APPLIED COGNITIVE PSYCHOLOGY
Appl. Cognit. Psychol. **16**: 937–945 (2002)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/acp.924

The Development of a Computer System to Collect Descriptions of Culprits

OTTO H. MacLIN,[1]* RYAN L. TAPSCOTT[1]
and ROY S. MALPASS[2]

[1]*University of Northern Iowa, USA*
[2]*The University of Texas at El Paso, USA*

SUMMARY

Recent research has demonstrated that verbal overshadowing occurs when a witness is forced to provide details of the culprit that are not readily available, but the effect does not occur when the witness is warned to provide only the information they are absolutely sure of (Meissner *et al.*, 2001). The present study attempts to replicate this effect and to further examine the instructional manipulation in the development of a computerized identification system, PC_Eyewitness (PCE). Among other things, PCE is designed to elicit verbal descriptions from witnesses. Overall, results from this study replicate those found by Meissner *et al.* (2001) in demonstrating lower identification performance for participants forced to provide descriptive details. However, no verbal overshadowing was observed for the computerized feature checklist presented as either verbal features or pictorial features. Implications for the development of a computerized system of eyewitness identification are discussed. Copyright © 2002 John Wiley & Sons, Ltd.

INTRODUCTION

Witnesses to a crime are at times repeatedly questioned and asked to provide verbal descriptions of the perpetrator. The quality of descriptions are important because it is the information that is used to see if the culprit can be located in the vicinity of the crime by other officers, for officers to determine if the description fits offenders in other cases they are (or have been) working on, for lineup construction (Luus and Wells, 1991), mock witness lineup evaluation (Malpass and Lindsay, 1999; Corey *et al.*, 1999; Valentine and Heaton, 1999), construction of facial composites (Cutler *et al.*, 1988; Koehn and Fisher, 1997), and/or to compare with the physical appearance of a suspect once apprehended. While verbal descriptions can aid in the law-enforcement process, they have also been shown to hinder subsequent recognition of the culprit. This effect, termed 'verbal overshadowing', is characterized by decreased recognition rates following a verbal

*Correspondence to: Otto H. MacLin, Department of Psychology, University of Northern Iowa, Cedar Falls, IA 50614. USA. E-mail: otto.maclin@uni.edu

Contract/grant sponsor: National Science Foundation.
Contract/grant number: SBR-9730937.

description of an individual. (Dodson *et al.*, 1997; Ryan and Schooler, 1998; Schooler and Engstler-Schooler, 1990; Schooler *et al.*, 1997). While verbal overshadowing has been found to be robust across many studies (see Meissner and Brigham, 2001 for a meta-analytic review) the effect is small and some labs have failed to consistently find the effect (Finger and Pezdek, 1999; Meissner *et al.*, 2001).

The research presented here examines the development of a computerized system designed to collect witness information, and to construct and administer lineups (MacLin *et al.*, under review). Specifically, we examine the aspect of the system that may be susceptible to verbal overshadowing, namely the collection of the witness's description of the culprit. When developing a system to collect witness descriptions, there are several methods one could use. The law-enforcement officer could interview the witness using a procedure such as a cognitive interview (Finger and Pezdek, 1999; Fisher and Geiselman, 1992; Fisher *et al.*, 1990). More simply, the witness could write the description down on a pad of paper or be administered a facial feature adjective checklist (e.g. FFAL; Ellis, 1986). Another method used by the Federal Bureau of Investigation (FBI) to elicit witness descriptions to aid in the construction of facial composites is to show the witness pictures representing facial characteristics similar to those that appear in the FFAL. Knowing how the different methods of collecting descriptions of the culprit might affect a witness's ability to identify the suspect is important.

Additionally, it is important to understand how recall instructions provided to the witness might affect the identification process (MacLin, this issue). In a recent series of studies conducted by Meissner and his colleagues (Meissner *et al.*, 2001; Meissner, this issue) an instruction criterion was manipulated to determine its affect on verbal overshadowing. It was found that when participants were forced to provide a detailed description of a culprit, compared to being warned to provide only the details they were absolutely certain of, identification accuracy of the suspect decreased.

The manipulation of instructional criterion may be important because feature checklists, compared to self-generated descriptions, are more exhaustive. The FFAL contains 53 items where the 'witness' or participant rates individual facial features and shapes (Ellis, 1986). Even though feature checklists often have the option to respond with 'Don't Know' or 'Not Applicable' the witness still has to read the verbal labels presented and thus be exposed to covertly verbalizing and considering the feature. Therefore, the checklists, by nature of the task, 'force' the witness to consider the feature and compare it to their visual image of the culprit. Wogalter (1991) found that participants using a checklist to describe faces performed poorer on a subsequent recognition task compared to participants in a free recall condition (also see Wogalter, 1996). It was concluded that the irrelevant descriptors contained in the checklist became incorporated into the participants' representation of the target face (Wogalter, 1996). Thus, an instruction that heightens the participant's criterion of responding on a checklist task may guard against this potential for misinformation.

A pictorial inventory, similar to one used by FBI composite sketch artists (United States Department of Justice, 1988), may also be presented to witnesses to extract their memory for the culprit. The FBI's feature inventory checklist (FIC) contains 13 facial feature categories with a total of 60 unique features. Each of the 60 unique features is represented on a page of the catalogue. Each page consists of 16 faces that have a similar appearance for a particular feature (e.g. protruding ears, connected eyebrows).

The present study was designed to replicate the findings of Meissner *et al.* (2001) using a different set of stimuli (video of a staged crime scene event and lineup) and varying the

type of description tasks that were presented. In addition to the traditional self-generated (recall) description task, subsets of participants were also administered either a feature checklist task similar to the FFAL or a pictorial ratings task (based on the FIC). These tasks differed from the traditional description task in that they were provided to participants via a computerized program, consistent with our attempts in developing PCE. As with Meissner *et al.*, participants in each description condition were randomly assigned to one of two instructional manipulation conditions: forced or warning. A no-description control condition was also included to assess the influence of verbal overshadowing in the description conditions. All participants viewed a video of a staged theft and completed a distracter task. Participants were then randomly assigned to one of the description conditions, or to a no-description control condition. Finally, all participants then completed the identification task.

Based on the findings of Meissner *et al.* (2001), we hypothesized that participants in the warning condition would not show evidence of overshadowing while those in the forced description condition would perform significantly worse than participants in the no-description control condition. We also hypothesized that participants in the verbal recall and the verbal FFAL checklist condition would show overshadowing while those in the pictorial task would not. Pictorial cues used in the description task should reduce but not eliminate the generation of verbal descriptions for the face in memory.

METHOD

Participants

Participants were 160 undergraduate students at the University of Northern Iowa. All participants received class credit for their participation.

Materials

Crime scene video

The video portrayed two men and one woman studying in a college library. The woman received a call on her cellular phone, exited the room, leaving her laptop computer behind. After her exit, one of the males left his chair and casually walked towards the woman's belongings at which point he stole her laptop computer. The other male confronted the thief before he left the room. The video lasted 1 minute and 30 seconds.

Digit search puzzle

The digit search puzzle was used for a distracter task. The puzzle consisted of two worksheets of simple mathematical equations. One worksheet contained 39 addition and subtraction equations with a space to provide an answer. The other worksheet was the numerical equivalent to a word-search task. Mathematical equations were hidden among numbers and operator signs. The puzzle contained both correct and incorrect addition and subtraction equations appearing horizontally and vertically in a 16 rows by 29 columns digit array. Space was provided to circle the correct equations hidden within the puzzle.

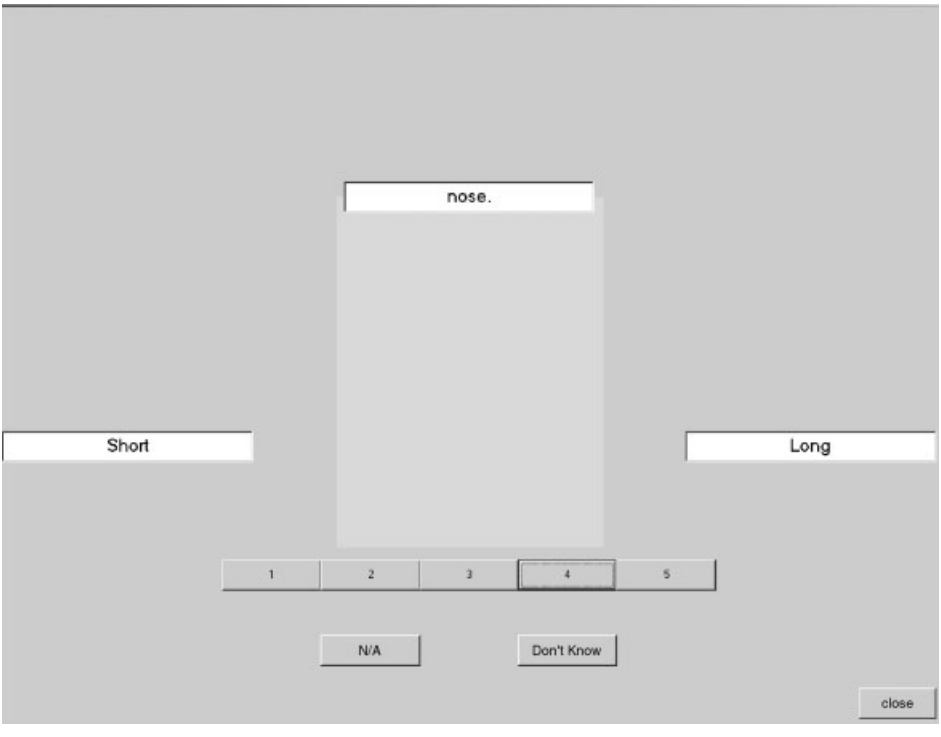

Figure 1. Screenprint of the computerized verbal checklist condition-warning condition

Description recall

A paper and pencil recall task was designed to be similar to that used by Meissner *et al.* (2001). Participants were presented with a response sheet containing lines numbered 1–25 for the participants to provide the verbal description.

Computerized checklist

Two versions of the computerized feature checklist were constructed: a verbal feature checklist (Figure 1) and a pictorial feature check list (Figure 2). The verbal checklist was loosely designed around the FFAL containing 31 facial features from 11 feature categories (e.g. nose, eyes, etc.). Attributes of these features (e.g. short versus long) functioned as anchors for a 5-point rating scale.

Analogous pictorial descriptions of those 31 features were used from the FIC to make the computerized pictorial feature checklist (see Figure 2).[1]Images from the FIC were used in place of the verbal anchors from the FFAL to create a pictorial feature checklist.[2] Because the verbal checklist requires the addition of the feature category with the

[1]While it is conceivable that the anchor pictures for the scales differ along more than one dimension, pilot data indicate approximately 85% agreement of feature dimensions across judges. These data have been reported elsewhere (Tapscott and MacLin, presentation at the 14th Annual Meeting of the American Psychological Society, New Orleans, June 2002).

[2]For example, the FFAL uses the adjectives 'short' and 'long' when rating the nose. Images of a face with a short nose and faces with long noses replaced the anchors in the pictorial checklist.

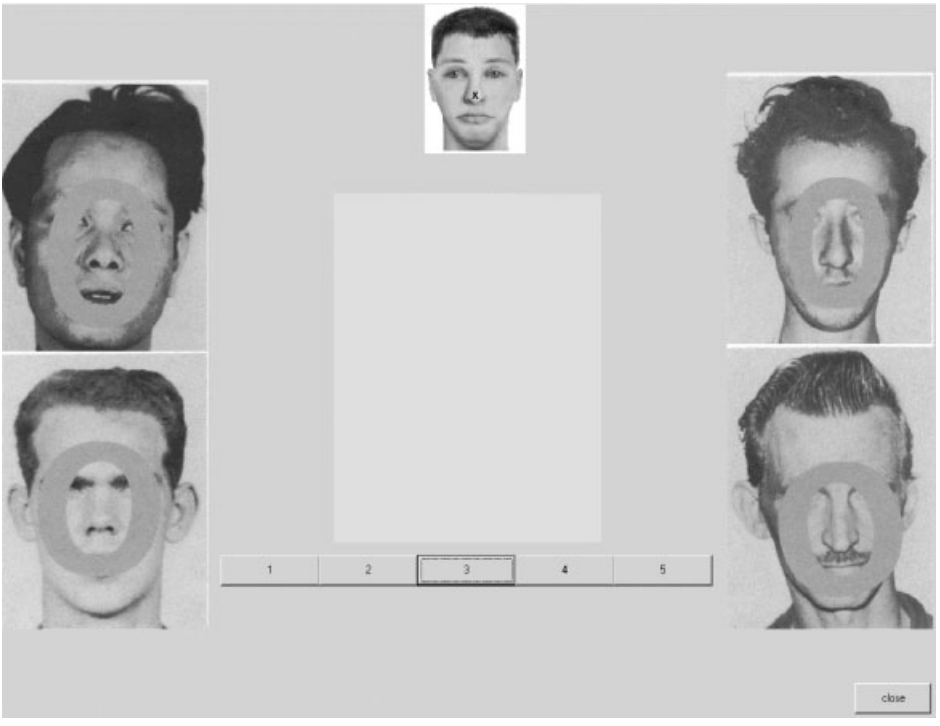

Figure 2. Screenprint of the computerized pictorial checklist-forced condition

descriptors (e.g. eyes: wide-set versus close-set), an icon was added to the pictorial checklist to indicate the feature category being examined (see Figure 1).

Two versions of each computer program were constructed so that both a warning and forced condition were produced. The program for the warning condition contained the addition of a 'N/A' (not applicable) and a 'Don't Know' button in addition to the five buttons to indicate a feature response (see Figure 1). The program for the forced condition did not have the two additional buttons (see Figure 2).[3]

Lineup and response sheet

A target-present lineup was used in the study and constructed in a 2×3 array. The suspect's position was counterbalanced in the lineup over all six positions to reduce position effects. A response sheet accompanying the lineup included six boxes in a 2×3 checkbox array. Participants were instructed to make a selection based on the position of the suspect. An additional option of 'target not present' was included to allow participants to reject the lineup. Only 'target present' lineups were used in this study because we were

[3]A pilot study was conducted to examine how participants would use these additional buttons. Judges in the pictorial warning condition used the used the 'N/A' option an average of 4.69 times (15.1%) and the 'Don't Know' option and average of 4.27 times (13.8%). Judges in the verbal warning condition used the 'N/A' option an average of 8.96 times (28.9%) and the 'Don't Know' option and average of 0.53 times (0.02%). Both the pictorial and verbal warning judges used the combination of the 'N/A' and 'Don't Know' button an equivalent number of times ($M = 8.96$ and $M = 9.11$ respectively).

attempting to replicate the results of Meissner *et al.* (2001) who only used target-present lineups.

Design

A 2 (Instruction Criterion: forced versus warning) × 3 (Description Task: recall, verbal checklist, or pictorial checklist) between-subjects design was used in which identification accuracy was the dependent variable. A no-description control condition was also included for comparison purposes.

Procedure

Participants were each tested at individual workstations in groups ranging in size from 1 to 12 participants per session. After filling out informed consent documents participants viewed the crime video. They were instructed to pay close attention to the video played on the workstation computer monitor, as they would be asked questions about it later. Participants were then asked to complete the distracter task (digit search puzzle) for 5 minutes prior to being randomly assigned to one of the three experimental conditions. All participants were given standard lineup instructions equivalent to Meissner *et al.* (2001). The standard instructions provided to participants in both computer checklist conditions are as follows:

> In the spaces below, please describe in as much detail as possible the face of the culprit you saw in the video. Use the lines below to provide details such that someone else could identify him on the basis of the description. As describing a face is often a difficult task, it is important that you concentrate and stay focused for the next few minutes.

In addition to the standard instructions given, participants in the warning condition were told:

> Prior research has also demonstrated the importance of striving for accuracy and reporting only that which you are certain you can remember. You do not have to fill in all of the lines, so be sure to report only those details you are confident of, and do not attempt to guess at any particular feature.

Participants in the forced condition were given the following instructions accompanying the standard instructions:

> Prior research has also demonstrated the importance of reporting everything you can remember about the culprit in the video. Try not to leave out any details about the face even if you think they are not important. You must fill in all of the numbered lines below with a description of the face, even if you start to feel that you are guessing.

Participants in the control condition were not asked to provide a verbal description of the perpetrator. They completed the distracter task and then were given the recognition task. Participants in the control condition were given extra time to complete the distracter task (approximately 3 minutes).

Table 1. Mean accuracy on identification task as a function of type of computerized task and type of instruction criterion

	Response criterion		
Description condition	Forced	Warning	Control
Recall	0.30	0.70*	
Verbal checklist	0.46	0.46	
Pictorial checklist	0.33	0.58*	
Control			0.54

*Significant $p < 0.05$.

Once the descriptions were obtained, participants were given the lineup response sheet and a folder containing the lineup. Participants were given fair lineup instructions specifying that the target may or may not be present. They were then directed to check the box corresponding to the suspect's position in the lineup, or circle the 'target not present' option. The study concluded and the participants were debriefed.

Results and discussion

Overall, participants in the warning condition had a higher rate of identifying the culprit (61.4%) than those in the forced condition (37.9%) ($\chi^2(1, n = 136) = 7.607$, $p < 0.01$). Planned comparisons indicated that participants in the description recall condition provided more accurate identifications when given the warning criterion than those with the forced criterion (see Table 1; $\chi^2(1, n = 40) = 6.04$, $p < 0.01$). A similar pattern was evident with the pictorial checklist ($\chi^2(1, n = 48) = 5.33$, $p < 0.05$); however, no significant difference was found for the verbal checklist condition. Recognition performance was 54.1% for the control condition. Comparisons were also conducted for recognition performance in the control group against the six experimental groups. Planned comparisons indicated that no significant differences existed between the control and the experimental conditions.

Recognition was significantly less when participants were forced to recall more details about the culprit's description than when they were admonished to only provide the information they are sure of. Meissner *et al.* (2001) found that the description conditions significantly differed in the number of correct, incorrect, and subjective details produced when describing the target face. Meissner *et al.* (2001) also observed that participants in the forced condition not only provided more correct details than those in the warning condition, but they also provided more subjective and incorrect details about the culprit than did those in the warning condition. These results validate the idea that instructional manipulations did significantly influence participants' overall description accuracy, thus affecting their subsequent accuracy in performing the identification task. While it is not in the scope of this study to examine the content of the written descriptions, it should be noted that participants in the forced condition provided a mean of 19.4 descriptions compared to a mean of 7.1 for the warning group. These results replicate findings from previous studies manipulating instructional criterion (Meissner *et al.*, 2001; Meissner, this issue).

Clearly, forcing the witnesses to provide verbal ratings for all of the features did not affect recognition performance for the verbal checklist condition as hypothesized. It was

expected that participants in the verbal forced checklist condition would perform similarly to those in the verbal recall condition who were forced to provide a verbal description. It has been demonstrated that when participants are forced to provide a description from recall, they report erroneous details. Meissner *et al.* (2001) found the generation of erroneous details as the major contributor to misidentification. However, when participants in the verbal checklist were forced to rate features they may have taken a conservative strategy and reported a value in the middle of the scale. Data from the pilot studies lend some support to this notion. The other question that remains is why there were fewer correct identifications made in the pictorial forced checklist condition than in the warning condition. If verbal overshadowing occurs as a result of converting a visual mental image in memory to a verbal code, then why would a decrement in recognition performance occur with pictorial stimuli? Sure, pictorial stimuli can be processed using both imagery and verbalizations (Paivio, 1986), and the verbal component may account for the overshadowing; however, no decrement was evident in the verbal feature checklist when participants were forced to provide descriptions. A parsimonious explanation would be that participants made errors when forced to rate the features in the pictorial task. These errors may have possibly interfered with the original mental image. Consistent with such a notion, it has been found that when witnesses view other facial images not belonging to the culprit, errors in identification increase (Brigham and Cairns, 1988; Gorenstein and Ellsworth, 1980).

When designing computer applications for use in law enforcement it is important to understand how they might be affected by psychological variables such as verbal overshadowing. Tasks involving descriptions should be designed to minimize interference with a witness's memory and subsequent ability to make a correct identification. At this juncture, it appears that instructing witnesses' to maintain a conservative output criterion when generating their own descriptions results in higher subsequent identification accuracy. However, this may be problematic given that participants in the warning recall condition generated as few as 7.1 details on the average. Furthermore, Wells and Hryciw (1984) found that participants often generated subjective details (e.g., personality and occupation characteristics) about the culprit. This may aid in face recognition, but subjective details constitute poorer verbal descriptions when used by a second party. Additionally, research has shown that verbal descriptions for information not typically verbalized, such as faces, are less accurate (Leibowitz *et al.*, 1993). When verbal descriptions are used by second-parties, such as law enforcement, the quality (and perhaps quantity) of verbal descriptors becomes more important. To explore the use of verbal descriptions by second-parties, we are currently examining the ability of a second party to the crime to identify a suspect based on the descriptions provided from the witnesses' free recall, the verbal checklist, or the pictorial checklist. Additionally, we are developing the pictorial checklist to use complete facial images, without the masking found in facial pictures used by the FBI in the FIC. Our goal is to compare complete facial images with masked images to examine the effects of verbal overshadowing.

ACKNOWLEDGEMENTS

The data presented here were presented at the 14th Annual Convention of the American Psychological Society in New Orleans. The authors would like to thank Kathy Pezdek and an anonymous reviewer for helpful comments on revisions of this manuscript. This

research was supported in part by a grant from the National Science Foundation awarded to the third author (#SBR-9730937).

REFERENCES

Brigham JC, Cairns DL. 1988. The effect of mugshot inspections on eyewitness identification accuracy. *Journal of Applied Social Psychology* **18**: 1394–1410.

Corey D, Malpass RS, McQuiston DE. 1999. Parallelism in eyewitness and mock witness identification. *Applied Cognitive Psychology* **13**: S41–S58.

Cutler BL, Stocklein CJ, Penrod SD. 1988. Empirical examination of a computerized facial composite production system. *Forensic Reports* **1**(3): 207–218.

Dodson CS, Johnson MK, Schooler JW. 1997. The verbal overshadowing effect: why descriptions impair face recognition. *Memory and Cognition* **25**(2): 129–139.

Ellis HD. 1986. Face recall: a psychological perspective. *Human Learning* **5**: 189–196.

Finger K, Pezdek K. 1999. The effect of cognitive interview on face identification accuracy: release from verbal overshadowing. *Journal of Applied Psychology* **84**: 340–348.

Fisher RP, Geiselman RE. 1992. *Memory-enhancing Techniques for Investigative Interviewing*. Charles C. Thomas: Springfield, IL.

Fisher RP, Quigley KL, Brock P, Chin D, Cutler BL. 1990. The effectiveness of the cognitive interview in description and identification tasks. In *Eyewitness Memory: Current Trends and Developments*, Ross D, Read J, Toglia M (eds). Cambridge University Press: Cambridge.

Gorenstein GW, Ellsworth PC. 1980. Effect of choosing an incorrect photograph on a later identification by an eyewitness. *Journal of Applied Psychology* **65**: 616–622.

Koehn CE, Fisher RP. 1997. Constructing facial components with the Mac-A-Mug Pro system. *Psychology, Crime and Law* **3**(3): 209–218.

Leibowitz HW, Guzy LT, Peterson E, Blake PT. 1993. Quantitative perceptual estimates: verbal versus nonverbal retrieval techniques. *Perception* **22**(9): 1051–1060.

Luus CAE, Wells GL. 1991. Eyewitness identification and the selection of distracters for lineups. *Law & Human Behavior* **15**: 43–57.

Malpass RS, Lindsay RC. 1999. Measuring line-up fairness. *Applied Cognitive Psychology* **13**: S1–S7.

Meissner CA, Brigham JC. 2001. A meta-analysis of the verbal overshadowing effect in face identification. *Applied Cognitive Psychology* **15**(6): 603–616.

Meissner CA, Brigham JC, Kelley CM. 2001. The influence of retrieval processes in verbal overshadowing. *Memory & Cognition* **29**(1): 176–186.

Paivio A. 1986. *Mental Representations: A Dual Coding Approach*. Oxford University Press: New York.

Ryan RS, Schooler JW. 1998. Whom do words hurt?: individual differences in susceptibility to verbal overshadowing. *Applied Cognitive Psychology* **12**: S105–S125.

Schooler JW, Engstler-Schooler TY. 1990. Verbal overshadowing of visual memories: some things are better left unsaid. *Cognitive Psychology* **22**: 36–71.

Schooler JW, Fiore RS, Brandimonte MA. 1997. At a loss from words: verbal overshadowing of perceptual memories. In *Handbook of Learning and Motivation*, Medin D (ed.). Academic Press: San Diego, CA; 293–334.

United States Department of Justice. 1988. *FBI Facial Identification Catalog*. Prepared by Graphics Design Unit, Special Projects Section, Laboratory Division.

Valentine T, Heaton P. 1999. An evaluation of the fairness of police line-ups and video identifications. *Applied Cognitive Psychology* **13**: S59–S72.

Wells GL, Hryciw B. 1984. Memory for faces: encoding and retrieval operations. *Memory & Cognition* **12**: 338–344.

Wogalter MS. 1991. Effects of post-exposure description and imaging on subsequent face recognition performance. *Proceedings of the Human Factors Society* **35**: 575–579.

Wogalter MS. 1996. Describing faces from memory: accuracy and effects on subsequent recognition performance. *Proceedings of the Human Factors and Ergonomics Society* **40**: 536–540.

