12-1976

Memory for persons, encounters and sex

William Sturgill
University of Nebraska at Omaha

MEMORY FOR PERSONS, ENCOUNTERS AND SEX

A Thesis

Presented to the

Department of Psychology

and the

Faculty of the Graduate College

University of Nebraska

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

University of Nebraska at Omaha

by

William Sturgill

December, 1976

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

Thesis Acceptance

Accepted for the faculty of the Graduate College, University of Nebraska, in partial fulfillment of the requirements for the degree Master of Arts, University of Nebraska at Omaha.

Thesis Committee Name	Department
Barbara Manning	Psychology
Kenneth Offenbacher	Psychology
John Loken	Physics

Evan Brown, Psych.
Chairman

Sept. 3, 1976
Date

Memory for Persons, Encounters and Sex

William Sturgill

A thesis

Running Head: Memory for Persons

Abstract

Recognition of previously seen persons and recall of the circumstances of their encounter were tested in a situation where subjects were unaware of the subsequent recognition task. Subjects encountered four persons, one of each sex in two separate encounters. Prior to a lineup one week later, only 51 subjects (N = 145) failed to recall either the number and/or the sex of the persons encountered, while only 28 correctly recalled both the number and sex and that it was two different persons in each encounter. Results from the lineup confirmed previous suggestions that subjects (N = 155) are better able to recognize persons than recall where they encountered them. The best recognition performance came from male subjects' recognition of female criminals, a finding that contradicts previous research. The best recall performance came from female subjects' recall of where they encountered male criminals. Prompted by considerable variation in the indictment rates and the recognizability of the individual suspects, the issues of representative sampling of stimuli and generalization are discussed as potential problems in facial recognition studies. Consideration was also given to the manner in which recall of the circumstances of encounter is typically calculated.

Memory for Persons, Encounters and Sex

It would seem that visual recognition memory of a previously seen face and the ability to recall accurately the circumstances of the previous encounter are both necessary for correct indictment of a criminal suspect. Recent research, however, suggests that subjects' abilities in these two memorial operations differ substantially. Whether these results may be generalized to identification of criminal suspects in real life is another question. Extant literature shows recognition of pictures of scenes and faces to be quite good, at least when tested under favorable conditions. However, when it has been tested, recall of the circumstances of encounter either has not been clearly separated from recognition or has been tested in designs that make it difficult to draw unequivocal generalizations to criminal indictment situations.

Shepard (1967) studied recognition, using self-paced presentations in a directed memory task (i.e., with subjects having knowledge of the subsequent recognition task). He had his subjects view 612 pictures of assorted scenes and objects which were chosen to be high in memorability and low in similarity. Subjects were then immediately tested for recognition with 68 "old-new" (previously seen-not previously seen) pairs. The task turned out to be quite

easy. Median correct recognition of the old member of the pair was 98.5%. It remained above 90% even for subjects tested after a one week delay. Standing (1973) made the recognition task more difficult by limiting viewing time to 5 sec per item with an interstimulus interval of about one-half sec. Under these less favorable conditions, he presented his subjects with 10,000 pictures over a four-day period, testing with 160 old-new pairs. Recognition accuracy was about 90%.

It is difficult to compare studies using pictures of scenes with those using pictures of faces because of differences in procedure in addition to differences in materials. Even so, recognition of faces generally seems to be rather respectable, if somewhat below that for scenes. Thus Hochberg and Galper (1967), using Shepard's procedure, presented subjects with 60 pictures of female college student faces, testing immediate recognition with 15 old-new pairs. Median accuracy was 90%. Yin (1969) limited viewing time to 3 sec per face and presented his subjects with as many as 64 slides of adult male faces. Immediate recognition accuracy on 24 old-new pairs was about 96%.

While the above studies generally involved subjects instructed to regard picture recognition as their main task, Bower and Karlin (1974) manipulated task instructions,

finding strong differences in resulting accuracy. While half their subjects were attempting to memorize the faces, the other twelve thought they were in a reaction-time experiment. The subjects studied 72 photographs of white college student faces (36 of each sex) for 5 sec each, making, for each, one of three binary judgments; either sex, likableness or honesty. Those subjects tested without knowledge of the subsequent recongition task (incidental memory) ranged in accuracy from 60% for faces whose sex had been identified to 81% for faces for whom honesty had been judged. Subjects in the directed memory task produced accuracy rates ranging from 56% for faces identified by sex to 80% for faces judged to be more likable than average. However knowledge affects judgments of likableness and honesty, these results suggest that it is not the intention to learn but the depth at which the input is processed that produces good recognition memory. That is, the depth at which the stimulus input is analyzed directly determines the probability of recognition. Judgments as to whether a face is more likable or honest than average require the input to be analyzed at a depth greater than that required for judgments of sex.

Though Bower and Karlin (1974) used sex of face as a variable in their study, they did not investigate memory for faces as a function of sex of subject. Though studies

doing so are not numerous and have many variations in procedure, there does seem to be a fairly consistent underlying pattern of results. Cross, Cross, and Daly (1971) had subjects study a matrix of faces, two faces from each of 12 age-sex-race categories, in an incidental learning task. Effort was made to simulate the real world experience of seeing many faces before encountering the need to recognize a previously seen face by giving the subjects an intervening task of judging 96 faces on their relative beauty. The subjects were then shown a second matrix containing 12 faces from the first matrix plus 12 new faces matched on the subject variables. Results showed that although overall recognition accuracy was only about 40%, the female subjects performed more accurately than the male subjects, and female faces were more often recognized. Closer inspection revealed that women recognized female faces (43%) more often than men (37%), but male faces (33%) less often than men (36%). The male subjects, on the other hand, recognized faces of both sexes with about equal facility.

While the small sample of faces of each type limits generalizations concerning facial variables, Cross et al. have received varied support for their recognition results. Several years earlier, Howells (1938) observed that women may be superior to men at face recognition. His subjects

studied a card containing three different pictures of the same face for 10 sec, then immediately attempted to find another card of three different pictures of the same face among a display of such cards. Though he showed faces of each sex and tested samples from several populations (c.g., college students, sales people and farmers), he apparently did not analyze for differences in recognition of each type of face. Across all faces, women tended to be more accurate than men. The finding reported by Cross et al. (1971) that is of particular interest, however, is the nature of the sex of subject by sex of photo interaction. Witryol and Kaess (1957) while reporting an overall female superiority also report a tendency for males to remember pictures of males better than pictures of females, and similarly for females to remember better pictures of their own sex. Ellis, Shepherd, and Bruce (1973) found that the overall superiority of the girls (12 and 17 years) was mainly due to their significantly better scores on recognizing pictures of females; the girls did not, however, do significantly better than the boys at recognizing faces of males. This latter finding of no sex difference in recognizing male faces has been recently reported also by Going and Read (1974). They showed their subjects 56 slides of college student faces (28 of each sex). The faces in the slides, which had previously been rated by other subjects for their uniqueness

on a 7-point scale of uniqueness, were chosen from the extreme groups of the uniqueness scale. That is, half had been rated high on the uniqueness scale while the other 28 (14 of each sex) had been rated as not being very unique faces. Results showed that regardless of the level of uniqueness of the stimumus face, women were overall more accurate in the recognition task and that female faces were the most often recognized. While the women recognizee female faces more often than male faces, they did not perform better on the male faces than did the men. Similarly, Shroder (Note 1), whose subjects studied 80 slides of faces that differed on race (Black; White) and sex (male; female), found her women subjects (her only Black subjects were Malawian males) to make the fewest recognition errors across all categories of faces. While these women did better on faces of their own race and sex, the White males did at least as well as the women on White male faces. This the pattern emerges that subjects recognize faces of their own sex better than those of the opposite sex with women performing better overall.

Thus, under a variety of experimental manipulations, recognition memory for previously seen pictures of faces appears to be quite robust. Recall of the circumstances of encounter, on the other hand, has not been shown to be comparable. Standing, Conezio, and Haber (1970) point this

out in one of their experiments. They first showed subjects pictures of scenes differing in orientation and then, after varying retention intervals, showed old-new pairs of pictures with instructions both to identify the old-pair-member as well as its previous orientation. While their subjects did reasonably well if tested soon after presentation, the ability to recall accurately the circumstances of encounter (previous orientation) declined markedly with longer delays. Recognition accuracy, however, showed no comparable decline. Comparable results have been recently demonstarted using pictures of faces. In a directed memory task, Brown, Deffenbacher, and Sturgill (in press, Experiment 1) had their subjects study 25 pictures of children's faces for 20 sec each in a particular room, then, two hours later, 25 more such pictures in a very dissimilar room. Two days later the subjects returned to a third room with instructions a) to select from the (old-new) pair the picture that was previously seen, and b) to place that picture in the pile corresponding to the room in which it was first encountered. Recognition accuracy was high, as one would expect, about 96%. Though at 58% and with scores that ranged from 0.44 to 0.68, recall of the circumstances of encounter (the room of encounter) was above chance--in the statistical sense, it was hardly impressive. Only five of

the 14 subjects were able to recall the circumstances of encounter in a statistically reliable fashion. In order to obtain an unbiased expression of subjects' performance and to facilitate interexperimental comparison, signal detection analyses were performed, yielding *d*'s of 2.48 and 0.40 for recognition and recall, respectively.

Thus in a standard laboratory task, only 36% of the subjects in Brown, Deffenbacher, and Sturgill (in press, Experiment 1) were able to recall the circumstances of encounter at above a chance rate. Similar performances would hardly be satisfactory were they to hold true in real-life identifications of criminal suspects. Perhaps witness performances in real-life is like that of these subjects, able to recognize having seen a face but with little memory of where that face was encountered. Brown et al. conducted two additional studies which simulate more closely than laboratory tasks actual encounters witnesses might be expected to have with criminals. Since the present study stems from some important questions raised by these two studies, they will be dealt with in some detail.

In Experiment 2 (Brown et al., in press) subjects were first presented with two groups of five "criminals" (paid vulunteers, all male) for 25 sec each in a directed memory task. About an hour later, 15 front-side view mugshots were presented with instructions to identify the criminals--

if any--who earlier had appeared at the front of the class. The mugshots were of five criminals, five suspects who would later be seen in a lineup and five fillers. In this more realistic situation where subjects were attempting to identify persons previously seen live from mugshots, their performance was much less accurate (d' = 0.71) than in Experiment 1 where the recognition was of the same picture presented twice.

The results of greatest forensic interest are those obtained one week later at the lineup phase of the experiment. As a result of the ways in which the suspects were arranged in the various phases of the experiment, each of the four lineups staged consisted of suspects from the following four conditions: Suspects who had been seen by the subjects both as criminals and in the mugshots (CMS); suspects who had been seen as criminals but who had not been seen in the mugshots (CNMS); suspects who had been seen in the mugshots only (MS); and suspects who were being seen for the first time, in the lineup only (LO). Results showed that indictment rates for the CMS, CNMS, MS and LO groups were 0.65, 0.51, 0.20, and 0.08, respectively ($ps < 0.0003$), with a mean indictment rate across all groups of 0.36. Comparisons of each criminal group with the noncriminal ones resulted in larger d's than the one

obtained in the mugshot phase, indicating that criminals were easier to recognize when they reappeared live than when they reappeared in mugshots.

The indictment rates at the lineup clearly show that a suspect was more likely to be indicted if he were a criminal, particularly if his mugshot also had been seen, than if he were an innocent. While this may seem to indicate an ability to recall the circumstances of encounter, note that an MS suspect stood a not terribly remote one-in five probability of being indicted and a CNMS suspect has as much as a one-in-two probability of escaping indictment. The main indication of an ability to recall the circumstances of encounter was evidenced by a doubling of the MS indictment rate in response to a second question asked at the lineup, namely, whether or not the suspect's mugshot had been seen (while this question was asked in both studies 2 and 3, the results were not reported because of space limitations). To this question the indictment rates for the CMS, CNMS, MS and LO groups were 0.59, 0.47, 0.40 and 0.19 respectively, with a mean rate of 0.41. Unfortunately, the increase in the MS indictment rate was accompanied by an increase in the LO and a decrease in the CMS indictment rates. The strongest evidence against an ability to recall the circumstances of encounter was that regardless of question asked, suspects were indicted in the same order of

decreasing magnitude. Suspects who had been seen both as a criminal and in the mugshots were indicted most often; suspects seen live but not in the mugshots were indicted next most often; third most frequently indicted were suspects previously seen only in the mugshots; and indicted least often were innocents making their first appearance. Indictments, then, seem to have been made primarily on recognition, where increases occured as a function of number of previous exposures, particularly live exposures.

Whereas Experiment 2 (Brown et al., in press) simulated a situation where witnesses were aware that a crime was occurring and attempted to memorize the criminals' faces, Experiment 3 simulated instead a situation in which witnesses were unaware that a crime was taking place and had, therefore, no obvious motivation to memorize the criminals' faces. In addition, there were only four criminals, two of whom were encountered only by half of the subjects and the other pair of criminals were encountered only by the other half of the subjects. Two or three days later mugshots of 10 fillers and two criminals were shown. Half the subjects saw mugshots of one criminal from each pair and the other two criminals were seen by the other half of the subjects. Thus of the two criminals encountered by any subject at the scene of the crime, only one appeared in the mugshots. One

week from the original encounter a lineup was staged consisting of the four criminals (all male), each of whom had been seen by approximately one-quarter of the subjects in one of the four encounter conditions--CMS, CNMS, MS and LO. In general the indictment rates were much lower in this nondirected memory experiment than in the previous directed one. The mean correct indictment rate of criminals in the mugshot phase was 0.28 while the mean false indictment rate of innocents was 0.15 (d' = 0.46, $p < 0.0001$, from chance, i.e., a d' of zero).

Indictment rates at the lineup for the CMS, CNMS, MS and LO conditions were 0.45, 0.24, 0.29 and 0.18 respectively, with a mean of 0.29. In contrast to Experiment 2, a criminal was more likely to be indicted than an innocent only if his mugshot had also been seen. Even then he (CMS) was indicted only 2.5 times as often as someone not previously encountered at all. A suspect who had been seen only at the scene of the crime (CNMS) was as likely to be indicted as a suspect in either of the noncriminal conditions. Results (unpublished) from the question regarding whose mugshot had been shown revealed that the mugshots considerably influenced the subjects' memories. Indictment rates for the CMS, CNMS, MS and LO conditions were 0.89, 0.63, 0.88 and 0.50 respectively. Again, an inability to recall accurately the

the circumstances of encounter is indicated. Regardless of whether the person present or mugshot present question was asked, indictments followed the same order of decreasing magnitude; CMS suspects were indicted most, MS next most, CNMS next most and LO the least. As in Experiment 2, it was a shift in the MS indictment rate that gave evidence of an ability to recall the circumstances of encounter. In this experiment, however, the shift was greater, probably due to the strong influence of the mugshots. The indication of a recall ability was best seen when the CMS condition was compared to the MS in the signal detection analyses. The d' between the two conditions involving mugshots when the question was changed is as would be expected if the circumstances of encounter were being recalled properly. However that may be, it still appeared that indictments were made primarily on the basis of recognition with increases being a function of number of previous exposures, particularly, in this case, mugshot exposures.

Thus, while in both experiments (Brown, Deffenbacher & Sturgill, in press, Experiments 2 and 3) indictments appeared to be based primarily on recognition with little memory for the circumstances of encounter. Which encounter (live or mugshots) most influenced recognition changed as a function of the type of crime being simulated, whether it

was a crime the witnesses knew was occurring or whether instead it was one in which the witnesses were ignorant of its occurrence. A possible interpretation of this change in influence from live (Experiment 2) to mugshot (Experiment 3) encounters might lie in the depth of processing hypothesis (Bower & Karlin, 1974), with live encounters being the more deeply processed in situations where the witnesses knew a crime was occurring, and mugshots when the witnesses did not have this knowledge. However depth of processing may effect recognition, its effect on recall of the circumstances of encounter in Experiments 2 and 3 is difficult to determine. Indeed, recall in the two experiments must be inferred indirectly from shifts in response rates resulting from changes in questions asked, from inquiries regarding live encounters to ones regarding mugshot encounters.

The present study sought to separate these two equally important--at least in the present context--memorial operations, i.e., recognition of faces previously seen and recall of the circumstances of their encounter. Additionally, it sought to test the generalizability of previous studies by using live stimulus persons of both sexes. Recognition of faces was tested in much the same manner as in Brown et al., Experiment 3. The subjects' task was to select persons they

had previously encountered from a lineup. The principle recall task was one based on previous (Brown et al., Experiment 1; Standing, Conezio, & Haber, 1970) methods of testing and measuring recall of the circumstances of encounter. The subjects were asked to indicate in which of two circumstances they had encountered the suspects they recognized. Two additional recall tasks were also incorporated into the experiment. One, given prior to the lineup, tested subjects' memory for the number and sex of the persons they had encountered at the scene of the crime. The other asked the specific activity of each of the two criminals. Thus the present experiment sought to investigate not only recognition of faces but recall of their encounter as a function of sex of subject and sex of stimulus person.

Method

Subjects were members of a large introductory psychology class. Though 237 class members made the initial encounter, only the responses from 155 (44 males and 111 females) class members were usable as 61 class members did not return for the lineup phase and 21 of those who did had incomplete protocols or otherwise failed to follow instructions, or claimed to know one or more of the criminals. One-hundred-fifty class members responded to the first questionnaire (concerning the number and sex of the persons encountered)

of which four protocols were incomplete and one was unscorable, leaving a total of 145 (107 female and 38 male) subjects. Initially the class members were unaware of their participation in a research project. However, prior to the collection of recognition and recall data, the class members were informed of what had occurred and were given the option to decline further participation.

Eight people agreed to act as "criminals": five graduate students (three women and two men), two male upper class undergraduates and a former student (female) presently employed outside the university system. Insofar as possible, these persons were selected so as to warrant their inclusion in the same lineup, at least within each sex, and so as to minimize the likelihood of previous contacts with the class members. Brown, Deffenbacher, and Sturgill (Experiments 2 and 3) successfully met these selection criteria, as evidenced by reasonably stable indictment rates within a condition, even within a fairly wide range of variables such as, e.g., hair color, skin pigmentation, build, etc.

As the class members entered the examination room to take their second midterm, they were given their test materials by the criminals. The examination room had two entrance corridors, and two criminals (one man and one woman) were stationed in each corridor; one handed out test questions and

the other IBM answer sheets. Owing to the narrowness of the entrance corridors (1.14m), the criminals were encountered in succession. Thus, sex of criminal and type of test material (test questions or answer sheets) first encountered by each subject were counterbalanced across entrance corridors. Class members entering a given corridor could not see the criminals in the other corridor. Attempting to get roughly half the class members encountering each set of criminals, the course instructor had given instructions to the class members to enter and exit a given corridor, based on their surnames. The IBM answer sheets were unobtrusively coded so that it could be determined which pair of criminals actually and been encountered by each class member.

A second criminal-witness encounter occurred after the class members had completed their examination and were leaving the room. During the examination, a second set of criminals (again, one man and one woman) unobtrusively replaced the first pair in each corridor. The exact location of the seocnd pair of criminals within each corridor was slightly different from that of the first pair of criminals. Also, the second pair was positioned behind a table whereas the first pair was not. These changes produced slight perceptual differences in the two encounters, such as backdrop, angle of regard and the table. The class members/witnesses left their

completed tests with this second set of criminals who coded the IBM answer sheets with respect to the corridor at which they were located. Thus it was possible to determine for each subject the set of criminals encountered both on the way into and out of the examination room.

One week later at the next scheduled meeting of the whole class, the class members were asked to respond to two questionnaires. The first, given prior to the lineup, asked the class members to indicate how many people they had encountered on their way into the classroom the preceding week, handing out test materials, and how many people they encountered on the way out, collecting test materials. In addition to asking the class members to recall the number of criminals in each encounter, they were also asked to indicate the sex of each criminal, whether it was the same people across both encounters, and, if not, to indicate the nature of the change.

A second questionnaire was distributed to the class members after the first one was collected, then a lineup was staged. The lineup consisted of all eight persons acting as criminals. However, for any one class member only four were actually criminals; the other four were distractors. Thus a person in the lineup was responded to both as a criminal and an innocent: Each suspect was his/her own control. The arrangement of the suspects within the lineup was balanced

on sex of suspect, encounter (in or out) and corridor (east or west). Each suspect in the lineup wore a large identification letter (A-H) around his/her neck and different clothing from that worn at the scene of the crime. The second questionnaire asked more specific questions about the people the class members encountered. The class members were told that they had encountered four people the preceding week, two handing out test materials and two collecting them, and that these four people were present in the lineup. Their task, should they choose to accept it, was to select from these eight people the four they had encountered previously and the four they had not seen. Then from the four selected as having been seen, they were to determine which pair handed them their test materials as they entered the room and which pair collected the tests as they were leaving. Finally, from the two selected as the ones encountered on the way into the room, which one was handing out test questions and which one the IBM answer sheets. On the way out, there was no distinction made as to which criminal collected test questions and which collected IBM answer sheets. Regardless of question asked in this second questionnaire for each response the class members were asked to indicate their confidence in that response by marking one of the following three choices: I am quite sure; I am moderately sure; I am not sure.

Results

Responses to the first questionnaire were grouped into three main categories (N = 145). Only 28 subjects (19%) were completely accurate (category 1). An additional 66 subjects (46%) were correct as to the number and sex of persons encountered but reported that they were the same two people in each encounter (category 2). The remaining 51 subjects (35%) incorrectly reported the number, the sex, or both (category 3). The proportion of subjects responding in each of these categories was independent of sex of subjects ($\chi^2(2) = 0.107$ ns).

The 51 subjects in category 3 were tabulated further according to the nature of their errors. Thirteen subjects reported the correct number of persons for each encounter, but were incorrect as to sex. Of these, only one reported that all persons encountered were of the same sex. This was the only subject responding (out of 145) who failed to report that the persons encountered were a mixed sex group. Among the 38 subjects reporting the wrong number of persons encountered, 29 reported too many, 24 claiming that there were five and the others that there had been six. The eight subjects reporting too few persons claimed that three had been encountered. Across category 3, 33 subjects reported that there were more men than women, 15 that there

were more women than men, and only 3 that the numbers of each sex were equal.

Two measures of recognition accuracy were obtained from the questionnaire given in conjunction with the lineup; the proportion of correct indictments (hits) and the proportion of incorrect indictments (false alarms). The hit rate, summed across all 155 subjects, was 0.54, while the false indictment rate was 0.44. These indictment rates correspond to an overall *d'* for recognition of 0.30, which by Marascuilo's (1970) one-signal significance test (derived from Gourevitch & Galanter, 1967) is significantly ($p < 0.0001$) greater than a *d'* of zero (chance discrimination). The rather small probability of such a small *d'* being attributable to random responding results from the large number of subjects encountering four criminals. The proportion of correct indictments were then grouped by encounter to determine whether subjects recognized better the persons they encountered when entering the room or when exiting the room. Results showed that across all subjects the indictment rate of the criminal pair handing out test materials (0.55) did not differ significantly from that of the criminal pair collecting test materials (0.56), using z-tests for correlated proportions ($z = 0.40$, ns). Similarly, when separated by sex of subject, there was no significant

(*zs* = 0.63 and 0.66, for males and females, respectively, *ps* > 0.05) tendency differentially to indict the criminals of one encounter.

Responses were pooled within sex of suspect as a function of sex of subject and *d*'s were calculated. As can be seen in Table 1, male subjects' recognition (*d*' = 0.46) of criminals was considerably better than recognition by females (*d*' = 0.20). In terms of a corresponding main effect for sex of criminal, recognition of female faces (*d*' = 0.38)

Insert Table 1 about here

---

was better than recognition of male faces (*d*' = 0.20). It can further be seen that the direction of both main effects can be accounted for by the male subjects' accuracy when indicting female criminals (*d*' = 0.69). Male subjects' recognition of female criminals was significantly better than female subjects' recognition of male criminals, using Gourevitch and Galanter's (1967) test for significance between two *d*'s (*G* = 2.25, *p* < 0.05), and was marginally significant (0.05 < *p* < 0.10) both over recognition of males by males (*d*' = 0.22) and recognition of females by females (*d*' = 0.28). Though female subjects recognized faces of their own sex better, this difference was not significant.

Responses were pooled across all subjects, and hit and false alarm rates and d's were calculated for each suspect in the lineup. As can be seen from Table 2, there was considerable variation in the indictments and d's of the suspects regardless of sex of subject. The mean of the

---

Insert Table 2 about here

---

proportion correct indictments as a function of suspects (Sum) was 0.54 with a SD of 0.18 and a range of 0.48. The variation among the false indictments was about the same: The mean of the proportion incorrect indictments as a function of suspects (Sum) was 0.46 with a SD of 0.14 and a range of 0.45. The mean of the d's by suspect was 0.30 while the SD was 0.47 and the range 1.55.

Three dependent t-tests were performed to determine whether subjects' confidence in their responses differed when making correct or incorrect responses, when responding to males or females and when responding to criminals or innocents. While all three t-tests failed to reach significance (ps > 0.20), the correlation of confidence within all three pairings was significant (ps < 0.001). The correlation of confidence in responses between the sexes was 0.72, between criminals and innocents 0.68 and

between correct and incorrect responses 0.67. On the other hand, there was no relationship between accuracy (correct indictments plus correct rejections) and mean confidence summed across all suspects (r = 0.04, ns).

To obtain a measure of recall accuracy, only responses to correctly indicted criminals were considered, following Brown et al. (in press) and Standing, Conezio, and Haber (1970). In order to use detection theory analysis, which Lockhard and Murdock (1970) consider feasible in a cued recall situation, the correct recognitions were converted into hits and false alarms for recall by taking a hit to be the correct association of criminal with encounter, and a false alarm to be the incorrect association of criminal with encounter. The hit rate, pooled across all subjects and criminals, was 0.54 and the false indictment rate was 0.46, corresponding to a recall d' of 0.20 ($p < 0.05$). Further analysis showed this ability to recall the circumstances of encounter not to be dependent upon the encounter. A z-test for correlated proportions indicated no differential tendency for suspects in one encounter to be recalled more than suspects in the other, whether handing out or collecting test materials. Thus on the average, the difference between correctly and incorrectly associating criminal with encounter was greater than would be expected if subjects were simply guessing.

Recall responses were pooled within sex of suspect as a function of sex of subject and d's were calculated. (See Table 3). Female subjects remembered where they encountered criminals (d' = 0.30) better than did male subjects (d' = 0.05), though this difference was not significant (G = 1.20, p > 0.25). On the other hand, encounters with male criminals (d' = 0.40)

Insert Table 3 about here

---

were better recalled than encounters with female criminals (d' = 0.05). This difference is marginally significant (G = 1.82, p < 0.07) by Gourevitch and Galanter's (1967) test. The d's comprising the cells of Table 3 were calculated by considering only the hits and false alarms on suspects with sex in common, i.e., hits on male criminals were placed against false alarms on male innocents and similarly for females. The accuracy of the female subjects (d' = 0.56) at correctly associating the male criminals with their respective encounters accounts for the direction of both main effects. The females recalled where they encountered male criminals significantly better than they recalled where they encountered female criminals (G = 1.99, p < 0.05) and significantly better than males recalled where female criminals were encountered (G = 2.10, p < 0.05). All other meaningful comparisons of recall d's were not significant.

Finally, recall responses were pooled across all subjects, and the proportion of correct recall was determined for each suspect in the lineup. The results of this analysis can be found in Table 4. While the variance of each distribution differs, the mean proportion of correctly indicted suspects who were also associated with the right encounter was 0.52 regardless of whether summing within or across sex of subject. The SD for the male subjects was

---

Insert Table 4 about here

0.14 and the range 0.42; for the female subjects the SD was 0.13 and the range 0.37; and summed across sex of subject the SD was 0.10 and the range 0.29. As with recognition, the accuracy with which the circumstances of encounter were recalled was independent of the subjects' confidence in their responses (r = 0.09, ns).

Subjects accuracy at recalling the activities of each criminal handing out test materials, whether he/she was the person distributing test questions or answer sheets, seemed to be improved if the criminals' circumstance of encounter had been correctly recalled. When subject performance was determined only from criminals who were correctly recognized as having been encountered, the probability was

0.34 that the criminals' activities at the encounter would be correctly recalled. On the other hand, when subject performance was determined from criminals who already had been correctly recalled as having indeed been among those distributing test materials, the probability was 0.67 that the test material the criminal distributed would be correctly recalled. These recall accuracies were dependent neither upon the particular activity of the criminal, i.e., whether the criminal was a test question or an answer sheet distributor (z = 1.07, ns), nor upon the subject's expressed confidence in their responses (r = 0.08, ns).

Discussion

Results of the present experiment support suggestions from previous research that recognition of a previously seen face is more accurate than recall of the circumstances of encounter. As compared with recognition accuracy rates from experiments using pictures of faces, the correct indictment rate in the present experiment is lowered in part for procedural reasons, namely, using a nondirected memory design as opposed to ones which assure the stimulus face is at least seen, together with any effects of greater retention interval. The recognition performance of these subjects was much like those subjects of Bower and Karlin's (1974) who identified faces by their sex, obtaining accuracy rates of

56% and 60% for directed and incidental tasks, respectively. With only one out of 145 subjects not reporting encounters with both sexes, subjects in the present experiment clearly made a note of the criminal's sex. Though procedures differed widely, perhaps Bower and Karlin's subjects (who identified faces according to the sex of the face) and the subjects in this experiment processed faces at a similar depth. This depth enabled them to identify previously encountered faces at greater than a chance rate but not with the accuracy of subjects directed to attend to the faces.

The recognition performance of the subjects in the present experiment also generally support the recognition results reported by Brown et al. (Experiment 3). Simulating the same nondirected memory situation, both studies obtained overall recognition *d*'s of 0.30. However, with all the variations between the two studies, it is debatable how much to make of such agreement. Where Brown et al. used four suspects, all male, the present experiment had eight suspects, four of each sex. In addition, there were obvious differences in suspect presentation. The Brown et al. suspects appeared once in each of four conditions (as a criminal whose mugshot was shown two days later to some subjects and three days later to some others; as a criminal without a mugshot and thus only seen once prior to the lineup; as an innocent seen only in mugshot form either four days for some subjects or

five days for others before the lineup; and as an innocent being seen for the first time), while the present experiment had only two conditions, criminals and innocents. Thus whether the effects of the different conditions and suspect presentations between the two experiments simply averaged out or whether recognition performance of subjects in nondirected memory is reliably described by something like a *d'* of 0.30 is difficult to determine. In both cases it is fairly clear that subjects typically gain little memory strength from encountering strangers distributing or collecting test materials, and for whom there is otherwise, except perhaps for the persons' sex, little motivation to remember the persons' face. In both cases, there was no relationship between accuracy and confidence of their judgment.

Recognition performance as a function of sex of subject was not in total agreement with previous experiments. Again, however, there is considerable difference in stimulus materials. The literature, with its pictures of faces, generally finds recognition best where the sex of the subject and of the face in the picture are the same, with women doing the better job across all subjects. Indeed, Brown et al. (Exepriment 3) support this pattern somewhat by finding their female subjects to have been the most

accurate. But Brown et al. only used male criminals. Adding female criminals to the design seems to have affected male subjects' recognition performance to be different than reviews of the literature would lead one to expect. Perhaps the male subjects paid more attention to the female criminals than they did to the male criminals and more than the female subjects paid to criminals of either sex. Extra attention may have worked through other processes, such as eye fixations (Loftus, 1972), to facilitate the recognition performance of the male subjects.

Finally, a caveat is in order regarding the interpretation of recognition results, not so much perhaps in the present experiment as in many reported in the literature: While it appears to be customary to generalize from the sample of stimulus persons to the population from which they were selected, this may not always be appropriate. Although these problems of generalization were treated extensively by Brunswik (1956) and are discussed in standard texts of research design (e.g. Plutchik, 1974), they seem often ignored or swept under the rug in facial recognition studies. The selection of stimuli in facial recognition experiments may be standardized to studies randomly selecting pictures of faces from college year books and studies matching live persons on demographic characteristics. While the latter type of studies usually contain persons representative of an

important criminal class (white males in their twenties), it may be hazardous to generalize results obtained from studies using pictures to situations where live persons are being remembered. Second, there is the problem of the number of stimuli: Can results from experiments using a small number of stimulus faces or persons properly generalize to the larger group? This is not only a problem for Cross et al. (1971) and most other experiments using pictures of faces, but for the present experiment and Brown et al. (Experiment 3) as well. With the latter two studies, however, arises the problem of feasibility of using a large number of live persons as stimuli. It may be difficult to have a large number of faces. Even if it can be done, it may not be feasible to get large numbers of persons to show up at a particular place at a specific time. (The question of feasibility of number should not apply to studies using pictures of faces.) Then, of course, in addition to the number of suspects, there are the other important questions in determining generalizability, such as control for indictability of the suspects, i.e., their tendency to collect indictments whether guilty or not (note, for instance, the high indictability of suspect F), variation in the recognizability (the d's) of the individual suspects and the representativeness of the sample. Within the

present experiment, the generalizability of the findings would seem to be a function of the number of suspects in a given cell: The d's on all eight suspects probably would be more stable than where recognition is of one sex versus the other.

The balanced presentation of the suspects to the subjects in both the present experiment and in Brown et al. (Experiment 3) helps reduce somewhat the problems of external reliability and over generalization. In these two studies, each stimulus person was responded to both as a criminal and as an innocent, each suspect serving as his/her own control. This procedure permits one to partial out any bias toward reporting a particular face as having been seen before. In contrast, the procedure followed by Cross et al. (1971) and most other studies of facial recognition do not permit comparison of hit and false alarm rates on the same face. Unless a forced-choice procedure is used, it is necessary to get a false alarm rate as well as a hit rate to determine any response bias as in any standard psychophysical experiment. However, for present purposes, it is not always necessary to get them on the same face. Rather, as in Brown et al. (Experiment 2), they can be obtained using other faces matched on demographic characteristics. However, since there is apt to be more variation in resulting recognition scores, this does not lead

to as good as representative sampling with small numbers of suspects as with each suspect being his/her own control,

While the recognition results were in general agreement with previous research using live persons, the results regarding recall of the circumstances of encounter can only be compared with studies using pictures. Where recall of the circumstances of encounter has been as directly assessed, it was done regarding either the previous orientation of pictures of scenes (Standing, Conezio, & Haber, 1970) or in which room a picture of a child was previously seen (Brown et al., Experiment 1). Although there are these differences in procedure, the results from these studies and the present experiment are in general agreement. Standing et al. (1970) for example, found that after a 24-hour retention interval, recall for the circumstances of encounter was around 70%. Brown et al. (Experiment 1) report that their subjects recalled the room of encounter of the pictures correctly at a mean proportion of 0.58. Though somewhat lower than Standing's et al. subjects' performance, Brown's et al. subjects were recalling after a retention interval of two days. The present study found recall of the circumstances of encounter to be a mean proportion of 0.54 after a retention interval of one week. Indeed, results of the present study and of Brown et al.

suggest witness performance in a criminal identification task seems to be determined largely by recognition performance with little recall of the circumstances of encounter.

The present experiment also obtained information regarding subjects' recall of the sex and number of the persons they previously encountered. Though subject performance at recalling their encounters with criminals was low, they were very accurate in recalling the number of persons encountered and, except for one subject, errorless in recalling having encountered a mixed sex group. The tendency to report seeing more men than women may be explained by the fact that all of the teaching assistants for this particular psychology course are male, and it is likely to be teaching assistants dealing with test materials. Thus, male would be a reasonable best guess. While recalling the number and sex of the persons encountered turned out to be an easy recall task, correct performance appears not to be facilitative to the other recall tasks tested, especially when recalling whether it was the same two people at each encounter, apparently a very difficult recall task. On the other hand, where subjects did recall the right circumstance of encounter for those criminals they correctly indicted, the chances were about two-in-three the criminal's activities would be recalled also.

Thus, recall of the circumstances of encounter may facilitate or at least be correlated with recall of the criminal's activities at the time of encounter.

While the results of the present experiment confirmed suspicions of a performance difference between recognition and recall in memory for faces, the results do not allow determination of the magnitude of this difference. The reason for this resides in the manner recall of the circumstances of encounter is typically assessed, i.e., on only correct recognition (cf., Brown et al., Experiment 1; Standing et al., 1970). In order to get a ratio measure of the difference between recognition and recall, a common baseline is required against which both recognition and recall may be compared. While analysis with signal detection theory provided a bias free index of the relative strengths of recognition and recall, it is believed that the noise distribution (i.e., the baseline) could not be the same for recall as it was for recognition. Basing the analysis of recall of the circumstances of encounter on only correct recognitions may ignore enough of certain kinds of information to change the noise distributions between the two, making comparisons indended to estimate the ratio of the difference impossible. With respect to their separate noise distributions, the two *d*'s (i.e., 0.30 and

0.20, for recognition and recall performance, respectively) show both recognition of previously seen faces and recall of the circumstances of their encounter to be quite weak in nondirected memory. Thus, subject performance in these tasks is being tested at the lower limits of memory.

While comparison of the recognition and recall *d*'s simply says that the average strength of discriminability of criminals from innocents is to some degree stronger than the average strength of the association of correctly indicted criminals with the right circumstances of encounter, some indication of performance differences may be seen when comparing indictment rates in both recognition and recall. A criminal stood a little less than a one-in-two probability of escaping indictment and an innocent stood about the same probability of being indicted. While subjects only recognized about half the criminals they encountered, they recalled where they encountered only about half the criminals they recognized. Since both the correct and incorrect indictment rates in both recognition and recall are about the same, similar looking *d*'s would be expected. Note, then that recall information is lost on about half the criminals the subjects encountered and, hence, lost in the calculation of the recall *d*'s as well.

Measuring recall of the circumstances of encounter on only the correct recognitions, however, is not unreasonable where real-life situations are concerned. With respect to forensic inquiries, it may indeed be the most suitable method of estimating recall performance: Courtrooms are notoriously defendent centered, never asking witnesses about persons they thought they didn't see.

Finally, regarding the interpretation of the results of recall of the circumstances of encounter as a function of sex of subject: They were symmetrically opposite those on recognition of previously seen faces as a function of sex. It might be concluded that male subjects' superior recognition of female criminals resulted from a greater depth of processing (Bower & Karlin, 1974) of female criminals by male subjects. It is debatable whether the depth of processing hypothesis helps to explain recall performance as a function of sex of subject. As in recognition, it may be that recall of the circumstances of encounter is more apt to be accurate in cases involving deeper processing. Results of the present experiment do not rule out that possibility, even though the poorest recall performance came from men recalling where they encountered female criminals. Besides the depth at which

critical identifying features are processed, memory performance may be determined to some extent by the total amount of information processed at the original encounter that is present at testing. Female subjects may have focused their attention to include aspects of the environment, fashions and other attributes of the persons they encountered that were more apt to be either absent or changed at the lineup, as well as (facial) features predictive of later recognition. Insofar as things changed about the criminals from encounter to lineup, they were perhaps greatest for the female criminals. (The female criminals, expressing concern about how to dress on both occasions, seemed to "dress-up" a bit more for the lineup than for the original encounter with the subjects, and a bit more than the male criminals for either encounter. The male criminals seemed to dress more nearly the same, though not exactly, on both occasions.) For stimuli where there was more change, there would be less information in which to generate an accurate association of criminal with encounter. Such being the base, the finding that women recognized the female criminals better, as the literature suggests they should, but recalled better the encounters with males does not contradict a depth of processing hypothesis.

Reference Note

1. Shroder, E. E. Recognition and attractiveness as a function of sex and race. Unpublished masters thesis, University of Nebraska at Omaha, February, 1975.

References

Banks, W. P. Signal detection theory and human memory. Psychological Bulletin, 1970, 74(2), 81-99.

Bower, G. H., & Karlin, M. B. Depth of processing pictures of faces and recognition memory. Journal of Experimental Psychology, 1974, 103, 751-757.

Brown, E., Deffenbacher, K., & Sturgill, W. Memory for faces and the circumstances of encounter. Journal of Applied Psychology, in press.

Brunswik, E. Perception and the representative design of psychological experiments. Berkeley: University of California Press, 1956.

Carterette, E., & Friedman, M. (Eds.) Handbook of perception: Historical and philosophical roots of perception. Vol. 1. New York: Academic Press, 1974.

Cross, J. P., Cross, J., & Daly, J. Sex, race, age and beauty as factors in recognition of faces. Perception & Psychophysics, 1971, 10, 393-396.

Ellis, H., Shepherd, J., & Bruce, A. The effects of age and sex upon adolescents' recognition of faces. Journal of Genetic Psychology, 1973, 123, 173-174.

Galanter, E. Textbook of elementary psychology. San Francisco: Holden-Day, 1966.

Going, M., & Read, J. D. Effects of uniqueness, sex of subject, and sex of photograph on facial recognition. Perceptual and Motor Skills, 1974, 39, 109-110.

Gourevitch, V., & Galanter, E. A significance test for one parameter isosensitivity functions. Psychometrika, 1967, 32, 25-33.

Hochberg, J., & Galper, R. E. Recognition for faces: An exploratory study. Psychonomic Science, 1967, 9, 619-620.

Howells, T. H. A study of ability to recognize faces. Journal of Abnormal Social Psychology, 1938, 33, 124-127.

Kimble, G., & Garmezy, N. Principles of general psychology. (3rd ed.). New York: The Ronald Press Company, 1968.

Lockhart, R. S., & Murdock, B. B., Jr. Memory and the theory of signal detection. Psychological Bulletin, 1970, 74(2), 100-109.

Loftus, G. R. Eye fixations and recognition memory for pictures. Cognitive Psychology, 1972, 3, 525-551.

Marascuilo, L. Extensions of the significance test for one-parameter signal detection hypotheses. Psychometrika, 1970, 35, 237-243.

Plutchik, R. Foundations of experimental research. (2nd ed.). New York: Harper & Row, 1974.

Shepard, R. N. Recognition memory for words, sentences and pictures. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 156-163.

Standing, L. Learning 10,000 pictures. Quarterly Journal of Experimental Psychology, 1973, 25, 207-222.

Standing, L., Conezio, J., & Haber, R. Perception and memory for pictures: Single trial learning of 2500 visual stimuli. Psychonomic Science, 1970, 19, 73-74.

Swets, J. (Ed.) Signal detection and recognition by human observers. New York: Wiley, 1964.

Witryol, S. L., & Kaess, W. A. Sex differences in social memory tasks. Journal of Abnormal and Social Psychology, 1957, 54, 343-346.

Yin, R. K. Looking at upside-down faces. Journal of Experimental Psychology, 1969, 81, 141-145.

Footnotes

[1]Introductions to signal detection theory may be found in Kimble and Garmezy (1968) or Galanter (1966). For more detailed accounts see chapters in Carterette and Friedman (1974) or articles in Swets (1964) and for applications to memory experiments see Banks (1970) and Lockhart and Murdock (1970).

Table 1

d' Scores for Recognition

Re Sex of Subject by Sex of Criminal

	Criminal		
Subjects	Male	Female	Sum
Male	.22	.69**	.46**
Female	.18	.28*	.20
Sum	.20*	.38**	.30****

*$p < 0.05$

**$p < 0.01$

****$p < 0.0001$

Table 2

Proportion Correct (Hits) and Incorrect (FAs) Indictments and d's for Recognition

	Male				Female			
Subjects	B_{IW}	C_{OW}	E_{OE}	H_{IE}	A_{OE}	D_{IE}	F_{IW}	G_{OW}
Male								
Hits	.55	.77	.18	.73	.55	.41	.77	.77
FAs	.50	.59	.32	.45	.27	.09	.64	.41
d'	.13	.51	-.44	.74	.74	1.11*	.38	.97*
Female								
Hits	.46	.62	.28	.72	.44	.28	.61	.72
FAs	.54	.49	.35	.51	.43	.29	.72	.54
d'	-.20	.33	-.20	.56*	.03	-.02	-.30	.48
Sum								
Hits	.48	.66	.25	.72	.48	.33	.65	.73
FAs	.52	.52	.34	.50	.39	.24	.69	.49
d'	-.10	.36	-.26	.58*	.23	.26	-.12	.64

[a]Subscripts of the suspects identification letters (A-H) indicate each suspect's circumstance of encounter, whether encountered on the way in or out of the room (I or O) and whether at the east or west corridor (E or W).

*$p < 0.05$

Table 3

d' Scores for Recall

Re Sex of Subject by Sex of Criminal

	Criminal		
Subject	Male	Female	Sum
Male	.15	-.05	.05
Female	.56***	.10	.30*
Sum	.40**	.05	.20*

*$p < 0.05$

**$p < 0.01$

***$p < 0.001$

Table 4

Proportion Correct Recalls

	Male				Female			
Subjects	B	C	E	H	A	D	F	G
Male	.33	.53	.75	.62	.42	.44	.41	.65
Female	.64	.64	.45	.57	.47	.27	.45	.63
Sum	.56	.61	.53	.59	.45	.35	.44	.64

Appendix A: Instructions

I am here to ask you to participate in an experiment, and then to conduct it with those of you who are willing to help me out. I am attempting to find out how well people do at remembering incidental or chance encounters with other people. To the best of my knowledge there has not been much of anything published on incidental memory for faces, but there has been quite a bit of work done with directed memory; that is, the subjects are told they will be asked to remember the material sometime later. If in these directed memory studies the material to be learned is photographs of faces or even live faces, I know from the literature and some work that I have done myself that people do quite well at picking out the faces or people they were asked to remember. What I don't know is, how good we humans are at remembering faces or people that we were not asked to remember in situations where we have no particular motivation to pay attention to them. Situations like this occur with some crimes, like getting your pocket picked, for example. The pickpocket may bump into you and take your wallet but you are not aware that this has occurred until sometime later, and thus have no particular motiv tion to notice the culprit at the time. To take

another example, you may encounter several persons committing a crime but be completely unaware that a crime is occurring, and pay very little attention to them. Sometime later you might be asked to identify these persons. While I doubt that people would do very well at correctly identifying any one of the culprits when later asked to do so--I don't know, I may be wrong about that--I would suspect, however, the witness could tell whomever was asking how many criminals there were and their sex; but I may be wrong about that, too. I don't know. It is questions like these that I would like those of you who are interested in this problem to help me answer.

As you may have surmised from those sheets of paper my helpers have been passing around to you, you had a chance encounter with a persons or persons last Monday. Now, today, for those of you who are willing to help me with my thesis, I would like you to answer some questions about the person or persons who either handed you your test materials as you came in or collected them as you were leaving. If you choose not to participate, simply leave your response sheets blank. Now, for those who are willing to help me out, let's go over the response sheet very carefully.....