

Face memory and spontaneous personality processing

JOHN H. MUELLER, TIM KELLER, and ALISON DANDROY
University of Missouri, Columbia, Missouri

Subjects studied photographs of faces preselected as likable or unlikable, one at a time, and then they were presented four-face arrays in which the target faces were either present or absent. The subjects were also identified in terms of their tendencies to analyze other people according to personality traits. Face memory was not related to either need-for-cognition scores or ability to form groups of people based on shared features. However, self-reported tendencies to analyze strangers in terms of personality rather than physical features did facilitate face memory to some extent, at least for target-absent cases, and also led to greater confidence about responses.

A number of studies have shown that face memory is affected by the type of study activity that subjects are induced to engage in. For example, Bower and Karlin (1974) showed that having subjects make judgments about the personality (e.g., honesty) of a stranger in a photograph led to better retention than judgments that directed their attention to physical features (e.g., gender). This outcome, which has been replicated many times (e.g., Mueller, Heesacker, Ross, & Nicodemus, 1983), seems to occur because personality decisions require more processing and thus make it more likely that subjects will find distinctive facial features (Courtois & Mueller, 1979; Winograd, 1981).

Such studies have explicitly required subjects to attend to various features that either improve or hinder face memory. In the present experiment, we were concerned with whether subjects differ in which features they encode spontaneously. That is, do subjects differ in the tendency to think about the personality features of other people in their interpersonal contacts, and can this characteristic predict which subjects will be good at remembering faces?

Various studies have examined individual differences in personality dimensions as factors in face memory, generally without finding large effects (see reviews by Ellis, 1975; Shepherd, Ellis, & Davies, 1982; Yarmey, 1979). However, the particular dimension of interest here—engaging in personality analysis and interpersonal judgments—would seem promising because of its connection with the well-established effects of inducing that encoding behavior explicitly. Furthermore, intrinsic differences in encoding activities have been used to explain performance differences in other domains, such as differences between young and elderly adults (e.g., Eysenck, 1974), and low- and high-anxiety subjects (e.g., Mueller & Courtois, 1980), for example. Thus the present hypo-

thesis was that just as experimenter-imposed attention to abstract traits encourages good face memory, so will spontaneous tendencies to attend to the personality traits of the people shown in photographs.

There is some evidence to indicate that people do attempt to assess various personality characteristics when confronted with photographs. For example, when inspecting mugbooks, observers often assess not only criminality but the type of criminal as well (e.g., Goldstein, Chance, & Gilbert, 1984), and there is also evidence that the perceived match between a face and a stereotype will affect its memorability (e.g., Klatzky, Martin, & Kane, 1982). Such data make it seem plausible that some factor analogous to implicit personality analysis may affect face memory, at least under some circumstances.

We employed three strategies to identify subjects who might engage in spontaneous personality processing. First, we devised a people-grouping task to assess subjects' ability to identify similarities among well-known people readily. The subjects were shown the names of nine familiar individuals and asked to form them into sets of at least three on the basis of some shared characteristic, as well as to identify that characteristic. In fact, this task was an adaptation of an Educational Testing Services subtest (French, Ekstrom, & Price, 1963) called "Making Groups (XU-3)," which requires things to be grouped and the grouping explained. For example, given the set (a) trout, (b) robin, (c) frog, (d) car, (e) boat, (f) bat, (g) airplane, a subject might put B, F, and G together because they fly, D, E, and G together because they are inanimate, and so forth. In our case, we provided a set of well-known people, so for the set (a) Dan Rather, (b) O. J. Simpson, (c) Abraham Lincoln, (d) Howard Cosell, (e) Wilt Chamberlain, (f) Barbara Walters, (g) Richard Nixon, and (h) Ben Franklin, the subject might group A, D, and F as journalists; B, D, and E as sports figures; C, G, and H as political figures; B, E, and G as retired, and so forth. The hypothesis was that subjects who are good at this task might be subjects who spontaneously attend to the beneficial deep traits associated with good face memory.

Correspondence may be addressed to John H. Mueller, Psychology Department, 210 McAlester Hall, University of Missouri, Columbia, MO 65211.

Second, we gave subjects the Need for Cognition Scale (Cacioppo, Petty, & Kao, 1984). This is an assessment of the extent to which a person engages in and enjoys effortful cognitive activity, and it may conceivably be a general, indirect index of how much a person thinks about interpersonal contacts. If so, the subjects who score high on this instrument might also show good face memory, because they would attend more effectively to the personality traits during the study phase.

Finally, we simply asked subjects how much they engaged in personality analysis when they met a stranger they didn't know. Presumably a high self-rating would be associated with good face memory.

In general it was expected that high scores on these assessments would correlate positively with face memory and with confidence in identifications. Furthermore, because some variables seem to have a more pronounced effect when the target is not actually present (Wells & Turtle, 1986), we included test conditions such that the target face was not always present in the test array (i.e., all new faces were present instead).

METHOD

Subjects

The subjects were 90 college students enrolled in introductory psychology courses, participating in return for extra credit in their course.

Materials

The initial pool of stimuli was composed of 160 black-and-white portrait slides of senior students from college yearbooks, 80 men and 80 women (all Caucasian, as were the subjects). Stimuli with unusual features (e.g., beards, glasses) were not used. These were presented one at a time at an 8-sec rate to 20 other subjects in a preliminary rating phase to obtain likability ratings. Each subject's task was to rate "likability" on a 7-point scale, ranging from 1 (unlikable) to 4 (neutral) to 7 (likable). To structure the task, the subjects were encouraged to consider whether "this is a person you would want as a roommate or friend, given the photograph as your only evidence." They were specifically told that they should not rate just attractiveness per se, but rather rate whether the person shown appeared to be "likable" in more general terms.

Final stimulus selection included only the upper and lower quartiles of likability. Overall, the average likability rating for all 160 faces was 3.72 ($SD = 0.86$). The 32 likable faces actually used had an average likability rating of 4.84 ($SD = 0.37$), and the 32 unlikable faces had an average of 2.53 ($SD = 0.36$).

Two sets of materials were formed. In Set A, four male and four female faces were randomly selected to serve as targets, and 2×2 arrays were formed by photographing each target with three likable same-sex distractors or three unlikable same-sex distractors. Eight arrays were thus produced, with one male and one female of each of the four types: likable target with likable distractors (LL), likable target with unlikable distractors (LU), unlikable target with likable distractors (UL), and unlikable target with unlikable distractors (UU). Set B was then similarly constructed, using different faces, to create a second set of eight arrays. The position of the target was systematically rotated through all four places in the 2×2 array.

Procedure

In the face-study phase, each subject saw eight faces—either Set A or Set B targets (counterbalancing across subjects); then the subject saw all 16 arrays on the test, both Set A and Set B. The study set was composed of two likable and two unlikable males, and two likable and two unlikable females, randomly ordered. The targets were presented for study at a 5-sec rate, with the subjects merely observing in anticipation of a memory test to follow.

Two questionnaires were completed between the study phase and the test phase. In the people-grouping task, subjects formed as many subsets as possible of people having something in common from a larger set of well-known people, and explained the basis for the grouping as outlined above. There were six sets of nine names, of historical and contemporary people, with 3 min allowed for each set. A grouping had to have at least three names to be accepted, groupings could not be repeated with an alternative reason, and groupings could not be made on the basis of names that rhymed or of any other superficial aspects of the names per se.

The next phase of the experiment was the Need for Cognition Scale (Cacioppo, Petty, & Kao, 1984). Here, 18 items, such as, "I really enjoy a task that involves coming up with new solutions to problems," are answered, each one according to a 5-point scale.

The face-memory test consisted of the presentation of 8 arrays with an old face present, plus 8 arrays with all four faces new. The 16 arrays were shown in random order at a 15-sec rate, and the subjects indicated both a choice and a confidence rating. The subjects were required to indicate which face was the old face in the array, and then to rate their confidence on a 7-point scale (1 = low). The answer sheet made provision for marking "none," but the subjects were not informed of the exact number of target-absent (or present) arrays. No penalties were indicated for guessing. The subjects were told to treat the test as analogous to a criminal lineup, but they had been informed that none of the people involved was really a criminal.

The final phase of the experiment was the self-report rating of personality analysis. The task was presented to the subject as follows: "When you encounter someone you don't know, how analytical are you about their personality; that is, to what extent do you try to figure out where they are from, what they do for a living, etc. Use a 7-point scale, where 1 means not analytical, focusing perhaps just on physical appearance instead, and a 7 means very analytical."

RESULTS

Identification accuracy and confidence were analyzed separately for the target-present and target-absent arrays. When the array actually contained a face from the study set, the subject might correctly select it (a hit), select one of the other three faces present (a miss), or select no face (respond "none"); these scores are summarized in Table 1, along with the average confidence rating associated with each type of response, by target-distractor likability. The data were analyzed in a 2×2 design, for target likability \times distractor likability.

Table 1
Recognition Accuracy (Maximum = 2) and Confidence Ratings (Maximum = 7), by Target-Distractor Likability, for Target-Present Arrays

	LL	LU	UL	UU
Recognition				
Hits	1.19	1.34	1.73	1.58
Misses	0.43*	0.26	0.18	0.12†
None	0.38	0.40	0.09	0.31
Confidence				
Hits	3.48*	4.28*	4.85*	4.21‡
Misses	3.82*	3.65	3.81	4.54
None	4.12*	4.41*	4.38	4.29‡

Note—LL = likable target with likable distractors, LU = likable target with unlikable distractors, UL = unlikable target with likable distractors, UU = unlikable target with unlikable distractors. Significant ($p < .05$) regressions are indicated by superscripts: *rated stereotyping; †number of groups of people; ‡need for cognition.

Target-Present Arrays

Recognition accuracy. There was a target-likability main effect for hits, since likable targets were harder to identify correctly than were unlikable targets [$F(1,82) = 35.87, p < .0001$], and a significant target-likability \times distractor-likability interaction indicated that this effect was greater when the distractors also were likable [$F(1,82) = 6.34, p < .02$]. The overall best performance occurred when the target and the distractors were unlikable (UU), and the worst when both target and distractors were likable (LL).

There were not many errors that involved explicit identification of a distractor as the target (a miss), but errors of this type were more probable when the target was likable rather than unlikable [$F(1,82) = 16.60, p < .0001$], or when the distractor was likable rather than unlikable [$F(1,82) = 5.98, p < .02$], with no significant interaction. Misses were most common overall when a likable target was arrayed with three likable distractors (LL).

The other type of error, indicating no target, also occurred infrequently, but it too showed main effects for target likability [$F(1,82) = 13.13, p < .001$] and distractor likability [$F(1,82) = 5.50, p < .02$], with no interaction. Overlooking the target was more likely when the array had likable targets (LL and LU vs. UL and UU), or when the distractor was unlikable (LU and UU vs. UL and LL).

Confidence in responses. The confidence ratings for hits mirrored the recognition accuracy data. There was a target-likability main effect [$F(1,82) = 15.10, p < .0001$], as subjects showed greater confidence with respect to unlikable targets. However, there was a target-likability \times distractor-likability interaction [$F(1,82) = 19.20, p < .0001$], which indicated that the target-likability difference was present only when the distractors were likable, with no target-likability difference when the distractors were unlikable. The subjects were least confident of their correct choice with respect to LL arrays.

There were no significant effects in the confidence ratings for either misses or "none" responses.

Individual differences. The results of regression analyses involving each of the three scores are shown in Table 1. It is clear that the three personality scores were essentially unrelated to recognition accuracy. Only two cells produced significant relationships, both involving misses. The subjects who generated more groupings on the people-grouping task made significantly more misses for the UU arrays [$r(90) = .21, p < .05$]. Otherwise, the only relationship was that subjects who self-reported themselves as being most prone to analyze strangers made significantly fewer misses on the LL arrays, [$r(90) = -.36, p < .001$ —and the other three array types, but nonsignificantly so [$r(90) = -.14, -.18, -.16$]. Need-for-cognition scores showed no significant relationships for accuracy.

In terms of confidence, self-report personality assessment showed several significant effects ($ps < .05$), particularly for the more difficult array types (involving likable targets). For LL, LU, and UL arrays, the subjects

who analyze strangers a lot were more confident of their correct responses (hits) than were subjects who reported less inclination to analyze strangers, and they were likewise more confident of their errors with respect to the more difficult arrays (LL misses, as well as LL and LU "none" responses). The subjects with high need-for-cognition scores were more confident of their responses for UU arrays (hits and "none" responses) than were those with low need-for-cognition scores. There were no effects on confidence due to number of groupings of people.

Target-Absent Arrays

For the arrays where all four faces were new—that is, where none were from the study set—the subjects could correctly respond "none" (a correct rejection), or incorrectly select one of the faces as having been in the study set (a false alarm). Because false alarms and rejections sum to a constant value here, and thus yield identical values for statistical tests, only the false alarm data will be presented, by likability of the face incorrectly chosen as an old face (see Table 2).

Recognition accuracy. There clearly was an effect of likability [$t(88) = 4.85, p < .001$], since unlikable faces were less likely to be misidentified as old than were likable lures.

Confidence in responses. Although they selected a likable face incorrectly more often than an unlikable face, the subjects nonetheless expressed more confidence in their errors when misidentifying a likable face than when misidentifying an unlikable face [$M_s = 4.68$ and 3.91 ; $t(88) = 2.56, p < .02$]. For comparison, the average confidence rating for correct rejections was 4.85. Thus, response confidence did not differ significantly for correct responses as opposed to incorrect selections of likable faces, both of which involved greater confidence than did incorrect choices of unlikable faces.

Individual differences. The subjects who reported that they analyzed encounters in terms of personality features showed significantly fewer false alarms, particularly a reduced tendency to identify an unlikable new face as old, than did the subjects who analyzed less in terms of personality. The correlation was significant for the number of false alarms involving unlikable faces, but nonsignificant for likable faces [$r(90) = -.33, p < .0001$, and $r(90) = -.15, p < .15$].

Furthermore, high self-raters showed greater confidence in their errors involving likable faces [$r(90) = .57$,

Table 2
False Alarms (Maximum = 4) and Confidence Ratings
(Maximum = 7), by Likability, for Target-Absent Arrays

	Likable	Unlikable
Number of false alarms	2.08	1.18*
Confidence in false alarms	4.68*	3.91

Note—Significant ($p < .05$) regressions are indicated by the superscripts: *rated stereotyping.

$p < .0001$]; and likewise, but not significantly so, for unlikable false alarms [$r(90) = .24, p < .10$]. In addition, high self-raters were also more confident of their correct rejections than were low self-raters [$r(90) = .50, p < .0001$].

DISCUSSION

In general, the effects of likability and target-presence were similar to what was found in an earlier study (Mueller, Heesacker, & Ross, 1984). That is, likable faces were harder to recognize than were unlikable faces, as if likable faces are less distinctive in their features. Unlikable faces, whether as targets or as distractors, led to better performance and greater confidence in decisions.

Of greater interest here is how the personality-assessment scores were related to recognition accuracy and performance. Previous studies of this relationship (Mueller & Thompson, 1986, 1988) have examined several procedures for assessing this characteristic and found only limited evidence to support any such relationship when face memory is assessed by a single-item recognition test.

In the present experiment, we used a multiple-item test format, but still found little support for such a relationship when individual differences in personality analysis were defined by an indirect assessment. That is, need for cognition scores did not predict face-memory performance, nor did scores on the people-grouping task. However, the simple global self-report measure was stronger. The subjects who identified themselves as being more likely to engage in personality analysis when confronting strangers were not necessarily more accurate in the target-present tests, but they were more accurate for all-new test arrays, or at least they were less likely to mistake an unlikable face as having been in the study set. Furthermore, the relationship between self-report analysis and response confidence was even more apparent in the target-absent arrays than in the target-present arrays, as per Wells and Turtle (1986).

Although the hypothesis that spontaneous individual differences in personality processing might predict face-memory differences is appealing, the evidence indicates that identifying such an individual difference dimension by means of conventional inventories is difficult. More complex procedures are available, such as the repertory test used in personal construct theory (e.g., Adams-Webber & Mancusco, 1983), but it isn't clear that the apparent magnitude of the relationship would warrant such efforts. In fact, the clearest effects were obtained here with the most straightforward assessment, by simply asking subjects for a global rating of their tendency to examine strangers in terms of personality attributes rather than physical features.

REFERENCES

- ADAMS-WEBBER, J., & MANCUSCO, J. C. (1983). *Applications of personal construct theory*. New York: Academic Press.
- BOWER, G. H., & KARLIN, M. B. (1974). Depth of processing of faces and recognition memory. *Journal of Experimental Psychology*, *103*, 751-757.
- CACIOPPO, J. T., PETTY, R. E., & KAO, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, *48*, 306-307.
- COURTOIS, M. R., & MUELLER, J. H. (1979). Processing multiple physical features in facial recognition. *Bulletin of the Psychonomic Society*, *14*, 74-76.
- ELLIS, H. D. (1975). Recognizing faces. *British Journal of Psychology*, *66*, 409-426.
- EYSENCK, M. W. (1974). Age differences in incidental learning. *Developmental Psychology*, *10*, 936-941.
- FRENCH, J. W., EKSTROM, R. B., & PRICE, L. A. (1963). *Kit of reference tests for cognitive factors*. Princeton, NJ: Educational Testing Service.
- GOLDSTEIN, A. G., CHANCE, J. E., & GILBERT, B. (1984). Facial stereotypes of good guys and bad guys: A replication and extension. *Bulletin of the Psychonomic Society*, *22*, 549-552.
- KLATZKY, R. L., MARTIN, G. L., & KANE, R. A. (1982). Influence of social-category activation on processing of visual information. *Social Cognition*, *1*, 95-109.
- MUELLER, J. H., & COURTOIS, M. R. (1980). Test anxiety and breadth of encoding experiences in free recall. *Journal of Research in Personality*, *14*, 458-466.
- MUELLER, J. H., HEESACKER, M., & ROSS, M. J. (1984). Likability of targets and distractors in facial recognition. *American Journal of Psychology*, *97*, 235-247.
- MUELLER, J. H., HEESACKER, M., ROSS, M. J., & NICODEMUS, D. R. (1983). Emotionality of encoding activity in face memory. *Journal of Research in Personality*, *17*, 198-217.
- MUELLER, J. H., & THOMPSON, W. B. (1986). Stereotyping and face memory. In H. D. Ellis, M. A. Jeeves, F. Newcomb, & A. Young (Eds.), *Aspects of face processing* (pp. 163-169). Dordrecht, The Netherlands: M. Nijhoff.
- MUELLER, J. H., & THOMPSON, W. B. (1988). Individual differences in face memory. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues* (pp. 71-76). New York: Wiley.
- SHEPHERD, J. W., ELLIS, H. D., & DAVIES, G. M. (1982). *Identification evidence: A psychological evaluation*. Aberdeen, Scotland: University of Aberdeen Press.
- WINOGRAD, E. (1981). Elaboration and distinctiveness in memory for faces. *Journal of Experimental Psychology: Human Learning & Memory*, *7*, 181-190.
- WELLS, G. L., & TURTLE, J. W. (1986). Eyewitness identification: The importance of lineup models. *Psychological Bulletin*, *99*, 320-329.
- YARMEY, A. D. (1979). *The psychology of eyewitness testimony*. New York: The Free Press.

(Manuscript received August 11, 1988.)