Laughery, K. R., Jensen, D. G., & Wogalter, M. S. (1988). Response bias with prototypic faces.

PRACTICAL ASPECTS OF MEMORY: CURRENT RESEARCH AND ISSUES

VOLUME 1

Memory in Everyday Life

EDITED BY

M. M. Gruneberg
University College of Swansea

P. E. Morris
University of Lancaster

R. N. Sykes
University College of Swansea

JOHN WILEY & SONS
Chichester · New York · Brisbane · Toronto · Singapore

RESPONSE BIAS WITH PROTOTYPIC FACES

Kenneth R. Laughery and Dean G. Jensen
Department of Psychology, Rice University
Houston, Texas, U.S.A.

Michael S. Wogalter
Department of Psychology, University of Richmond
Richmond, Virginia, U.S.A.

ABSTRACT

Three experiments employed a facial recognition task where target faces did not appear in a five or six-item test set. The test set consisted of a prototype face that differed by one feature from each of the distractors, which in turn differed by two features from each other. Subjects ranked the faces on the likelihood they were a target face. Results showed the prototype was ranked significantly above chance, indicating the procedure resulted in a response bias. The findings have implications for lineup and photospread construction.

This paper presents the results of three experiments that addressed the issue of response bias in a facial recognition task. This issue concerns the situation in which the recognition test set contains a face that is a prototype of the others in the set. For example, suppose the set is made up of faces each of which is a variation on one particular face, the prototype, that is also included in the set. One might expect that such a set would result in a response bias favoring the identification of the prototype face. There is some support for this prototype notion in a study by Solso and McCarthy (1981). Using a recognition memory paradigm, they constructed distractor faces from the features of faces that had been presented. Their subjects were more confident in recognizing (incorrectly) the distractor faces than the faces they had seen. Additionally, Wogalter and Jensen (1986) have demonstrated a bias towards a prototype in a recognition task using nonfacial stimuli.

The practical issue with which this work is concerned is fairness, or its opposite -- bias, in law enforcement lineups and photospreads. A biased lineup is one where persons who were not witnesses to a crime are more likely than chance to pick the suspect. Malpass and Devine (1983) and Wells (1978) have discussed lineup bias and have noted that the suspect must not be distinctive in comparison with other members (the distractors). Malpass and Devine (1983) reported an experiment in which similarity between a suspect photograph and the other photospread members was manipulated. Their results showed an increase in fairness with increasing suspect-distractor similarity. Hence, in constructing lineups or photospreads, law enforcement agencies would be advised

to select lineup or photospread distractors who are similar to the suspect. But such an approach creates the possibility of bias due to the suspect being a prototype of the lineup or photospread faces. That is, the suspect may have more features in common with the distractors than the distractors share with each other. In this regard the suspect may be distinctive, and in situations where the suspect is not the target person (criminal), the lineup or photospread may fail a crucial criterion in that the likelihood of the suspect being chosen is greater than chance.

The present experiments employ a recognition paradigm. In Experiment 1 subjects saw a single target face before examining a group of photographs -- a photospread. In Experiments 2 and 3, subjects saw a large number of faces, and then examined a series of photospreads. In Experiment 1 the target face appeared in some of the photospreads, whereas in Experiments 2 and 3 it did not. Each of the photospreads was made up of a prototype face (not the target) and distractors that were more similar to the prototype face than they were to each other. The hypothesis is that the likelihood is greater than chance that the prototype will be identified as a target face. In addition to differences in some procedural details, the experiments also differed in the stimulus materials used: Identi-Kit faces, Mac-a-Mug Pro faces and photographs of real faces. A sample Mac-a-Mug photospread is shown in Figure 1.

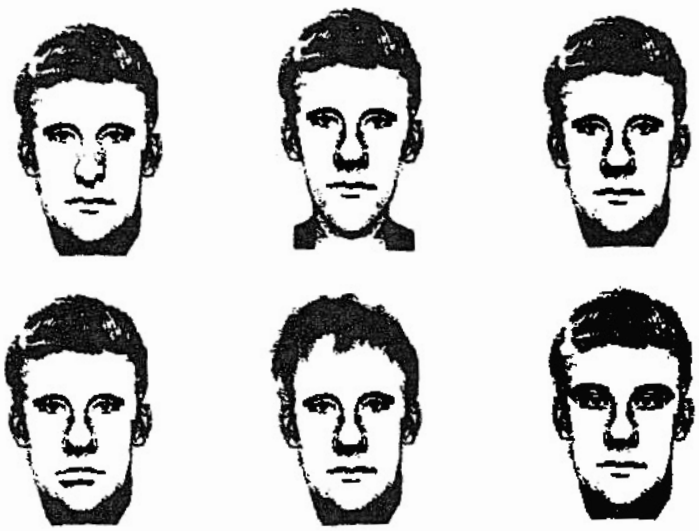

Figure 1. Sample Mac-a-Mug photospread, top-right face is the prototype.

EXPERIMENT 1

In the first experiment subjects went through a series of trials in which they were shown a target face, given a distractor task, and then asked to rank the members of a photospread in terms of their similarity to the target. All stimuli were male faces constructed with the Identi-Kit.

Method

Subjects were 25 undergraduate students at Rice University. The faces were constructed from the 1960 edition of the Identi-Kit facial construction system. They consisted of six features: hair, eyes, eyebrows, jaw, lips and nose. Two different exemplars were used for each feature, except the nose which was the same for all faces. The Identi-Kit features had actually been stored in a computer and the target faces were presented on a monitor.

Subjects were run individually and were told that they were to be tested on their memory for faces. They then were given five trials consisting of five different targets and a different photospread for each target. Immediately after each target presentation, subjects worked for three minutes on an anagram distractor task. They then viewed a photospread that consisted of six faces presented on a single page in a booklet. They ranked the faces where one was most like the target and six was least like the target.

The photospreads were constructed in the following manner. Four of the five photospreads for each target consisted of the target, a prototype face, and four distractors. The target and each of the distractors differed from the prototype face by one feature and from each other by two features. The fifth photospread for each target did not contain the target but rather consisted of a prototype face and five distractors each differing from the prototype by one feature. Target exposure duration was also varied. For the four trials on which the target appeared in the photospread the durations were 17ms, 2s, 4s or 16s. A 17ms exposure of a face that had no features in common with the prototype or the distractors was used for trials where the target did not appear in the photospread. The orders of exposure durations and targets were balanced between subjects, and the photospreads associated with the different targets were varied across subjects.

Results

The mean rankings for the prototype and target faces are shown in Table 1. A lower ranking indicated the face was perceived as more like the face presented at exposure. As expected, increased exposure time produced better recognition of the target, $F(3,72)=4.78$, $p<.01$. Although the prototype faces in the photospreads received better rankings than the targets when exposure time was 17ms or 2s, there was no significant difference between the prototypes and targets, nor did this difference significantly interact with exposure time.

TABLE 1
Mean Ranking Scores for Target and Suspect Faces

Target Exposure Time	Target	Suspect
17 ms	Absent	2.52
17 ms	3.28	2.84
2 s	2.64	2.32
3 s	1.80	2.48
16 s	1.80	2.20

One result that speaks directly to the issue of primary interest in this paper, concerns the trials where the target did not appear in the photospread. The mean ranking of the prototype faces in the target-absent condition was significantly less than the 3.5 chance value, $t(24)=4.69$, $p<.001$.

EXPERIMENT 2

In this experiment subjects were exposed to a large number of facial photographs after which they examined a series of ten photospreads, each containing six faces. Their task was to rank each face in each of the photospreads as to the likelihood of it being a face they had seen in the earlier set. None of the photospread faces appeared in the earlier set. All photospread faces were constructed using Mac-a-Mug Pro software.

Method

Subjects were 25 undergraduate students from the University of Richmond. Faces presented in the exposure phase were 449 white male photographs that were taken from a 1972 University of Richmond yearbook. Photospread faces were constructed using a Macintosh Plus computer and the Mac-a-Mug Pro software. This software is a computer-assisted face construction program that uses a large file of digitized feature exemplars. Five features were manipulated: hair-head, eyes-eyebrows, nose, mouth-lips and chin.

Subjects were given 10 minutes to examine a set of 449 numbered photographs. Their task during this exposure period was to record the numbers of any faces that looked familiar and to rate the degree of familiarity on a three point scale. This phase of the experiment was merely a means of exposing subjects to a large number of faces so that they would believe that the unexpected photospread task that followed involved their memory for these faces. They then ranked each face in a series of 10 photospreads as to the likelihood that the face had appeared in the earlier sequence. Each photospread was contained on a page in a booklet.

The photospreads were formed by constructing 10 prototype faces by randomly selecting ten exemplars for each feature. Once these prototype faces were developed, five distractor faces for each were constructed such that each differed from the prototype by a single feature and from each other by two features. The substitute feature exemplars were also selected randomly. The faces in any one photospread were randomly ordered on that page of the booklet.

Results

A mean ranking for the prototype faces was computed by collapsing across the 10 photospreads and 25 subjects. The mean, 2.9, was significantly lower than the 3.5 chance value, $t(24)=4.79$, $p<.0001$. Additional analyses examined the effect for each of the 10 photospreads. Seven of the 10 prototypes had rankings significantly lower than expected by chance ($p<.05$). The other three prototypes, while not statistically significant, were in the expected direction.

EXPERIMENT 3

This experiment was similar to Experiment 2 except actual photographs were used to construct the photospreads.

Method

Subjects were 25 undergraduate students from Rice University. They were shown a series of 234 photographs of male faces presented on photo-album pages. This phase of the experiment was merely a means of exposing subjects to a large number of faces so that they would believe that the photospread task that followed involved their memory for these faces. The subjects then completed a half-hour distractor task. They subsequently ranked each face in a series of 10 photospreads as to the likelihood that the face had appeared in the earlier set. Each photospread consisted of five photographs mounted on a large page.

The photospreads were formed by first selecting photographs of 10 adult white males. All 10 were free of facial hair, glasses and any unique markings. Once these prototype faces were selected, four distractor faces for each were constructed such that each differed from the prototype by a single feature and from each other by two features. Four features were manipulated: hair, eyes-eyebrows, nose, and mouth. The substitute feature exemplars were also selected from a large set of male faces in the same age range. The substitutions were then made using a Minolta Montage Synthesizer and photographing the composite. The faces in the photospreads were randomly arranged. None of the photographs in the photospreads were in the original set of 234 photographs.

Results

A mean ranking for the prototype faces was computed by collapsing across the 10 photospreads for the 25 subjects. The overall mean, 1.88, was significantly lower than the 3.0 chance value, $t(24)=8.49$, $p< .0001$. Additional analyses examined the effect for each of the photospreads. Each of the 10 prototypes had rankings significantly lower than expected by chance ($p< .05$). The means ranged from 1.6 to 2.4.

Table 2 presents the mean rankings for the prototype faces and for each type of distractor face (hair different, eyes different, etc.) for Experiments 2 and 3. In Experiment 2 the nose, mouth, and chin distractors do not appear to differ from the expected chance value. The faces that differed by the hair or eyes were ranked higher (less likely). In experiment 3 hair and nose do not appear to differ from chance while eyes and mouth were ranked less likely. Generally, these results are consistent with findings on feature salience in that if salient features differ from the prototype, that distractor is less likely to be selected. An exception is the hair-different distractor in Experiment 3 which is probably due to the minor hair differences that were used.

TABLE 2

Mean Face Rankings	Mac-a-Mug (Exp.2)	Photograph (Exp.3)
Prototype	2.9	1.9
Hair	3.8	2.9
Eyes	3.9	3.7
Nose	3.3	3.0
Mouth	3.6	3.5
Chin	3.5	---
(Expected Values)	(3.5)	(3.0)

DISCUSSION

These experiments show rather clearly that in a facial recognition task where the target-absent test set consists of a face that is the prototype of the others, there will be a bias towards identifying that prototype face. Furthermore, the effect holds across a substantial number of faces, including faces that are artificial and real.

Questions remain, of course, about the type and degree of prototype-distractor differences. In these experiments the differences were defined in terms of feature manipulations, and the alternate features were selected to be discernable but not extreme. To date, however, no attempt has been made to scale these differences. Malpass and Hughes (1986) have recently explored some of the issues of forming facial prototypes and present a very good discussion of the subject.

In our view these findings have implications for the construction of lineups and photospreads in law enforcement settings. A high priority in such situations is not to bias the identification task towards the suspect, and suspect-distractor similarity is an important factor in this regard. The message from these experiments is that certain kinds of systematic suspect-distractor relationships may result in the suspect becoming a prototype which in turn could lead to bias.

REFERENCES

Malpass, R.S. and Devine, P.G. (1983). 'Measuring the Fairness of Eyewitness Identification Lineups', in Evaluating Witness Evidence (Eds. S.M.A. Lloyd-Bostock and B.R. Clifford), pp. 81-102, Wiley, Chichester.

Malpass, R.S. and Hughes, K.D. (1986). 'Formation of Facial Prototypes', in Aspects of Face Processing (Eds. H.D. Ellis, M.A. Jeeves, F. Newcombe and A. Young), pp. 154-162, Nijhoff, Dordrecht.

Solso, R.L. and McCarthy, J.E. (1981). 'Prototype Formation of Faces: A Case of Pseudo-memory', British Journal of Psychology, 72, 499-503.

Wells, G.L. (1978). 'Applied Eyewitness Testimony Research: System Variables and Estimator Variables'. Journal of Personality and Social Psychology, 36, 1546-1557.

Wogalter, M.S. and Jensen, D.G. (1986). 'Most Similar is Different: Response Bias in Lineups', Proceedings of the Human Factors Society 30th Annual Meeting, 30(1), 725-728.