Effects of Post-exposure Description and Imaging on Subsequent Face Recognition Performance

Michael S. Wogalter
Department of Psychology
Rensselaer Polytechnic Institute
Troy, NY 12180

ABSTRACT

After viewing a crime, witnesses are frequently asked by police investigators to give a verbal description of the perpetrator. At a later time, witnesses may be asked to try to recognize the perpetrator in a lineup or a mugfile. The purpose of this research was to determine whether performance on a later recognition test is influenced by an earlier verbal description test of the target. Participants viewed six target faces, and after each, performed one of four post-exposure activities. Two were verbal description tasks. The descriptor checklist task had participants indicate, using a list of potential adjectives, those descriptors that best described the most recently seen face. The descriptor generate task had participants recall descriptors that best described the most recently seen face. Both description tasks were accomplished using a questionnaire that was identical except for the presence or absence of specific adjectives. Other participants were instructed to image the target face or they performed an irrelevant task during the post-exposure period. In a later recognition test, participants tried to find the targets in a set of faces containing 134 distractors. The results showed that participants using the checklist produced lower performance on the recognition test than participants who generated their own descriptors. Imaging produced the highest recognition performance but was not significantly different from the irrelevant condition. The recognition decrement of the descriptor checklist is explained in terms of interference caused by the presence of irrelevant face descriptors that added noise to participants' memory of the targets. Implications for testing witness memory by police investigators are discussed. For example, if verbal descriptions of faces are requested from eyewitnesses, descriptor recall may be preferred over a descriptor recognition because the former task does not degrade later recognition of the perpetrator.

INTRODUCTION

After viewing a crime, witnesses are frequently asked by police investigators to give a verbal description of the perpetrator. At a later time, witnesses may be asked to participate in other tests of memory, for example, recognition of the perpetrator in a lineup or a mugfile. The purpose of the present research was to determine whether performance on a later recognition test is influenced by an earlier verbal description test.

Previous research is equivocal on whether verbal description influences subsequent recognition. Target-related description has been reported to degrade (Dent & Stephenson, 1979; Schooler & Engstler-Schooler, 1990; Williams, 1975), to facilitate (Chance & Goldstein, 1976; Read, 1979), and to have no effect (Hall, 1977; Mauldin & Laughery, 1981; Wogalter, Laughery, & Thompson, 1989) on subsequent face recognition performance. Unfortunately, each of these studies used different techniques of eliciting the descriptions. Thus, it is unclear whether the results differed because of the questionnaire instruments they used (e.g., how many and what kinds of verbal cues witnesses were given during the description process) or other aspects of the particular methodologies used in each study. To date empirical work directly comparing different verbal description techniques has been virtually nonexistent, except for one study. Goulding (1971) had police officers make cued or free descriptions to target faces. The free description method was found to produce better descriptions than the cued description method. However, the effect of these two description methods on subsequent recognition was not tested. The present research directly compared a description technique that provided specific potential descriptors (descriptor checklist) to a technique that only provided a general list of feature headings in which participants produced their own terms (descriptor generate) and examined their effect on a later recognition test.

Another activity the witness might perform before a recognition test is visual imaging. Few studies have examined the effects of imaging, but those that have tend to support the notion that covert visual rehearsal facilitates subsequent recognition performance (Graefe & Watkins, 1980; Read, 1979; Read, Hammersley, Cross-Calvert, & McFadzen, 1989; Sporer, 1988; Wogalter, Cayard, & Jarrard, in press). However, the effect of imaging in these studies was quite small. Other research has found no effect (Schooler & Engstler-Schooler, 1990) or a negative effect (Read, et al., 1989) of imaging on subsequent recognition. The possible effect of post-exposure imaging on recognition is reexamined, and is compared to the verbal description conditions and an irrelevant control activity.

METHOD

Participants and design

Eighty-seven Rice University undergraduates participated for course credit in their introductory psychology course. The between-subjects independent variable was post-exposure task in which participants were assigned randomly to one of four conditions. Participants either: (1) marked a feature descriptor checklist (n=23), (2) generated adjective descriptors for the features (n=21), (3) performed an irrelevant letter-search task (n=22), or (4) tried to hold a mental image of the face in their minds (n=21).

Stimulus materials

The materials and apparatus included a slide projector and a total of 146 35mm face photographs (two slides for the six targets and 134 distractors). The slides contained frontal poses of Caucasian males who were approximately 20 years old. Slides were taken from a large pool of faces. Selection avoided pictures of faces with facial hair, and having distinctive clothing, hair styles, and facial gestures. Six targets faces were randomly selected from the resulting pool and were shown as color slides at study. All photographs in the recognition test were black-and-white, and consisted of the six targets and 134 distractors. The target depictions at study and test were similar but not identical. Target pictures were taken by two different cameras within a few minutes of each other.

The materials for the four post-exposure tasks included the following:

(1) For the *Descriptor Checklist* condition, six identical pages were provided in a booklet. Each page contained six feature headings: hair-hairline, eyebrows, eyes-eyelashes, nose, mouth-lips, and chin. Under each heading was 10 adjective descriptors taken from a facial feature descriptor dictionary compiled by Laughery (1977). The adjectives were collected from a previous face recall project (see Laughery & Fowler, 1980) and were among the most frequently used to describe the features.

(2) For the *Descriptor Generate* condition, the six pages of the booklet consisted of only the feature headings. Thus, it was otherwise identical to the Descriptor Checklist except it lacked the specific descriptors. Space was available under each heading for participants to write their own descriptors of the faces.

(3) For the *Irrelevant* condition, six pages filled with typewritten capitalized letters were provided. On each page two different letters were circled. The participant's task was to search for other instances of these two letters.

(4) For the *Image* condition, participants were instructed to hold a mental image of the previous face during the post-exposure interval. No post-exposure task materials were provided.

A recognition response sheet was provided at test. Target slides appeared in positions 52, 71, 79, 103, 120, and 135 of the 140 slide sequence. The target faces were sequenced randomly with respect to their appearance in the study series.

Procedure

Participants were told that they would be shown a sequence of six faces for five seconds each, and following each exposure there would be a pause of 60 seconds where they would be performing another activity. They were told that they would be tested on their memory of the faces later in the session. Further, they were informed that the quality of the photographs as well as the facial expression and pose might change from study to test so that it was important to study the persons' faces and not other aspects of the picture.

Each group was told how to fill out their post-exposure worksheets. Participants in the *Descriptor Checklist* condition were instructed to check all adjectives on the checklist that describe the most recently viewed face. Participants in the *Descriptor Generate* condition were instructed to list as many appropriate adjectives as they could under the feature headings to describe the most recently viewed face. Participants in the *Irrelevant* condition were told that a second purpose of the study was to measure fast they could do a visual scanning/perceptual speed task. They were told to put a slash through all other instances of two circled letters in a large matrix of letters starting from the top of the page. Different pairs of letters were searched on each page. Participants in the *Image* condition were instructed to try to hold an image of the most recently viewed face by "seeing" it in their mind during the post-exposure period.

After these instructions, the study slides were viewed for five s each with a 60 s post-exposure interval following each face presentation. They were to complete their post-exposure task within the one min interval.

Following the study sequence, all participants received a questionnaire asking them to report the kinds of study strategies they used and the features that they found most difficult to remember. Specifically, they were asked "What strategies did you use to help you remember the face?" Participants were asked to give percentages of time they used for each strategy so that they added to 100. The following strategies were listed: (1) looking at the individual features, (2) looking at relationships between features and groups of features, (3) imaging the entire face, and (4) naming or labeling features. These study strategies were the most frequently reported by participants in prior work (Laughery, Duval, & Wogalter, 1986). The questionnaire also asked how difficult it was to remember the features: hair-hairline, eyebrows, eyes-eyelashes, nose, mouth-lips, and chin. Participants responded on an 11-point scale anchored from (0) easiest to (10) most difficult. Five minutes was allowed for completion of the questionnaire.

Immediately following the questionnaire, the recognition test procedure began. Participants were told that in the test sequence: (1) the faces they viewed earlier may or may not appear, (2) they should not expect identical study-to-test pictures of the targets, (3) they should answer according to whether they saw the person before, not whether they had seen the particular picture before. Participants were also instructed how to mark the recognition response sheets. For every test face, they were told to put a "Y" (yes) or "N" (no) on the response sheet according to whether it was presented earlier and to indicate their degree of certainty using a three-point scale where 1 = guessed, 2 = probably correct, and 3 = certain. Test slides were presented at a six-second rate.

RESULTS

Recognition performance

Six measures of recognition performance were examined. Two measured target hits: the proportion hit (PH), and the hit-miss (HM) scores. Two measured false alarms: the proportion false alarm (PFA), and the false alarm-correction rejection (FACR) scores. The two proportion scores, PH and PFA, simply denote the level of yes's (scored as 1) and no's (scored as 0) to the targets and distractors, respectively. HM and FACR were derived by combining the yes-no responses with the confidence ratings onto a single six-point scale (N3 = 1, N2 = 2, N1 = 3, Y1 = 4, Y2 = 5, & Y3 = 6). Thus, HM and FACR reflect recognition confidence to the targets and distractors, respectively. Two discrimination (sensitivity) measures were also used: the difference between the HM and FACR scores (HM-FACR), and the mean z-score for the targets after standardizing each participant's responses to all test photographs (SHM).

TABLE 1

Recognition performance for the post-exposure conditions

	Descriptor Checklist	Descriptor Generate	Irrelevant	Image
Hits				
PH	.45	.64	.52	.53
HM	3.20	4.05	3.56	3.56
False Alarms				
PFA	.23	.27	.21	.15
FACR	2.30	2.50	2.29	1.92
Discrimination				
HM-FACR	.90	1.55	1.27	1.64
SHM	.55	.98	.95	1.29

Better recognition performance is reflected by high scores for the two hit and two discrimination scores, and low scores on the two false alarm measures. Table 1 shows these mean scores as a function of post-exposure condition.

A one-way between-subjects analysis of variance (ANOVA) on the PH scores was not significant, $F(3, 83) = 1.94$, $MSe = .072$, $p > .05$. The ANOVA on the HM scores was marginal but not significant, $F(3, 83) = 2.32$, $MSe = 1.129$, $p = .08$. However, a planned comparison between the two description conditions was significant, $t(42) = 2.36$, $p < .05$. Table 1 shows that participants in the Descriptor Generate condition had significantly greater HM scores than participants in the Descriptor Checklist condition.

Both ANOVAs on the false alarm scores showed reliable main effects, $F(3, 83) = 3.48$, $MSe = .016$, $p < .05$ for PFA, and $F(3, 83) = 3.41$, $MSe = .355$, $p < .05$ for FACR. Subsequent comparisons on both sets of means showed that Image participants produced significantly fewer false alarms than Descriptor Checklist and Descriptor Generate participants (*p*s < .05).

ANOVAs were also significant for the discrimination measures, $F(3, 83) = 2.96$, $MSe = .836$, $p < .05$ for HM-FACR, and $F(3, 83) = 5.11$, $MSe = .392$, $p < .01$ for SHM. Subsequent comparisons showed that Descriptor Checklist participants had significantly lower HM-FACR than Descriptor Generate and Image participants (*p*s < .05). For SHM, the significant comparisons were similar to FACR except Descriptor Checklist participants also showed significantly lower discrimination than the Irrelevant condition participants (*p*s < .05). The difference between the Image and Irrelevant participants was marginal but not significant ($p < .09$).

The effect of six targets was analyzed using a 4 X 6 mixed-model ANOVA. Using the HM, there was a large main effect of target, $F(5, 415) = 27.23$, $MSe = 2.14$, $p < .0001$. This is not unusual in face memory research and it simply means some faces are easier to remember than others. No interaction with post-exposure condition was found, $F < 1.0$. The same pattern was shown using the discrimination scores.

Questionnaire responses

Study strategies. A questionnaire was used to assess the kinds of study strategies that participants reported they used. The mean time percents using different strategies is shown in Table 2. A mixed-model ANOVA using strategy as the within-subjects variable and the post-exposure condition as the between-subjects variable showed a significant main effect of strategy, $F(3, 249) = 38.03$, $MSe = 343.50$, $p < .0001$. The ANOVA also showed a significant interaction of strategy and post-exposure condition, $F(9, 249) = 6.54$, $MSe = 343.50$, $p < .0001$. Post-hoc comparisons were made using Fisher's Least Significant Difference test (LSD = 11.1 at $p = .05$). Participants in the Descriptor Checklist condition reported the greatest time looking at individual features, followed by naming or labeling features, while the gestalt-like strategies of imaging the entire face and looking at relationships between features and feature groups were given relatively less time. Participants in the Descriptor Generate condition, like those in the Descriptor Checklist condition, spent the greatest time looking at individual features. They spent the least time looking at relationships between features and feature groups, while naming or labeling features and imaging the entire face were intermediate. This pattern is in contrast with the strategies reported by participants in the Irrelevant and Image conditions. Participants in these two conditions reported using most time to image the entire face, followed by looking at individual features, looking at relationships between features and feature groups, and naming or labeling features.

Feature ratings. The questionnaire also asked the difficulty of remembering six face features. Table 3 shows the mean ratings as a function of post-exposure condition. Higher values indicate greater difficulty. A mixed model ANOVA with face feature as the within-subjects variable and post-exposure condition as the between-subjects variable showed a significant main effect of feature, $F(5, 415) = 23.6$, $MSe = 5.20$, $p < .0001$, and of post-exposure condition, $F(3, 83) = 5.58$, $MSe = 6.93$, $p < .01$. The interaction was also significant, $F(15, 415) = 3.68$, $MSe = 5.20$, $p < .0001$. The pattern of means in Table 3 is

TABLE 2

Mean Percentage Time using Study Strategies as Reported by Participants

	Descriptor Checklist	Descriptor Generate	Irrelevant	Image
Individual features	45.4	44.1	32.0	32.1
Relating features	6.1	10.6	16.4	14.8
Imaging entire face	18.7	24.9	41.6	42.4
Naming or labeling	29.8	20.4	10.0	10.7

TABLE 3

Mean Rated Difficulty of Remembering Features

	Descriptor Checklist	Descriptor Generate	Irrelevant	Image
Hair-hairline	1.96	1.67	4.41	3.09
Eyebrow	3.70	3.43	6.50	6.05
Eyes-eyelashes	6.65	6.71	5.91	4.48
Nose	5.74	5.48	5.82	4.43
Mouth-lips	5.35	4.71	6.00	4.76
Chin	5.91	6.48	6.45	5.33

Note. *Higher scores indicate greater difficulty.*

complex, but it can be seen that participants generally reported the features in the lower portions of the face as more difficult to remember.

Quantity of descriptors. For the two verbal description conditions, the number of descriptors checked by Descriptor Checklist participants and generated by Descriptor Generate participants were counted. Table 4 shows the mean number of words for each target feature and description method. An ANOVA showed a significant main effect of post-exposure condition, $F(1, 42) = 8.44$, $MSe = 1.33$, $p < .01$. Descriptor Checklist participants checked more descriptors than Descriptor Generate participants produced. Face feature also showed a significant main effect, $F(5, 210) = 41.10$, $MSe = .151$, $p < .0001$. More words are generated or checked for the upper portions of the face than the lower portions of the face (LSD = .12 at $p = .05$). This pattern is similar to the difficulty ratings in Table 3. No reliable interaction was noted ($F < 1.0$). Other analyses explored the possibility of a relationship between recognition performance and the word counts. No significant correlations were noted ($ps > .05$).

DISCUSSION

The results show that recognition performance following the Descriptor Checklist task was reliably lower than the Descriptor Generate task. This result is somewhat surprising given the fact that participants only had to check off appropriate descriptors. Why was recognition degraded after the Descriptor Checklist? Two possibilities can be offered. First, lowered recognition might be due to the use of high-frequency descriptors. The adjectives were among the most frequently produced by subjects in an earlier set of studies (Laughery & Fowler, 1980; Laughery et al., 1986). Given that the adjective descriptors were selected on the basis of production frequency and not on the basis that they encompassed the potential dimensions of faces, the descriptors might have lacked the specificity needed to discriminate the target faces from the distractors. It is therefore possible that a different set of descriptors might not show recognition decrement.

TABLE 4

Mean Words Checked or Generated for Targets as a Function of Description Method and Face Feature

	Descriptor Checklist	Descriptor Generate
Hair-hairline	2.49	2.19
Eyebrows	2.04	1.59
Eyes-eyelashes	1.63	1.44
Nose	1.66	1.08
Mouth-lips	1.72	1.27
Chin	1.60	1.10

Second, poor performance by the Descriptor Checklist participants appears to be similar to the interference reported in other research by Loftus and Greene (1980) and Jenkins and Davies (1985). In these studies, misleading information presented after study lowered performance on a subsequent test. By its nature, the checklist provided extraneous descriptors; some adjectives were not descriptive of the particular face they had just viewed. By considering and processing these erroneous terms, participants possibly incorporated this information into their memories of the targets, and thus becoming confused about the features the targets possessed and lowering their ability to discriminate the targets from the distractors. However, the other description method, Descriptor Generate, did not interfere with recognition. This technique allowed participants to recall verbal descriptions without memorial confusion because irrelevant descriptors were not present and were less likely considered.

Strong support for the benefit of imaging faces on recognition was not found. While imaging led to fewer false alarms than either description condition and higher discrimination than the checklist condition, recognition was never reliably better than the irrelevant task condition. Only one marginally significant effect was noted (with SHM). If imaging improves face recognition, the effect is small.

Participants reported the lower features of the face the most difficult to remember. These results are consistent with a large body of research showing that the upper-most face features are the most salient (see Shepherd, Davies, & Ellis, 1981). In addition, the study strategies reported by participants were consistent with the kinds of conditions they were in. Participants in the verbal description conditions reported looking at individual features and naming the parts of the face, just as the demands of the description tasks would suggest. Participants in the image and irrelevant conditions reported relatively more imaging and study of the relationships between features. They also showed approximately the same performance on the recognition test. Apparently if not directed to use a specific strategy, as in the Irrelevant task condition, people are more likely to use a gestalt-like strategy.

Counts of the descriptors showed that participants checked off or generated an average of 11.1 and 8.7 descriptors, respectively. This is comparable to the quantity of descriptors emitted by subjects in other research. Ellis, Shepherd, and Davies (1980) report 9.4 descriptors, and Shepherd, Davies, and Ellis (1978) report 7.5 after a brief view followed immediately by recall, and 5.7 in describing colleagues from memory. These results also concur with other research indicating that people have difficulty describing faces (Ellis et al., 1980). One possible reason is that the English language lacks the vocabulary to communicate many characteristics of faces (Laughery et al., 1986).

This research has practical implications for test and evaluation of witness memory by law enforcement. Deffenbacher and Horney (1981) suggest that when gaining facial information from a witness, free recall description should be used rather than (or before) cued description because spontaneously mentioned details are more accurate, and that free recall would be less likely to contaminate memory. Moreover, Neisser (1987) has argued that free recall is less likely to produce distorted memorial accounts than cued recall and recognition because the latter two tests constrain or limit what can be reported. The present research supports both of these notions. When capturing face descriptions, free recall methods are preferred over methods that rely on descriptor recognition like checklists because descriptor generation does not degrade performance on a subsequent recognition test, and it produces the best quality descriptions (Goulding, 1971).

REFERENCES

Chance, J. E., & Goldstein, A. G.. (1976). Recognition of faces and verbal labels. *Bulletin of the Psychonomic Society, 7*, 384-386.

Deffenbacher, K., & Horney, J. (1981). Psycho-legal aspects of face recognition. In G.M. Davies, H.D. Ellis, & J.W. Shepherd (Eds.), *Perceiving and remembering faces* (pp. 221-226). London: Academic Press.

Dent, H. R., & Stephenson, G. M. (1979). An experimental study of the effectiveness of different techniques of questioning child witnesses. *British Journal of Social and Clinical Psychology, 18*, 41-51.

Ellis, G. D., Shepherd, J. W., & Davies, G. M. (1980). The deterioration of verbal descriptions of faces over different delay intervals. *Journal of Police Science and Administration, 8*, 101-106.

Goulding, G. J. (1971). Facial description ability. *Police Research Bulletin, 19*, 42-44.

Graefe, T. M., & Watkins, M. J. (1980). Picture rehearsal: An effect of selectively attending to pictures no longer in view. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 156-162.

Hall, D. F. (1977). *Obtaining eyewitness identifications in criminal identifications: Two experiments and some comments on the Zeitgeist in forensic psychology.* Paper presented at the meeting of the American Psychology-Law Society.

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

Laughery, K. R. (1977). *Adjective descriptors used in generating sketches and Identi-kit composites (Report No. UHMUG-8).* Unpublished manuscript, University of Houston.

Laughery, K. R., Duval, G., & Wogalter, M. S. (1986). Dynamics of face recall. In H. D. Ellis, M. A. Jeeves, F. Newcombe, & A. W. Young (Eds.), *Aspects of face processing* (pp. 373-387). Dordrect, Netherlands: Martinus Nijhoff.

Laughery, K. R., & Fowler, R. F. (1980). Sketch artist and Identi-kit procedures for recalling faces. *Journal of Applied Psychology, 65*, 307-316.

Loftus, E. F., & Greene, E. (1980). Warning: Even memory for faces may be contagious. *Law and Human Behavior, 4*, 323--334.

Mauldin, M. A., & Laughery, K. R. (1981). Composite production effects on subsequent facial recognition. *Journal of Applied Psychology, 66*, 351-357.

Neisser, U. (1988). Time present and time past. In M. M. Gruneberg, P.E. Morris, and R.N. Sykes (Eds.), *Practical Aspects of Memory: Current Research and Issues*, Vol. 2 (pp. 545-560). Chichester: Wiley.

Read, J. D. (1979). Rehearsal and recognition of human faces. *American Journal of Psychology, 92*, 71-85.

Read, J. D., Hammersley, R., Cross-Calvert, S., & McFadzen, E. (1989). Rehearsal of faces and details in action events. *Applied Cognitive Psychology, 3*, 295-311.

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology, 22*, 36-71.

Shepherd, J. W., Davies, G. M., & Ellis, H. D. (1978). How best shall a face be described? In M.M. Gruneberg, P.E. Morris, & R.N. Sykes (Eds.), *Practical Aspects of Memory.* London: Academic Press.

Shepherd, J. W., Davies, G. M, & Ellis, H. D. (1981). Studies of cue saliency. In G. M. Davies, H. D. Ellis, & Shepherd, J. W. (Eds). *Perceiving and Remembering Faces.* London: Academic Press.

Sporer, S. L. (1988). Long-term improvement of facial recognition through visual rehearsal. In M. M. Gruneberg, P. E. Morris, and R. N. Sykes (Eds.), *Practical Aspects of Memory: Current Research and Issues*, Vol. 1 (pp. 182-188). Chichester: Wiley.

Williams, L. (1975). *Application of signal detection parameters in a test of eyewitnesses to a crime.* Center for Responsive Psychology, Report number CR-20, Brooklyn, NY: Brooklyn College, CUNY.

Wogalter, M. S., Cayard, J. A., & Jarrard, S. (in press). Massed vs. distributed exposure and imaging of faces: Changing the test view. *Current Psychological Research and Reviews.*

Wogalter, M. S., Laughery, K. R., & Thompson, B. G. (1989). *Eyewitness identification: Effects of composite construction on subsequent recognition performance.* Unpublished manuscript, Rensselaer Polytechnic Institute, Troy, New York.