1990).
96. See, e.g., Ronald C. Kessler, The Interplay of Research Design Strategies and Data Analysis Procedures in Evaluating the Effects of Stress on Health, in STRESS AND HEALTH: ISSUES IN RESEARCH METHODOLOGY (Stan V. Kasl & Cary L. Cooper eds., 1987). Verbal labels (e.g., very stressed, somewhat stressed, mildly stressed, not at all stressed, etc.) are no less ambiguous.
97. On a five-point scale, a value of “1” may correspond with the anchor “not at all aroused,” while a “5” may correspond with “extremely aroused,” for example.
lier level of arousal. People often inaccurately report the emotional aspects of previous experiences; such reports are subject to distortion. If this same inaccuracy is present in witnesses' retrospective reports of stress levels, then such self-reports serve as unreliable estimates of the degree of stress that they actually experienced.

B. PHYSIOLOGICAL MEASURES

Researchers also study stress and arousal through measuring physiological correlates of stress. When individuals perceive their environment as potentially threatening, their bodies undergo a series of physiological changes meant to prepare them to deal with the stressor, a concept historically referred to as "fight or flight" (i.e., stress) responding. Stress responses include increased heart rate, sweating, changes in skin conductance (i.e., galvanic skin response), and changes in the digestive and reproductive systems. Approximately twenty to forty minutes after an individual's exposure to a stressor, the individual's body also reaches a peak level of the hormone cortisol, which, among other mechanisms, signals glucose secretion to provide the body with energy to recover from the stressful experience. Researchers can measure some of these physiological changes, such as heart-rate or cortisol changes, reasonably accurately.

98. See generally Ulric Neisser, Snapshots or Benchmarks?, in MEMORY OBSERVED: REMEMBERING IN NATURAL CONTEXTS 68, 69 (Ulric Neisser & Ira E. Hyman, Jr. eds., 2d ed. 1995).
100. While the word "perceive" implies an awareness of one's surroundings, there is evidence that the brain automatically scans the environment for threatening situations without any conscious thought. See Robert M. Sapolsky, WHY ZEBRAS DON'T GET ULcers: AN UPDATED GUIDE TO STRESS, STRESS RELATED DISEASES, AND COPING (1998). A threatening situation can be either physically threatening (e.g., being mugged) or psychologically threatening (e.g., arguing a case in front of a jury).
102. Id.
105. Others, such as sweating or changes in the digestive and reproductive systems, are harder to quantify.
While physiological measures are less common than self-report measures in studies of stress and eyewitness memory, several eyewitness memory researchers have utilized physiological measures of stress, most often heart rate. Researchers studying stress and basic memory processes have also utilized measures of cortisol. Researchers examine cortisol levels using samples of participants' saliva obtained either by holding a cotton swab in the participant's mouth or by having participants drool through a straw into a vial. Although some researchers may have the facilities to measure salivary cortisol themselves, most researchers do not have such facilities, and instead they contract with companies that specialize in salivary sample analysis.

1. Pros of Physiological Measures

As most researchers assume that stress-induced changes in memory have physiological bases, measuring stress through physiological measures seems appropriate. Physiological measures also allow researchers to avoid the weaknesses associated with self-report measures. Physiological measures require no cognitive awareness of changes and utilize objective measurement. In addition, using physiological measures of stress responding is more consistent with most current stress research, which emphasizes physical as opposed to psychological responding. Furthermore, physiological measures allow researchers to compare the predictive value of different measures on


107. By “basic memory processes” we mean laboratory studies of relatively simple stimuli (e.g., digits or word lists), as opposed to studies of face recognition or more naturalistic events. See, e.g., Tony W. Buchanan et al., Impaired Memory Retrieval Correlates with Individual Differences in Cortisol Response but not Autonomic Response, 13 Learning & Memory 382 (2006); Sabrina Kuhlmann et al., Impaired Memory Retrieval after Psychosocial Stress in Healthy Young Men, 25 J. Neuroscience 2977 (2005); N.Y.L. Oei et al., Psychosocial Stress Impairs Working Memory at High Loads: An Association with Cortisol Levels and Memory Retrieval, 9 Stress 133 (2006).

108. Elizabeth A. Shirtcliff et al., Use of Salivary Biomarkers in Biobehavioral Researcher: Cotton-Based Sample Collection Methods can Interfere with Salivary Immunoassay Results, 26 Psychoneuroendocrinology 165 (2001). Cortisol levels can also be obtained from blood samples, but the collection of saliva samples is easier and less invasive.


111. See supra Part IV.A.2.

memory. For instance, Buchanan and colleagues reported that individuals, who underwent a stressful laboratory task and saw increases in heart rate but not in cortisol had no differences in memory compared to those individuals in a control condition.113 However, individuals who exhibited increases in both heart rate and cortisol had poorer memory performance than the individuals in the control condition.114 This research suggests that heart rate is associated more with the orienting type of response characteristic of arousal, whereas cortisol is associated more with the defensive type of response characteristic of extreme stress.115

2. Cons of Physiological Measures

Unlike self-report measures, however, physiological measures are often costly and can require specialized equipment and training. For some physiological measures, the equipment is a one-time investment (e.g., heart-rate monitors), whereas for others the cost is ongoing (e.g., cortisol analysis). In a research context, a paper-and-pencil measure of arousal can be easily administered to a room full of participants, but physiological measurement is limited by the amount of available equipment, such as heart rate monitors.

Physiological measures of stress and arousal also lack real-world utility in working with actual witnesses. In an actual eyewitness situation, witnesses would generally not be wearing heart-rate monitors or providing samples of saliva. Even if police had the necessary equipment to take such physiological measures in the field, by the time they were able to use this equipment, the witnesses’ stress levels would likely have diminished considerably. In contrast, police investigators can readily obtain a retrospective report of a witness’s subjective arousal level. Thus, it might make more sense for researchers to use measures that would be practical and appropriate in the field, namely self-report measures.116

113. Tony W. Buchanan et al., Impaired Memory Retrieval Correlates with Individual Differences in Cortisol Response but not Autonomic Response, 13 Learning & Memory 382, 383 (2006). The participants underwent a cold-pressor task, which requires a participant to submerge an arm into a bath of cold water for several minutes. The task elicits a stress response in most individuals.

114. Id.


116. As noted in Part IV.A.2, supra, these measures have their own problems.
V. CONCLUSION

The foregoing section illustrates the complicated nature of measuring and studying the effects of negative emotional states on eyewitness memory. Yet, numerous findings from multiple methodologies demonstrate that the effect of eyewitness arousal on memory is generally, though not uniformly, negative. Thus, it is important for those persons working with eyewitnesses, such as police and lawyers, and those persons evaluating their testimony, such as judges and jurors, to consider eyewitnesses' stress levels at the time of a crime. In this sense, eyewitness stress or arousal is an important estimator variable. A better understanding of stress effects on memory could help reduce overreliance on eyewitness testimony, and ultimately, the prevalence of false convictions.

Researchers struggle to single out which factors affecting eyewitness memory are the most important to study and understand, for either theoretical or practical purposes. We are loath to undertake that task ourselves, but we would argue nonetheless that witness arousal level deserves a spot on any such list. Our justification for this contention is that the effects of eyewitness arousal on memory performance are relevant to psychological theory (e.g., the cognitive mechanisms by which arousal exerts its effects); there are significant physiological concomitants of arousal that help explain those effects; and witnessing a crime is often an arousing situation—at times highly so. Like many other areas of eyewitness research, the topic has important psychological and legal implications.¹¹⁷ We are encouraged by the amount of recent attention devoted to the topic of the effect of stress on eyewitness memory from both psychological and legal perspectives, as well as by the advent of new technologies for measuring stress. As legal psychologists with one foot in each discipline, we eagerly look forward to the next generation of research on eyewitness memory and arousal.

¹¹⁷ See generally 1 THE HANDBOOK OF EYEWITNESS PSYCHOLOGY: MEMORY FOR EVENTS (Michael P. Toglia et al eds., 2006); 2 THE HANDBOOK OF EYEWITNESS PSYCHOLOGY: MEMORY FOR PEOPLE (R.C.L. Lindsay et al eds., 2007).