Law and Human Behavior, Vol. 28, No. 6, December 2004 (© 2004)
DOI: 10.1007/s10979-004-0565-x

A Meta-Analytic Review of the Effects of High Stress on Eyewitness Memory

Kenneth A. Deffenbacher,[1,4] Brian H. Bornstein,[2] Steven D. Penrod,[3] and E. Kiernan McGorty[2]

In the past 30 years researchers have examined the impact of heightened stress on the fidelity of eyewitness memory. Meta-analyses were conducted on 27 independent tests of the effects of heightened stress on eyewitness identification of the perpetrator or target person and separately on 36 tests of eyewitness recall of details associated with the crime. There was considerable support for the hypothesis that high levels of stress negatively impact both types of eyewitness memory. Meta-analytic Z-scores, whether unweighted or weighted by sample size, ranged from −5.40 to −6.44 (high stress condition–low stress condition). The overall effect sizes were −.31 for both proportion of correct identifications and accuracy of eyewitness recall. Effect sizes were notably larger for target-present than for target-absent lineups, for eyewitness identification studies than for face recognition studies and for eyewitness studies employing a staged crime than for eyewitness studies employing other means to induce stress.

KEY WORDS: meta-analytic; eyewitness memory; high and low stress.

The performance of eyewitnesses under conditions of heightened stress is of particular forensic interest. When witnessing a crime of violence, the response of the eyewitness is almost always one of generating a stress response to the stressor imposed by the crime. The stress response is actually the defensive response set studied in some detail by psychophysiologists (e.g., Klorman, Weissberg, & Wiesenfeld, 1977). This defensive reaction is the physiological response (acceleration in heart rate, increased blood pressure and muscle tone) that results when the *activation* mode of attention control is dominant (Tucker & Williamson, 1984). The activation mode is one of two neural control systems for regulating response to environmental demands. It is characterized by a tonic readiness for action, a bias against stimulus change, and processing under tight attention controls. Tasks eliciting activation

[1]University of Nebraska, Omaha, Nebraska.
[2]University of Nebraska, Lincoln, Nebraska.
[3]John Jay College of Criminal Justice, New York, New York.
[4]To whom correspondence should be addressed at University of Nebraska, Department of Psychology, 6001 Dodge, Omaha, Nebraska 68182-0274; e-mail: kdeffenbacher@mail.unomaha.edu.

mode dominance include any task serving to increase cognitive anxiety (worry) and/or somatic anxiety (conscious perception of physiological activation), including vigilance, escape, avoidance, or "pressure" tasks (Deffenbacher, 1994).

When assessing any effect of a condition of heightened stress, one must of course compare it with a condition demonstrably lower in stress or even one free of stress, the latter condition being one wherein the *arousal* mode of attention control is dominant (Tucker & Williamson, 1984). Here the physiological response is that of the orienting response (Lacey & Lacey, 1974), a deceleration of heart rate, lowered blood pressure and muscle tone, and an increase in skin conductance with different temporal characteristics than is the case when the activation mode is dominant. The arousal mode of neural control functions to support alert wakefulness and responsiveness to environmental change and novel stimulation. Attention is allocated to the most informative aspect of the stimulus array, rather than being restricted to a specific semantic or motivational content, as is typical when the activation mode of attention control is dominant. Tasks involving simple perceptual intake have been shown to elicit the arousal mode of attention control (Deffenbacher, 1994). Presumably nonthreatening eyewitness events would elicit the arousal mode.

It is clearly important to know just how heightened stress impacts the fidelity of an eyewitness's memory. Since the renaissance of research on eyewitness testimony began in the early 1970s, a scientific literature has accumulated concerning the effects of heightened stress on the fidelity of eyewitness memory. Nevertheless, 30 years of data have not as yet yielded a clear picture of whether heightened stress has a positive, negative, or null effect on eyewitness memory. The principal goal of the present review is to ascertain which of these three possible relations actually obtains.

The first systematic review of the literature relating the effects of heightened stress to eyewitness memory was conducted by Deffenbacher (1983). Typical for the time, Deffenbacher made the assumption that all stressors act to increase general arousal, whether they be high intensity white noise, electric shock, ego-involving instructions, seriousness of a viewed crime, or violence level of a viewed crime. The generally accepted theoretical explanation of the stress-performance relationship was that variations in stressor intensity affected performance level according to an inverted-*U* function, the function described by the Yerkes–Dodson law (1908). For tasks of at least moderate complexity, and eyewitness identification tasks would appear to qualify, the Yerkes–Dodson law states that performance improves with increases in arousal up to some optimal point and then declines with further increases. In his review of 21 relevant published and unpublished studies, Deffenbacher noted that 10 had produced results which suggested that higher arousal levels increased eyewitness accuracy or at least did not decrease it. The remaining 11 studies produced results showing lowered memory accuracy with increases in arousal. Deffenbacher argued that the studies showing facilitation of memory by arousal increases were likely dealing with arousal increases within the range encompassed by the ascending portion of the inverted-*U* curve; studies showing memory debilitation with arousal increases were likely operating in the range encompassed by the descending portion of the Yerkes–Dodson curve.

A decade later, Christianson (1992) again reviewed the now burgeoning literature relating what he referred to as *emotional stress* and eyewitness memory. He came to rather different conclusions than did Deffenbacher (1983). First, he argued that there was not much evidence to support the notion that emotional stress debilitates eyewitness memory. Second, he proposed that the Yerkes–Dodson law is not an appropriate description of the relation of emotional stress to the fidelity of eyewitness memory. Third, he concluded that in general memory for negative emotional events is better than that for neutral events, at least for central details; typically, however, memory for noncentral details is worse for negative emotional events than for neutral ones. He suggested that the better memory for central details was due to negative emotional events causing greater focusing of attention and increased elaboration of the details within that focus.

Thus far we have established that the studies assessing the effect on memory of what has been variously referred to as heightened stress, anxiety, arousal, or negative emotionality have yielded all possible effects on memory performance, positive, negative, and null. The two reviewers of this literature (Christianson, 1992; Deffenbacher, 1983) arrived at different empirical generalizations that would characterize its body of findings, though admittedly, these two snapshots of the literature occurred a decade apart and included a different mix of research methodologies. Is there a way of resolving this apparent muddle?

We propose that a theoretical alternative to the Yerkes–Dodson law can assist in clearing up at least some of the muddle. In a more recent review, Deffenbacher (1994) revisited the concept of an unidimensional continuum of arousal and concluded that it could no longer be sustained. He likewise concluded that the Yerkes–Dodson law was no longer a useful explanatory construct. Deffenbacher then presented an integrative theoretical alternative to unidimensional arousal theory, a synthesis of Tucker and Williamson's (1984) asymmetric neural control systems model and Fazey and Hardy's (1988) catastrophe model of anxiety and performance, a model that has made some very specific predictions that have been empirically confirmed (e.g., Hardy & Parfitt, 1991). The latter model is a three-dimensional model including two predictor variables, *cognitive anxiety* (worry) and physiological activation, the conscious perception of which has been labeled *somatic anxiety*; the dependent variable is performance. Fazey and Hardy (1988) had concluded from their review of the anxiety-performance literature that any satisfactory model had to be at least three-dimensional. Fazey and Hardy also noted that their model accounted for four different relationships between anxiety and performance found in their literature review. As Deffenbacher (1994) has pointed out, the most interesting prediction from their model is the prediction that at relatively high levels of cognitive anxiety, continuous gradual increases in somatic anxiety (physiological activation) will at first result in continuous, gradual increases in performance, followed at some point by a catastrophic, discontinuous drop in performance. Thus acting in concert, cognitive anxiety and physiological activation produce nonlinear effects on performance. As Deffenbacher (1994) has also noted, a close examination of the data of at least two studies of eyewitness memory (Bothwell, Brigham, & Pigott, 1987; Peters, 1988) confirms the prediction of a catastrophic drop in memory performance at high levels of cognitive anxiety and physiological activation.

Thus by Deffenbacher's (1994) integrative theoretical formulation, if a task elicits the arousal mode of attention control, then memory will be enhanced for the most informative aspects of the stimulus display, those aspects on which the orienting response is focused. If, on the other hand, a task elicits the activation mode of attention control, then memory will either be modestly enhanced or drastically reduced, depending on the relative amounts of cognitive anxiety and physiological activation present.

We are now in a position to begin clearing up the aforementioned muddle of findings. Since 1984, a substantial number of studies have been published showing that increases in what was referred to as *negative emotionality* not only did not adversely impact memory for central details of a scenario but actually improved memory relative to that for central details of a scenario significantly lower in negative emotionality. The only adverse impact on memory by increased negative emotionality was on the less important peripheral details. These studies were a major focus of Christianson's (1992) review and certainly make quite understandable his claim that there was not much evidence to support the notion that emotional stress debilitates eyewitness memory.

We believe that these studies (e.g., Burke, Heuer, & Reisberg, 1992; Christianson, 1984; Christianson, Loftus, Hoffman, & Loftus, 1991; Heuer & Reisberg, 1990; Libkuman, Nichols-Whitehead, Griffith, & Thomas, 1999, Safer, Christianson, Autry, & Osterlund, 1998) were generating facilitation of eyewitness memory for central details, because their principal experimental manipulations likely generated an orienting response (arousal mode of attention control) to stimulating conditions, rather than the defensive response (activation mode of attention control) typically produced by a successful manipulation of stress or anxiety (Deffenbacher, 1994, 1999). The implicit assumption by these investigators appears to have been that higher ratings of negative emotionality for experimental condition stimulus materials (e.g., a modestly gruesome accident or surgery scene) as compared to ratings of control condition stimuli signified a successful manipulation of an emotional state that was akin to stress or anxiety. Another implicit assumption was that by increasing physical exertion not directly relevant to the viewed scenario (e.g., riding an exercise bicycle), a successful manipulation of a physiological state akin to that comprising a defensive response had been attained. As Deffenbacher (1994, 1999) has argued, however, the key experimental manipulation in these studies almost certainly elicited orienting responses, rather than defensive responses.

For one thing, these studies were carried out in laboratory settings where both cognitive anxiety and physiological activation should have been relatively low. Baseline heart rates in these studies averaged 68–82 beats per min (b.p.m.), within the normal range of resting heart rates for young adults. Second, the tasks presented were those of simple perceptual intake or perceptual intake plus instructions emphasizing the need to attend closely to the central and peripheral details of an external event. The negative emotional content of events depicted on key slides was neither a threat to the bodily integrity nor to the self-esteem of the observer. The content was gruesome enough, however, to have elicited an orienting response (Hare, 1972; Hare, Wood, Britain, & Frazelle, 1971; Hare, Wood, Britain, & Shadman, 1970;

Klorman et al., 1977). For example, Hare et al. (1970, 1971) showed that there was not only not a defensive response but in actuality a stronger orienting response to unretouched, color slides of homicide victims than to slides of everyday objects. Hare et al. (1970, 1971) showed heart rate deceleration of 3–5 b.p.m., an important index of an orienting response, as did several of the aforementioned investigators of increases in negative emotionality. Christianson (1984) also noted an increase in skin conductance to stimuli of greater negative emotionality, another index of an orienting response to a stimulus display. Interestingly, Lang, Greenwald, Bradley, and Hamm (1993) found that interest ratings and duration of time an observer chooses to view a visual display both load on the same factor as does the magnitude of the skin conductance response.

A straightforward prediction would be that if a task elicits an orienting response, then memory will be enhanced for the most interesting and informative aspects of the stimulus display, those aspects enjoying the beam of attention provided by the orienting response. Indeed, such memory enhancement was observed in many of the studies involving a manipulation of negative emotionality. Thus, Christianson (1992) was correct that the better memory for central details in these studies was due to a greater focusing of attention on them. However, the relevant mechanism was a qualitatively different one than the one he supposed, the orienting response of the arousal mode of attention regulation, rather than the defensive response characteristic of the activation mode of attention regulation. Hence the results of these studies are not relevant to assessing the effect of heightened stress on the fidelity of eyewitness memory.

Thus, we are limiting our focus to studies whose experimental manipulations were actually productive of a difference in stress response level. Now given that neither the seminal Shapiro and Penrod (1986) meta-analysis of the eyewitness literature nor any subsequent one has addressed the effect of heightened stress on eyewitness memory, it would be desirable to have a firm estimate of effect size and direction, both in regard to accuracy of face identification and recall of details. As mentioned earlier, this is the primary goal of the present review. Other goals include identifying variables that might moderate any consistent effect of stress on the fidelity of memory, identifying any methodological or theoretical shortcomings in the body of relevant literature, and to consider possible directions for future theoretical development and research.

METHOD

Sample

Inasmuch as the present review was part of a comprehensive meta-analysis project, a thorough search of social science citation retrieval systems was conducted. These systems included PsycINfO, Educational Resources Information Center (ERIC), *Sociological Abstracts*, *Dissertation Abstracts International*, Dissertations on-line (http://www.contentville.com/content/dissertations.asp), Medline, and Social Scisearch (the *Social Science Citation Index*). These computer database searches were supplemented with more traditional search methods, including use of bibliographic citations in published research and in social science convention proceedings and contacting leading researchers, in order to identify the most recent published research.

No unpublished studies were included, because the legal standards for proffered scientific testimony established by the US Supreme Court in *Daubert v. Merrell Dow Pharmaceuticals* (1993) have strengthened the preference by the legal system for meta-analytic conclusions based on a body of well conceived, well executed, and easily retrievable studies. In order to be included, a published study must have met three criteria: (a) a statistical test of the effect of heightened stress or anxiety on one or more measures of eyewitness memory accuracy must have been provided; (b) either stress/anxiety must have been manipulated directly or have been included as a quasi-experimental independent variable; and (c) manipulation checks must have been provided showing that putative manipulations of stress/anxiety or perceived level of violence had been successful.

There were in addition three specific exclusion criteria for studies wherein the claim was made that the relation between stress and eyewitness memory had been successfully tested. First, ratings of witness stress or anxiety, whether self-ratings or by others, had to be concurrent, as soon after encoding of the target person(s) as possible, not retrospective. Some studies of children's memory for medical procedures have involved retrospective ratings that were delayed by periods of a week up to as much as a year or more (e.g., Peterson & Bell, 1996; Quas et al., 1999). Second, measures of memory had to be from an initial assessment, a measure of memory unsullied by previous attempts at identification or recall. Again, some studies of children's memory for medical procedures have focused on long-term recall after two or more previous assessments of recall accuracy (e.g., Burgwyn-Bailes, Baker-Ward, Gordon, & Ornstein, 2001; Peterson & Whalen, 2001). Third, measures of recall accuracy that were included were only of straightforward efforts at either free recall or interrogative recall (cued recall). Measures of recall accuracy after attempts at misleading postevent suggestion were not included (e.g., Bruck, Ceci, Francoeur, & Barr, 1995).

There were two final study samples. The first sample included 16 published papers, providing a total of 27 independent estimates of effect size for heightened stress on accuracy of *face identification*. This sample included work published between 1974 and 1997, with a total of 1727 participants involved in relevant tests of the stress/anxiety effect. Sample sizes across the 27 tests of the effect ranged from 18 to 165 ($M = 64.0$).

In order to conduct a companion meta-analysis of the effects of heightened stress on accuracy of *eyewitness recall* (of perpetrator characteristics, crime scene details, and actions of central characters), another sample of 18 published papers meeting the aforementioned criteria for inclusion was collected. Ten of these published papers were also included in the prior sample, studies which had included measures both of face identification accuracy and accuracy of recall. This latter sample provided 36 independent estimates of effect size and yielded a total of 1946 participants in the various tests of the effect of stress/anxiety on accuracy of eyewitness recall. Across the 36 tests, sample sizes varied from 18 to 249 ($M = 54.1$).

Study Characteristics

The previously mentioned comprehensive meta-analysis is intended to update and to extend the one conducted by Shapiro and Penrod (1986). The current comprehensive meta-analysis, from which the current study springs, ultimately has encompassed coding of approximately 450 existing studies of face recognition from both the eyewitness and laboratory face recognition memory traditions. Several dozen independent variables were coded for each of these studies. These variables included stable (e.g., sex and race) and malleable (e.g., disguise) characteristics of both participants and targets, situational (e.g., exposure duration), and procedural (e.g., lineup presentation) factors. Variables worthy of specific mention for their usefulness in the present review included type of study (eyewitness identification study or laboratory face recognition study), whether or not the study employed a staged crime, whether the lineup included the target-present (TP) or target-absent (TA), number of participants, age of participants, and most important, whether anxiety, stress, or violence level was manipulated.[5] Dependent variables recorded were proportion correct, hit and false alarm rates, if provided, for TP lineups, correct rejection and false alarm rates, when provided, for TA lineups, and the signal detection measures, *d*-prime and beta.

Statistics

In order to test the statistical reliability of an estimate of the typical effect size found in any particular meta-analysis, we have adopted the Stouffer method (Rosenthal, 1995). Here a meta-analytic $Z(Z_{ma})$ was calculated by combining Z-scores associated with individual tests of the hypothesis that heightened stress negatively impacts eyewitness memory. The resulting algebraic sum, when divided by $\sqrt{k}$, where k is the number of independent estimates of the effect size, yields the meta-analytic Z. The probability associated with the meta-analytic Z is the overall probability of a Type I error associated with the observed pattern of results. Inasmuch as Z_{ma} provides an unweighted estimate of the overall probability level, a meta-analytic $Z(Z_{mn})$ was also calculated which weighted individual Z-scores by sample size of the study; this allows estimation of population parameters with greater emphasis on larger samples and their more reliable parameter estimates.

It should be noted that whenever recovery of sample sizes and proportion of correct identifications per condition permitted, the Z-score entered into the meta-analysis was one calculated for the difference between proportions. When an exact Z-score could not be calculated for a given effect size estimate, a Z-score associated with the p value for the estimate was entered, 1.65 for $p = .05$, for instance. When a test of the hypothesis was reported as not significant, but no statistics were cited, the conservative procedure of entering $Z = .00$ was followed (Rosenthal, 1995).

[5]To assess the reliability of coding study independent variables, two raters generated separate codings for each of 80 variables across a randomly selected 50% of the studies included in the present meta-analysis. Rate of agreement across all variables and 14 studies averaged 93%.

All studywise differences between proportions for high and low stress conditions were converted to the effect size *h*, inasmuch as *h* is the coefficient recommended by Cohen (1988) when testing for differences between proportions. Hence, mean effect size for any set of studies and its associated 95% confidence interval is expressed in terms of *h*.

Finally, given that we have included only published studies in the present meta-analysis, it is quite clear that our sample of studies is not a random sample of all studies that may actually have been conducted. As Rosenthal (1995) has pointed out, it is rather likely that published studies have reported lower probabilities of a Type I error than have those studies "squirreled away in file drawers." The concern in regards to this "file drawer problem" is that a sufficient number of such studies averaging null results could threaten a meta-analytic conclusion. We have therefore employed Rosenthal's (1995) suggested procedure for calculating a fail-safe $N(N_{fs})$ in order to determine the number of unknown or not retrieved studies averaging null results required to increase the probability of a Type I error to the just significant level of $p = .05$. Actually, inasmuch as this number is typically a whole number plus a fractional number of studies, we have adopted the rule of rounding to the next higher number. Thus, most values of the fail-safe *N* that we report in connection with a meta-analytic *Z*, represent the number of additional null results studies required to increase our probability of a Type I error to a value slightly greater than .05. Clearly, the fail-safe *N* represents a "tolerance for future null results" (Rosenthal, 1995). We would propose that at an absolute minimum the fail-safe *N* must be at least as large as the number of independent estimates of effect size that went into calculating the meta-analytic *Z*.

RESULTS

Meta-Analysis 1: Identification Accuracy

All Tests

We first sought to determine the overall status of the hypothesis that heightened stress debilitates eyewitness memory for faces. For this analysis proportion correct for the low stress condition was subtracted from that of the high stress condition. This has the virtue of producing a positive meta-analytic *Z* should high stress facilitate eyewitness memory and a negative value should high stress debilitate eyewitness memory. In this instance, overall proportion of correct identifications for the high stress condition was .42; for the low stress condition, it was .54. Mean effect size, *h*, for this analysis was −.31 (95% *CI*: −.04 to −.58); median effect size was −.27. The meta-analytic *Z* (Z_{ma}) was −6.44, $p < .0001$, $N_{fs} = 390$. Weighting each of the 27 independent tests of the hypothesis by sample size yielded $Z_{mn} = -6.03$, $p < .0001$, $N_{fs} = 336$ studies. These analyses provide clear support for the hypothesis that heightened stress has a negative impact on eyewitness identification accuracy. In subsequent analyses, we tested for moderator variables which might account for the considerable variability in effect sizes of individual tests of the hypothesis (effect size range: −3.02 to + .52; $s = .68$; see Table 1). Table 2 contains a summary of all effect sizes associated with Meta-Analyses 1 and 2.

Table 1. Identification Accuracy Effect Sizes (Studies Ordered by Date)

Author	Date	*N*	Overall (*h*)	TP (*h*)	TA (*h*)
Buckhout et al.	1974	48	−3.02	−3.02	
Mueller et al.	1979	96	.04	.04	
Nowicki et al., Exp. 1	1979	93	−.35	−.35	
Bailis & Mueller	1981	120	.13	.13	
Clifford & Hollin	1981	60	−.16	−.16	
Brigham et al.	1983	20	−.46	−.46	
Bothwell et al., Neurotics[a]	1987	35	−.74		
Bothwell et al., Stables[a]	1987	36	.52		
Cutler et al.[a]	1987	165	−.08		
Tooley, Brigham, Maas, and Bothwell	1987	96	.14	.14	
Peters, TP	1988	106	−.51	−.51	
Peters, TA	1988	106	−.18		−.18
Maas & Kohnken[b]	1989	86	−.43		−.43
Hosch & Bothwell, Exp. 1[a]	1990	39	.49		
Kramer et al., Exp. 1[b]	1990	64	−.58	−.58	
Goodman et al., Exp. 1	1991	18	−.47	−.47	
Goodman et al., Exp. 2	1991	47	−.27	−.27	
Goodman et al., Exp. 3	1991	34	.24	.24	
Peters, Exp. 1, TP	1991	36	−.57	−.57	
Peters, Exp. 1, TA	1991	35	−.04		−.04
Peters, Exp. 2, TP	1991	34	−.61	−.61	
Peters, Exp. 2, TA	1991	33	.32		.32
Peters, Exp. 3, TP	1991	32	−1.32	−1.32	
Peters, Exp. 3, TA	1991	32	.39		.39
Peters, Exp. 4[a]	1991	96	−.12		
Peters, Exp. 1[a]	1997	64	−.42		
Peters, Exp. 2[a]	1997	96	−.31		

[a]Only overall proportion correct reported.
[b]Weapon visibility totally confounded with anxiety level; thus reportedas a stress effect.

Lineup Type

Because of a comment by Peters (1988) that he had not found the same statistically reliable difference between high and low stress conditions for TA lineups that he had found for TP lineups and because of the forensic implications of TA and TP lineups, we decided to code both conditions as independent estimates of the effect of heightened stress on eyewitness identification for all those studies manipulating lineup type as a between subjects variable (see Table 1). Lineup type clearly emerged as the most powerful moderator of the impact of stress on face identification accuracy (hits and correct rejections). TP lineups ($N = 15$) generated a mean effect size $h = -.52$, 95% *CI*: −.08 to −.96. Here $Z_{\text{ma}} = -7.08$, $p < .0001$, and $N_{\text{fs}} = 264$. On the other hand, TA lineups ($N = 5$) generated a negligible mean effect size, $h = +.01$ (95% *CI*: −.39 to + .41). In this case, $Z_{\text{ma}} = -.56$, which was clearly not statistically reliable. Yet another way to examine the different impact of heightened stress on face identification accuracy as a function of lineup type is to note that mean proportions correct for TP lineups under high and low stress conditions were .39 and .59, respectively. Corresponding mean proportions correct for TA lineups were .34 and .34. Thus, the overall negative impact of heightened stress on accuracy of face identification was due entirely to a substantial effect on hit rate for TP lineups. The correct rejection rate for TA lineups was unaffected by stress level.

Table 2. Meta-Analysis Effect Sizes

Type of Analysis	Effect size *h*	Effect size *d*
Identification accuracy		
All tests[a]	−.31	
All tests[b]	+ .22	
TP lineups[a]	−.52	
TA lineups[a]	+ .01	
TP lineups[b]	+.37	
TA lineups[b]	.00	
Identification paradigm[a]	−.36	
Recognition paradigm[a]	−.10	
Staged crimes[a]	−.58	
Other stressors[a]	−.28	
Adult witnesses[a]	−.34	
Child witnesses[a]	−.27	
Recall accuracy		
All tests		−.31
Interrogative recall		−.34
Narrative recall		−.20
Adult witnesses		−.44
Child witnesses		−.06
Staged crimes		−.45
Other stressors		−.16

[a]Overall proportion correct, including hit and correct rejection rates.
[b]False alarm rate.

Lineup type moderated the effects of heightened stress on the rate at which faces were falsely identified, as well. In this instance, mean false alarm rates for TP lineups under high and low stress conditions were .34 and .19, respectively. Comparable false alarm rates for TA lineups were .66 and .65. Accordingly, it is not surprising that TP lineups ($N = 7$) generated a mean effect size $h = +.37$, 95% *CI*: + .05 to + .69. Though $Z_{\mathrm{ma}} = +3.36$, $p = .0004$, the fail-safe N was only 23 additional null-results studies. In parallel with the results for proportion correct, the false alarm rates for TA lineups ($N = 5$) were not differentially affected by stress levels, with the result that mean effect size was negligible, $h = .00$ (95% *CI*: −.43 to +.43). The meta-analytic Z_{ma} was + .56, certainly not significant. Clearly, the overall tendency of heightened stress to increase false alarm rates in a face identification task, $h = +.22$ (95% *CI*: −.02 to +.46), $Z_{\mathrm{ma}} = +2.92$, $p = .0018$, $N_{\mathrm{fs}} = 26$, was due entirely to a substantial effect for TP lineups.

Research Paradigm

A somewhat smaller, though still sizeable moderator effect was found for the variable of research paradigm, whether the study was conducted as a standard face recognition task in the tradition of cognitive psychology or whether it was conducted in the context of the eyewitness identification paradigm. The face recognition task exposes observers to a relatively large number of target faces (at least 24 in studies included in the present meta-analysis). A recognition memory test following exposure to the target faces usually includes twice as many faces, the targets plus an equal number of unfamiliar distracter faces. Observers are exposed to the test faces one at a time and are instructed to respond "yes" or "no" as to whether a given face had been exposed previously. Studies conducted in the eyewitness identification paradigm usually expose witnesses to just one or two target faces, the perpetrator(s), and memory for each target's face is tested either by embedding his/her face in a 5–9-person simultaneously or serially presented live lineup or photo spread (TP lineup) or else by substituting someone else who is a match to the perpetrator's description (TA lineup). Witnesses are asked to identify the perpetrator or to indicate that he/she is not in the lineup.

For the face recognition studies included in our sample ($N = 5$) the mean proportion correct under high stress conditions was .56 and was .58 under low stress conditions. Not surprisingly, the mean effect size in this instance was only −.10 (95% *CI*: −.45 to +.25). Even though the meta-analytic Z was significant $Z_{\text{ma}} = -2.46$, $p = .0069$, the fail-safe N was only seven additional null results studies.

Mean proportions correct were .39 and .53 under high and low stress conditions, respectively, for witnesses in the 22 studies executed in the more ecologically valid eyewitness identification tradition. This difference resulted in a mean effect size $h = -.36$, 95% *CI*: −.04 to −.68. The debilitating effect of heightened stress on eyewitness memory for studies conducted in the eyewitness identification paradigm was a statistically reliable one, $Z_{\text{ma}} = -6.00$, $p < .0001$, a conclusion not likely to be overturned by unknown null results studies, $N_{\text{fs}} = 269$. Clearly, heightened stress is much more likely to have a debilitating effect on memory for the human face when encoding and memory testing occur under the requirements of the eyewitness identification paradigm than when encoding and testing occur under conditions of the face recognition memory paradigm.

Presence/Absence of a Staged Crime

Within the 22 eyewitness identification studies, six manipulated stress in the context of a staged crime, and 16 manipulated stress by some other means, threat of an injection, for instance. Mean proportions correct under high and low stress conditions were .33 and .50, respectively, for the staged-crime studies. For the studies manipulating stress by other means, the comparable means were .56 and .69. Even though the adverse effect of heightened stress on eyewitness memory was statistically reliable for both sets of studies, $Z_{\text{ma}} = -3.82$, $p < .0001$, $N_{\text{fs}} = 27$ for the staged-crime studies and $Z_{\text{ma}} = -4.68$, $p < .0001$, $N_{\text{fs}} = 113$ for the other studies, there was a pronounced difference in mean effect sizes generated. Mean effect sizes were $h = -.58$, 95% *CI*: −1.88 to + .72, for the staged crime studies and for the studies manipulating stress by other means, a smaller $h = -.28$, 95% *CI*: −.02 to −.54. One study (Buckhout, Alper, Chern, Silverberg, & Slomovits, 1974) was responsible for most of this difference in effect sizes, however. Nevertheless, it should be noted that the study of Buckhout et al. (1974) was a rather realistic, *live* staged crime, rather than a filmed one.

Age

Given that 15 of our independent estimates of the effect size of heightened stress on face identification were produced by adult witnesses/observers and that 12 were produced by children (ages ranged from 3 to 10 years), we decided to test age as a moderator of effect size. It was also of interest to assess age as a possible moderator, given the concern about competency of child witnesses (e.g., Goodman, Hirschman, Hepps, & Rudy, 1991). At least for face identification accuracy, witness age appears to have contributed little to the variability in effect size. Mean proportions correct for children under high and low stress conditions were .42 and .55, respectively, while for adults the proportions were .42 and .54. These differences generated average effect sizes that were comparable, $h = -.27$, 95% *CI*: −.57 to + .03, for the children and $h = -.34$, 95% *CI*: −.80 to + .12, for the adults. The debilitating effects of heightened stress on identification accuracy were statistically reliable in both instances. For the children, $Z_{ma} = -3.43, p < .0003, N_{fs} = 41$; for the adults, $Z_{ma} = -5.61, p < .0001, N_{fs} = 159$ studies.

Meta-Analysis 2: Accuracy of Eyewitness Recall

All Tests

We should first note that it was not possible to calculate the effect size h across all 36 tests of the hypothesis that heightened stress debilitates eyewitness recall. In only five of these instances did investigators report proportion of details correctly recalled as a function of stress level; for these five studies $h = -.25$, corresponding to an average proportion correctly recalled of .52 in the high stress condition and .64 in the low stress condition. Consequently, d was adopted as a substitute measure of effect size. Again, Table 2 summarizes all effect sizes reported for Meta-Analysis 2.

Both meta-analytic Zs were statistically significant, $Z_{ma} = -5.40$, $p < .0001$, $N_{fs} = 355$ studies, and $Z_{mn} = -6.06, p < .0001, N_{fs} = 453$ studies. Calculation of the mean effect size yielded $d = -.31$, 95% *CI*: −.14 to −.48. Clearly, heightened stress produces the same debilitating effect on accuracy of eyewitness recall as it does on identification accuracy. In the remaining analyses, we tested for variables that might have moderated this effect on recall.

Type of Recall

Eight estimates of effect size were associated with narrative recall (free recall); the remaining 28 estimates involved some form of interrogative recall (specific questions). The meta-analytic Z for narrative recall was not statistically reliable, $Z_{ma} = -1.17$ .Average effect size was $d = -.20$ (95% *CI*: −.68 to + .28), all the effect being generated by a single study (Clifford & Scott, 1978). For interrogative recall, $Z_{ma} = -5.50, p < .0001, N_{fs} = 288$. Here d was −.34, with a 95% *CI* extending from −.15 to −.53. Heightened stress would appear to impact interrogative recall much more negatively than narrative or free recall.

Age

In this sample of effect size estimates, there were 23 tests with adult witnesses and 13 tests with children as witnesses (age range again 3–10 years). Unlike the situation with identification accuracy, here age emerged as an important moderator. For adults, $d = -.44$ (95% *CI*: −.19 to −.69), while for children, *d* was a negligible value, −.06 (95% *CI*: −.16 to + .04). There was a significant meta-analytic *Z* for adult eyewitnesses, $Z_{\text{ma}} = -6.05, p < .0001, N_{\text{fs}} = 286$ studies. For child eyewitnesses, on the other hand, $Z_{\text{ma}} = -1.01$, a not statistically reliable value. Surprisingly enough, heightened stress debilitated eyewitness recall for adults, but not for children. However, before concluding that the null hypothesis that $Z_{\text{ma}} = 0.00$ might have validity for children, one should consider that in this instance the *counternull hypothesis* (Rosenthal, 1995), $Z_{\text{ma}} = -2.02$, is just as likely to be true.[6] Even were the counternull hypothesis true, however, it would still be the case that witness age is an important moderator of the effect of heightened stress on eyewitness recall.

Presence/Absence of a Staged Crime

Though the distinction between face recognition and eyewitness identification research paradigms is not applicable to eyewitness recall, the distinction between presence and absence of a staged crime is indeed relevant. In our sample of studies assessing eyewitness recall as a function of stress level, there were 18 independent estimates of effect size that included a staged crime, on film or live, and 18 that did not. Meta-analytic *Z*s were statistically significant in both instances, $Z_{\text{ma}} = -4.67$, $p < .0001$, $N_{\text{fs}} = 127$ for the studies including a staged crime and $Z_{\text{ma}} = -3.01$, $p = .0013$, $N_{\text{fs}} = 42$ for those investigations not including a staged crime. Despite both conditions producing statistically reliable decrements in recall under higher levels of stress, the presence of a staged crime would appear to have generated a somewhat greater decrement. In support of this assertion, we may note that the effect size generated by the staged crime studies ($d = -.45$; 95% *CI*: −.17 to −.73) was more than twice that generated by the studies employing other means to induce stress ($d = -.16$; 95% *CI*: −.35 to + .03).

DISCUSSION

By adopting our particular inclusion criteria for our two samples of studies, we sought to limit our focus to experimental manipulations productive of defensive responses to stimulating conditions. In so doing, we have adduced considerable support for the hypothesis that high levels of stress negatively impact both accuracy of eyewitness identification as well as accuracy of recall of crime-related details. For eyewitness identification, the average effect size *h* was −.31, with a 95% *CI* that did not include zero. Whether unweighted or weighted by sample size, the meta-analytic *Z*s were associated with fail-safe *N*s of 300–400 studies. Thus, the current

[6]Here the counternull value of the effect size of heightened stress is found by doubling the obtained effect size ($Z_{\text{ma}} = -1.01$) and subtracting the effect size expected under the null hypothesis, .00.

meta-analytic conclusion regarding the negative effect of stress on eyewitness identification accuracy is unlikely to be overturned any time soon by unknown findings averaging null results. The conclusion that heightened stress debilitates eyewitness recall (average effect size $d = -.31$), too, is at least as safe, with fail-safe *N*s of more than 350 studies associated with the overall meta-analytic *Z*s.

In addition to ascertaining the direction and magnitude of the effect of heightened stress on accuracy of eyewitness memory, we had sought to identify variables that might moderate any consistent effects of stress on the fidelity of memory. We found two principal moderators of the negative effect of heightened stress on eyewitness identification accuracy, lineup type and research paradigm. Certainly there is precedent for the importance of lineup type as a moderator of accuracy in meta-analyses of the eyewitness identification literature (e.g., Steblay, 1997). Steblay found a moderate size effect for unbiased instructions to increase accuracy in TA lineups but to have no effect on accuracy in TP lineups. The moderator effect for lineup type found here was the reverse of the one found by Steblay. Here TP lineups generated an effect size ($h = -.52$), such that face identification accuracy was much reduced under conditions of heightened stress as compared with low stress conditions, a difference of .20 in mean proportion correct/hit rate (.39 versus .59, respectively).[7]

Quite possibly, encoding of a target person under conditions of heightened stress reduces the veridicality of a witness's memory representation of him sufficiently to decrease the probability of a match between that representation and the target when present (TP lineups). Such an effect of stress would certainly serve to reduce the hit rate. The proportion correct measure for TA lineups, the correct rejection rate, might be similarly affected, inasmuch as reduced quality of a witness's memory representation of the target would not provide as good a basis for rejecting a lineup that does not contain his face. On the other hand, members of a fair TA lineup only roughly resemble the target person; their faces are consistent with just a witness's prior verbal description of the target, not necessarily with a high-quality *visual* representation of him. Perhaps, therefore, differences between the relatively nondegraded and stress-degraded visual memory representations of the target are insufficient to affect the basis for deciding that the target face is not present in the TA lineup. Indeed there was a negligible effect size of heightened stress on accuracy in TA lineups; proportion correct/correct rejection rates were .34, regardless of stress level. Inasmuch as correct rejection rates and false alarm rates for TA lineups must sum to 1.00, there would likewise be no difference expected in false alarm rates for TA lineups as a function of differences in stress at encoding, .66 in both cases.

We should note in passing that lineup type moderated the effect of heightened stress on the false alarm rate, as well. TP lineups generated an effect size ($h = +.37$) that was considerably larger than the negligible effect generated by TA lineups ($h = .00$). This TP effect corresponded to false alarm rates of .34 under high

[7] Another and perhaps more meaningful way to interpret a difference in proportion correct of this magnitude has been suggested by G. L. Wells (cited in Steblay, 1997). Consider, for instance, 1000 TP lineups conducted over a period of time. Given the accuracy difference we have obtained, we would expect TP lineups to generate 200 more correct identifications of perpetrators witnessed under low stress conditions than of perpetrators witnessed under high stress conditions.

stress conditions and .19 under low stress conditions. The aforementioned stress-induced memory degradation could account for the increased false alarm rate, as well, if indirectly. Having reduced the probability of a match to the target in a TP lineup, the stress-induced loss of memory fidelity would then serve to increase the probability of a "match" to some other lineup member, in direct proportion to the motivation of the eyewitness to choose someone from the lineup. In the present instance, much of the stress-induced lowering of the hit rate (.20) was likely "transferred" to an increase in the false alarm rate (.15); the remaining decrease in the hit rate (.05) might have resulted in an increase in the false rejection rate for a TP lineup, the latter rate rarely reported in the literature. Obviously, in a TA lineup there is no possibility of a similar transfer of witness choices from one person, the target, to other possible choices. At any event, our results portend that a greater proportion of high-stress witnesses than low-stress witnesses will choose a foil from a TP lineup and will thus have their subsequent credibility as a witness undermined.

We should also note that even though the stress effect resides in TP arrays, plausible arguments can be made that the mix of guilty and innocent persons identified by witnesses will change as a consequence, even if the proportion of positive identifications from TA lineups does not. For simplicity's sake, let us assume that of all lineups conducted by police, half are TP and half are TA. Given the .59 hit rate in low-stress TP lineups and the .66 false alarm rate obtained in our sample of TA lineups, this means that 59 guilty perpetrators would be identified from every 100 TP lineups and 66 mistaken identifications would be made from every 100 TA lineups. If we were also to assume that all lineups were perfectly fair 6-person arrays, an innocent suspect embedded in a TA lineup would be mistakenly identified 11 times (66/6) per 100 arrays. Hence, the resulting pool of identifications would be 59 guilty and 11 innocent, an accuracy rate of .84 for choosers of suspects—of course, we should not forget that 41 guilty perpetrators would not be identified. Extending our argument to the high-stress situation, our results imply that 100 TP arrays would generate identifications of 39 guilty perpetrators and that 100 TA arrays would generate 66 mistaken identifications. An innocent suspect in the TA lineups would again be chosen 11 times; the resulting pool of identifications would be 39 guilty and 11 innocent, an accuracy rate of .78 for choosers of suspects.

The mix of guilty and innocent suspects would change even further, if the lineup arrays were to be as biased as those studied by researchers to date (Penrod, 2003). The bias is such that we would expect that the innocent suspect to be chosen 2–3 times as often as the average foil. If the actual multiple were 2.5, then .33 of choices from a 6-person TA lineup would be the innocent suspect ($2.5/2.5 + 1 + 1 + 1 + 1 + 1$). This means that the innocent suspect would be chosen by 22 witnesses from a viewing of 100 TA lineups. Consequently, the resulting pool of identifications would be 59 guilty plus 22 innocent for low-stress witnesses (73% accuracy) and 39 guilty plus 22 innocent (64% accuracy) for high stress witnesses. In short, even if heightened stress does not impact TA lineup false alarm rates, its impact on TP lineup hit rates can change materially the mix of guilty and innocent suspects identified by witnesses and the mix of correct and incorrect identifications presented to jurors.

Perhaps not surprisingly, nature of the research paradigm was also an important moderator of the effect of stress level on eyewitness identification accuracy. Mean effect size for stress level ($h = -.36$) was more than three times as great when studies were executed under the more ecologically valid conditions of the eyewitness identification paradigm than when executed within the parameters of a standard laboratory face recognition task ($h = -.10$).

Two variables were likewise identified as important moderators of the effect of heightened stress on accuracy of eyewitness recall, type of recall and witness age. Though it is not obvious why, it is clear that high stress levels impact interrogative recall much more negatively than they do narrative or free recall. Possibly the negative impact of a heightened stress level is moderated by the witness in a narrative recall situation having control over what to report and in what order. It is likewise not obvious why witness age should have emerged as a substantial moderator of the effect of heightened stress on recall, when it did not act as a moderator of stress effects on accuracy of face identification. However, there is support for the notion that measures of facial recognition and facial recall are independent, uncorrelated (Bothwell et al., 1987; Jenkins & Davies, 1985; Pigott & Brigham, 1985; Pigott, Brigham, & Bothwell, 1990). For instance, Bothwell et al. (1987) found an interaction between manipulated stress level and neuroticism on accuracy of facial identification but did not obtain a similar interaction for a measure of facial recall, description accuracy.

Our final goals for the present review were to identify any methodological or theoretical shortcomings in the body of relevant literature and to consider possible directions for future theoretical development and research. As indicated earlier, there was quite likely a major methodological difficulty with a number of studies showing apparent facilitation of memory by increases in arousal or what was defined as negative emotionality (e.g., Burke et al., 1992; Christianson, 1984; Christianson et al., 1991; Heuer & Reisberg, 1990; Libkuman et al., 1999; Safer et al., 1998). The negative emotionality manipulation in these studies generated an orienting response (arousal mode of attention control) to stimulating conditions, rather than the defensive response (activation mode of attention control) typically produced by a successful manipulation of stress or anxiety (Deffenbacher, 1994, 1999). Hence, investigators of stress effects on memory need to be concerned as to whether their experimental manipulations are eliciting the arousal mode of attention regulation or the activation mode.

There is yet another matter of methodological and theoretical importance that investigators should consider in future research, the issue of individual differences, whether they be differences in state or trait anxiety, neuroticism, specific fears, or physiological reactivity. These differences turn out to be very important. Very different patterns of response to the same stimulus situation may be shown. If an investigator were not aware of this possibility, two quite different patterns of response may cancel each other, leading to the unfortunate conclusion that an increased stress level had no demonstrable impact on memory performance. Consider an instance provided by a study (Bothwell et al., 1987) included in both of the present meta-analyses. Bothwell et al. split witnesses at the median on a scale of neuroticism, with neurotics scoring above the median and stables (emotionally stable) scoring below. Neurotics are theorized to have very low thresholds for emotional arousal and to be predisposed to perceive a wide range of objectively nondangerous situations as threatening and to respond with autonomic activation. Stables, on the other hand, would tend to be less anxious and physiologically reactive. The effect of stress level in their study varied dramatically with the level of neuroticism. As stress level increased from low to moderate to high, stables showed an increased level of identification accuracy, proportion correct increasing from .50 to .62 to .75, respectively. In dramatic opposition to this pattern of results was that of the neurotics; as stress level increased, proportions correct were .68, .68, and .32.

Consider one more example from another study included in both our meta-analyses, Peters' (1988) study of university students getting inoculated at a Department of Health clinic. Students were asked for physical descriptions and photo lineup identifications of both the inoculating nurse and of a second person who took their pulse 2 min later. Heart rate averaged 88 b.p.m. at inoculation versus 71 b.p.m. 2 min later. Identification accuracy overall was 66% for the second person but only 41% for the nurse. Individual differences in physiological reactivity had a profound effect on identification accuracy for the inoculating nurse. The 20 most physiologically reactive witnesses (39 b.p.m. average difference between inoculation and two minutes later) demonstrated an identification accuracy level of 31%; the 20 least physiologically reactive witnesses (3 b.p.m. difference), on the other hand, displayed an identification accuracy level of 59% for the nurse. The former witnesses clearly defined the inoculation situation as one requiring vigilance, if not actually escape or avoidance. The latter appeared to have defined the situation as more nearly one of informative perceptual intake and did not suffer the catastrophic drop in memory accuracy of the more physiologically reactive witnesses.

Hence, researchers should pay particular attention to the nature of their task. It may well be defined differently by different observers, whether it be one of simple perceptual intake or one of vigilance, for example. Whether the task be one that ordinarily produces an orienting response or a defensive response, physiological and self-report data must be examined carefully for the presence of the alternative response set in individual observers.

Thus, the modest size of the debilitating effect of heightened stress on the accuracy of eyewitness memory obtained in these meta-analyses may well be due to the averaging of its effects on two categories of witnesses, with those more anxious and physiologically reactive persons suffering a more serious drop in accuracy than those more emotionally stable persons. However, we would be remiss if we did not issue a further caveat concerning the modest obtained effect size. Whether the stress manipulation was a realistic and unexpected theft, a particularly violent video, or perhaps the threat of an injection or mild electric shock, all stress manipulations in the literature that we have examined quite likely do not reach the stress-inducing levels of extra-laboratory violent crime scenes. Thus an effect size of −.31 is perhaps a serious underestimate of the debilitating effects of stress engendered by violent crime.

For instance, results of a recent study by Ihlebaek, Love, Eilertsen, and Magnussen (2003) demonstrated that witnesses to a live staged robbery reported fewer details about the criminal event and with less accuracy than did witnesses viewing a video recording of the same event, even though the pattern of memory errors was similar in both conditions. We agree with Ihlebaek et al. (2003) that results of laboratory studies may be an overestimate of eyewitness memory performance, especially we might add, when eyewitnesses are in a state of heightened stress.

Indeed, Morgan et al. (2004) have provided strong support for this latter caveat. They studied eyewitness capabilities of more than 500 active-duty military personnel enrolled in a survival school program. After 12 hr of confinement in a mock prisoner of war camp, participants experienced 4 hr apart, both a high-stress interrogation with real physical confrontation and a low-stress interrogation without physical confrontation; interrogations were 40 min in length. The interrogators in each instance were different individuals, with order of interrogation being counterbalanced across participants. A day after release from the prisoner of war camp, and having recovered from food and sleep deprivation, participants viewed a 15-person live lineup, a 16-person photo-spread, or a serial-presentation photo lineup of up to 16 persons. Regardless of testing method, memory accuracy for the high-stress interrogator suffered the same catastrophic decline from the level displayed for the low-stress interrogator, the same sort of catastrophic decline noted by Bothwell et al. (1987) and Peters (1988). Consider just the results from the live lineup condition: For the low stress condition, the hit and false alarm rates were .62 and .35, respectively, but comparable rates for the high stress condition were .27 and .73.

REFERENCES

References marked with an asterisk indicate studies included in the meta-analyses.

*Bailis, K. L., & Mueller, J. H. (1981). Anxiety, feedback, and self-reference in face recognition. *Motivation and Emotion*, *5,* 85–96.

*Bornstein, B. H., Liebel, L. M., & Scarberry, N. C. (1998). Repeated testing in eyewitness memory: A means to improve recall of a negative emotional event. *Applied Cognitive Psychology*, *12,* 119–131.

*Bothwell, R. K., Brigham, J. C., & Pigott, M. A. (1987). An exploratory study of personality differences in eyewitness memory. *Journal of Social Behavior and Personality*, *2,* 335–343.

*Brigham, J. C., Maass, A., Martinez, D., & Wittenberger, G. (1983). The effect of arousal on facial recognition. *Basic and Applied Social Psychology*, *4,* 279–293.

Bruck, M., Ceci, S. J., Francouer, E., & Barr, R. (1995). "I hardly cried when I got my shot!" Influencing children's reports about a visit to their pediatrician. *Child Development*, *66,* 193–208.

*Buckhout, R., Alper, A., Chern, S., Silverberg, G., & Slomovits, M. (1974). Determinants of eyewitness performance on a lineup. *Bulletin of the Psychonomic Society*, *4,* 191–192.

Burgwyn-Bailes, E., Baker-Ward, L., Gordon, B. N., & Ornstein, P. A. (2001). Children's memory for emergency medical treatment after one year: The impact of individual difference variables on recall and suggestibility. *Applied Cognitive Psychology*, *15,* 25–48.

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

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

Christianson, S.-A. (1992). Emotional stress and eyewitness memory: A critical review. *Psychological Bulletin*, *112,* 284–309.

Christianson, S.-A., Loftus, E. F., Hoffman, H., & Loftus, G. R. (1991). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17,* 693–701.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

*Clifford, B. R., & Hollin, C. R. (1981). Effects of the type of incident and the number of perpetrators on eyewitness memory. *Journal of Applied Psychology*, *66,* 364–370.

*Clifford, B. R., & Scott, J. (1978). Individual and situational factors in eyewitness testimony. *Journal of Applied Psychology*, *63,* 352–259.

*Cutler, B. L., Penrod, S. D., & Martens, T. K. (1987). The reliability of eyewitness identification: The role of system and estimator variables. *Law and Human Behavior*, *11,* 233–258.

Daubert v. Merrell Dow Pharmaceuticals (1993). 509 US 579.

Deffenbacher, K. A. (1983). The influence of arousal on reliability of testimony. In S. M. A. Lloyd-Bostock & B. R. Clifford (Eds.), *Evaluating witness evidence* (pp. 235–251). Chichester, England: Wiley.

Deffenbacher, K. A. (1994). Effects of arousal on everyday memory. *Human Performance*, *7,* 141–161.

Deffenbacher, K. A. (1999, July). *Negative emotionality and everyday memory: Disentangling the effects of defensive and orientation response sets*. Paper presented at the meeting of the Society for Applied Research in Memory and Cognition, Boulder, CO.

Fazey, J. A., & Hardy, L. (1988). *The inverted-U hypothesis: A catastrophe for sport psychology* (British Association for Sports Sciences Monograph No. 1). Leeds, England: National Coaching Foundation.

*Goodman, G. S., Hirschman, J. E., Hepps, D., & Rudy, L. (1991). Children's memory for stressful events. *Merrill Palmer Quarterly*, *37,* 109–158.

Hardy, L., & Parfitt, G. (1991). A catastrophe model of anxiety and performance. *British Journal of Psychology*, *82,* 163–178.

Hare, R. (1972). Cardiovascular components of orienting and defensive responses. *Psychophysiology*, *9,* 606–614.

Hare, R., Wood, K., Britain, S., & Frazelle, J. (1971). Autonomic responses to affective visual stimulation: Sex differences. *Journal of Experimental Research in Personality*, *5,* 14–22.

Hare, R., Wood, K., Britain, S., & Shadman, J. (1970). Autonomic responses to affective visual stimulation. *Psychophysiology*, *7,* 408–417.

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

*Hosch, H. M., & Bothwell, R. K. (1990). Arousal, description and identification accuracy of victims and bystanders. *Journal of Social Behavior and Personality*, *5,* 481–488.

Ihlebaek, C., Love, T., Eilertsen, D. E., & Magnussen, S. (2003). Memory for a staged criminal event witnessed live and on video. *Memory*, *11,* 319–327.

Jenkins, F., & Davies, G. M. (1985). Contamination of facial memory through exposure to misleading composite pictures. *Journal of Applied Psychology*, *70,* 164–176.

Klorman, R., Weissberg, R. P., & Wiesenfeld, A. R. (1977). Individual differences in fear and autonomic reactions to affective stimulation. *Psychophysiology*, *14,* 45–51.

*Kramer, T. H., Buckhout, R., & Eugenio, P. (1990). Weapon focus, arousal, and eyewitness memory: Attention must be paid. *Law and Human Behavior*, *14,* 167–184.

Lacey, J., & Lacey, B. (1974). On heart rate responses and behavior: A reply to Elliott. *Journal of Personality and Social Psychology*, *30,* 1–18.

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral actions. *Psychophysiology*, *30,* 261–273.

Libkuman, T. M., Nichols-Whitehead, P., Griffith, J., & Thomas, R. (1999). Source of arousal and memory for detail. *Memory & Cognition*, *27,* 166–190.

*Loftus, E. F., & Burns, T. E. (1982). Mental shock can produce retrograde amnesia. *Memory & Cognition*, *10,* 318–323.

*Maas, A., & Kohnken, G. (1989). Eyewitness identification: Simulating the "weapon effect." *Law and Human Behavior*, *13,* 397–408.

Morgan, C. A., Hazlett, G., Doran, A., Garrett, S., Hoyt, G., Thomas, P. et al. (2004). Accuracy of eyewitness memory for persons encountered during exposure to highly intense stress. *International Journal of Law and Psychiatry*, *27,* 265-279.

*Mueller, J. H., Bailis, K. L., & Goldstein, A. G. (1979). Depth of processing and anxiety in facial recognition. *British Journal of Psychology*, *70,* 511–515.

*Nowicki, S., Winograd, E., & Millard, B. A. (1979). Memory for faces: A social learning analysis. *Journal of Research in Personality*, *13,* 460–468.

Penrod, S. D. (2003). Eyewitness identification evidence: How well are witnesses and police performing? *Criminal Justice, Spring*. 36–47, 54.

*Peters, D. P. (1988). Eyewitness memory in a natural setting. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues: Vol. 1. Memory in everyday life* (pp. 89–94). Chichester, UK: Wiley.

*Peters, D. P. (1991). The influence of stress and arousal on the child witness. In J. Doris (Ed.), *The suggestibility of children's recollections* (pp. 60–76). Washington, DC: American Psychological Association.

*Peters, D. P. (1997). Stress, arousal, and children's eyewitness memory. In N. L. Stein, P. A. Ornstein, B. Tversky, & C. J. Brainerd (Eds.), *Memory for everyday and emotional events*. Mahwah, NJ: Erlbaum.

Peterson, C., & Bell, M. (1996). Children's memory for traumatic injury. *Child Development*, *67,* 3045–3070.

Peterson, C., & Whalen, N. (2001). Five years later: Children's memory for medical emergencies. *Applied Cognitive Psychology*, *15,* 7–24.

Pigott, M. A., & Brigham, J. C. (1985). Relationship between accuracy of prior description and facial recognition. *Journal of Applied Psychology*, *70,* 547–555.

Pigott, M. A., Brigham, J. C., & Bothwell, R. K. (1990). A field study on the relationship between quality of eyewitnesses' descriptions and identification accuracy. *Journal of Police Science and Administration*, *17,* 84–88.

Quas, J. A., Goodman, G. S., Bidrose, S., Pipe, M., Craw, S., & Ablin, D. S. (1999). Emotion and memory: Children's long-term remembering, forgetting, and suggestibility. *Journal of Experimental Child Psychology*, *72,* 235–270.

*Read, J. D., Yuille, J. C., & Tollestrup, P. (1992). Recollections of a robbery: Effects of arousal and alcohol upon recall and person identification. *Law and Human Behavior*, *16,* 425–446.

Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin*, *118,* 183–192.

Safer, M. A., Christianson, S.-A., Autry, M. W., & Osterlund, K. (1998). Tunnel memory for traumatic events. *Applied Cognitive Psychology*, *12,* 99–117.

Shapiro, P. N., & Penrod, S. D. (1986). A meta-analysis of facial identification studies. *Psychological Bulletin*, *100,* 139–156.

*Shrimpton, S., Oates, K., & Hayes, S. (1998). Children's memory of events: Effects of stress, age, time delay and location of interview. *Applied Cognitive Psychology*, *12,* 133–143.

*Siegel, J. M., & Loftus, E. F. (1978). Impact of anxiety and life stress upon eyewitness testimony. *Bulletin of the Psychonomic Society*, *4,* 191–192.

*Stanny, C. J., & Johnson, T. C. (2000). Effects of stress induced by a simulated shooting on recall by police and citizen witnesses. *American Journal of Psychology*, *113,* 359–386.

Steblay, N. M. (1997). Social influence in eyewitness recall: A meta-analytic review of lineup instruction effects. *Law and Human Behavior*, *21,* 283–297.

*Tooley, V., Brigham, J. C., Maas, A., & Bothwell, R. K. (1987). Facial recognition: Weapon effect and attentional focus. *Journal of Applied Social Psychology*, *17,* 845–859.

Tucker, D. M., & Williamson, P. A. (1984). Asymmetric neural control systems in human self-regulation. *Psychological Review*, *91,* 185–215.

*Zanni, G. R., & Offermann, J. T. (1978). Eyewitness testimony: An exploration of question wording upon recall as a function of neuroticism. *Perceptual and Motor Skills*, *46,* 163–166.