

Accuracy of Eyewitness Identification is significantly associated with performance on a standardized test of face recognition

Charles A. Morgan III^{a,b,*}, Gary Hazlett^c, Madelon Baranoski^b,
Anthony Doran^d, Steven Southwick^{a,b}, Elizabeth Loftus^e

^a Department of Psychiatry, Yale University School of Medicine, United States

^b Section of Law and Psychiatry, Yale University School of Medicine, United States

^c JFK Special Warfare Training Center and School, Fort Bragg, N.C., United States

^d FASOTRAGRULANT, Naval Air Station, Brunswick, ME, United States

^e Department of Psychology and Social Behavior, University of California, Irvine, United States

Abstract

Objectives: This study assessed the relationship between Eyewitness Accuracy regarding a person met under conditions of high stress and performance on a standardized, neutral test of memory for human faces.

Method: Fifty-three U.S. Army personnel were exposed to interrogation stress. Forty-eight hours later, participants were administered the Weschler Face Test and then asked to identify the one interrogator who they had encountered 48 h earlier.

Results: A significant positive relationship was observed between performance on the Weschler Face Test and performance on the Eyewitness task. Inaccurate eyewitnesses exhibited more False Negative errors when performing the Weschler Face Test.

Discussion: Trait ability to remember human faces may be related to how accurately people recall faces that are associated with highly emotional circumstances. Detection probability methods, such as ROC curve analyses, may be of assistance to forensic examiners, the police, and the courts, when assessing the probability that eyewitness evidence is accurate.

© 2007 Elsevier Inc. All rights reserved.

1. Introduction

The vast majority of studies designed to assess eyewitness memory have been conducted in the laboratory and have used videotapes or live simulations of crime events; only a handful have been field studied involving genuine victims and eyewitnesses of real crimes. While informative, laboratory studies have limited applicability to the real world because crime simulations do not involve the high degree of personal “threat” or “alarm” experienced during actual life threatening events (Penrod, Fulero, & Cutler, 1995).

Military survival training offers a unique opportunity to study Eyewitness Accuracy after a naturally occurring, controlled, stressful event. In past research using this methodology, we examined the accuracy of eyewitness memory in soldiers for an “Interrogator/Perpetrator” who subjected them to high stress 48 h earlier during their confinement in

* Corresponding author. National Center for PTSD, 116A, VA Connecticut Healthcare Systems, West Haven, CT, 06516, United States. Tel.: +1 203 932 5711x2464; fax: +1 203 937 3886.

E-mail address: charles.a.morgan@yale.edu (C.A. Morgan).

mock captivity (Morgan et al., 2004). In that study, we used three established law enforcement methods for identifying crime suspects (the live line up; the photo-spread technique; the sequential photo method) and for comparing the accuracy of participants' eyewitness recognition when asked to identify the instructors who conducted their interrogations. We also assessed differences in accuracy in Eyewitness Identification for cued photographs (mug shots of "suspects"[interrogators] in uniform) and un-cued photographs (mug shots of "suspects"[interrogators] in civilian clothing). Accuracy (True Positive identifications) of eyewitness recognition was higher when using the serial photo presentation method compared to the line up or photo-spread methods (48% vs. 34% and 28%, respectively). Eyewitness recognition was also more accurate if participants were shown a serial presentation of cued photographs of the interrogators taken at the time of exposure to interrogation stress (65% vs. 48% respectively)(2).

In the above noted study, no significant relationship was observed between demographic variables (such as age, rank, time in the service, history of trauma) and Eyewitness Accuracy. Likewise, no significant relationship was observed between Eyewitness Accuracy and the variable of stress-induced symptoms of dissociation. Thus, in our previous study, the factors contributing to one person being a better eyewitness compared to another remained unidentified.

Although many studies have investigated a number of demographic, personality and cognitive factors which might account for individual differences in Eyewitness Accuracy, surprisingly few have been found to be predictive of a person's general eyewitness performance (Schooler & Loftus, 1993). One variable that has been found to be predictive of Eyewitness Accuracy is general observational ability (Boice, Hansley, Shaughnessy, & Gansler, 1982). In that study, participants were shown three videotapes: One depicting a crime of purse snatching; the second, a class room lecture; and the third a casual request for a date, respectively. Participants who were accurate in recalling the details of the videotape depicting the "crime" (of purse snatching) also recalled more details from the other two videotapes. These results suggest that there may be some general observational ability that helps mediate Eyewitness Accuracy.

To our knowledge no studies have explored the degree to which performance on a standardized test designed to assess how a person's ability to remember and recognize human faces may be related to that person's accuracy as an eyewitness. Based on the findings of Boice et al. (1982) one might hypothesize that the ability to remember and recognize human faces on a standardized recognition test would be significantly related to the ability to perform accurately during an Eyewitness Identification Task for personally relevant events. The present study was designed to test this hypothesis.

2. Methods

2.1. Participants

Participants were 53 consecutively recruited male active duty military personnel who were enrolled in military survival school training. The mean age of participants was 29 (SD=3), and the average number of years in the service was 4.2 (SD=4). As designated by their military branch, all participants were active duty Special Operations personnel. None of the participants enrolled in the study wore glasses or contact lenses. Each participant completed in-processing into the survival training course prior to enrollment in the study. Recruitment of participants was conducted by the principal investigator (CAM). All participants gave written, informed consent and understood that participation (or lack of participation) in the study would not influence their status in the survival school course in any manner. As per survival training course requirements, all participants provided documentation of physical examination, medical and psychiatric clearance prior to enrollment. All participants were free of illicit substances.

2.2. The study site: U.S. Military Survival School

The phases of survival school training have been described in detail elsewhere. (Morgan, Wang, Mason, et al., 2000a; Morgan, Wang, Southwick, et al., 2000b; Morgan, Hazlett, et al., 2001a; Morgan, Wang, et al., 2001b). Briefly, survival school training is divided into two general phases. The first phase is didactic; the second phase is experiential. In the didactic phase, participants undergo one week of classroom instruction designed to prepare them for the experiential phase. In the experiential phase, participants are confined in a mock prisoner of war camp (POWC). This phase is designed to offer one of the most challenging training experiences that active duty participants will ever experience while in the military. In the POWC, each participant is exposed to isolation and interrogation stress. The stress exposure is designed to test the ability of participants to apply what they have learned during the didactic phase.

As we have previously reported, the stress experienced during the confinement phase (and especially during exposure to interrogation stress) is intense and provides a valid environment for the study of uncontrollable stress in human participants. Hormones such as cortisol, norepinephrine, and epinephrine are significantly elevated and gonadal steroids, such as testosterone, significantly reduced below fertility levels in response to POWC stress. The hormone alterations are comparable to those observed in humans exposed to real world, threat-to-life events (Morgan, Wang, Mason, et al., 2000a; Morgan, Hazlett, et al., 2001a; Morgan, Wang, et al., 2001b). Previous studies in this setting have also consistently shown that trainees rate the subjective stress of survival school as being 9.5/10 where 10 = the most stress ever experienced by participants many of whom have already experienced combat. Thus, survival school offers a model wherein there is: 1) exposure to realistic, personally relevant acute stress; 2) uniform application of a stressor across participants; 3) objective data about the “perpetrator” (i.e., the interrogator) against which eyewitness reports provided by participants can be compared; and finally, 4) a uniform timeframe for the post-stress assessment of memory for perceived events.

After the didactic phase and prior to placement in the POWC, all participants experienced approximately 48 h of partial food and sleep deprivation. Once in the POWC the application of food deprivation, sleep deprivation and stress was uniform across participants. In order to facilitate an understanding of the current data, an abbreviated time line of events will be provided: Once placed in the POWC all participants were separated from one another and placed in isolation. Approximately 8 h after being placed in the POWC, participants were exposed to interrogation stress. During interrogation stress participants remained alone in a room with a survival school instructor who was not known to them. The room in which interrogations were conducted was well illuminated and participants were able to see and hear the instructor. After approximately 30 min of exposure to the instructor, participants were placed in isolation. Over the subsequent 48 h participants experienced food and sleep deprivation. Upon release from the POWC, all participants were given access to food. Thirty minutes after release from the POWC, all participants assembled in the classroom and participated in testing.

3. Assessment of facial memory ability

3.1. Authors note

Because researchers in the field of eyewitness and memory studies use specific terms that may be unfamiliar to the reader, we explain the following terms and their specific meaning in this manuscript: “Target” is the term we use when referring to the person (i.e., the interrogator) or a face to which a person was exposed and is asked to recognize (i.e., specific faces participants are told to ‘remember’ on the [Wechsler Face Test \(1997\)](#)). The term “Target Present” is used to indicate that when the participants’ memory is being assessed, the Target is included among the other faces or people shown to participants. The term “Target Absent” is used when the Target is not included in the array of faces or people shown to participants. Finally, we use two tests in this study, both of which have an Accuracy score. To reduce the likelihood of confusion, these are referred to as “Accuracy on the Eyewitness Identification Task” and “Accuracy on the Wechsler Face Test.”

Once assembled in the classroom, participants were administered the [Wechsler Face Test \(GAH\)\(1997\)](#). This test is comprised of standardized color photographs of non-personally relevant human faces. By “non-personally relevant” we mean that participants have never seen the faces before and there is no known cognitive or emotional association between the standardized faces on the Wechsler Face Test and the experience of participants during survival school training. Testing was conducted by showing the participants the 24 “target” faces (the faces they were instructed to remember), one at a time. To administer the test, a large projection screen was lowered in the front of the room on which the standardized Wechsler Face Test photographs were sequentially presented using a projection system in a manner identical to the manual method. The projection system rendered a 6 ft × 8 ft image of high color quality. During the presentation of each face, participants were given the instruction “Remember this face.” After participants had been shown each of the 24 faces, immediate recognition for these faces was tested by next presenting participants an array of 48 faces, one at a time, among which the original 24 photographs of faces were interspersed. As each of the 48 photographs was presented, participants were asked whether the face on the screen was one of the 24 “target” faces they had just seen.

Thirty minutes after completing the immediate recall assessment, participants participated in the delayed recall testing phase of the Wechsler Face Test. Once again, participants were presented, one photograph at a time, an array of

48 photographs. This array contained the original 24 faces and 24 new faces that were not included in the previous administration of the test. Due to significant differences in the nature of the delay recognition task compared to the immediate phase or the Eyewitness Identification Task (see Data Analysis Section), the delay data are not presented in the current study.

4. Assessment of Eyewitness Identification (eyewitness memory for interrogation stress)

After completing the immediate phase of the Weschler Face Test, all participants participated in the assessment of their own “eyewitness” accuracy by attempting to identify the person who had conducted their stressful interrogation while in the POWC. Because participants were individually exposed to interrogation stress while in the POWC and because the participants were not exposed to the interrogators during any other phase of the POWC, participants had no knowledge about the identity of another participant’s interrogator. During testing of Eyewitness Accuracy participants remained seated in the classroom and were instructed to remain silent and not speak to one another during the testing so as not to influence or distract fellow participants. The method used for testing Eyewitness Identification was the sequential photo presentation method (Guide, USDOJ). Based on our previous study (Morgan et al., 2004) which showed eyewitness accuracy was best for cued photographs, the photographs used in the sequential array were those taken at the time of interrogation stress exposure of participants.

Prior to the presentation of the photographs, participants were given the following instructions: “You have each been given a sheet of paper on which there are spaces for up to 16 responses. Next to each of the numbers on the page you will see the words “yes” and “no.” To the right of these words you will see a row of numbers ranging from 1 to 10. You will be asked, when viewing a photograph projected on the screen whether or not the face on the screen is the individual who conducted your high stress interrogation. You may select “Yes,” “No,” or “Not Present.” Once you have selected a response, please circle the number to the right of your answer that best corresponds to your level of confidence in the accuracy of your answer. A response of 10 indicates you have no doubt in your mind; a response of 1 indicates that you are guessing. Once you have selected a photograph of an individual who you believe to be the person who was your interrogator when you were in the POWC, your participation in the test is complete. You will not be permitted to review the photographs once we have shown you all of them. It is important for you to remember that it is possible that none of the photographs in this test will include the face of the person who interrogated you.”

At this point, the slide presentation began. Once again, the large projection screen was lowered in the front of the room. A series of 10 photographs was sequentially presented using a projection system. Five of the 10 photographs in the sequential array were of Survival School instructors who had actually participated in the interrogation of some of the participants. The remaining five photographs of “interrogators” included Survival School instructors who were on leave at the time of the study and therefore, had never been seen by participants. The photographs of these instructors were taken during a separate iteration of training at another time point in the year when they were at the POWC and conducting interrogations. The arrays of photographs were compiled such that for 42% of participants the picture of their interrogator did not appear in the array. For 58% of the participants, the picture of the interrogator was present. The omission of actual interrogator/instructors for 42% of participants was deliberate. For these participants to be classified as an accurate eyewitness, they had to provide a “True Negative” type response to the test — and reject all of the photographs. This was done to more accurately mirror a real world situation wherein the serial photo arrays used by the police may not contain the mugshot of the actual perpetrator. For the 58% of participants whose interrogator’s photograph was included in the array, a classification as an accurate eyewitness was possible only if the participant correctly selected the photograph (a “True Positive” identification response). Because participants did not know whether their “target” would be included in the array [i.e., they did not know if they were in the ‘target present’ or ‘target absent’ condition], the challenge for all participants viewing the photographs was the same: each participant had to decide whether the person depicted in the slide was or was not their interrogator.

Although there were only 10 photographs presented, the score sheet that participants were given included 16 response slots. This discrepancy between the number of “possible” responses on the score sheet and the number of photographs that were actually presented during the sequential photo presentation was also deliberate and was designed to prevent the participants from knowing in advance the exact number of photographs that they would be shown when performing the eyewitness test. This was done to more accurately mirror real world conditions experienced by witnesses who, when administered a sequential photo array by the police, are ideally not informed about the exact number of photographs they will be shown (10). Finally, as was true for the administration of the Weschler Face Test,

the projection system rendered a 6 ft × 8 ft image of high color quality. Participants were given 1 min to view each slide, provide their response and either quit or continue the test. Response sheets were passed to the research staff and the participants who had finished the test remained seated in the room.

5. Data analysis

5.1. Accuracy scores for performance on the *Weschler Face Test*

As per the scoring instructions for the *Weschler Faces* test, sub-scores of True Positive responses, True Negative responses, False Positive responses and False Negative responses were tallied for the immediate recall phase of testing. A total “Accuracy” score for the *Weschler* immediate recall condition was created by summing the “True Positive” and “True Negative” responses provided by the participant in the immediate recall phase.

As noted above, the results from the “delay” phase of testing cannot be exactly compared to the eyewitness data as was done with the results from the “immediate” phase of the *Weschler Face Test*. This is due to the fact that by the time participants are asked to perform the delay test, the “targets” (the photographs that they were initially told to remember) have now been viewed twice by participants [at the initial viewing and during the immediate recall phase of the *Weschler Face Test*]. This means that participants have had an additional opportunity to view the target and to remember it. Thus the delay phase of the *Weschler* test for facial recognition is not comparable to the *Eyewitness Identification Task* that the participants performed. As a result, the data from the delay condition are not presented.

5.2. Analysis of *Eyewitness Accuracy*

Eyewitness Accuracy scores were created to reflect whether or not the participant was able to accurately identify his or her interrogator from the POWC. A score of 1 was assigned for a correct response; a score of 0 was assigned to an incorrect identification. A separate coding for the *Eyewitness Identification* scores was also created in order to indicate whether “accurate” eyewitness participants had provided True Positive or True Negative identification responses and whether “inaccurate eyewitnesses” had provided False Positive and or False Negative type identification responses. Note: Given the design of the photo array, it was possible for 58% of participants – for whom the array was a “target present” array – to overlook the target when viewing the sequential photo array [and thus commit a False Negative type error] and then subsequently select an individual from the array who was not their interrogator [a false positive type error]. Confidence scores for *Eyewitness Identifications* were scored as the numerical value (from 1 to 10) endorsed by the participant on the *Eyewitness Identification* sheet.

6. Relationship between *Weschler Face Test* Scores and Accuracy of *Eyewitness Identification*

In order to assess the relationship between participants’ performance on the standardized face recognition memory test (the *Weschler Face Test*) and the accuracy of eyewitness memory for the person who had conducted their stressful interrogation (as assessed by the sequential photo array), separate cross tab (eta) analyses were performed. This was done using the variables *Eyewitness Accuracy* (correct ID for interrogator; incorrect ID for interrogator) and *Weschler Face Test Score* (the total Accuracy score on the immediate phase of the *Weschler Face Test*). Separate cross tab (eta) analyses were also conducted in order to evaluate the relationship between *Eyewitness Accuracy* and the sub-scores (true positive, true negative, false positive, false negative) of the *Weschler Face Test*. Additional Chi-squared analyses were conducted in order to detect whether the types of errors committed by participants in the *Eyewitness Identification Task* (false positives, false negatives) differed significantly between the group of participants whose interrogator was present in the serial photo presentation array (a “target present” line up) and those whose interrogator was not (a “target absent” line up). Finally, separate univariate analyses of variance were used to assess whether the scores on the *Weschler Face Test* were significantly associated with demographic variables such as age, rank, or years in the military.

Logistic regression analyses were conducted using Status (accurate/inaccurate) as an *Eyewitness* as the dependent variable and using the four sub-scales on the *Weschler Faces Test* (true positive; true negative; false positive; false negative) as the independent variables. The sub-scores of false positive and of false negative did not contribute to the model and were removed from the analysis. An Enter method was used when conducting the regression.

7. Receiver operator characteristics curves

To understand how well performance by participants on the standardized Weschler Face Test would predict which individuals would be accurate eyewitnesses when asked to identify the person who conducted their stressful interrogation, receiver operator characteristics (ROC) curves were created (Metz, 1978, 2000; Zweig & Campbell, 1993). ROC curves were created by using the Weschler Accuracy scores for immediate recall (the test variable) in order to predict the likelihood that a participant would be correctly classified as an accurate eyewitness (the state variable, where the value of the state variable = 1 [correct eyewitness]). The area under the curve as well as coordinate points for the curve were calculated (SPSS 11.5). The null hypothesis assumption was that the true area under the curve = 0.5. With regard to the parameters for the standard distribution of error, the distribution assumption was nonparametric and the confidence interval was 95%.

8. Results

8.1. Standardized face recognition testing results (*The Weschler Face Test*)

Of the 53 participants, seven did not complete the immediate recall phase of the Weschler Face Test. Thus, the analyses correlating performance on the Weschler Face Test with performance on the Eyewitness Identification Task are based upon the data from 46 participants.

8.2. Immediate Weschler Face Test Scores

For the group as a whole, the mean score for Immediate True Positive responses was 15 (SD=5.3). The mean score for Immediate True Negative responses was 16 (SD=4.0). The mean score for Immediate False Positive responses was 14 (SD=7.3). The mean score for Immediate False Negative responses was 24 (SD=8.2). For the group as a whole, the total Accuracy score for the Immediate Weschler test (the sum of True Positive and True Negative responses) was 31.4 (SD=6.89).

8.3. Accuracy of Eyewitness Identification

Of the 53 participants, 33 (62%) accurately identified the presence or absence of their interrogator. Twenty participants (38%) were incorrect in this task. For the group as a whole, twenty four participants (45%) provided True Positive identifications; 19 participants (36%) provided False Positive identifications; 9 participants (17%) provided True Negative identifications and 1 participant (2%) provided a False Negative identification. Of note, the seven participants who did not complete the Weschler Face Test did not differ in their performance on the Eyewitness Identification Task, compared to participants who did complete the Weschler Face Test. As a result they are included in the analysis of Eyewitness Accuracy for the group as a whole.

The following break down of the data regarding accurate versus inaccurate eyewitnesses according to the condition (target present or target absent) provides a more specific understanding of the data: Of the 31 participants for whom the array was “target present,” 24 participants (77%) provided True Positive identifications. Of the 22 participants for whom the array was “target absent,” 9 (41%) provided True Negative responses while the remaining 13 participants provided a False Positive eyewitness identification. Chi-square analyses indicated that the difference in the number of accurate eyewitness identifications in the two groups of participants (“target present” vs. “target absent”) was significant (Chi-square=7.3; $df=1$, asymptotic significance (2 sided) $p<0.007$; exact significance= $p<0.01$ (Fisher’s Exact Test).

In order to better understand the behavior of participants who gave False Positive responses, we examined where in the array of photographs such errors were made. We found that of the 19 participants in the study who committed False Positive type errors when performing their Eyewitness Identification Task: 7 participants incorrectly selected the first photo in the array; 1 participant incorrectly selected the second photo in the array; 7 participants incorrectly selected the third photo in the array; 1 participant incorrectly selected the fourth; 1 participant incorrectly selected the sixth; 1 participant incorrectly selected the seventh; and 1 participant incorrectly selected the eighth photo in the array. Thus, the majority of participants who provided False Positive responses did so early in the array (i.e. prior to the 5th photograph) — suggesting that they may have been “eager” to select someone.

Table 1
 Weschler Face Test Scores for accurate and inaccurate eyewitness groups

Accurate eyewitness group ($N=29$)	Inaccurate eyewitness group ($N=17$)
True Positive: 17.3 (SD=7) *	True Positive: 12.2 (SD=7)
True Negative: 16.5 (SD=3)	True Negative: 15.1 (SD=5)
False Positive: 17.4 (SD=3)	False Positive: 7.3 (SD=4)
False Negative: 6.7 (SD=3) **	False Negative: 11 (SD=6)

* $p < 0.003$; ** $p < 0.006$.

8.4. Confidence in Eyewitness Identification

Nearly all participants were confident in their responses. The mean confidence rating for the group as a whole was 8.4 (SD=3) on the 10-point subjective confidence scale. No significant differences were observed in confidence scores between participants who were accurate versus inaccurate eyewitnesses ($p=.3$). Similarly, no significant differences were observed between the subjective confidence ratings provided by participants for whom the photo array was “target present,” compared to those for whom the array was “target absent.” ($p=0.2$). Also, no significant differences in confidence were observed when comparing individuals who provided False Positive responses to photographs early (prior to the third photograph) in the eyewitness array compared to those who provided False Positive responses late (after third photograph) in the array ($p=0.4$). Univariate analyses of variance using Confidence as the dependent variable and type of response (true positive, true negative, false positive, false negative on the eyewitness task) as the fixed factor did not reveal any significant differences in the Confidence levels provided by participants ($F(2, 51)=2.2$; $p=0.12$).

Table 1 depicts immediate the Weschler Face Test Scores for the two major sub-groups of eyewitness participants: those who were correct in identifying their interrogator and those who were not. As noted in the table, multivariate analyses of variance indicated that accurate eyewitnesses differed significantly from inaccurate eyewitnesses in True Positive and False Negative responses on the Weschler Face Test [$F(1)=10$; $p < 0.003$; $F(1)=8.5$; $p < 0.006$, respectively].

9. Relationship between Eyewitness Accuracy and Performance on the Weschler Face Test

Cross tab (eta) analyses revealed a significant, positive relationship between Accuracy of Eyewitness Identification and the Weschler Accuracy scores for the immediate recall phase of the Weschler Face Test (nominal by interval value [eta]=.42; approximate $T=2.5$; $p < 0.02$). Cross tab analyses between Eyewitness Identification Accuracy and the four sub-scales scores for the immediate recall phase of the Weschler Face Test indicated there was a significant, positive relationship between Eyewitness Accuracy and True Positive responses (eta=.47; approximate $T=3.5$; $p < 0.001$), and False Negative responses only (eta=.41; approximate $T=-2.9$; $p < 0.006$).

10. Relationship between types of responses on the Weschler Face Test and types of responses provided by participants in the Eyewitness Identification Task

Univariate analyses of variance using the sub-scores (true positive, true negative, false positive, false negative) from the immediate phase of the Weschler Face Test as the dependent variables, and “Type” of response (true positive, false positive, true negative, false negative) provided by participants when performing the Eyewitness Identification Task as the “Fixed Factor,” revealed statistically significant differences for Weschler Face Test True Positive responses ($F(2,44)=5.4$; $p < 0.008$ and Weschler Face Test False Negative responses ($F(2,43)=3.7$; $p < 0.03$) only. Post hoc analyses (Tukey) indicated that these significant differences were due to the fact that participants who had provided False Positive identifications in the Eyewitness Identification Task provided significantly fewer True Positive responses (mean difference=5.6; $p < 0.02$) and significantly more False Negative responses (mean difference=4.7; $p < 0.03$) on the Weschler Face Test.

Results of the Logistic Regression indicated that the model was significant. Omnibus tests of model coefficients, step one (True Positive score) Chi-square was 10.2; $p < 0.006$. The Cox & Snell R square was .20. The classification table percentage was 73.3%. The standardized beta coefficient for the variable True Positive score was .24; the Wald

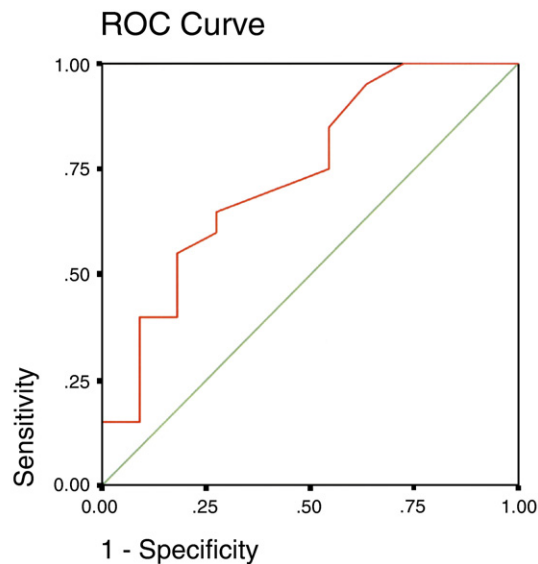
statistic was 0.0, suggesting the variable True Positive score made a meaningful contribution to the model. The variable True Negative score was not significant.

Receiver Operating Characteristic (ROC) Curve Data: When an ROC curve is created, the area may take values between 1 and zero. A value of 1 or zero would indicate that the test is always right or always wrong, respectively. If the test performs no better than chance at detecting the state variable (for example, status of being an accurate eyewitness) the area under the curve would be 0.5. As noted in Fig. 1 the area under the curve was 0.736; the nonparametric standard of error was 0.096; the asymptotic significance was $p < 0.03$.

Table 2 lists the coordinate points of the ROC graph and indicates, for a given [immediate] Accuracy score on the Weschler Face Test, the probability of being right (sensitivity) or of being wrong (1 minus the specificity) in detecting whether the person obtaining such a score on the Weschler Face Test would be an accurate eyewitness.

11. Discussion

When shown a serial photographic array of cued photographs and asked to identify the “perpetrator” who had conducted their stressful interrogation, more than one out of every three participants (38%) was an inaccurate eyewitness. These data replicate our previous findings of substantial identification errors when people try to make an identification associated with an event that is highly personal and very stressful (Morgan et al., 2004). Sixty-two percent of the participants, however, were accurate eyewitnesses. The present data extend the findings of our previous report by providing specific clues about how and why individuals differed in Eyewitness Accuracy.



Diagonal segments are produced by ties.

Area Under the Curve

Test Result Variable(s): accuracy for immediate (in pos plus true neg)

Area	Std. Error ^a	Asymptotic Sig. ^b	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.736	.096	.032	.548	.925

The test result variable(s): accuracy for immediate (in pos plus true neg) was at least as good between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Fig. 1.

Table 2
Coordinates of the curve

Test result variable(s): accuracy for immediate (tru posplustrue neg)		
Positive if greater than or equal to ^a	Sensitivity	1 — Specificity
– 1.0000	1.000	1.000
11.5000	1.000	.909
24.5000	1.000	.818
26.5000	1.000	.727
27.5000	.950	.636
28.5000	.850	.545
29.5000	.750	.545
30.5000	.650	.273
31.5000	.600	.273
33.0000	.550	.182
35.0000	.400	.182
36.5000	.400	.091
37.5000	.300	.091
38.5000	.150	.091
40.5000	.150	.000
42.5000	.050	.000
44.0000	.000	.000

Test result variable(s): accuracy for immediate (tru posplustrue neg) has at least one tie between the positive actual state group and the negative actual state group.

^a The smallest cutoff value is the minimum observed test value minus 1, and the largest cutoff value is the maximum observed test value plus 1. All the other cutoff values are the averages of two consecutive ordered observed test values.

Consistent with our primary hypothesis, participants' performance on a standardized test of facial recognition was significantly associated with how well they were able to identify the face of the person who had interrogated them 48 h earlier. The faces participants were asked to view (and remember) when performing the Weschler Face Test were novel and neutral in nature and were not associated with a threatening experience. Further, the faces contained in the Weschler Face Test reflected both genders and a wide range of age and race. Thus, the significant relationship between performance on the Weschler Face Test and performance on the Eyewitness task suggests that a person's general observational capacity for remembering generic human faces may be a trait that significantly predicts his ability to accurately recognize faces that are associated with personally relevant, emotional distress.

Participants who were accurate eyewitnesses exhibited a greater number of True Positive responses, and fewer False Negative responses on the Weschler Face Test compared to participants who were inaccurate eyewitnesses. The two groups (accurate vs. inaccurate eyewitnesses) did not differ in the number of False Positive response errors on the Weschler Face Test. Thus, the type of response error that best accounted for the difference between the two groups of eyewitnesses was the False Negative response. At the present time, the nature of the relationship between the two types of errors committed by inaccurate eyewitnesses (False Negative errors on the Weschler Face Test and False Positive errors on the Eyewitness Identification Task) is unknown.

The present data may have practical implications for professions dealing with issues of eyewitness testimony for serious crimes. This report reinforces the idea that despite being confident when providing an identification regarding a "suspect," a significant number of eyewitnesses – even if shown a photograph taken at the time of the crime – will not be accurate. There has long been an interest in the confidence that witnesses express in their memories. Jurors and members of the legal profession believe that confidence is a good indicator of accuracy (Brewer, Keast, & Rishworth, 2002; Deffenbacher & Loftus, 1982) Mock jurors who render verdicts rely strongly on confidence (Bradfield & Wells, 2000). Yet many studies have revealed that confidence is not a good indicator of accuracy (Kassin, Tubb, Hosch, & Memon, 2001). While one could argue that many if not most of the prior eyewitness studies did not expose participants to genuine stress and thus the weak relationship might not appear in stressful settings, the argument is undermined by the present results. We have replicated the findings from our previous eyewitness study (Morgan et al., 2004) and found no significant relationship between confidence and accuracy in a highly stressful setting. Further, we find that false identifications are expressed with very high levels of confidence.

An individual's ability to remember standardized human faces appears to be related to his or her accuracy as an eyewitness and may provide an objective means of critically evaluating information provided by eyewitnesses. For example, if one were asked to predict whether a person is likely to be an accurate eyewitness based on the Weschler Face Test Score, the probability of being wrong in making this prediction (i.e. in stating that the person will be a correct eyewitness) is shown on the ROC Curve (Fig. 1). The coordinate points which are indicated in Table 2, show that if one uses a cutoff score of greater than 11.5 on the Weschler Face Test, this would result in the inclusion of all accurate eyewitnesses but would also include 90% of inaccurate eyewitnesses. Therefore, using this cutoff score would result in the situation wherein a professional making the prediction as to who is an accurate eyewitness would have a 90% chance of being wrong. Alternatively, if one were to use a cutoff score of 40, while this would assure that the person achieving such a score was an accurate eyewitness, this cutoff level would eliminate many people (85% of individuals) who would also be accurate eyewitnesses. However, if one used a cutoff score of 36, 40% of accurate eyewitnesses would be identified and there would only be a 9% chance of the professional misidentifying who would be an accurate eyewitness.

Clearly, the decision as to which cutoff score is acceptable would depend on the task at hand. As a hypothetical example, if the police are trying to apprehend a sniper and are interviewing potential witnesses, they might elect to use a low cutoff score on the Weschler Face Test in order to include an accurate eyewitness in the pool of potential witnesses (i.e., the low cutoff score on the Weschler reduces false negatives and acts as a "screening test" that would allow the police the chance to detect people more likely to provide True Positive identifications). However, during a legal proceeding the burden of proof to convict is more rigorous. In order to protect a defendant's presumed innocence, hypothetically, one might select a cutoff score on the Weschler Face Test that would eliminate some accurate eyewitnesses but that would enhance the probability that the eyewitness whose testimony is to be considered, is accurate (i.e., the high cutoff score on the Weschler eliminates individuals who are more likely to provide False Positive identifications).

Obviously, prior to applying this methodology, much additional research is needed in order to clarify and refine an understanding of the current data. For example a greater number and more diverse population of participants is needed prior to any practical application. At the present time, it would be premature (and ill-advised) to use the Weschler Face Test as a predictor of eyewitness testimony or reliability in forensic evaluations. In the future, it is, however, possible that forensic experts may have access to a useful methodology that might help them distinguish between individuals whose memory for faces is more likely to be accurate from those whose memory is not. Knowledge of this type would potentially allow the expert to give an opinion based on scientific data about the likelihood that his or her opinion regarding Eyewitness Accuracy is correct or incorrect. An opinion rooted in data about which there are known rates of accuracy and reliability, would meet the criteria set by the Supreme Court of the United States (USSC) regarding the reliability and accuracy of expert opinions. Studies such as this may help generate the type of forensic data required by the USSC under *Daubert* (1993; *Kumho*, 1999).

The present study has a number of limitations: First, participants in this study had been exposed to significant sleep and food deprivation between the time of stress exposure and the time that Weschler Face Test and Eyewitness Identification Task were assessed. Although such factors may have influenced the overall rates of accuracy on both tests, it is unlikely that these factors account for the present finding of a significant relationship between the Weschler Face Test and the Eyewitness Identification Task due to the fact that the deprivation of food and sleep was kept uniform across all participants. A second limitation of the present study is that the participants were members of special military units and, as a result, may exhibit performance characteristics on both tests that do not generalize to the civilian population. It is possible that the participants in this study – who have prior experience with highly stressful situations – may have performed better than a group of stress naive civilians in a similar situation. Finally, the present design did not permit us to examine the relationship between performance on the Weschler Face Test and eyewitness memory for different types (or intensities) of stressful events. As noted in our previous study (Morgan et al., 2004) eyewitness memory was significantly better for less stressful events compared to high stress events. Thus the degree to which performance on the Weschler Face Test might predict Eyewitness Accuracy for events that are less stressful (or perhaps more stressful) than the experience of participants in this study remains to be explored. This said, it is important to underscore the fact that previous investigations have provided robust evidence that the stress experienced by participants in the survival school is of similar intensity to that of real world, threat to life events. Thus it is likely that the findings of this study will be applicable to real world scenarios.

In summary, the present data suggest that there may be enduring traits or capacities related to an individual's ability to remember human faces that contribute to the accuracy of eyewitness recognition. An enhanced understanding of such capacities may assist in the future development of a body of knowledge by which forensic examiners, the courts and juries may more scientifically appraise eyewitness evidence as well as the testimony of forensic experts.

Acknowledgements

Funding for this study was provided by a grants from the U.S. Army (CAM), the Office of Naval Research, (CAM) and from U.S.A.R.I.E.M (CAM). Support for this study was also provided by the National Center for PTSD, VA Connecticut. The views and opinions expressed herein do not reflect the views of the United States Government but solely reflect the views of the authors.

References

- Brewer, N., Keast, A., & Rishworth, A. (2002). The confidence–accuracy relationship in Eyewitness Identification. *Journal of Experimental Psychology: Applied*, 8, 44–56.
- Bradfield, A. L., & Wells, G. L. (2000). The perceived validity of Eyewitness Identification testimony: A test of the five Biggers criteria. *Law and Human Behavior*, 24, 581–594.
- Boice, R., Hanley, C. P., Shaughnessy, P., & Gansler, D. (1982). Eyewitness Accuracy: A general observational skill? *Bulletin of the Psychonomic Society*, 204, 193–195.
- Daubert v. Merrell Dow Pharmaceuticals, Inc., (1993), 509 U.S. 579.
- Deffenbacher, K. A., & Loftus, E. F. (1982). Do jurors share a common understanding concerning eyewitness behavior? *Law and Human Behavior*, 6, 15–30.
- Guide. U.S. Department of Justice, Office of Justice Programs, 810 Seventh Street, NW, Washington, DC 20531. Document can be downloaded at the Internet address: <http://www.ojp.usdoj.gov/nij/pubs-sum/178240.htm>
- Kassin, S. M., Tubb, V. A., Hosch, H. M., & Memon, A. (2001). On the “general acceptance” of eyewitness testimony research: A new survey of the experts. *American Psychologist*, 56, 405–416.
- Kumho v. Carmichael, (1999), 526 U.S. 137.
- Metz, C. E. (1978). Basic Principles of ROC analysis. *Seminars in Nuclear Medicine*, 8, 283–298.
- Metz, C. E. (2000). Fundamental ROC analysis. In J. Buetel, H. Kundel, & R. Van Metter (Eds.), *Handbook of medical imaging Physics and psychophysics, Vol. 1* (pp. 751–769), Bellingham, WA: SPIE Press.
- Morgan III, C.A., Hazlett, G., Wang, S., Richardson, G., Schnurr, P., & Southwick, S. M. (2001a). Symptoms of dissociation in humans experiencing acute uncontrollable stress: A prospective investigation. *American Journal of Psychiatry*, 158(8), 1239–1247.
- Morgan III, C.A., Hazlett, G. A., 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.
- Morgan III, C.A., Wang, S., Hazlett, G., Rasmussen, A., Anderson, G., & Charney, D. S. (2001b). Relationships among cortisol, catecholamines, neuropeptide Y and human performance during uncontrollable stress. *Psychosomatic Medicine*, 63, 412–442.
- Morgan III, C.A., Wang, S., Mason, J., Hazlett, G., Fox, P., Southwick, S. M., et al. (2000a). Hormone profiles in humans experiencing military survival training. *Biological Psychiatry*, 47, 891–901.
- Morgan III, C.A., Wang, S., Southwick, S. M., Rasmussen, A., Hauger, R., & Charney, D. S. (2000b). Plasma neuropeptide-Y in humans exposed to military survival training. *Biological Psychiatry*, 47, 902–909.
- Penrod, S. D., Fulero, S. M., & Cutler, B. L. (1995). Expert psychological testimony on eyewitness reliability before and after Daubert: The state of the law and the science. *Behavioral Sciences & the Law*, 13, 229–259.
- Statistical Procedures for the Social Sciences (SPSS) spreadsheet, version 11.5.
- Schooler, J. W., & Loftus, E. F. (1993). Multiple mechanisms mediate individual differences in eyewitness accuracy and suggestibility. In J. M. Puckett & H. W. Reese (Eds.), *Mechanisms of everyday cognition* (pp. 177–203). Hillsdale, NJ: Erlbaum.
- Weschler Memory Scale III. (1997). *Faces I*. The psychological corporation. San Antonio, TX: Harcourt Brace & Company.
- Zweig, M. H., & Campbell, G. (1993). Receiver operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, 39, 561–577.