Eyewitness Memory of Police Trainees for Realistic Role Plays

John C. Yuille, Graham Davies, Felicity Gibling, David Marxsen, and Stephen Porter

One hundred and twenty police recruits (probationer constables) being trained at the Metropolitan Police Training Centre in Hendon, England, participated in 1 of 2 occupational simulations, either nonstressful or stressful. Twenty of the recruits in each condition were active participants in the event, whereas the other 40 were observers. Half of the recruits were interviewed after a delay of 1 week and again after 12 weeks. The others were interviewed only once after a period of 12 weeks. Stress decreased the amount recalled but improved both accuracy and resistance to decay over time. Participants recalled more details than observers. Delay of recall and the number of recall trials also affected performance. The results are discussed in terms of the advantages of a research context that permits much of the control of the laboratory while retaining many of the characteristics of actual forensic contexts.

As with any new area of scientific inquiry, the first century of eyewitness memory research (e.g., Binet, 1900) has been characterized by debate and an evolution of method. During the past decade, the debate over the ecological validity of the research has taken center stage, as it did in the early part of this century (cf. Wigmore, 1909). The current version of this debate (e.g., Loftus, 1991; Yuille & Wells, 1991) focuses on the relationship between memory as it operates in the laboratory and memory in other contexts. Some have argued that memory is best studied in the laboratory where maximum control and precision are possible (e.g., Banaji & Crowder, 1989). Although there are several problems associated with this argument (see commentaries edited by Loftus, 1991), one salient problem is the inability to produce certain circumstances in the laboratory. For example, for obvious ethical reasons, strong emotional reactions or high levels of stress cannot be induced in laboratories. Anyone interested in the relationship between emotions or stress and memory must, at some point, leave the laboratory to seek research venues in which such reactions occur naturally.

This article reports the results of our first study of eyewitness memory at a police-training college in Hendon, England. Probationer constables of the London Metropolitan Police Force spend the first 20 weeks of their career at this college in "initial training." A major component of the training is a series of role plays, assessed throughout the course. A community microcosm has been constructed including a pub, homes, and a police station that constitutes a realistic context for the role plays. The recruits, in groups of three to five, are sent on "patrol" in the training area. One recruit serves as the constable on patrol while the rest are observers, including one person carrying a video camera. Each patrol encounters an actor or actors and a situation that reflects a difficult problem for police officers. The situation is open-ended, and the scenarios are as realistic as possible.

The cooperation of the training college allowed us to introduce manipulations into the design of the scenarios to permit the study of three central issues in eyewitness memory. One focus of this study, of clear forensic importance, is the potential effect of eyewitness involvement on the amount and accuracy of the information recalled about an event. Most previous adult and child (see Leippe, Romansczyk, & Manion, 1991) eyewitness research has examined the memories of bystanders who are not particularly involved in the eyewitness event. However, the uninvolved bystander is not a common feature in forensic contexts (Yuille & Tolhurst, 1992). Two studies that have addressed the issue of involvement used eyewitness-identification ability as the dependent variable of interest. In the first study by Hosch and Cooper (1982), undergraduates were assigned to one of three levels of personal involvement (victimization): no-theft control, impersonal calculator theft, or personalized watch theft. The participants in the watch-theft condition were subsequently the most accurate in identifying the perpetrator. However, they were not significantly more accurate than the witnesses to the less personal calculator theft. Still, the uninvolved witnesses fared less well in eyewitness performance.

In an investigation by Kassin (1984), pairs of undergraduates participated in what they believed to be a study of risk taking. In each session, the contest was interrupted by a confederate who entered the room and proceeded to steal one of the participant's (the victim's) game money while the other (the bystander) watched from across the table. Eight of the fifteen bystanders (53%) subsequently identified the perpetrator from a photograph lineup, whereas none of the victims were able to do so, in contrast to the Hosch and Cooper (1982) finding.
Obviously, further research is required on the issue of eyewitness involvement. Examining other independent variables such as the quantity and quality of the information contained in accounts of involved and uninvolved witnesses. It was possible to investigate the variable of involvement in this study by the nature of the training drills used at the college. One trainee in each role-play group is the officer on patrol who participates in any situation encountered. The trainees become very involved in these situations because they must satisfy their instructors of their practical expertise before they can graduate. The remaining members of the group are observers and are thus comparable to uninvolved bystanders. This arrangement permitted a comparison of the memory of participants and observers for the same event.

Stress has been of central interest in the study of eyewitness memory (see Christiansen, 1992; Deffenbacher, 1983). In forensic contexts, witnesses often experience considerable stress, and there is an unresolved issue about the consequences. In a review of the literature, Deffenbacher (1983) noted that 10 studies found no effect on subsequent recall or increased accuracy. In contrast, 11 studies were reported that demonstrated a detrimental effect of stress on memory (e.g., Siegel & Loftus, 1978). Elsewhere, Cheek (1980) went so far as to state that “victims of crime and witnesses to violence may be so horrified that they repress all conscious memory for important details” (p. ix). Many investigations addressing the impact of stress on eyewitness memory have defined stress as white noise or electric shock (e.g., Deffenbacher, 1983) that is independent of the event to be recalled by the participant (see Yuille & Tollestrup, 1992). However, in most criminal cases, the source of stress is precisely what the witness will be required to recall. Also, many of these studies have used slides or films (e.g., Clifford & Scott, 1978) or live, staged events (e.g., Leippe, Wells, & Ostrom, 1978) depicting violence as the to-be-remembered stimulus. Unfortunately, such methodologies may not attain the realism needed to generalize to forensic contexts. One study examining eyewitnesses to an actual shooting incident found that higher stress levels at the time of the event had no negative effect on memory (Yuille & Cutshall, 1986).

This study represents an attempt to exploit a situation in which stress is normally induced and yet sufficient control exists to permit the kind of research more characteristic of the laboratory. The police-training college setting permitted a manipulation of stress in a fairly realistic context. The Metropolitan Police consider it appropriate to introduce stressful events into the scenarios to assess the ability of the recruits to respond to the kind of conditions they will later encounter in their job. Hence, this is a “naturally” occurring situation that permits a degree of stress and subject involvement not generally attainable in the laboratory setting. Two similar events were used in the role plays: One was characterized by tension and anxiety, and the other was relatively unstressful. In addition, the videotaping of the events by one recruit provided a complete record of what transpired during the simulations.

Finally, the effect of delay on eyewitness recall was investigated in this study. Many factors, particularly court delays, often result in a lengthy passage of time between an offense and the recall of the event by a witness. Although considerable research has been conducted on the effect of delay on eyewitness memory, conclusions have been varied. Some studies have found a deterioration of memory over time (e.g., Lipton, 1977; Flin, Boon, Knox, & Bull, 1992), whereas others have found little effect of time on memory for remarkable events or faces. For example, an often-cited study by Bahrick, Bahrick, and Wittlinger (1975) demonstrated that memory for high school classmates’ faces is resistant to decay, being accurate after a mean of 48 years. Yuille and Cutshall (1986) found that the memories of witnesses to an actual shooting were highly accurate after a delay of several months. In this study, delay of recall was examined by testing the memory of half the subjects 1 week following the event and all the subjects after 12 weeks. Aside from any main effects delay might exhibit, we were also interested in the possibility that the delay variable would interact with levels of stress or involvement. Similarly, it was possible that an early recall trial would facilitate memory at a later time.

Method

Subjects

A total of 120 recruits (29 women) from the Metropolitan Police Training Centre at Hendon, England, took part in the study. Their ages ranged from 19 to 45 years (mean of 23.8). A new “intake” of about 200 recruits occurs every five weeks. The subjects were from two consecutive intake. Each intake is divided into 10 classes, and each class is further subdivided into four smaller groups for the role plays. Five classes from each intake participated in the study, and three recruits from each of the 40 groups served as subjects. The data of 10 trainees were discarded because of their unavailability for the delayed-recall test. The final sample contained 110 recruits.

Design

We used a $2 \times 2 \times 2$ design. Three recruits from each role play were interviewed: the officer dealing with the incident (participant) and two of the observers. Thus, one third of the subjects participated actively in the event, while two thirds were nonparticipant observers. For half of the participants and observers (those from the first intake), the role simulation was nonstressful. For the remainder, a more stressful situation was staged. Within each intake, half of the recruits were interviewed twice (1 week and 12 weeks after the event), and the remainder were interviewed only once (after 12 weeks). Because of practical limitations, all the recruits from an intake had to be assigned to the same condition regarding stress, resulting in a quasi-experimental design.

Role-Play Incidents

The role simulations involved a standard “stop-and-search” scenario that was the focus of the theoretical and legal classes the week before. In the basic situation, the patrol officer was called over by a member of the public (the “witness”) who informed the officer that a gentleman (the “suspect”) had been acting very suspiciously around a couple of cars in a parking lot. The suspect’s story was that he was a sales representative for Canon and was on his way to a business meeting with the managing director of a company to try to sell him some computer equipment. He carried a briefcase that contained some manuals and brochures to corroborate his story. It was the officer’s task to stop and question the man to determine whether he owned any of the cars in the lot. Although the man was innocent, the officer had to establish this fact by a thorough investigation. Both the suspect and witness were played by individuals who were unknown to the recruits. In the nonstressful version of the event, the suspect was early for his meeting and very polite and cooper-
Results

Stress and Difficulty Ratings

The recruits rated the stressful event as more stressful ($M = 4.40$) than the nonstressful event ($M = 2.45$; $t = 7.02$, $p < .01$). The participants and observers did not differ in their ratings. A similar pattern was found for the rated difficulty of the event ($M = 5.15$ for the stressful event and $M = 2.77$ for the nonstressful one; $t = 9.67$, $p < .01$). Again the participants and observers did not differ, which indicated that the two different role-play scenarios were indeed perceived by both participants and observers as being differentially stressful and difficult experiences.

Amount Recalled

Table 1 provides a summary of the total mean amounts of information recalled by the officers who gave two separate recall accounts. An analysis of variance (ANOVA) was conducted on this criterion. There was a significant effect of stress, $F(1, 58) = 12.64$, $p < .01$; with a greater amount of recall associated with the nonstressful event ($54.00$ vs. $42.21$ details). The witnesses recalled more information after 1 week ($51.54$ details) than after 12 weeks ($45.43$ details), $F(1, 58) = 22.74$, $p < .01$. As the means in Table 1 suggest, these two variables interacted, $F(1, 58) = 9.71$, $p < .01$, indicating that the loss of information over time occurred only with the stressful event.

Table 2 indicates the mean amounts of information (number of details) recalled 12 weeks after the event occurred. These means demonstrate the impact of an early (1 week) recall opportunity on subsequent recall ability. The trainees who recounted the event after one week remembered many more details ($44.71$) after 12 weeks than those for whom the 12 week recall trial was their first ($35.20$), $F(1, 106) = 11.42$, $p < .01$. There was also a strong effect of stress, $F(1, 106) = 25.24$, $p < .01$. The recruits in the stressful role simulation recalled fewer details ($33.82$) than those witnessing the less stressful event ($46.88$). Finally, the participants recalled more details ($44.94$) than the observers ($37.42$), $F(1, 106) = 8.32$, $p < .01$.

The information recalled consisted of approximately $42\%$ actions, $42\%$ person descriptions, and $16\%$ object descriptions. The effects of stress, number of recall trials, and delay reported above were replicated with all three types of information.

Accuracy of Recall

Table 3 indicates the mean accuracy scores (percentage of details recalled that were correct) for the trainees who had two recall trials. An ANOVA conducted on these results yielded only one significant effect: the recruits in the nonstressful role plays were less accurate ($67.5\%$) than those in the stressful role plays ($74.3\%$), $F(1, 58) = 8.51$, $p < .01$. The means in Table 4 contrast the accuracy of recall after a 12-week interval of the trainees who had recalled 1 week following the role plays and those who did not. There was a main effect of stress, $F(1, 106) = 23.32$, $p < .01$, with the trainees who witnessed the stressful scenario displaying higher accuracy ($74\%$) than those who had witnessed the nonstressful event ($62\%$). There was also an effect of the number of recall trials, $F(1, 106) = 6.57$, $p < .05$. The trainees...
who provided an account after 1 week were more accurate after 12 weeks (71%) than those who did not (65%).

Again, the pattern of effects found with the overall accuracy rates was replicated for action, object, and person details.

Discussion

In recent years, the importance of studying the behavior of eyewitnesses under more naturalistic conditions has been recognized. This study represented an examination of the effects of stress, involvement, and retention interval on adult eyewitness ability in a setting somewhat analogous to a forensic context. The results obtained make several contributions to our body of knowledge concerning eyewitness behaviors.

The variable of stress showed complex effects. In the interviews 12 weeks after the events, the subjects recalling the stressful situation were more accurate than those recalling the nonstressful situation. This occurred regardless of whether they had been interviewed previously. One explanation for this effect on accuracy and retrieval practice may be what Yuille and Cutshall (1989) have termed remarkable memories. This simply means that people are more likely to ruminate and remark on a unique experience (e.g., a stressful eyewitness experience) than on a mundane one. This is distinguished from "flashbulb memories" (Brown & Kulik, 1977) chiefly in terms of the impact of the event and the encoding mechanism. A flashbulb memory refers to the creation of highly accurate images in memory for the circumstances in which the person had heard about a newsworthy event of high emotional impact (e.g., the death of a president) at the time of initial encoding. Remarkable events, alternatively, are personally meaningful simply because of their novelty and interesting qualities. Yuille and Cutshall (1989) proposed that people can exhibit quite accurate and detailed memories for remarkable events because they have had practice in accessing those memories (retrieval practice). Similarly, rehearsal occurs through reflection on the event and remarking on the incident to others. It is likely in this study that the stressful encounter was intrinsically interesting and unique to the recruits, which resulted in frequent accessing and recounting of the event.

An opposite effect was evidenced for the amount of information recalled. The police cadets in the higher stress condition consistently provided less information than their peers. This occurred despite the fact that the mean duration time of the stressful event was longer than that of the nonstressful one. One explanation is Easterbrook’s (1959) hypothesis that stress narrows the focus of attention; the recruits who experienced the stressful event may have responded by narrowing the scope of their attention to those aspects producing the stress. This narrowing would reduce the amount of information encoded and later reported.

As noted above, this lesser degree of detail in no way compromised the accuracy of the subjects’ accounts. Taken independently, the greater accuracy of the subjects in the stressful condition suggests that they were near the top of the conventional inverted U-shaped function of the relationship between memory and stress. The moderate mean stress rating given to the stressful event (4.4 on a 7-point scale) certainly supports this contention. However, the greater accuracy of the stressed subjects was accompanied by a reduction in the quantity of information being recalled. These findings on eyewitness performance cannot be explained by a simple inverted-U function. They are, however, in accord with Yuille and Tollestrup’s (1992) suggested model concerning the effects of emotion (including stress) on memory. One of their proposed mechanisms by which stress affects memory is that some witnesses focus directly on the source of stress, attending much more to core details than to peripheral ones. Some of these witnesses then proceed to recount the event to themselves and others. Thus, memory for the event will be accurate but limited in scope. As previously mentioned, either criteria (qualitative or quantitative) could be the essential factor in a particular criminal investigation.

Another manipulation in this study was the dichotomy of subjects into participants and observers. Although this variable did not have a substantial effect, the effect it did evidence was

<table>
<thead>
<tr>
<th>Group</th>
<th>Nonstress</th>
<th>Stress</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1st trial</td>
<td>2nd trial</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participants</td>
<td>45.6 (11.4)</td>
<td>38.7 (7.1)</td>
</tr>
<tr>
<td>Observers</td>
<td>53.3 (11.6)</td>
<td>45.3 (13.2)</td>
</tr>
</tbody>
</table>

Table 2

Mean (SD) Total Details Recalled (Correct and Incorrect) After 12 Weeks

Table 1

Mean (SD) Total Details Recalled (Correct and Incorrect) for Subjects With Two Recall Tests

<table>
<thead>
<tr>
<th>Group</th>
<th>Nonstress</th>
<th>Stress</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1-week delay</td>
<td>12-week delay</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participants</td>
<td>59.3 (8.7)</td>
<td>51.1 (8.9)</td>
</tr>
<tr>
<td>Observers</td>
<td>53.3 (11.6)</td>
<td>45.3 (13.2)</td>
</tr>
</tbody>
</table>
complex. Participation interacted with retention interval; the participants remembered more than the observers in the 12-week interview. Otherwise the accounts were not quantitatively or qualitatively different. Thus, it appears that higher levels of participation by eyewitnesses are associated with more enduring memory traces for the event. Although the participants and bystanders seem to have encoded comparable information regarding the event, the memories of the participants were more intact after a 12-week interim.

The recall performance of the recruits indicated a strong retrieval practice effect. Twelve weeks after the event, the recruits who had been interviewed 11 weeks previously recalled more information and were more accurate than those for whom the 12th-week interview elicited a first recall account. This replication of Roediger and Payne's (1982) finding is in itself interesting because the probationer constables of this experiment are quite a different population from the undergraduate students traditionally used in eyewitness memory studies. This finding supports the generalization of the practice retrieval effect to forensic contexts. Providing victims with an early opportunity to describe their experience may substantially improve their later ability to recall the event (e.g., at trial). The benefits of an immediate recall do not appear to be mitigated by the stress of an event (this study) or by alcohol in moderate amounts (Yuille & Tollestrup, 1990). Also, there is some evidence that this effect occurs with children (Dent & Stephenson, 1979).

In this study, stress and participation levels strongly affected eyewitness ability in a complex fashion. It should be noted that it is on these two dimensions that findings in laboratory research differ from the reporting behavior of actual eyewitnesses. Because these are interactional effects, it is difficult to state which laboratory-based findings are relevant to the forensic context.

There are a number of restrictions that must be placed on the generalizability of the present findings. The first concerns the differences between the nonstressful and stressful role simulations. Although the events were equated as much as possible, the duration and structure of the events were necessarily different.

Thus, any overall findings concerning the effects of stress may be confounded by these factors. Alternatively, the fact that recruits in the shorter nonstressful event recalled more information seems to contradict this hypothesis. Also, the findings within the role-play groups (i.e., concerning the participants vs. observers) are unlikely to be affected by these factors. Second, the recruits in this study may not have been entirely representative of eyewitnesses in forensic settings in terms of their levels of personal involvement. For example, these subjects were in a situation in which their performance was being evaluated for their abilities to observe and handle a suspect effectively. In addition, they were aware beforehand of the possibility that their memories might be tested. The nonparticipant observers may not have been as detached from the event as some eyewitnesses who view an offense: in this case, the observers held the knowledge that their comrade was being evaluated and that they would be similarly evaluated at a later time. Finally, the stress induced in this study did not attain a level that is frequently experienced by a victim of crime. For example, at extreme levels of stress, eyewitness performance may deteriorate in accordance with the Yerkes-Dodson model. Future studies should explore naturalistic situations in which higher levels of stress are present (e.g., Goodman and her colleagues have interviewed children about high stress medical procedures).

In light of these issues, it is conceivable that some of the obtained effects were artifacts of the rather unique environment used in this study. Because of lack of control inherent in such nontraditional research venues, these variables will have to be examined in a variety of ways before more can be stated conclusively. Because of ethical and practical constraints, much of this research will necessarily consist of quasi-experimental field and case studies. However, if we were to rely solely on such paradigms, the picture we would obtain of human memory would be as incomplete as the one derived from purely laboratory-based research. Laboratory studies, field research, and archival work are all required. Each of these endeavors, as Davies (1990) has argued, makes "a different compromise between ecological validity and methodological rigour" (p. 21). Only when we have

---

Table 3

<table>
<thead>
<tr>
<th>Group</th>
<th>Nonstress</th>
<th>Stress</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1-week delay</td>
<td>12-week delay</td>
</tr>
<tr>
<td>Participants</td>
<td>68.1 (10.4)</td>
<td>63.1 (8.7)</td>
</tr>
<tr>
<td>Observers</td>
<td>68.0 (14.3)</td>
<td>68.2 (11.2)</td>
</tr>
</tbody>
</table>

---

Table 4

<table>
<thead>
<tr>
<th>Group</th>
<th>Nonstress</th>
<th>Stress</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1st trial</td>
<td>2nd trial</td>
</tr>
<tr>
<td>Participants</td>
<td>57.3 (10.3)</td>
<td>63.1 (8.7)</td>
</tr>
<tr>
<td>Observers</td>
<td>56.2 (8.2)</td>
<td>68.0 (11.2)</td>
</tr>
</tbody>
</table>
converging results from such diverse methods of inquiry should we feel confident in generalizing to the forensic context, where any such assertions may have serious ramifications.

References


Received September 4, 1992
Revision received May 16, 1994
Accepted May 18, 1994