Law and Human Behavior, Vol. 19, No. 6, 1995

Using Similarity Judgments to Conduct a Mugshot Album Search*

A. M. Levi,† N. Jungman,† A. Ginton,† A. Aperman,† and Gavriel Noble‡

Mugshot searches fail because the witness must be shown too many photos. Current methods to reduce the number rely too heavily on verbal reports of individual features. This research reports three tests of a new system that augments the presently used tools by having the witness choose photos that are subjectively similar to the target. Each photo thus chosen increases the ranking of every photo in the album that is similar to it, as determined by a similarity network in which the album photos are embedded. Because the ranking of the target itself is usually thus increased, it is soon displayed. The tests used familiar targets, an incidental one, and unfamiliar videoed ones. From 76% to 84% of the "witnesses" reached the target, having viewed 2.4 to 8.5 times fewer photos than they would have had they simply leafed through photos as is traditionally done.

The daunting task facing police when conducting a mugshot search has been compared to finding a needle in a haystack (Clifford & Davies, 1989). Albums have grown beyond witnesses' capacity to effectively search through them. Even scrutinizing only 100 to 150 photos may reduce a person's capacity to identify a target (Davies, Shepherd, & Ellis, 1979a), yet witnesses are required to search through far more.

Researchers (Clifford & Davies, 1989; Laughery, Rhodes, & Batten, 1981) agree that witnesses must be helped to choose a smaller subset of the album. Early efforts (Rhodes, Laughery, Batten, & Bargainer, 1979) tried physical measurements of facial features to achieve such a subset. However, witnesses seem unable to be as accurate with these measures, either because of imprecise memory or

* Send correspondence to Avraham Levi, Division of Identification and Forensic Science, Israel Police Headquarters, Jerusalem 91906, Israel.

† Division of Identification and Forensic Science, Israel Police.

‡ Academy for Senior Officers, Neurim.

difficulty communicating their remembered image in terms of physical measurements.

Police instead often select a few characteristics, primarily facial features (e.g., hair and eye color), to determine a subset of photos for the witnesses. Witnesses use these categories in providing a verbal description of the suspect, and the appropriate photos are chosen for viewing. Because witnesses can err in their reports, this method has proven useful only for small albums or for unusual targets. In other cases the target still falls within a subset too large for effective scanning.

These methods rely primarily on memory of individual facial features. However, as Clifford and Davies (1989) point out, this is based on a logical rather than a psychological analysis: It does not seem to be the major way that people remember faces. Rather, we need greater emphasis on a global approach, memory of the entire facial gestalt (Davies, 1981).

A variety of studies have established that memory of the total face is not simply the sum of memory of the parts. The total face takes precedence (Davies & Christie, 1982; Davies, Ellis, & Shepherd, 1977; Harmon, 1973; Purcell & Stewart, 1988; Tanaka & Farah, 1993). Witnesses can provide only vague and limited details of even well-known faces (Phillips, 1978; Shepherd, Davies, & Ellis, 1978). The recognition advantage of familiar faces is not related to better memory of the facial parts (Klatsky & Forrest, 1984).

The latest use of verbal descriptions is the British FRAME (Face Retrieval and Matching Equipment) system (Ellis, Shepherd, Flin, Shepherd, & Davies, 1989; Shepherd, 1986), based on 47 facial feature parameters plus age. A computer ranks the album photos in terms of their fit to the witness's description, and photos are presented in the order of the ranking.

In one test of this method, participants were told to try to remember a face they would see and were shown a slide slightly larger than life for 10 s (Ellis, et al., 1989). They then described the target in detail and did the computer search. Targets used were either "typical" or "distinctive." Participants viewed up to four screens of six photos. From an experimental album of 1,000 photos, typical faces appeared among the first 24 photos 78% of the time, distinctive faces 91%.

However, would these results generalize to field conditions? The witnesses described the target right after viewing him, but witnesses forget very quickly most verbal facial detail (Ellis, Shepherd, & Davies, 1980). Also, witnesses were warned to pay attention to the target in order to remember him. Because witnesses can provide only poor detail of even well-known faces, which are learned through incidental learning, we suspect that forewarning leads witnesses to pay greater attention to specific features.

Finally, albums are much larger than the 1,000 photos used in their test. Verbal categories may fail, although they are useful, because they are not useful enough with existing album sizes. We have therefore added a dimension based on the total facial gestalt. Our mugshot search has witnesses use their subjective estimate of the similarity of viewed album photos to the target. Photos similar to the target are used to focus in more quickly on the target.

We first created a similarity network between the album photos. Researchers differ in methods for establishing such networks (Davies et al., 1979b; Laughery, Fessler, Lenerovitz, & Yoblick, 1974). We preferred gathering judgments of photos as a whole. We needed, however, a method requiring relatively few judgments to code each photo. Previous studies with intuitive judgments used no more than 200 photos. Our experimental album was six times larger, and we were looking ahead to coding operational albums many times that size.

We coded each album photo in terms of its dissimilarity to reference photos. Our reasoning was that if two photos were judged dissimilar to the same reference photos, they could be expected to be similar to each other. Eighty reference photos were chosen intuitively, differing as much as possible one from the other. They were divided into 10 sets of eight photos each. Because dissimilarity choices from sets of eight photos were easier and far quicker to make than similarity choices, we used the former.

The similarity network is an 80-dimensional space, one dimension for each reference photo. Because each rated photo was either similar or dissimilar to each reference photo (a binary decision) each dimension has two values, 1 when a reference photo has been chosen as dissimilar to the scaled photo, and 0 otherwise. In scaling a photo the senior author chose, from each of the 10 sets of eight reference photos, the three or four most dissimilar to the photo being scaled. The resulting set of about 35 reference photos received the value of 1 for the scaled photo, the remaining 45 getting the value of 0.

Similarity judgments are purely subjective. The success of our search depends on relatively high consensus between the coder and the witness. Fortunately, many studies confirm that similarity judgments of faces are relatively consensual (e.g. Davies, Shepherd, & Ellis, 1979b), though there is also evidence of individual differences (Hirschberg, Jones, & Haggerty, 1978). Experiment 1 tested this consensus also.

For the mugshot search, witnesses first entered into a PC computer a description of the target using the small set of categories (age, ethnic group, hair color, eye color, darkness of face, facial hair, blemish, eyeglasses, body build), choosing from a fixed set of descriptors (e.g., for eye color—dark or light). They also gave a confidence rating of 1 to 10 to each category.

The computer weighted each album photo in terms of its compatability to the description. The value of 1 was assigned for total fit (e.g., an age within 5 years of the description, or the precise body build), the value of ½ for partial compatibility (an age over 5 years but within 10 years, or to a medium body build when the suspect was described as thin), and a value of 0 for incompatability. Each value was multiplied by the confidence rating given to that category, and the results summed over all the categories.

The computer ranked the resulting weights and displayed the 24 photos with the best fit. At this point, the method is similar to Ellis et al. (1989)—the chief difference being the greater detail of their 48 categories—but from here on, the methods part company. Their method displays six photos at a time, using the fixed rank order to determine precedence. Our method has witnesses use similarity judgments, choosing up to five most similar photos to the target from the 24.

The computer compares the vector of each chosen photo with that of all other album photos. Each album photo's weight is increased by 4 for each value of 1 that is identical to that of the chosen photo (for both, the same reference photo had been chosen as dissimilar during scaling). An album's photo can be increased, for each chosen photo, by a maximum of $35 \times 4 = 140$ if exactly the same set of 35 reference photos had been chosen as dissimilar during scaling (the value of 4 was added in order to balance the effect of multiplying the verbal categories by their confidence rating. Otherwise, similarity judgments would have a negligible effect).

Photos more similar to the chosen photo have their weights increased more. When the album is reranked using the new weights, those photos, including the target, rise in the ranking. Again, the photos with the largest weight are displayed. The process continues until the target is displayed or an arbitrary number of sets is shown.

EXPERIMENT 1

Our system requires that witness and coder similarity judgments be relatively consensual. We tested this by limiting effects of memory, following the example of the first test of FRAME (Shepherd, 1986), in using well-known faces.

Our research design was affected by the small size (1,200) of our album compared to operational ones. Using verbal categories with only 1,200 photos would often result in the target being displayed among the first screens, as did Ellis et al. (1989). This would leave little room for additional improvement for the similarity network.

We created an analogue of an operational search by having the computer put the target on the average in the 808th rank after the verbal category stage: The initial weight of the target was switched with the 808th photo. In our analogue, a rank of 808 resulted from using the verbal categories with an album numbering some tens of thousands of photos. The similarity method then took over to achieve further improvement.[1]

In the traditional method, witnesses scan all 808 photos. Ellis et al. (1989) found that respondents leafing through an album identified only ⅛ of typical faces placed in the 800th position. The similarity method would be considered a success if meaningfully less photos were scanned.

Ellis et al. (1989) terminated their computerized search if the target did not appear by the 24th face. That criterion is far too strict for an actual mugshot search. Our respondents came to the experiment expecting that it would least about 30 min. An average witness took about that time to view 24 screens, so we set the cutoff at 24 screens.

The cutoff point allowed witnesses to view 24 screens $\times$ 24 photos per screen, or 576 photos. This is much more than Laughery, Alexander, and Lane

[1] The variability in target placement in Experiment 1 was caused by a technical difficulty later resolved.

(1971) and Davies et al. (1979a) conclude is feasible. However, Ellis et al. (1989) and Lindsay, Nosworthy, Martin, and Martynuck (1993) found better identification rates than the earlier studies, with conditions more similar to mugshot searches. Indeed, setting a stringent cutoff point is not realistic, given the size of many operational albums. Police would stop conducting mugshot searches, because witnesses would so rarely reach the suspect.

The fear of viewing many photos causing more false identifications (choosing someone who is not the target) concerns us even less. The studies are in agreement that viewing more photos does not increase such errors (Davies et al., 1979a; Ellis et al., 1989; Laughery et al., 1971; Lindsay et al., 1993).

Our experimental album of 1,200 photos consisted of a random selection of males from the national police album, which includes all convicted criminals of the past 10 years. It is thus fairly representative of the male population.

Method

Respondents

Twenty-five men and 25 women serving at police national headquarters agreed to participate, when visited in their office. They had varied jobs and experience, with ages ranging from 18 to 55. (Not all those asked were able or willing to participate)[2]

Targets

Fifteen photographs of male public figures were selected from newspapers and added to our album. Each target was used a maximum of six times. The quality of these photos was poorer than the original album, as was the quality of other photos that had been added in various pretests.

Procedure

Participants were read the list of public figures and were requested to choose the one with whom they were most familiar. The mugshot search was then conducted as previously described.

Results and Discussion

Thirty-eight (76%) of the respondents reached the target, the remaining 12 having the session terminated beforehand. With the null hypothesis at the cutoff point of 24 screens, this result is significant, χ^2 (1, N = 50) = 12.5, $p < .001$ (All chi squares with 1 degree of freedom were corrected for continuity.) Thirty-four

---

[2] The functions of most headquarters staff are no different from any other organization, so there is no reason to expect results different from typical witnesses in the general population. In all events, attempts to date to demonstrate greater accuracy of police regarding eyewitness issues have failed (Ainsworth, 1981; Tickner & Poulton, 1975; Yarmey & Jones, 1982). This should not be too surprising, considering the general lack of success in training for improved face recognition (Malpass, 1981).

(68%) of the respondents had better than 50% improvement, χ^2 (1, N = 50) = 5.78, $p < .025$).

The distribution for those who reached the target resembles a normal curve truncated at the lower end: Three respondents (8%) reached the target in the minimum of two screens. The modal response was six screens (7 respondents, 18%). The median was eight screens. This is a 4.3 times improvement over the median initial position of 34 screens, the number of photos that would have been scanned by the traditional method of leafing through them.

The targets were not equally successfully reached. Six of the targets, accounting for more than half of the searches, never had a failure.

One might be tempted to compare our results with those of Ellis et al. (1989). However, we propose similarity judgments in order to supplement verbal descriptions, not supplant them. Indeed, they are an integral part of our method.

Of the 38 witnesses who reached the target, 30 (79%) identified him. The poorer quality of the target photos may have reduced identifications. However, our prime dependent variable is not identifications but the speed in which the target reaches the screen. We could have printed the name of the target on the screen instead of presenting a photograph.

EXPERIMENT 2

Experiment 1 used famous people as targets to test the system under good memory conditions. In reality, however, witnesses must identify a target after seeing him or her once. Also, employing a cutoff point prevented obtaining full data on all witnesses. The 4.3 times improvement related only to the 76% most successful witnesses. Were the remaining witnesses even worse off than they would be if they simply had leafed through the photos?

In Experiment 2, police officers in a lecture incidentally witnessed a male target briefly and were then requested to search for him. Technological improvements (moving from a 386 to a 486 PC, and faster software to implement the similarity judgment algorithm) decreased computer time, enabling all witnesses to complete the search within 30 min.

Each photo acquires a weight based on verbal categories, and this weight increases each time the witness chooses a photo similar to it. In Experiment 1, the initial weighting was small relative to the increase from similarity judgments. In this study, we added a condition that gave larger weight to the verbal categories. Choices of photos similar to the target affected the target's ranking less, requiring the viewing of more photos to reach it. Poor choices had a smaller negative influence. We expected this conservative effect to reduce variability. The question was how the average speed in reaching the target would be influenced.

Method

Respondents

Fifty police officers participating in a course at a police academy agreed to search for the target. None had a previous aquaintance with the target. No one refused.

Procedure

The target entered a regular class and, standing near the door, announced that those whose names he was about to read were to leave with him to participate in an experiment. He then named most of the class. The "surprised" lecturer announced a recess. In the ensuing confusion, the target disappeared, and the commander of the course entered and explained that they would be requested to search for the target in our album. The target was present for about a minute. All witnesses said that they saw him, but with little attention and at times from a suboptimal angle. The mugshot search was conducted as previously described.

Design

On one computer, the confidence ratings of the verbal categories were identical to the previous experiment, a range of 1 to 10. On the other, the range was 10 times larger, from 10 to 100.

Results and Discussion

The search took a minimum of 10 min and a maximum of 27. The first search began on the first day, half an hour after the class incident, and the last 31 hours later. A median split chi-square on the effect of the interval between the exposure to target and the test on the number of screens viewed yielded χ^2 $(1, N = 50) = 5.12, p < .03$. For larger intervals, performance was worse.

All witnesses reached a screen with the target. Only 1 of the 50 viewed as many screens, 34, as would be required by the normal method of simply leafing through the photos. Using the 24-screen cutoff point, 84% of the witnesses reached the target, χ^2 $(1, N = 50) = 21.8, p < .001$.

The median number of screens viewed on each computer was almost identical (20 screens using the confidence ratings of 1 to 10, 19 using ratings from 10 to 100). The variance, as predicted, was very different, F $(27,21) = 2.6, p < .02$, with less variance using the new method. The old method accounted for 75% of search failures by the 24 screen cutoff, and therefore comparisons with Experiment 1 will be with this method only.

Of those, 78% reached the target within 24 screens, compared to 76% in Experiment 1. However, whereas 56% of the witnesses succeeded in identifying the target in Experiment 1 (79% of the 38 who reached the target), only 18% did in the present one. Though the 76% who reached the target in Experiment 1 did so 4.3 times faster than if they had leafed through the photos, the comparable 78% in the present experiment achieved a median improvement rate of 2.3 (a median of 15 screens compared to 34).

Experiment 2, then, replicated Experiment 1, within the limitation of the poorer viewing conditions. The low identification rate reflects the viewing conditions. It is a credit to the similarity judgment method that it succeeded as it did in reducing the number of photos the witness viewed.

The target also likely caused the poorer results. The best rate of improvement by a witness was 4.0, short of the median improvement of Experiment 1, despite his seemingly excellent similarity choices. The rate of 4.3 in Experiment 1 hides the poor improvement rate of some targets: The rate varied from 9.8 to an inability to estimate the (low) rate owing to a predominance of failures to reach the target. Part of the variability, then, seems related to the target.[3]

EXPERIMENT 3

One puzzling aspect of the two studies was the very similar percentages of witnesses who reached the target by the cutoff point. Why didn't the far greater familiarity of the targets in Experiment 1 result in improved performance, as was found with the other measures?

Perhaps 25% of the population are poor at making similarity judgments. This would have practical implications. We found in Experiment 2 that giving greater weight to the initial description decreased variability, resulting in 9% instead of 22% of the witnesses failing to reach the target within 24 screens. If it were possible to quickly test witnesses to discover those poor at making similarity judgments, the relative weight of the initial verbal judgments could be increased for them.

There are alternate explanations. The 76% in Experiment 1 who reached the target within 24 screens reflected the average performance of targets that highly differed from each other. The similarity to the 78% in Experiment 2 could be chance, the factor being the effect of target on performance. For example, the uniqueness of the target might affect the ease of reaching it (Ellis et al., 1989).

Further, the advantage of familiar targets may have been balanced by an increased difficulty in choosing similar photos. Better memory should result in improved discrimination, and witnesses with a strict criterion for similarity would then have greater difficulty in finding other photos that they could judge as similar. They might then either more often refuse to choose, or choose on the basis of a specific feature. Either strategy would hurt performance.

In addition, witnesses in Experiment 1 chose a familiar target from a set. Some witnesses may have mistakenly chosen a target that they remembered poorly.

This experiment tested these alternate possibilities, within the overall practical goal of exploring the feasibility of giving witnesses a pretest. If witnesses do well (or poorly) on one mugshot search, will they also do well (or poorly) on the next? Therefore, in the present study, witnesses participated in two consecutive mugshot searches.

Because we were testing for skill in judging similarity, we minimized differences in memory by creating good viewing conditions. Witnesses were warned that they would be conducting a mugshot search before they viewed the target. They were given unlimited time to study the target, and they searched for him

---

[3] The target effect in Experiment 1 is confounded, however, by a number of factors. The witness chose the target from a given set, the targets were likely not equally familiar, and newspaper photos of variable quality were used.

immediately. However, because we were testing person, not photo identification (Read, Vokey, & Hammersley, 1990), the targets viewed were not the mugshots themselves, but videos of the men talking while being interviewed.

Conducting two mugshot searches with each respondent was analogous to how testing of witnesses might be achieved in the field. The first search would be the test, using some standard target, and the second one would be for the offender.

The first target viewed was either familiar or unfamiliar to test for familiarity effects. In practice, only the first target could be either familiar or unfamiliar, because it would be the test. The actual mugshot search, the second target, would by definition have to be unfamiliar. Because Experiments 1 and 2 indicated that the particular target could strongly affect the results, we used four people as the first ("test") target (two familiar and two unfamiliar) and five as the second ("actual") target.

Differences between targets could result in better performance on either the first or second search. This could also be caused by order effect in a within-subjects design (learning effects, etc.). We therefore added a condition in which the order of the unfamiliar targets was reversed.

Method

Respondents

Ninety people (57 female) were recruited from the same subject pool and in the same way as Experiment 1.

Targets

Seven unfamiliar and two familiar targets were sampled from a videotape of a TV broadcast about a year previous to the experiment. Stills were sampled from the video of each target and added to the our album. They again were of poorer quality than the original album photos.

Procedure

Respondents were told that they would participate in two mugshot searches. Before each, they would view the target on video for as long as they liked. After viewing the first target, respondents searched for him. They then viewed the second, and searched for him. The video segments were from 10 to 65 s in length. If the segment seemed too short, respondents could view it more than once.

Design

The 90 respondents were randomly assigned to three conditions. Each condition had two test targets and two actual targets, with 3 respondents in each cell: (1) One of two familiar targets was the first (test) and one of five unfamiliar targets was the second (actual). (2) One of two unfamiliar targets was the first (test), and one of the same five unfamiliar targets was the second. (3) The targets in Condition 2 were displayed and searched for in reverse order: first one of the five, and then one of the two unfamiliar targets.

Results and Discussion

Eighty-four percent of the searches reached the target within 24 screens. The median number of screens to reach the target was 4 for those who succeeded, an improvement of 8.5 times over the standard method of leafing through the photos.

Of the 25 witnesses who failed to reach the target, only 3 failed to reach both targets. There was, then, no individual consistency in witnesses failing to reach the target and therefore no support for initial testing of witnesses.

There was a clear target effect. One photo was always reached, whereas another was reached only 66% of the time. One target was always reached in less than the median number of screens, but for another, only 17% of the witnesses reached the target in less than that median. The median test between targets was significant, χ^2 (8, N = 90) = 34.515, $p < .01$.

There was no difference between the familiar and unfamiliar test targets. Perhaps the advantage of the familiar targets was negated by the increased difficulty in finding photos similar to them. Although the test targets tended to be easier (they were reached quicker), this was not due to a position effect, because the unfamiliar test targets were reached as quickly when they were searched for last.

Respondents who made unsuccessful searches failed to choose any photo a median of 56% of their screens, compared to 6% for the successful ones. The median test between successful and unsuccessful searches was significant, χ^2 (1, N = 90) = 16.258, $p < .005$.

Failing to choose a photo hurts the search. Because the computer is given no new data, it has no basis for reranking the photos. The 24 photos that were displayed are removed, and therefore, the rank of the target is improved by only 24, just as would occur by leafing through the photos. In comparison, because the median number of screens needed by successful witnesses to reach the target was 4, and the target was placed arbitrarily in the 800th rank, such witnesses improved the rank of each screen by an average of 800/4 = 200 photos.

Perhaps, then, witnesses should be required to choose a photo from each screen. The danger is that witnesses may be forced to make poor choices and thus lower the target's rank, worse than an improvement of only 24. This could occur if the witness was presented with a screen with no good photo.

This was tested by choosing randomly 10 searches of successful witnesses who failed at least once to choose a photo. Their search was replicated to the first screen where they failed to choose. At that point, the senior author made a choice. The improvement of rank (24) of the original searches was compared to that achieved by the senior author's choices. In 9 out of the 10 searches the latter improvement was greater ($p < .01$ by binomial). The median improvement of the author's choices was 81, with a range of 1 to 735. The successful witnesses had, then, good photos when they failed to choose.

GENERAL DISCUSSION

Three experiments, each using different eyewitness conditions, verified the value of adding similarity judgments to the traditional mugshot search. The fewer photos viewed could make the difference between identifying a suspect and failing to do so. Using the poorer improvement rate of 2.3 of Experiment 2, a witness required to view 800 by the traditional method would need only about 350 photos to reach the suspect.

However, this is only a feasibility study. We are implementing a field test in one of our small country's larger police stations, with a full operational album and in actual field conditions.

Coding the experimental album into the similarity network was so time consuming that we were forced to make do with a single coder. Coding an operational album would be beyond the capacity of one person, and issues of quality control (interrater reliability) could become a major problem.

We have pilot tested an automated coding method by computer, using an algorithm for facial identification such as described by Turk and Pentland (1991). We require about 15 s per photo to manually mark the location of the eyes and mouth. The PC computer then establishes an initial core of analyzed photos. After that, any additional photo is added automatically to the similarity network once it is marked.

Fourteen witnesses searched twice for the same target, once using the original data base and once using the automated one, two witnesses per target in counterbalanced order. Nine witnesses (64%) reached the target faster with the automated method. This result is not statistically significant, but we conclude that the automated is as good as the original coding system, and we are using it in our field test.

A second issue is false identifications. Lindsay et al. (1993) showed that when the mugshot album consists of similar photos (if they have been chosen to fit a certain verbal description), there will be more false identifications than when the album consists of normally divergent photos. Our method, by using both verbal categories and similarity judgments, causes the witness to view similar photos. Therefore, we should expect false identifications.

Lindsay et al. (1993) propose allowing witnesses more than one choice. The mugshot search is an investigative tool, providing leads. The police are better off with more than one potential suspect, if the witness chooses more than one, than with no suspect at all (if the witness chooses no one).

We have argued (Levi & Jungman, in press) that even police lineups should allow more than one choice and that the lineup should be very similar to our mugshot search. We claim that calling the witness's choice of a suspect an "identification" is an error. Witnesses are unable to distinguish between people who are very similar. Yet their choice of more than one person in a mugshot album or lineup provides valuable information.

Our research began on an improved system for making police composites. We plan to combine the two tools. If witnesses fail to choose a suspect, a composite will be created by superimposing the photos they chose as similar to the target (Weil, 1982). This will simplify the task of constructing a composite, an important goal (Bennet, 1986).

We thus hope to provide an improved investigative system. Witnesses should be able to reach suspects faster in the mugshot album, thus increasing chances of successful identifications. When they fail, only a few minutes more effort should produce a composite.

REFERENCES

Ainsworth, P. B. (1981). Incident perception by British police officers. *Law and Human Behavior, 5*, 231–236.

Bennett, P. (1986). Face recall: A police perspective. *Human Learning, 5*, 197–202.

Clifford, B. R., & Davies, G. (1989). Procedures for obtaining identification evidence. In D. C. Raskin (Ed.), *Psychological methods in criminal investigations and evidence* (pp. 47–95). New York: Springer.

Davies, G. M. (1981). Face recall systems. In G. Davies, H. Ellis, & J. Shepherd (Eds.), *Perceiving and remembering faces* (pp. 227–250). London: Academic Press.

Davies, G. M., & Christie, D. (1982). Face recall: An examination of some factors limiting composite production accuracy. *Journal of Applied Psychology, 67*, 103–109.

Davies, G. M., Ellis, H., & Shepherd, J. (1977). Cue saliency in faces as assessed by the "Photofit" technique. *Perception, 6*, 263–269.

Davies, G. M., Ellis, H. D., & Shepherd, J. W. (1978). Face recognition accuracy as a function of mode of presentation. *Journal of Applied Psychology, 63*, 180–187.

Davies, G. M., Shepherd, J. W., & Ellis, H. D. (1979a). Effects of interpolated mugshot exposure on accuracy of eyewitness identification. *Journal of Applied Psychology, 64*, 232–237.

Davies, G. M., Shepherd, J. W., & Ellis, H. D. (1979b). Similarity effects in face recognition. *American Journal of Psychology, 92*, 507–523.

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

Ellis, H. D., Shepherd, J. W., Flin, R. H., Shepherd, J., & Davies, G. M. (1989). Identification from a computer-driven retrieval system compared with a traditional mug-shot album search: a new tool for police investigations. *Ergonomics, 32*, 167–177.

Harmon, L. D. (1973). The recognition of faces. *Scientific American, 229*, 71–82.

Hirscherg, N., Jones, L. E., & Haggerty, M. (1978). Whats in a face: Individual differences in face perception. *Journal of Research in Personality, 12*, 488–499.

Klatzky, R. L., & Forrest, F. H. (1984). Recognizing familiar and unfamiliar faces. *Memory & Cognition, 12*, 60–70.

Laughery, K. R., Alexander, J. F., & Lane, A. R. (1971). Recognition of human faces: Effects of target exposure time, target position, pose position, and type of photograph. *Journal of Applied Psychology, 55*, 477–483.

Laughery, K. R., Fessler, P. K., Lenerovitz, D. R., & Yoblick, D. A. (1974). Time delay and similarity effects in facial recognition. *Journal of Applied Psychology, 59*, 490–496.

Laughery, K., Rhodes, B., & Batten, G. (1981). Computer-guided recognition and retrieval of facial images. In G. Davies, H. Ellis, & J. Shepherd (Eds.), *Perceiving and remembering faces* (pp. 252–269). London: Academic Press.

Levi, A. M., & Jungman, N. (In press). The police lineup: Basic weaknesses, radical solutions. *Criminal Justice & Behavior.*

Lindsay, R. C. L., Nosworthy, G. J., Martin, R., & Martynuck, C. (1994). Using mugshots to find suspects. *Journal of Applied Psychology, 79*, 121–130.

Malpass, R. S. (1981). Training in face recognition. In G. Davies, H. Ellis & J. Shepherd (Eds.), *Perceiving and remembering faces* (pp. 271–285). London: Academic Press.

Phillips, R. J. (1978). Recognition, recall and imagery of faces. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 270–222). London: Academic Press.

Purcell, D. G., & Stewart, A. L. (1988). The face-detection effect: Configuration enhances detection. *Perception & Psychophysics, 43,* 355–366.

Read, J. D., Vokey, J. R., & Hammersley, R. (1990). Changing photos of faces: Effects of exposure duration and photo similarity on recognition and the accuracy–confidence relationship. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 870–882.

Rhodes, B. T., Laughery, G. M., Batten, G. M., & Bargainer, J. D. (1979). *Summary report for a research project: "A man–computer system for solution of the mug file problem."* University of Houston.

Shepherd, J. W. (1986). An interactive computer system for retrieving faces. In H. D. Ellis, M. A. Jeeves, F. Newcombe, & A. W. Younge (Eds.), *Aspects of face processing* (pp. 398–409). Dordrecht: Martinus Nijhoff.

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* (pp. 278–285). London: Academic Press.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology, 46A,* 225–245.

Tickner, A., & Poulton, E. (1975). Watching for people and action. *Ergonomics, 18,* 35–51.

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience, 3,* 71–86.

Weil, P. (1982). *About face: Computergaphic synthesis and manipulation of facial imagery.* M.A. Thesis, MIT.

Wells, G. L., & Hryciw, B. (1984). Memory for faces: Encoding and retrieval operations. *Memory and Cognition, 12,* 338–344.

Wells, G. L., & Turtle, J. W. (1988). What is the best way to encode faces? In M. W. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 163–168). New York: Wiley.

Yarmey, D., & Jones, H. T. (1982). Police awareness of the fallibility of eyewitness identification. *Canadian Police College Journal, 6,* 230–239.