JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2011, Vol. 56, No. 5
doi: 10.1111/j.1556-4029.2011.01896.x
Available online at: onlinelibrary.wiley.com

PAPER

PSYCHIATRY & BEHAVIORAL SCIENCES

Charles A. Morgan III,[1,2] *M.D., M.A.; Kevin Colwell,*[3] *Ph.D.; and Gary A. Hazlett,*[4] *Psy.D.*

Efficacy of Forensic Statement Analysis in Distinguishing Truthful from Deceptive Eyewitness Accounts of Highly Stressful Events*

ABSTRACT: Laboratory-based detecting deception research suggests that truthful statements differ from those of deceptive statements. This nonlaboratory study tested whether forensic statement analysis (FSA) methods would distinguish genuine from false eyewitness accounts about exposure to a highly stressful event. A total of 35 military participants were assigned to truthful or deceptive eyewitness conditions. Genuine eyewitness reported truthfully about exposure to interrogation stress. Deceptive eyewitnesses studied transcripts of genuine eyewitnesses for 24 h and falsely claimed they had been interrogated. Cognitive Interviews were recorded, transcribed, and assessed by FSA raters blind to the status of participants. Genuine accounts contained more unique words, external and contextual referents, and a greater total word count than did deceptive statements. The type-token ratio was lower in genuine statements. The classification accuracy using FSA techniques was 82%. FSA methods may be effective in real-world circumstances and have relevance to professionals in law enforcement, security, and criminal justice.

KEYWORDS: forensic science, detecting deception, false claims, cognitive interview, military stress

The criminal justice system places a great deal of faith in eyewitness testimony despite psychological research showing that eyewitness reports may be very inaccurate (1–6). Further, the degree of confidence displayed by a witness contributes to his or her believability in the eyes of the jury and the court (7,8). Eyewitness statements are often the *only* evidence available to authorities. Taken together, the literature suggests that judges and juries often make decisions based on the demeanor and perceived credibility of the eyewitness.

A similar situation exists within the community of government agencies responsible for national security. Officers in such agencies must often rely heavily on information provided to them by human "sources." These sources often claim to have first-person knowledge about information of value to the U.S. government. In the majority of instances, U.S. officials working with such sources have little objective means by which they might determine whether the source providing such information is valid and reliable or invalid and deceptive. As in the criminal justice system, most U.S. intelligence or national security officials tend to believe human sources who appear confident in what they report. Thus, as in the justice system, case officers tasked with making decisions related to national security often do so based on the only "evidence" available to them: the statements from a human source.

With regard to the criminal justice system, this state of affairs has led to support for forensic research designed to develop systematic approaches to be used in discriminating honest from fabricated reports. Although a variety of methods for assessing eyewitness accounts have been developed, nearly all are underpinned by the hypothesis that witnesses who are highly motivated to lie (i.e., highly motivated to ensure that their deceptive statements will be believed) will exhibit noticeably different patterns of speech, thought, and behavior (9). Broadly speaking, the majority of findings in current literature supports this view and suggests that compared to statements given by truthful persons, statements provided by deceptive witnesses are shorter in length and contain fewer relevant details (10). It must be acknowledged that within the literature, discrepant findings do exist and show that differences in statement length and level of detail between truthful and deceptive individuals are not always as described earlier (11–13). Whether the discrepant data are because of differences in modes of communication in studies (oral, written, translated) or some other factor remains to be elucidated.

It has also been contended by some researchers that statements provided by deceptive witnesses exhibit less "lexical diversity" than those of truthful eyewitnesses. This construct is often referred to as the "type-token ratio (TTR)" and is calculated by dividing the number of distinct words (i.e., "types"), by the total number of words (i.e., "tokens") contained in the person's statement (14). For example, the sentence, "One small step for man, one giant leap for mankind" has eight nonrepetitive words and 10 total words. To calculate the TTR for this statement, one divides the unique words (i.e., eight) by the total number of words contained in the statement

[1]Department of Psychiatry, Yale University School of Medicine, 234 Church Street, Suite #301, New Haven, CT 06510.
[2]National Center for PTSD, VA Connecticut Healthcare Systems, 950 Campbell Avenue, West Haven, CT 06516.
[3]Department of Psychology, University of Georgia, 1500 N. Patterson Street, Valdosta, GA 31602.
[4]Woodard-Cody Specialty Consultants, Inc., 4115 Hulon Drive, Durham, NC 27705.

*Funded with a grant from the U.S. Department of Defense (CAM), support from the National Center for PTSD, and by the Center for Research and Development, Inc., New Haven, CT.

Received 5 April 2009; and in revised 13 April 2010; accepted 17 April 2010.

(i.e., 10). This results in a TTR of 0.8. It has been suggested that low anxiety (and presumably truth-telling) will hypothetically lead to higher TTR, whereas high anxiety (presumably associated with deception) will lead to lower TTR (15,16). Some investigators have suggested that increases in TTR are the result of attempts at impression management, whereas deception-related reductions in TTR stem from a motivational impairment effect (17). Regardless of theory, extant data from laboratory-based deception studies support the idea that significant differences in TTR may be observed between statements provided by deceptive and truthful individuals.

At present, two significant factors limit our understanding about the degree to which the findings from forensic statement analysis (FSA) studies may be applied to real-world cases confronted by law enforcement and national security officials: First, laboratory-based studies do not entail levels of stress or threat comparable to that experienced by people experiencing real-world legal jeopardy or highly stressful events. Research findings from psychobiological studies designed to assess the impact of realistic stress on humans show that such stress may result in significant alterations in cognition, perception, and memory (18–20). Therefore, it is reasonable to hypothesize that eyewitness statements about one's experience during highly stressful events may also be adversely affected. Thus, at present, it remains an open question as to whether (or to what degree) current findings in the literature will apply to eyewitness statements associated with personally relevant, high stress events.

A second issue limiting our current understanding of the applicability of extant FSA research is the absence of "ground truth" in field studies. This is to say that in published field studies, neither the investigators nor the police have knowledge about the true number of "false negative" subjects (i.e., the number of individuals who provided false statements but who avoided being detected). This prevents a determination of "accuracy" in field studies because the calculation of accuracy for any technique is dependent upon information about the number of true positive and false negative classifications. With regard to deception studies, this means it is essential to know for certain which individuals were deceptive and which were truthful.

The present study was designed to address these limitations by assessing, under controlled conditions of realistic, personally relevant stress, the efficacy of two specific FSA techniques (one automated, one manual) in differentiating genuine from deceptive eyewitness statements related to personally relevant, highly stressful events. The statements were obtained by using a Cognitive Interview technique that consisted of two components: (i) the initial, open-ended question: "Please describe everything you remember from the time you entered the interrogation booth until the time you exited the interrogation booth" and (ii) the mnemonic prompt question designed to elicit more memory recall. This consisted of asking participants to "take a moment and think about anything else you may have seen, smelled, touched, tasted or felt during this experience. Once you have done this, please start at the beginning once more and tell us everything you remember and include anything new you remember." The main hypothesis(ses) of the present study were related to the automated method for statement analysis and were as follows:

- *Genuine* eyewitness statements would be characterized by a greater response length (i.e., the actual number of uttered words in the recalled experience will be greater) than that observed in deceptive eyewitness statements.
- *Genuine* eyewitness statements would be characterized by an increased number of unique words in response to the mnemonic prompt.
- *Genuine* eyewitness accounts would be characterized by a higher TTR than deceptive eyewitness accounts.

Secondary hypotheses in this study were related to our use of a relatively new manual scoring method (21) and were as follows:

- Genuine eyewitness statements would be characterized by a greater amount of detail as measured by the number of external, contextual, and internal referents contained in the accounts.
- Genuine eyewitness statements would be characterized by longer statements as measured by line length (see Methods section for definition).
- Manually scored measures would correlate significantly and positively with automated measures of unique word count, response length and TTR.

Methods

Thirty-five male active duty military personnel were the participants of this study; the data of one participant randomized to the truthful condition were dropped from analysis because of damage to the videotape recording which prevented its transcription. *All* subjects provided written, informed consent and were instructed that participation (or lack of participation) in the study would not influence their status in, or eligibility for, Survival School in any manner (positively or negatively). Participants were randomly assigned to one of the two test conditions (Genuine Eyewitness [N = 19]/Deceptive Eyewitness [N = 15]).

Description of the Survival School Venue

Military Survival School (i.e., SERE) training is one of the most rigorous forms of training experienced by special operations military personnel (22). The training is based on the U.S. Military Code of Conduct and is designed to prepare war-fighters for the possibility that they might be trapped in enemy territory, pursued by enemy forces, and captured and detained as prisoners of war (POWs). The methodology employed in this study has been reported in extensive detail elsewhere (23–26). However, a brief description will be given to facilitate an understanding of the data.

The training is rigorous, realistic, and modeled after the experiences of American POWs from WWII, and the Korean, Vietnam, and Gulf Wars. The course is designed to provide individuals at risk for capture and exploitation by the enemy with specific skills and armed forces code of conduct training so as to enhance their chances of surviving behind enemy lines and to return home with honor. The stress experienced by subjects during the confinement phase of SERE training produces neurobiological alterations on a par with those documented in individuals exposed to real world, threat to life experiences (23–25).

The course is divided into a didactic and an experiential phase. The didactic phase is comprised of classroom lectures and demonstrations. During these classroom activities, students are instructed on how to find food and water, how to build shelters, how to navigate, and methods designed to adhere to the U.S. Armed Forces Code of Conduct if ever captured and held as a POW (26). Once they have completed the didactic phase, students participate an experiential phase of the training course to put what they have learned into practice. The experiential phase is comprised of an evasion exercise and a confinement experience in a mock POW camp. During the evasion phase, students are pursued by mock-enemy forces and must avoid detection; after the evasion phase, students are placed in mock internment. During this phase, they experience physical and psychological stress similar to those experienced by former POWs. Students experience food and sleep deprivation and are also exposed to stressful types of interviews or

interrogations. During this experience, they are expected to put into practice what they have learned in the classroom phase of the course. At the conclusion of the experiential phase, students participate in a debriefing day where they are given the opportunity to review and learn from their individual and group experiences. Following this, they graduate the course.

As noted in numerous scientific publications, SERE stress experienced by students in mock captivity results in psychobiological changes similar to those observed in real-world threat to life events (18,20,23–25).

Genuine Eyewitnesses

Participants assigned to the genuine eyewitness condition were interviewed *c.* 4 h after they had been released from the experiential phase of SERE (i.e., the mock POW camp. Each participated in a 20-min semi-structured interview (Cognitive Interview) during which they were asked to recall a specific event they had experienced while in mock captivity: that of a highly stressful interrogation. This interrogation stress event was selected because previous research conducted at SERE has shown this event to be an ethologically valid analogue of real-world, high-stress events (14–16,18). All interviews were videotaped and used to generate a verbatim transcript of the interview, so that the eyewitness statements could be submitted for FSA. Once the participants provided their eyewitness account, their participation in the study was complete.

Deceptive Eyewitnesses

Participants randomized to the deceptive condition were given genuine eyewitness transcripts to study to provide them with information that might facilitate their ability to be credible, deceptive eyewitnesses. The contents of these transcripts provided detailed genuine descriptions of the types of experiences to which genuine participants were exposed while in mock captivity. This was done to provide a more realistic assessment of how well FSA methods may work in real-world situations, wherein deceptive individuals are not spontaneously lying but instead have prepared their lies to be more believable. These transcripts were combined to provide deceptive participants with a detailed account of interrogation stress; the transcripts given to deceptive participants were selected by SERE instructors and judged by them to be representative in detail, length and lexical complexity (response length: 1021; unique words: 357; TTR: 0.349) to other genuine eyewitness accounts from students at SERE.

Participants assigned to the deceptive eyewitness condition were given 24 h to study the transcript containing information about experiences in the mock captivity phase of SERE. Each was told that they would need to learn the material such that when asked about this information, they would appear to be genuinely reporting their own experience. Participants were informed that they would be interviewed by an Investigative Interviewer who did not know whether they had truly attended SERE. Each was encouraged to appear as honest and forthcoming so as to lead the Investigative Interviewer to believe that they were a genuine "witness" about the experience of being subjected to interrogation stress while at SERE. The duration of all interviews was *c.* 20 min. Once the deceptive participants provided their "eyewitness" account, their participation in the study was complete.

The Modified Cognitive Interview

In a recent investigation, Colwell et al. (27) compared the relative efficacy and accuracy of three main forensic interviewing techniques currently used in oral statement analysis research [i.e., the Cognitive Interview (28); the Structured Interview (29–31), and the Inferential Interview (27)]. The results of this study provided robust evidence that the Cognitive and Inferential Interview techniques most effectively discriminated truthful from false testimony did not differ significantly from one another and were superior to the Structured Interview technique (27). Based on this, we elected to use the Cognitive Interview technique in this study. However, because of restrictions of time (20 min) imposed on our research team by the SERE training staff, we modified the Cognitive Interview such that participants were given only two—and not four—opportunities to tell their story. Participants were given the open-ended recall prompt to describe their experience. Once they had completed their account, participants were given the first mnemonic prompt of the Cognitive Interview after which they provided their eyewitness account for the second time. Once participants had provided their second eyewitness account, the interview process was complete.

The Method of Statement Analysis

Printed text transcripts were created for each of the eyewitness accounts provided by participants. These were used for the automated and manual scoring of the dependent variables of interest. The automated analysis method was used to calculate response length (i.e., the actual number of words uttered in responses), the number of unique words, and the TTR (i.e., the total number of unique words divided by the response length).

Manual scoring of the eyewitness accounts was performed using the Assessment Criteria Indicative of Deception (21). In this method, detail is assessed by counting the number of internal, external, and contextual indicators in each of the two accounts provided by participants. To assess the impact of the mnemonic prompt, only *new* internal, external, and contextual elements were counted when assessing the second eyewitness accounts provided by participants.

Transcripts were coded by three trained raters who were blind to the status (truthful/deceptive) of the participants in the experiment. Raters were trained in one 3 h group meeting with the first author and with a second member of the research team (KC). During the first meeting, each of the three classes of detail was defined, and raters were provided with standard scoring sheets to ensure consistent operational definitions of the variables. External details were defined as information regarding the event in question that was gained from the senses (e.g., describing who, what, and where). For example, the statement "A *tall man* wearing a *black uniform*" contains four external details. Contextual details were defined as those which described relationships among objects and/or actors (e.g., temporal, spatial relationships). For example, the statement "There was a phone *on top of* the desk" contains one contextual detail. Internal details were defined as information regarding the subjective mood, experiences, or cognitive processes of the respondent, as well as any information that referenced the respondent's history rather than the event in question. "I was *nervous*" contains one internal detail. Importantly, raters were trained only to code each detail the first time it appeared within a statement. A detail mentioned in response to two different questions was only tallied in response to the first one. The result is that the amount of novel detail elicited in response to each question of the interview was available for assessment. The inter-rater agreement for the three detail categories was as follows: external = 0.85, contextual = 0.89, and internal = 0.79.

The manual method used to calculate response length consisted of counting the number of full or partial lines for each eyewitness account when the accounts were printed in 12-font on 8 1/2 by 11

sheets of paper. Full lines of text or lines that passed the midline of the sheet of paper were awarded one point. Lines of text that extended just to the half-way point on the page, or less, were assigned 0.5 points. Total line length was the sum of these points (21,27).

Data Analysis

Descriptive statistics and standardized scores were generated for each of the variables. Statistical analyses were performed using a commercially available statistical software package (SPSS 11.5, spreadsheet version 10.1.19; IBM, Armonk, NY). Preplanned bivariate Pearson's correlation coefficients were calculated to examine the relationships among the variables generated by automated and manual techniques. Independent *t*-tests were used to assess whether the dependent variables differed significantly between the two groups. Following this, general linear model analyses of variance were used to assess whether there was a significantly different response to the mnemonic prompt in the two eyewitness groups. Stepwise forward logistic regression analyses were conducted to construct and test models that might best predict membership of the subjects in each of the groups.

Receiver Operator Characteristics Curves

To understand how well the dependent variables would predict which individuals were genuine or deceptive eyewitnesses, receiver operator characteristics (ROC) graphs were created by using dependent variables (response length, unique words) from the first and second eyewitness accounts (the test variable) to predict the likelihood that a subject would be correctly classified as a truthful or genuine eyewitness (the state variable, where the value of the state variable is equal to 1 [genuine eyewitness]). The area under the curve as well as coordinate points for the curve was calculated (SPSS 11.5). The null hypothesis assumption was that the true area under the curve is equal to 0.5. With regard to the parameters for the standard distribution of error, the distribution assumption was nonparametric and the confidence interval was 95%.

Results

As noted in Table 1, genuine eyewitness accounts were characterized by a greater response length, more unique (distinct) words and a *lower* TTR compared to deceptive eyewitness accounts. Independent *t*-tests indicated that these differences were significant. However, after controlling for differences in response length, no significant difference in the number of unique words in the first eyewitness accounts was observed between the two groups: (genuine: 258 [SD = 95]; deceptive: 152 [SD = 97] [$F_{1,31} = 2.5$; $p < 0.13$]). By contrast, after controlling for differences in response length in the second eyewitness account (i.e., the account provided *after* participants had been exposed to the mnemonic prompt), a significantly higher unique word count was observed in the accounts provided by genuine, compared to deceptive eyewitness (genuine: 181 [SD = 115]; deceptive: 92 [SD = 54] [$F_{1,31} = 5.7$; $p < 0.02$] [Fig. 1]).

As noted in Table 2, both the first and second eyewitness accounts provided by genuine eyewitnesses contained significantly more external and contextual elements than did those provided by deceptive eyewitnesses. Similarly, line length was significantly greater in genuine, compared to deceptive, eyewitness accounts (Table 2). Controlling for differences in line length did not reveal significant differences between eyewitness groups in external or contextual referents. Similarly, no significant differences were observed in the number of internal referents in account of the two eyewitness groups.

TABLE 1—*Automated analysis of eyewitness statements.*

	Genuine	Deceptive	*t* (Sig.)
Unique words			
1st account	257.6 (SD = 95.1)	152.0 (SD = 97.2)	3.1($p < 0.003$)
2nd account	180.6 (SD = 115.0)	92.1 (SD = 53.9)	2.7($p < 0.01$)
Response length			
1st account	837.3 (SD = 497.5)	399.8 (SD = 398.8)	2.8($p < 0.01$)
2nd account	481.3 (SD = 524.0)	184.3 (SD = 149.2)	2.1($p < 0.04$)
Type-token ratio			
1st account	0.348 (SD = 0.07)	0.469 (SD = 0.10)	−3.9($p < 0.0001$)
2nd account	0.465 (SD = 0.11)	0.598 (SD = 0.11)	−3.5($p < 0.001$)

Correlations Between Automated and Manual Methods of Analysis

As noted in Table 3, robust relationships were observed among the automatically scored variables (i.e., response length, unique word count, TTR) and the variables derived from the manual scoring method (i.e., external, contextual, internal referents, line length). As noted in the table, a shift in the strength of the correlation between unique word count and TTR in the first account was observed in the second account. At present, the reasons for this are not known.

Binary Logistic Regression Analyses

Separate binary logistic regression analyses were conducted for each of the eyewitness accounts (first account, second account) using status (genuine, deceptive eyewitness) as the dependant variable and predictor variables created by the automated (response length, unique word count, TTR) and manually calculated variable (line length, external, internal, and contextual referents) methods. With regard to the first eyewitness account, the best fit occurred

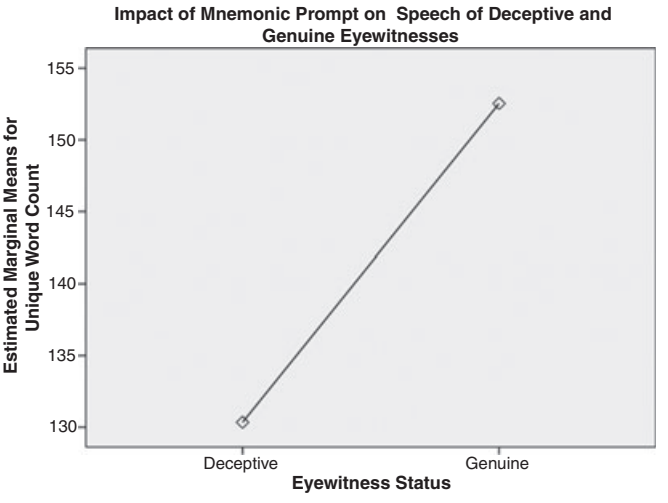

FIG. 1—*Impact of mnemonic prompt on speech of deceptive and genuine eyewitnesses.*

TABLE 2—*Manual analysis of eyewitness statements.*

	Genuine	Deceptive	*t* (Sig.)
External			
1st account	142.6 (SD = 63.6)	87.8 (SD = 43.9)	2.8 ($p < 0.01$)
2nd account	56.1 (SD = 40.5)	29.2 (SD = 18.9)	2.4 ($p < 0.02$)
Contextual			
1st account	45.4 (SD = 19.4)	27.3 (SD = 16.9)	2.9 ($p < 0.01$)
2nd account	23.9 (SD = 22.7)	9.7 (SD = 7.4)	2.3 ($p < 0.03$)
Internal			
1st account	12.2 (SD = 7.3)	8.3 (SD = 6.5)	1.6 (N.S.)
2nd account	10.7 (SD = 13.2)	5.6 (SD = 5.8)	1.4 (N.S.)
Line length			
1st account	40.9 (SD = 23.5)	23.0 (SD = 20.2)	2.3 ($p < 0.02$)
2nd account	29.2 (SD = 28.5)	11.3 (SD = 9.5)	2.3 ($p < 0.03$)

TABLE 3—*Relationships among automated and manually scored variables.*

	External	Contextual	Internal	LL	UW	RL	TTR
First eyewitness account							
UW	0.79*	0.79*	0.61*	0.62*	1.00	0.96*	−0.83*
RL	0.74*	0.77*	0.56*	0.91*	0.96*	1.0	−0.85*
TTR	−0.65*	−0.68*	−0.51*	−0.76*	−0.83*	−0.85*	1.0*
External	1.0	0.94*	0.55*	0.87*	0.79*	0.74*	−0.65*
Contextual	0.94*	1.0	0.70*	0.94*	0.79*	0.77*	−0.68*
Internal	0.55*	0.70*	1.0*	0.75*	0.61*	0.56*	−0.51*
LL	0.87*	0.94*	0.75*	1.0	0.62*	0.91*	−0.76*
Second eyewitness account							
UW	0.85*	0.87*	0.87*	0.96*	1.00	0.96*	−0.11
RL	0.79*	0.84*	0.91*	0.99*	0.96*	1.0	−0.10
TTR	−0.67*	−0.68*	−0.67*	−0.77*	−0.11	−0.10	1.0*
External	1.0	0.94*	0.55*	0.87*	0.86*	0.77*	−0.67*
Contextual	0.94*	1.0	0.70*	0.94*	0.79*	0.77*	−0.68*
Internal	0.55*	0.70*	1.0*	0.75*	0.61*	0.56*	−0.67*
LL	0.87*	0.94*	0.75*	1.0	0.62*	0.91*	−0.77*

*$p < 0.0001$ ($N = 35$).
UW, unique words; RL, response length; LL, line length; TTR, type-token ratio.

with a Forward Conditional model using the predictor variables of response length, unique words and of contextual referents ($\chi^2 = 9.6$; d.f. = 2; $p < 0.008$); combining these variables resulted in an overall classification accuracy of 82% (correct classification of deceptive eyewitnesses = 87%; correct classification of genuine eyewitnesses = 79%). With regard to the second eyewitness account, the best fit occurred with a Forward Conditional approach using predictor variables of external and contextual referents ($\chi^2 = 13.9$; d.f. = 3; $p < 0.003$); the correct overall classification accuracy was 81.5% (correct classification of deceptive eyewitnesses = 80%; correct classification of genuine eyewitnesses = 82.4%).

ROC Curve Data

When an ROC curve is created, the area may take values between 1 and 0. A value of 1 or 0 would indicate that the test is always right or always wrong, respectively. If the test performs no better than chance at detecting predicting the state variable (e.g., status of being a genuine eyewitness), the area under the curve would be 0.5. Based on our findings (noted above) that genuine eyewitnesses were more likely to have longer accounts and, after a mnemonic prompt, contain a greater number of unique words, we created ROC graphs using the line length from the first and second eyewitness accounts and the unique word count from the second eyewitness account. As noted in Fig. 2 for these variables (line length 1, line length 2, unique words 2), the area under the curve

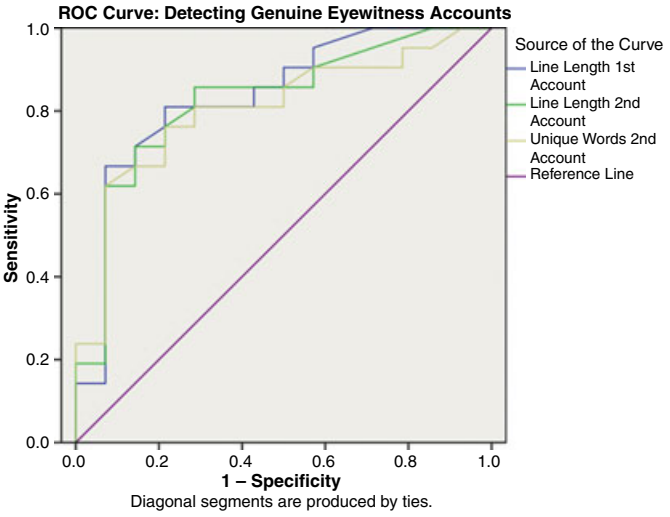

FIG. 2—*Receiver operator characteristics curve: detecting genuine eyewitness accounts.*

was 0.81, 0.79, and 0.80, respectively. The nonparametric standard of error for each of the three variables was 0.1; the asymptotic significance for each was $p < 0.003$.

Table 4 lists the coordinate points of the ROC graph and indicates, for a given line length or unique word count in an eyewitness statement, the probability of being right (sensitivity) or of being wrong (1 *minus* the specificity) in concluding whether the person providing a statement is a genuine eyewitness. For example, if the unique word count is ≥144, the probability of being wrong in concluding the eyewitness is genuine is 7%. Similarly, if the line length is equal to or greater than a value of 35 lines, the probability of being wrong in classifying an eyewitness as genuine is 7%.

Discussion

Automated and manual FSA methods performed significantly better than chance in discriminating between genuine and deceptive eyewitness accounts regarding personally relevant, highly stressful events. Indeed, when combining these variables, classification rates were 82%. The present data are consistent with the extant literature and demonstrate that eyewitness accounts for real-world experiences contain more detail and are longer in length than those provided by individuals feigning exposure to such experiences. These data provide robust support for the idea that FSA methods may represent an objective, scientifically valid manner by which one may assess the credibility of eyewitness accounts—even under circumstances when the eyewitness accounts are being provided by deceptive individuals who have had access to genuine information and an opportunity to rehearse their lies.

Unlike previous laboratory-based research groups, we did not observe an increase in the TTR in genuine, compared to deceptive eyewitnesses accounts; instead, we found the TTR to be significantly *lower* in genuine eyewitnesses. Because the TTR has been typically thought to index "lexical diversity," the present finding warrants comment in that the findings appear to support the idea that lexical diversity is diminished in truth-tellers. Given the unique aspects of this study (the incorporation of realistic stress for genuine eyewitnesses and a rehearsal phase for deceptive eyewitnesses), we believe such a conclusion is premature. First, it may be helpful to bear in mind that lexical diversity (as measured by TTR) is calculated by dividing the number of unique words by the total

TABLE 4—*Receiver operator characteristics columns and reference data for sensitivity and 1 − specificity.*

Coordinates of the Curve			
Test Result Variable(s)	Positive if Greater Than or Equal To	Sensitivity	1 − Specificity
(Lines of text 2nd account)	−0.500	1.000	1.000
	1.000	1.000	0.933
	2.750	1.000	0.867
	4.750	0.947	0.733
	6.500	0.947	0.600
	7.750	0.895	0.600
	8.250	0.895	0.533
	9.000	0.895	0.467
	10.000	0.895	0.400
	10.750	0.895	0.333
	11.500	0.842	0.333
	12.500	0.789	0.333
	14.250	0.737	0.333
	16.000	0.737	0.267
	16.750	0.684	0.267
	17.500	0.632	0.267
	19.000	0.632	0.200
	20.638	0.579	0.200
	22.388	0.421	0.067
	24.750	0.368	0.067
	27.000	0.316	0.067
	30.000	0.263	0.067
	32.750	0.211	0.067
	34.500	0.158	0.067
	40.000	0.158	0.000
	59.500	0.105	0.000
	101.000	0.053	0.000
	128.500	0.000	0.000
(Unique Words 2nd account)	4.00	1.000	1.000
	23.50	1.000	0.933
	42.50	0.947	0.867
	50.00	0.947	0.800
	58.00	0.947	0.733
	60.50	0.947	0.667
	71.00	0.947	0.600
	80.50	0.895	0.533
	82.00	0.842	0.533
	85.00	0.842	0.467
	89.50	0.842	0.400
	94.00	0.842	0.333
	103.00	0.789	0.333
	115.00	0.737	0.333
	124.00	0.684	0.333
	134.79	0.684	0.267
	141.79	0.632	0.133
	144.00	0.579	0.067
	148.50	0.526	0.067
	152.50	0.474	0.067
	164.00	0.421	0.067
	179.00	0.368	0.067
	191.00	0.316	0.067
	198.50	0.263	0.067
	209.00	0.211	0.067
	220.00	0.211	0.000
	240.50	0.158	0.000
	336.00	0.105	0.000
	466.00	0.053	0.000
	521.00	0.000	0.000

The test result variable(s): SMEAN(LINES2), SMEAN(UWB) has at least one tie between the positive actual state group and the negative actual state group.

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.

number of words uttered in a statement; thus, the reduced lexical diversity may simply be a reflection of poststress "talkativeness" (greater response length) on the part of the SERE students. The genuine eyewitness accounts were obtained within the 8 h of the students' release from the mock POW camp. During their time in the camp, the students did not have the opportunity to speak freely with fellow students. SERE training staff and medical support staff have consistently observed that students exhibit increased "talkativeness" after their release from the stressful portion of their training (personal communications to the first author from GAH and SERE staff member G. Steffian, unpublished observations).

This increased talkativeness may be due to efforts on the part of students to process their experience. It may also reflect the continued neurobiological activation experienced by SERE participants after their exposure to the training stress is complete. In our previous neurobiological investigations of SERE participants, we have documented that following exposure to SERE stress, students exhibit increased levels of norepinephrine and epinephrine for at least 24 h after their exposure to SERE stress (15). It is possible this increase in adrenergic hormones contributed to increased talkativeness of subjects.

After controlling for differences in response length, we did not observe a significant difference in the number of unique words contained in the *first* eyewitness accounts provided by the two eyewitness groups. This too is at apparent odds with the findings of previous studies which have shown an increase in the number of unique words in genuine compared to deceptive accounts. It is possible that aspects of the present study's design account for this difference. Unlike previous studies in which deceptive individuals give "spontaneous" lies (i.e., no rehearsal time, and/or no access to specific information), deceptive participants in this study had 24 h to prepare for their interview as well as access to genuine data about the experience at SERE about which they were to lie. It is likely this opportunity to rehearse as well as access to genuine material about the SERE experience enhanced the recall more relevant unique words and, as a result, minimized the differences between the groups in the first eyewitness account. Thus, the present findings may not be inconsistent with the findings of previous studies assessing FSA methods for spontaneous lies.

Consistent with the findings of previous detecting deception studies, we did observe a significantly greater effect of the memory prompt in genuine, compared to deceptive eyewitnesses. After controlling for response length of the second eyewitness accounts (i.e., the account produced after exposure to the mnemonic prompt), we observed a significantly higher number of unique words in the second accounts of genuine, compared to deceptive eyewitnesses. This finding supports the view that memory prompts are less effective in deceptive, compared to genuine eyewitnesses. One explanation put forward in the literature about the "lack of efficacy" of such prompts in deceptive individuals is that such individuals attempt to "stick to their story" to be believed and as a result provide fewer details, alterations or additions to the story they have already provided. Giving too many details about which they might be inconsistent at a later time might put them at risk for being caught in their lies.

The manual scoring methods used in this study were strongly correlated with the automated analyses. Whereas some circumstances may lend themselves to automated analyses, such analyses may not always be possible. That an easily calculated variable like "line length" was so highly correlated (i.e., $r = 0.96$) with the total word count (i.e., "response length") suggests these hand-scored systems can be used to achieve very similar results if professionals do not have access to automated methods or equipment. This said, it should be underscored that the best classification models were achieved using a combination of the automatic and manually scored variables—suggesting that future research examining the relative advantages of each method would be fruitful.

The classification accuracy rates of the FSA methods used in this study are encouraging and suggest that these methods may be very helpful to professionals tasked with performing credibility assessments of individuals who claim to have witnessed or experienced events of interest to the criminal justice system or to national security. Clearly, in order for these forensic methods to be as useful as possible tools for real-world investigations, future studies designed to establish normative databases regarding eyewitness statements need to be conducted. Once done, these databases for specific populations and situations of interest would provide an enhanced ability to perform ROC curve analyses—and theoretically assist in better discrimination of genuine and deceptive statements. In addition, and to be most effective in producing results that would best meet the diverse needs of professionals, future studies must include non-English speaking populations as well as the impact of using a translator on the accuracy of FSA methods.

There are a number of limitations to the present study. First, whereas some researchers have applied their FSA methods to written statements provided by participants, others (like us) have analyzed the transcripts of oral statements. It is possible that the predictor variables identified in this study will not be as effective when applied to written statements.

Second, all of the deceptive eyewitnesses had the opportunity to rehearse their stories which prevented an assessment of how accounts based on spontaneous lies would compare to those based on lies that have been rehearsed. This limitation notwithstanding, we believe that the present data are directly relevant to real-world situations confronted by special agents in which they must assess the credibility of people who are highly motivated to lie, who have had access to genuine information and who have had time to rehearse their story. A third limitation in the present study was the lack of a stressed-exposed deceptive group of eyewitnesses. This prevented us from testing the degree to which stress may have induced "talkativeness" in the genuine eyewitnesses. Because SERE training requires all students to experience the same events during the training, it was not possible to include a stress-exposed group of participants who had not been exposed to interrogation stress. Future studies designed to test the impact of stress exposure on false eyewitness accounts are currently under way. We anticipate that this type of information will assist in determining how well FSA methods can be applied to situations in which individuals genuinely exposed to potentially traumatic (or traumatic events) lie about their experience by substituting the genuine trauma for another "traumatic" event that they have not experienced. FSA methodologies may then be helpful to forensic examiners who assess individuals with genuine posttraumatic stress disorder (PTSD) but who lie about the true nature of the index traumatic event that caused their PTSD.

Finally, time constraints were imposed on the research team such that the Cognitive Interviews were limited to 20 min. Within this limited time frame, it was not possible to administer the four traditional mnemonic prompts. To maintain consistency across interviews, only one mnemonic prompt was used. Thus, it is possible that the classification accuracy noted in this study (i.e., 82%)—which is somewhat lower than in other previously published speech content analysis studies—might have been closer to that noted in previously published studies (i.e., 85–90%) (23). This said, the present classification accuracy is higher than that of 72% reported in a review of speech content studies (32) and suggest that content analysis of oral statements may be a useful tool for government officials working in law enforcement, national security and in intelligence. Future studies examining the efficacy of these methods in languages other than English and in the translated (English) speech of an interpreter are currently under way.

Acknowledgments

The authors wish to thank Dr. Lisa Steffian, Ph.D., Amanda Rose, and all the members of the SERE training staff, FASOTRAGRULANT, Naval Air Station, Brunswick, ME, without whose support, insight and assistance this research would not have been possible.

References

1. Cutler BL, Penrod SD. Mistaken identification: the eyewitness, psychology and the law. New York, NY: Cambridge University Press, 1995.
2. Loftus EF, Ketcham K. Witness for the defense: the accused, the eyewitness and the expert who puts the memory on trial. New York, NY: St. Martin's Press, 1991.
3. Memon A. The cognitive interview. In: Hargie O, editor. The handbook of communication skills, 3rd edn [e-book]. New York, NY: Routledge, 2006;531–50.
4. Morgan CA III, Hazlett GA, Rasmusson A, Zimolo Z, Southwick SM, Charney DS. Relationships among plasma dehydroepiandrosteron sulfate and cortisol levels, symptoms of dissociation and objective performance in humans exposed to acute stress. Arch Gen Psychiatry 2004;61:819–25.
5. Morgan CA III, Hazlett GA, Doran T, Garrett S, Hoyt G, Baranoski M, et al. Accuracy of eyewitness memory for persons encountered during exposure to highly intense stress. Int J Law Psychiatry 2004;27(3):265–79.
6. Sporer SL, Malpass RS, Koehnken G. Psychological issues in eyewitness identification. Mahwah, NJ: Lawrence Erlbaum, 1996.
7. Cutler BL, Penrod SD. Juror decision making in eyewitness identification cases. Law Hum Behav 1988;12:41–55.
8. Penrod SD, Cutler B. Witness confidence and witness accuracy: assessing their forensic relation. Psychol Public Policy Law 1995;1:817–45.
9. Ruby CL, Brigham JC. The usefulness of the criteria based analysis technique in distinguishing between truthful and false allegations: a critical review. Psychol Public Policy Law 1997;3:705–37.
10. Vrij A. Verbal characteristics of deception. In: Davies G, Bull R, editors. Detecting lies and deceit: the psychology of lying and the implications for professional practice. New York, NY: Wiley, 2000;103–63.
11. Hancock JT, Curry L, Goorha S, Woodworth M. Automated linguistic analysis of deceptive and truthful synchronous computer-mediated communication. In: Sprague RH, editor. Proceedings of 36th Hawaii International Conference on System Sciences; 2003 Jan 6–9; Big Island, HI. Washington, DC: IEEE Computer Society, 2005;23b.
12. Sporer SL. The less traveled road to truth: verbal cues in deception detection in accounts of fabricated and self-experienced events. Appl Cog Psychol 1997;11:373–97.
13. Morgan CA III, Rabinowitz Y, Christian J, Hazlett GA. Detecting deception in Vietnamese: efficacy of forensic statement analysis when interviewing via an interpreter. J Intelligence Community Res Dev 2009;1:1–22.
14. Osgood CE. Some effects of motivation on style of encoding. In: Sebeok T, editor. Style in language. Cambridge, MA: MIT Press, 1960;293–306.
15. Porter S, Yuille JC. Credibility assessment of criminal suspects through statement analysis. Psychol Crime Law 1995;1:319–31.
16. Carpenter R. The statistical profile of language behavior with Machiavellian intent or while experiencing caution or avoiding self-incrimination. Ann NY Acad Sci 1990;606:5–16.
17. Vrij A, Akehurst L. Verbal communication and credibility: statement validity assessment. In: Memon A, Vrij A, Bull R, editors. Psychology and law: truthfulness, accuracy and credibility. New York, NY and Maidenhead, UK: McGraw-Hill, 1998;77–97.
18. Morgan CA III, Hazlett G, Wang S, Richardson G, Schnurr P, Southwick SM. Symptoms of dissociation in humans experiencing acute, uncontrollable stress: a prospective investigation. Am J Psychiatry 2001;158:1239–47.
19. Morgan CA III, Wang S, Hazlett G, Rassmusson A, Anderson G, Charney DS. Relationships among cortisol, catecholamines, neuropeptide y and human performance during uncontrollable stress. Psychosom Med 2001;63:412–42.
20. Morgan CA III, Doran A, Steffian G, Hazlett G, Southwick SM. Stress induced deficits in working memory and visuo-constructive abilities in special operations soldiers. Biol Psychiatry 2006;60(7):722–9.
21. Colwell K, Hiscock-Anisman C, Memon A, Rachel A, Colwell L. Vividness and spontaneity of statement detail characteristics as predictors of witness credibility. Am J Forensic Psychol 2007;25(1):5–30.

22. Doran AP, Hoyt G, Morgan CA III. Survival, evasion, resistance, and escape (SERE) training: preparing military members for the demands of captivity. In: Kennedy CH, Zillmer EA, editors. Military psychology: clinical and operational applications. New York, NY: Guilford Press, 2006; 241–61.
23. Morgan CA III, Wang S, Mason J, Hazlett G, Fox P, Southwick SM, et al. Hormone profiles in humans experiencing military survival training. Biol Psychiatry 2000;47:891–901.
24. Morgan CA III, Wang S, Southwick SM, Rasmusson A, Hauger R, Charney DS. Plasma neuropeptide-y in humans exposed to military survival training. Biol Psychiatry 2000;47:902–9.
25. Morgan CA III, Rassmusson A, Wang S, Hauger R, Hazlett G. Neuropeptide-y, cortisol and subjective distress in humans exposed to acute stress: replication and extension of a previous report. Biol Psychiatry 2002;52(2):136–42.
26. Steffian G, Bluestein B, Ogrisseg J, Doran A, Morgan CA. Code of conduct and the psychology of captivity: training, coping and reintegration. In: Adler AB, Castro CA, Britt TW, editors. Minds in the military: psychology and life in the armed forces, volume 2: operational stress. Westport, CT: Praeger Security International, 2006;124–250.
27. Colwell K, Hiscock CK, Memon A. Interview techniques and the assessment of statement credibility. Appl Cog Psychol 2002;16:287–300.
28. Geiselman RE, Fisher RP. Ten years of cognitive interviewing. In: Payne D, Conrad F, editors. Intersections in basic and applied memory research. Mahwah, NJ: Lawrence Erlbaum, 1997;291–310.
29. Yuille JC, Hunter R, Joffe R, Zaparniuk J. Interviewing children in sexual abuse cases. In: Goodman G, Bottoms B, editors. Child victims, child witnesses: understanding and improving children's testimony. New York, NY: The Guilford Press, 1993;95–115.
30. Yuille JC, Cutshall J. Analyses of the statements of victims, witnesses and suspects. In: Yuille JC, editor. Credibility assessment. Dordrecht: The Netherlands: Kluwer Academic Press, 1989;175–91.
31. Farr VL, Yuille JC. Investigating and assessing the credibility of children's allegations of sexual abuse: issues for professionals. J Prevent Sex Abuse 1988;1(1):8–12.
32. Vrij A. Criteria-based content analysis: a qualitative review of the first 37 studies. Psychol Public Policy Law 2005;11(1):3–41.

Additional information reprint requests:
Charles A. Morgan III, M.D., M.A.
National Center for PTSD, 151E
VA Connecticut Healthcare System
950 Campbell Avenue
West Haven, CT 06516
E-mail: charles.a.morgan@yale.edu