MASSEY UNIVERSITY

Instructions:

(1) Please complete two of these forms, by providing title details and striking out one option from each of the sections 1-3.

(2) If you select option b of any section you must include a specific time period. The maximum permitted period is 24 months.

(3) Include the two forms with the copies of your thesis submitted to your supervisor.

(4) The Library must receive the original of your thesis.

(5) We strongly recommend that you provide the Library with two copies of your thesis if it has colour plates.

Massey University Library. Thesis Copyright Form

Title of thesis: The Effect of Photofit-Type Faces on Recognition Memory

✓ (1) (a) I give permission for my thesis to be made available to readers in the Massey University Library under conditions determined by the Librarian.

(b) I do not wish my thesis to be made available to readers without my written consent for ________ months.

✓ (2) (a) I agree that my thesis, or a copy, may be sent to another institution under conditions determined by the Librarian.

(b) I do not wish my thesis, or a copy, to be sent to another institution without my written consent for ________ months.

✓ (3) (a) I agree that my thesis may be copied for Library use.

(b) I do not wish my thesis to be copied for Library use for ________ months.

LaMontagne

Signed [signature]

Date 21/12/89

NAME AND ADDRESS DATE

THE EFFECT OF PHOTOFIT-TYPE FACES

ON RECOGNITION MEMORY

A thesis presented in partial fulfillment

of the requirement for the degree

of Master of Arts in Psychology

at Massey University

Hilary LaMontagne

1989

Dedicated to the memory of my

deceased father

Kirby Damian LaMontagne

27.9.33 to 29.9.87

ACKNOWLEDGEMENTS

First, I want to thank my mother, brothers and the rest of the family (including Robert) for their constant encouragement given to me from thousands of miles away.

I also want to thank my supervisor, John Podd, for his patience and help in guiding me through to the completion of this thesis. Special acknowledgement must be given for John's invaluable assistance with respect to the sections in the thesis concerning Signal Detection Theory.

ABSTRACT

Two attempts were made to replicate the results of Ellis, Davies, and Shepherd (1978) who showed that the addition of simulated photofit lines and randomly placed lines on photographs of faces caused a decrease in recognition memory for those faces. In the first experiment, three groups of subjects were shown 20 slides each of faces with no lines, photofit-type lines or random lines. Immediately afterwards they were shown the same faces mixed with 20 distractors, their task being to indicate whether a face had been previously seen. The addition of lines had no statistically significant effects on memory. In the second study, the number of faces initially shown was increased from 20 to 35 and subjects had to identify the previously seen faces from a set of 70 faces either immediately or following a three week delay. Again, the addition of lines to the faces produced no significant decrements in recognition rates, but there was a main effect for delay. However, trends seen in the recognition measures used for both studies suggested that the addition of lines may have a small effect on recognition memory but not enough to always reach statistical significance in single studies. The implications of the results for the use of the photofit-kit in recognition memory studies are discussed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Identification of faces is an important attribute in our society being absolutely essential for our day to day living. Misidentification can result in a variety of lost opportunities and in embarrassment when we fail to identify correctly friends, fellow workers and acquaintances.

The consequences of failing to recognise a person, or misidentifying a person, can be more serious than mere embarrassment. For example, face recognition and identification play a vital role in the criminal justice system both for law enforcement and the witnesses of crime. Also, traumatic family upheavals, or even the break-up of the family, may occur when a person suffers from prosopagnosia. This is a neurological disorder whereby the sufferer is unable to recognise familiar faces of friends, relatives, children, spouse, and in some cases, even the self (Bruyer, 1989). As Loftus (1979) points out, misidentification can result in the conviction of an innocent person (or the acquittal of a guilty person) in the criminal justice system, perhaps with the ensuing loss of freedom and social stigma attached to having been implicated in criminal activity. Therefore, research into how we recognise or identify a face has important practical implications.

There has been a wide variety of research on faces. Examples include research on recall (Davies, 1981, 1986), emotion (Salzen, 1981), developmental aspects (Carey, 1981), training (Malpass, 1981; Woodhead, Baddeley, and Simmonds, 1979), neuropsychology of face recognition (Benton, 1980; Hecaen,1981) and social factors in face recognition (Shepherd, 1981). The present study was limited to recognition studies done inside the laboratory.

TYPICAL RECOGNITION STUDY

A typical face recognition study is divided into two phases: the study (or inspection) phase and the recognition (or test) phase. In the study phase, subjects view a series of faces called target faces. The number of target faces has been varied between one (e.g., Davies, Ellis, and Shepherd, 1978a) and 100 (Light, Kayra-Stuart, and Hollander, 1979). However, the number of targets used is typically about 20 (Shepherd, 1983). Subjects may be told to try to memorise the targets because they have to recognise them later (e.g., Shepherd and Ellis, 1973), or this instruction may be omitted (Brigham, Maass, Snyder, and Spaulding, 1982). The inclusion or omission of such instructions does not seem to make a difference to the outcome of the study (Courtois and Mueller, 1981).

In the recognition phase, the target faces are randomly interspersed with new faces, called decoys or distractors. The subject's task is to choose which faces are old (previously seen) or new (not previously seen). The number of faces shown in the recognition phase varies from study to study. For example, Baddeley and Woodhead (1982) showed 50 faces at study and 50 at test whereas Davies, Shepherd, and Ellis (1979) showed their subjects 10 faces at study and 24 at recognition. While it is true to say that the number of faces shown in both the study and recognition phases varies widely, it is usually the case that the number of targets and distractors are kept equal.

The period between the study and recognition phase (delay period or retention interval) can be minutes, hours, days, or even months. For example, a retention interval of 10 minutes was used by Gehring, Toglia, and Kimble (1976), 48 hours by Chance, Goldstein, and McBride (1975), seven days by Chance and Goldstein (1987), five weeks by Shepherd and Ellis (1973), and 11 months in a study by Shepherd, Ellis, and Davies (1982). However, most studies use relatively short retention intervals of up to eight weeks (Shepherd, 1983).

Delay is used primarily to study the rate of decay of the facial engram over time. Short retention intervals of 20 minutes (Yarmey,1971) and 48 hours (Chance et al., 1975) showed little or no deterioration in facial memory whereas two week delay periods show mixed results. For example, Deffenbacher, Carr, and Leu (1981) reported no decline in face memory after two weeks whereas Podd (in press) found that the same delay did produce a small effect. Yarmey (1979) and Shepherd and Ellis (1973) noted a decrease in recognition accuracy after 30 and 35 days respectively and Egan, Pittner, and Goldstein (1977) found that hit rate remained the same after eight weeks but false alarms (reporting a face as a target when it is a distractor) increased markedly. Shepherd et al. (1982) found no significant decrease in recognition accuracy from one week to 90 days, but hit rate decreased after 11 months. The topic of delay will be further discussed later, but the few studies mentioned here serve to illustrate that recognition memory for faces under laboratory conditions seems remarkedly resilient to deterioration.

RECOGNITION MEASURES

Except for the early work on face recognition, most studies assess face recognition performance using measures derived from Signal Detection Theory (SDT: see Green and Swets, 1966; McNicol, 1972). The three measures most frequently used are hits (the proportion of "target" responses for faces previously seen in the study phase), false alarms (the proportion of "target" responses for faces not previously seen in the study phase), and d'. The d' measure is a criterion-free estimate of the subject's ability to discriminate 'old' (previously seen) from 'new' (not previously seen) faces. It is thus a relatively pure measure of recognition being able to vary independently of the effects of response bias. It is defined as the z-score of the false alarm rate minus the z-score of the hit rate (e.g., see Banks, 1970). Another criterion-free index of recognisability is A_g, the area under the Receiver Operating Characteristic (ROC) curve (e.g., see McNicol, 1972). A_g seems to have been reported in few studies of face recognition (but see Podd, in press). The d' index is a parametric estimator which assumes underlying normal-normal, equal variance distributions, whereas A_g is a non-parametric estimator whose validity is unaffected by the nature of the underlying distributions assumed by the recognition model.

It has become customary in the face research literature to report hits and false alarms as well as a recognition index which is based on both these measures (e.g., d', area under the ROC curve). This is attributable to researchers' interest in whether recognition accuracy is affected more by forgetting distractor faces (misses) or by calling distractor faces targets (false alarms). This practice may aid analysis but must be used with caution because hits (or false alarms) alone may change as a result of changes in response bias and may indicate no effect on recognition accuracy. Therefore, it is important to use criterion-free recognition indices derived from SDT rather than simply using hit rate or some other estimator which is confounded by response bias (Banks,1970; McNicol,1972).

VARIABLES STUDIED

It is not the author's intention to review all the variables studied in facial research using the recognition methods and measures previously discussed. However, some of the major variables studied will be briefly described in order to demonstrate the wide range of factors that seem to affect face recognition and to illustrate the versatility of the "target-distractor" method. Some of the variables studied are pose, target distinctiveness, similarity, levels of processing and memory load.

The majority of face recognition studies entail the presentation of full face views of photographs which have to be later identified in similar form. However, in real life, faces are observed from different positions. The effects of pose and pose change on facial recognition are unclear, a number of studies having produced inconsistent results. Davies, Ellis, and Shepherd (1978b) and Laughery, Alexander, and Lane (1971) found that there was no change in recognition performance when full and three-quarter face views were presented as targets and the views switched at recognition. Other research has shown that three-quarter pose produces the best performance followed by frontal pose and then profile (Baddeley and Woodhead, 1983; Krouse, 1981; Patterson and Baddeley, 1977).

Target distinctiveness has been investigated by varying the degree to which people report a face as 'unusual' or 'distinctive'. Distinctive faces are generally better recognised than their average counterparts. For instance, a study by Light et al. (1979) found that recognition memory was better for atypical (distinctive) faces than typical (or average) looking faces regardless of the encoding time (from three seconds to 15 seconds per face) and delay interval (from three hours to 24 hours). Similarly, Cohen and Carr (1975), Going and Read (1974), and Winograd (1981) found that distinctive faces led to higher recognition performances than non-distinctive faces. Therefore, atypicality should be taken into consideration when analysing results.

Similarity can be defined as the number of physical features appraised to be homogeneous across a group of people. The degree of similarity in facial features between targets and distractors affects recognition performance. For example, Davies et al. (1979) mixed target faces with faces rated according to their similarity by cluster analysis. Hit rates stayed the same but false alarms increased when the distractor faces were from the same cluster as the target faces. Similarly, Laughery, Fessler, Lenorovitz and Yoblick (1974) and Patterson and Baddeley (1977) found that the greater the target-distractor similarity, the poorer the recognition performance. Thus, similarity of targets and distractors should be taken into account when reporting results.

Craik and Lockhart (1972) postulated that the higher the degree of semantic interpretation placed on a stimulus (deep processing) the better the retention compared to dealing with the stimulus by itself (shallow processing). Bower and Karlin (1974) investigated the effects of levels of processing on facial recognition. Their subjects had to judge the gender (shallow processing) and sincerity (deep processing) of a series of faces shown to them. Performance was better for judgements of sincerity than for gender. Bower and Karlin's findings suggest that ascribing personal characteristics to faces rather than paying attention to the physiognomic features may lead to better recognition memory.

As Goldstein and Chance (1981) point out, some important variables have received very little attention. For example, it would be useful to know how recognition accuracy varied as a function of the number of target and distractor faces. However, until the study of Podd (in press) there was no systematic research on memory load. Podd used composite faces from the photofit to examine the effects of memory load and delay on facial recognition. Not unexpectedly, his results showed that recognition accuracy for an ensemble of target faces decreased with increasing load. Nonetheless, the decline in accuracy was largely the result of a decrease in hit rate, false alarm rate being relatively unaffected.

In summary, the results of a facial recognition study can be influenced by any of the above mentioned variables. Therefore, researchers should bear in mind the effects of these variables, amongst others, before coming to any conclusion about their results.

USE OF PHOTOFIT IN RECOGNITION STUDIES

A few studies of facial recognition have used composite faces, the faces usually being drawn from the photofit-kit (Sergent, 1984). The photofit-kit consists of five facial features: chin and cheeks, mouth, nose, eye and eyebrows, and hairline and ears. The large numbers of each separate feature allow for the construction of several thousand composite faces (e.g., see Davies, 1981). The different features are fitted on a special frame to form a face, individual features of which can be interchanged until the best possible resemblance is attained. Further elaboration of the face can be achieved by the addition of spectacles, beard or moustache, hats and sideburns (Davies, 1983). These can be drawn on a clear cover and fitted over the original image.

The photofit-kit has been developed as a tool principally for use by the police. The kit's large range of features allows for an enormous number of novel, composite faces to be generated. Thus, in principle, it seems an invaluable tool for researchers in laboratory studies of facial recognition because an easy source of several million faces is available each of which can be systematically varied, feature by feature.

The photofit-kit has been used mainly in studies of recall and reconstruction of faces (see Davies, 1981; Davies, 1986; Laughery, Rhodes, and Batten, 1981 for reviews). The kit has been used in only a few studies of facial recognition (Christie and Ellis, 1981; Ellis, Davies and Shepherd, 1978; Sergent, 1984; Podd, in press).

Christie and Ellis (1981) compared verbal descriptions of faces to photofit constructions. Subjects gave a verbal description of a target face seen for 60 seconds, then constructed the face from photofit. Independent judges recognised more of the verbal descriptions than the photofits.

Ellis et al. (1978) found that recognition was poorer for photofit faces than for "real" faces (that is, ordinary black and white photographs). Likewise, recognition was poorer for real faces which had lines drawn on them to simulate the boundaries of photofit.

In the previously described study by Podd (in press) using composite faces, it was found that there was a decrease in recognition with an increase in memory load. Podd used the photofit-kit rather than ordinary photographs so that in further related studies he could systematically vary facial features to control facial similarity amongst targets and distractors.

Sergent (1984) used the photofit because of the ease with which the separate features could be changed without drastically changing the whole face. She found that the face can be processed both wholistically and in a piecemeal manner.

The literature is replete with criticisms of the make-up and use of the photofit-kit. These include the limited number of features, especially the lack of features for the young and old, and the absence of the latest hairstyles (Venner, 1969). The composite faces also lack realism. This is probably because of the lines demarcating the five features and 'absence of skin texture and colour information normally present in pictures of faces' (Ellis et al., 1978, p.467).

The lack of life-like qualities in the composite faces make them appear very artificial. The artificiality of the faces may not tap the real processes underlying facial recognition. Thus, composite faces may produce results which are different from studies using real faces (Bruce, 1982, 1989; Sergent, 1984; Podd, in press). The initial concern over the use of composite (especially photofit) faces was raised by Ellis et al. (1978) who found differences in recognition accuracy for real and composite faces. This led Ellis (1981) to point out that using composite faces in laboratory studies of facial recognition may cause researchers to draw incorrect conclusions from their results. In view of the fact that several studies have used photofit faces to study variables believed to affect recognition memory for faces, it seems important to examine the Ellis et al. study in some detail.

ELLIS, DAVIES, and SHEPHERD (1978) STUDY

Ellis et al. (1978) conducted three experiments comparing recognition memory between (1) photographs of real faces and photofit faces; (2) real faces with faces with lines drawn on them simulating photofit boundaries; (3) normal faces, simulated photofit faces and faces with lines drawn randomly on them. These lines were of the same number (eight) and of the same length as those found in photofit.

In Experiment I, 60 subjects were randomly assigned to four groups. At study, they were shown 18 normal faces and 18 photofit constructions of normal faces. The recognition test followed five minutes later. They were shown 36 photographs consisting of 18 targets and 18 distractors. Each slide was shown for five seconds. The subjects who saw real faces at study saw real faces at test. Similarly, those who saw photofit faces at study saw the same at test. The subjects had to respond 'YES' for faces seen previously and 'NO' for faces not previously seen. Memory for real faces was significantly better than for photofit faces, though the differences were quite small.

A second experiment was conducted to study the effect that photofit-type lines have on recognition memory. The same 36 faces were used but instead of photofit faces, real faces were used with lines drawn on them to simulate the edges of the photofit features. Forty female subjects (20 in each group) were randomly assigned to view either real faces or simulated photofit faces. The subjects were individually shown each of 18 slides for five seconds each. Five minutes later they were shown the distractors (18) together with the targets. Therefore, the subjects saw 36 faces at recognition. The subjects responded 'YES' for the old faces (previously seen) and 'NO' for faces not previously seen.

Subjects performed significantly better on the unlined faces than the simulated photofit faces, suggesting that the lines which break the face into component features affect memory.

A third experiment was conducted to examine whether lines placed randomly on the face will have the same effect on memory as the lines that break the face into specific photofit-like features. Extra prints were made of the faces in the two previous experiments and random lines drawn on them (eight) of the same length found in photofit (see the method section of the first experiment of the present study for a full description of the procedure).

Eighty one subjects (27 in each group) were randomly assigned to one of three conditions: (1) real faces with no lines drawn on them; (2) simulated photofit faces; (3) randomly lined faces. In each condition the subjects were shown 18 targets in the study phase and 36 faces (18 targets, 18 distractors) in the recognition phase. Each face was again shown for five seconds. (It is worth noting that the sequence of presentation of target and distractor faces were the same in all three experiments.) The subjects again performed best on the unlined faces. There was no significant difference between performances for photofit faces and the randomly lined faces.

The implications of the results of Ellis et al. (1978) are far reaching. The photofit system has the potential to be a useful tool in laboratory studies of face recognition. The kit's large range of features allows for certain number of features to be interchanged with the rest of the face remaining the same. It also allows for the construction of faces that can be systematically varied for their degree of similarity. However, the results of Ellis et al. suggest that such use of the photofit-kit should be avoided.

Their results also have practical implications. Police have to memorise large numbers of faces in line with their work. The presence of lines in the photofit-kit may lead to inferior memory for faces, or even false identifications.

In summary, if the results of Ellis et al. (1978) are valid then it seems that photofit faces should not be used in recognition studies because the lines interfere with recognition and may force subjects to view the face in a piecemeal rather than a wholistic fashion. That is, photofit faces may cause the subjects to use an unnatural or less efficient coding strategy (Bruce, 1982). In addition, their results raise the interesting issue of what stage in the memory process is being affected by lines: the encoding, storage or retrieval stage.

In view of the practical and theoretical implications of the results of Ellis et al. (1978) it would seem important to assess the method they employed and their mode of analysis.

CRITIQUE OF THE ELLIS ET AL. (1978) STUDY

The first study where composite faces were constructed to the likeness of the real faces must be viewed with scepticism because it is known that composite faces never attain the desired likeness and life-like qualities of real faces (Ellis, Shepherd, and Davies, 1975), a point noted by Ellis et al. (1978). However, the researchers appeared to have used the first study merely to assess the size of the difference to be expected between photofit and real faces.

The method used for experiments II and III was in the standard recognition format (see earlier description) with 18 targets and 18 distractors. However, the small number of targets and distractors used may have created a problem. Despite calculating hit rate and false alarm rate, Ellis et al. (1978) did not use d' as the recognition measure. Instead they used a most unorthodox statistic, hits minus false alarms [this will be abbreviated as (H-FA)]. Presumably, this measure was used in place of d' because some subjects attained 100% hits and/or zero false alarms. Because of these ceiling and floor effects, the z-score of the hits and false alarm rates would yield a d' value of infinity, making it impossible to obtain average d' values for the experimental groups. A measure such as (H-FA) does not have the infinite range of d' and, therefore, the same problem does not arise.

However, the differences in recognition memory obtained in Experiments II and III for ordinary photographs and the photofit simulations may have been the result of using (H-FA) as the accuracy index. It can easily be shown that for a given hit rate (H-FA) and d' produce different graphs when plotted as a function of false alarms.

Figures 1(a) and 1(b) show d' (left ordinate) and (H-FA) (right ordinate) plotted as a function of the false alarm rate for hit rate values of 0.60 and 0.90 respectively.

FIGURE 1

(H-FA) (right ordinate) and d-prime (left ordinate) plotted as a function of the false alarm rate for fixed hit rates of 0.60 (Figure 1a) and 0.90 (Figure 1b). It can be seen that (H-FA) increases linearly but d-prime produces an accelerating function.

Note: increasing false alarm rate values on the abscissa are from right to left.

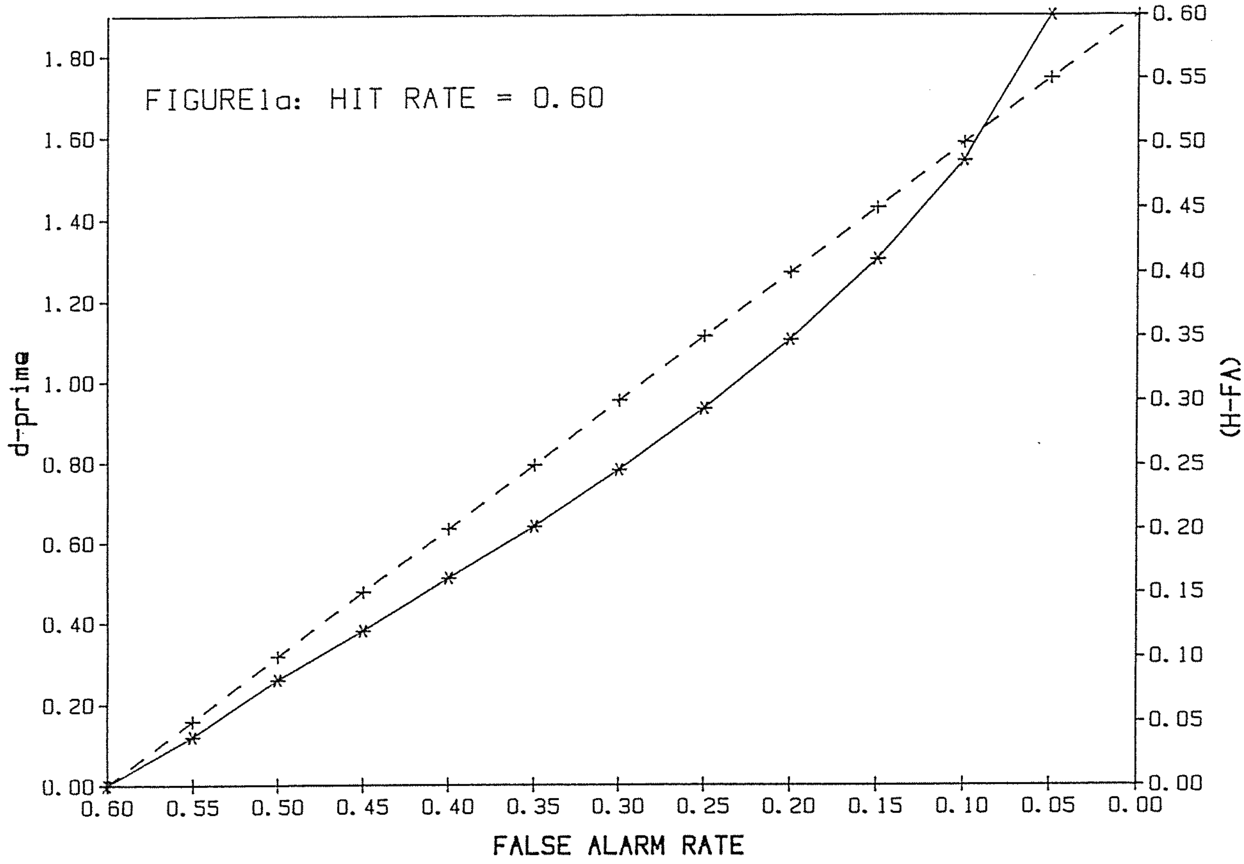


FIGURE1a: HIT RATE = 0.60
d-prime
(H-FA)
FALSE ALARM RATE
1.80
1.60
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
0.60
0.55
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00

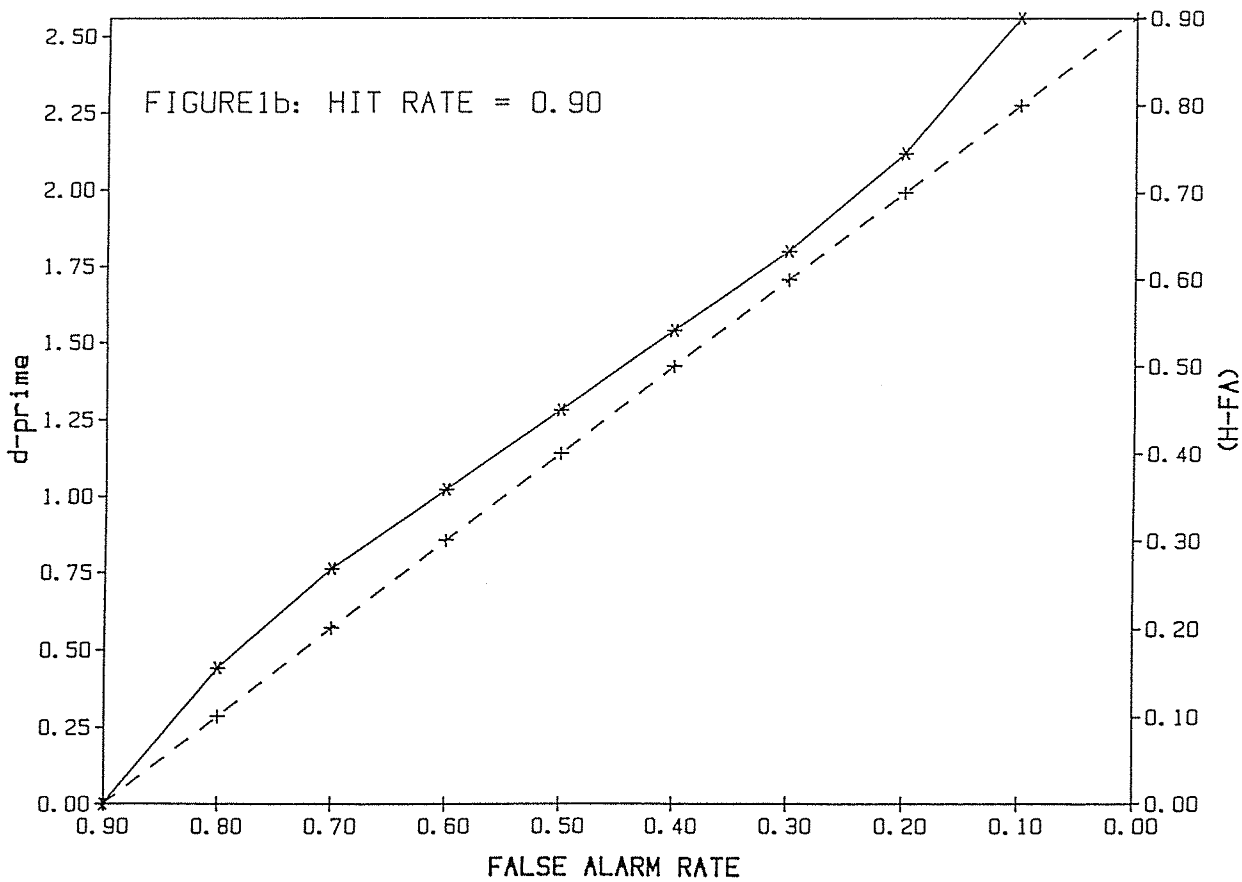


FIGURE1b: HIT RATE = 0.90
d-prime
(H-FA)
FALSE ALARM RATE
2.50
2.25
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10

The two functions differ in two quite obvious ways. First, (H-FA) is a linear function of false alarms while d' is not. Second, as a result of d' being an accelerating function of the false alarm rate, the rate of change in d' is greater at low false alarm rates than at high false alarm rates. This is not the case for (H-FA). It must be noted that (H-FA) is most definitely not an index derived from SDT whereas d' is. From a strict SDT view point, (H-FA) is an invalid measure because it can be affected by response bias whereas (at least in theory) d' cannot be.

Given the doubts about the validity of the Ellis et al. (1978) experiments created by the use of (H-FA), and the fact that their results effectively rule out the use of photofit-like faces in recognition studies, it seems most important to attempt a replication. The major aim of the present investigation, then, was to attempt to replicate the results obtained by Ellis et al.

If breaking up a face with lines affects recognition accuracy, then it is of interest to ask at what stage in the memory process the lines are having their effect. If the lines are affecting memory at the encoding stage, or while the faces are stored in memory, subjects who see lined faces at study and the same faces with no lines at test, should perform worse than subjects who see unlined faces at both test and recognition. However, if the lines are interfering with the retrieval process, then viewing the same lined and unlined faces should yield the reverse result. That is, subjects viewing unlined faces at test and lined faces at recognition should show poorer recognition rates than subjects seeing unlined faces in both the test and recognition phases. Although Ellis et al. (1978) suggest such a study, none appears to have been published at the present time. Therefore, a subsidiary aim of the current research was to investigate at what stage lines on faces seem to affect the memory process. This investigation was to be carried out as the second experiment following the replication attempt.

In summary, the main aim of the present research was to attempt to replicate the results of Ellis et al. (1978). Their methodology was closely adhered to, but because of the doubtful validity of their main recognition index, (H-FA), it was hoped to use more appropriate indices, such as d' and A_g.

METHOD

SUBJECTS

The subjects were 60 Caucasian first year students drawn from a pool of volunteers from various Faculties at Massey University and the local Teachers Training College. There were 42 females and 18 males with an age range of 17-20 years and a mean of 18 years. The subjects were randomly assigned to one of three groups with the constraint that each group consisted of the same proportion of sexes (fourteen females and six males).

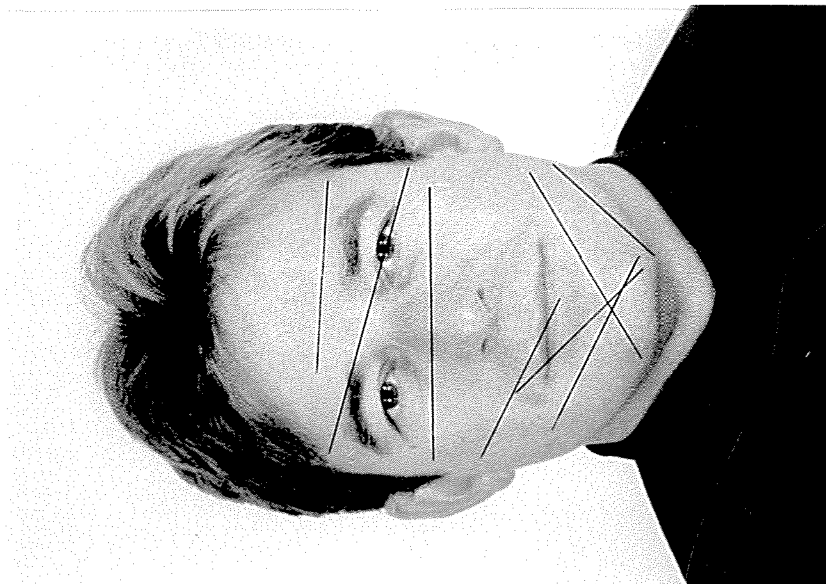
MATERIALS AND DESIGN

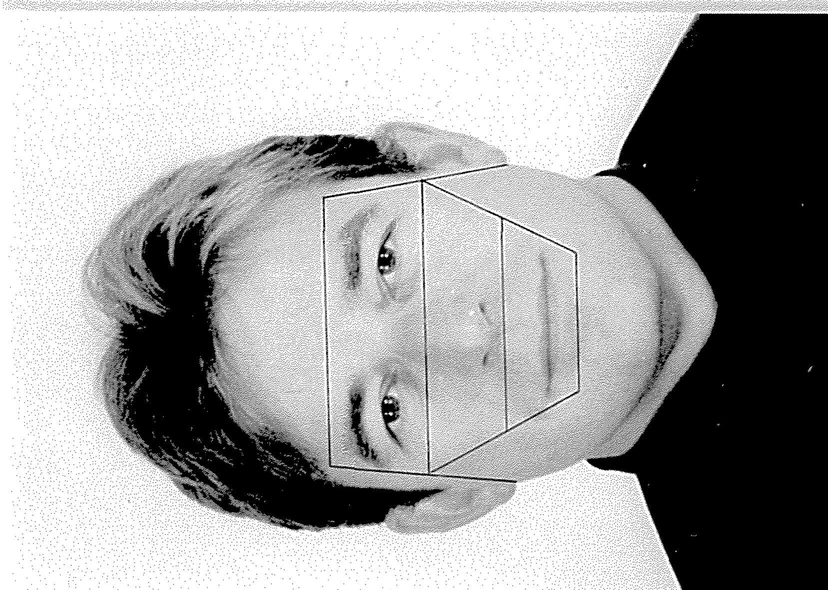
Stimuli consisted of 40 faces randomly selected from a catalogue of 100 male faces held in the library of faces in the Department of Psychology at Massey University. The catalogue consisted of full face photographs of males taken under standard studio conditions. A hairdresser's cape was used to eliminate differences in clothing. The age range of the males depicted in the photographs used for the present study was between 18 to 35 years. None of the faces had any unusual features, facial hair, or adornments such as ear-rings.

Three sets of 40 black and white prints (enlarged to 9 x 12.5 cm) were prepared from the original colour prints. For one set, eight lines (as found in photofit) were drawn on each face (in black ink) to simulate the boundaries of the photofit-kit (see Figure 2

FIGURE 2

Examples of the three types of facial stimuli used. The first photograph shows an unlined face, the second with photofit boundary lines, and the third with these boundary lines drawn randomly across the face.





and Ellis et al., 1978 for examples).

Eight lines of the same length and thickness as the photofit lines were positioned haphazardly on a second set of faces using the following procedure. An angle (to the nearest degree) was randomly selected along with a randomly selected point on the face. This point was used to determine the midpoint of any given line to be drawn on the face at the predetermined angle. Where an angle and its corresponding midpoint did not allow a line to fit completely on the face, they were discarded and a new angle and midpoint were randomly determined. In addition, if a line drawn appeared too close to an already existing line, the line was discarded and a new angle and midpoint determined. An example of the end result of this procedure can be seen in Figure 2 (see also Ellis et al., 1978).

No lines were drawn on the third set of photographs (see Figure 2). All three sets of photographs were then produced as slides. Each of the three groups of 20 subjects was shown a different set of 40 slides: no lines, photofit lines, or random lines.

Each stimulus face was assigned an arbitrary number (1-40) and 20 face numbers randomly selected. The associated faces were used as target faces for all three groups of subjects in the study phase. The 20 target faces and the remaining 20 (distractor) faces were randomly ordered and used in the recognition phase with the restriction that not more than three target or distractor faces could appear in succession. The same sequence of 40 faces was used for all 40 subjects. Therefore, the same faces were presented using the same random sequences in the study and recognition phases for all three experimental groups. The only between-group factor was the pattern of lines on the stimulus faces (no lines, photofit lines, and random lines). The same pattern of lines was used for each group in both the study and recognition phase. In other words, subjects who saw unlined faces in the study phase saw unlined faces in the recognition phase, and so on.

The slides were projected with a Kodak Carousel projector onto a white screen. The projector was equipped with an automatic timer which was set to display each face for five seconds with three seconds between each slide. The display time and interstimulus interval were kept constant for both the study and recognition phase. The distance between the subject and the screen was kept constant for both the study and recognition phase. The distance between the subject and the screen varied between 2.0 and 3.0 metres, according to seating position. The subjects used the same sitting position in both the study and recognition phases of the experiment. The faces were presented twice life-size in a dark room with sufficient light to read instructions and write on the response sheets (see Appendix 1)

PROCEDURE

The stimulus faces were presented to the subjects in small groups of three to six. This was to ensure that the viewing angle and distance from the screen was kept within reasonable bounds. For the study phase, subjects were given the following instructions:

> 'The purpose of this experiment is to see how well you can remember faces. Shortly I will show you a number of faces and your task in this part of the experiment is to look at the faces and do your best to remember them. You will be asked to identify these faces in a later stage of the experiment. Any questions?"

Immediately after the study phase, the subjects were handed the instructions for the recognition phase:

> "This part of the experiment will test how well you can recognise the faces you saw in the first part of the experiment. Faces you have already seen are

called OLD FACES. These are mixed in with faces you have not previously seen. These are called NEW FACES. Your task is to write on the response sheet "YES" (Y) if you think it is old (previously seen), or "NO" (N) if you think it is new (not previously seen). On each trial the presentation of an old or new face is equally likely; that is 50/50. Each face will be presented for five seconds followed by a three second interval during which you can record your response. The first five trials are for practice purposes only. These trials with blank slides will give you an idea of how long each slide will be presented and the time you have to respond."

Prior to the recognition phase, subjects were given five practice trials to give them some idea of the viewing time for each slide and the time available to make a response. The practice trials were presented by running the projector with no slides. (The screen was dark for three seconds and illuminated by the projector lamp for five seconds). In principle, the subjects had eight seconds to make a decision (viewing time plus the interstimulus interval); however, subjects were encouraged to use the full five seconds to study the slide and use the three second interstimulus interval to respond. The reading of instructions and the presentation of the five practice trials took between four and five minutes. Therefore, the interval between the end of the study phase and the beginning of the recognition phase was standardised at five minutes.

After the completion of the recognition phase, subjects were debriefed and given the opportunity to return at a later date to obtain feedback on their individual results and on the overall study. The subjects were also asked if they knew any of the people whose faces had been used as stimuli. Five subjects thought they knew one or more of the faces they had seen. The data of these subjects were discarded and replacement subjects found.

RESULTS

Hits (saying 'target' when a target face was presented) and false alarms (saying 'target' when a distracter face was presented) were totalled across all 40 trials for each subject. The measure (H-FA) (hits minus false alarms), used by Ellis et al. (1978) was also calculated for each subject. Means and standard deviations for each condition and for each of these measures are presented in Table 1.

Table I

Means and standard deviations for hits, false alarms,

(H-FA) and A' (see text for details).

Condition	Hits	False Alarms	H-FA	A'
No lines				
M	0.83	0.13	0.69	0.90
SD	0.10	0.11	0.14	0.05
Photofit lines				
M	0.78	0.10	0.68	0.91
SD	0.12	0.07	0.12	0.04
Random lines				
M	0.77	0.12	0.65	0.89
SD	0.09	0.11	0.14	0.05

Note: Due to rounding errors, (H-FA) may not equal the average hits minus false alarm values given in the table.

It was hoped that the index d' could be used as a recognition index. However, eight subjects obtained either 100% hits or zero false alarms or both. The index, d' (which can be defined as the z-score of the false alarm rate minus the z-score of the hit rate), could not be calculated for these subjects. Therefore, an alternative (non-parametric) index, A' (Grier, 1971), was used. A' is an estimate of the area under the ROC curve, but it is based on a single set of hit and false alarm values. It is defined as:

$$A' = \frac{1}{2} \quad \frac{(y - x)(1 + y - x)}{4y(1 - x)}$$

where x = probability of a false alarm

and y = probability of a hit

A' was calculated for each subject and the group means and standard deviations are shown in Table 1. (A complete set of data for each subject for all dependent measures is given in Appendix 2a to 2c).

It can be noted from Table 1 that there is a 5% decrease in hits from no lines to photofit lines and a 6% decrease from no lines to random lines. There was also a 1% drop in the recognition index (H-FA) between no lines and photofit lines and a 4% decrease between no lines and random lines. These changes are quite small but parallel those obtained by Ellis et al. (1978). However, the addition of lines on the photographs had little or no effect on A', a more appropriate index of recognition accuracy.

In order to assess the statistical significance of the mean differences across the three conditions, a one-way analysis of variance was carried out on each of the measures shown in Table 1. The results of these are shown in Table 2. The addition of lines to the photographs had no effects on hits $F(2, 57) = 1.91$, $p = 0.16$ (see Table 2a). Similarly there were no significant effects for false alarms: $F(2, 57) = 0.65$, $p = 0.53$ (Table 2b), (H-FA): $F(2, 57) = 0.64$, $p = 0.53$ (Table 2c) or A': $F(2, 57) = 0.49$, $p = 0.62$ (Table 2d).

Table 2

Summary tables of one-way ANOVA for (A) Hits;

(B) False alarms; (C) (H-FA) and (D) A'

A

Source	SS	df	MS	F	F-Prob
A (Hits)	0.0413	2	0.0207	1.91	0.16
S/A (Error)	0.6155	57	0.0108		
Total	0.6568	59			

Table 2 continued over

B

Source	SS	df	MS	F	F-Prob
A (False Alarms)	0.0126	2	0.0063	0.65	0.53
S/A (Error)	0.5507	57	0.0097		
Total	0.5633	59			

C

Source	SS	df	MS	F	F-Prob
A (H-FA)	0.0236	2	0.0118	0.64	0.53
S/A (Error)	1.0432	57	0.0183		
Total	1.0668	59			

D

Source	SS	df	MS	F	F-Prob
A (A')	0.0022	2	0.0011	0.49	0.62
S/A (Error)	0.1307	57	0.0023		
Total	0.1329	59			

As previously discussed in the introduction, the use of hit rate alone as a measure of recognition accuracy must be done with caution. The present results show that hit rate decreased by about 5% or 6% with the addition of lines to the photographs but the recognition accuracy index, A', did not change. This suggests that hit rate may have changed partly as a result of a change in response bias. To check this, a nonparametric estimate of response bias suggested by Grier (1971) was used. Grier defines bias as:

$$B'' = \frac{y\,(1 - y) - x\,(1 - x)}{y\,(1 - y) + x\,(1 - x)}$$

where x = probability of a false alarm

and y = probability of a hit

B" can take values between -1 and +1. In the present context, values of -1 and +1 represent extreme biases towards responding to a face as old or new respectively, and a value of zero represents no bias. B" was calculated for each subject and means and standard deviations were obtained for each condition. These are shown in Table 3.

Table 3

Means and standard deviations for B"

Condition	Means	Standard deviations
No lines	0.18	0.40
Photofit lines	0.28	0.50
Random lines	0.37	0.37

As can be seen in Table 3, B" increased with the addition of lines to the faces paralleling the changes in hit rate. While the changes in B" were not significant, $F(2, 57) = 1.05$, $p = 0.36$, they suggest that the small changes in hit rate can be partly accounted for by a shift in response bias. That is, lines on the faces induced subjects to make slightly more 'new' responses. This change in bias can be seen directly in the average percentage of 'new' responses for each condition (percent new responses for no lines = 52%, 56% for photofit lines, and 57% for random lines).

In summary, although hits and (H-FA) show a similar trend to the results obtained by Ellis et al. (1978), the addition of both photofit and random lines to the photographs produced no statistically significant effects on any of the measures. An analysis of response bias changes across the three conditions suggest that a small change in bias may have brought about the equally small changes in hit rate, and possibly in (H-FA).

DISCUSSION

The present study failed to replicate the results of Ellis et al. (1978). While there was a decrease in hits and (H-FA) with the addition of lines to the faces, the changes were not statistically significant. Moreover, an analysis of response bias suggested that the changes in hit rate [and possibly (H-FA)] were partly due to a change in response bias rather than recognition accuracy. This view is supported by the stability of A' across all three conditions.

Unfortunately, Ellis et al. (1978) did not report response bias data. Nevertheless, the present results suggest that the measure of recognition accuracy, (H-FA), may be affected by changes in response bias as well as changes in recognition accuracy. For example, a change in bias toward reporting more faces as new decreases the proportion of hits more than the proportion of false alarms, assuming normal-normal equal variance underlying distributions for the recognition model (see Banks, 1970). Therefore, the measure (H-FA) is not really appropriate for assessing pure recognition accuracy since accuracy and bias may be confounded. However, Ellis et al. showed that there was a concomitant increase in false alarms with the decrease in hits. This suggests that the addition of lines did bring about some decrease in recognition accuracy. However, it is not possible to assess the effects of response bias from their data. All that can be said is that the apparent small change in recognition accuracy as assessed by (H-FA) in the present study is most conservatively accounted for partly in terms of criterion change as well as recognition accuracy.

The present results are somewhat surprising in light of the fact that every attempt was made to carefully replicate the study of Ellis et al. (1978). The same procedure was used for drawing lines on the faces, the display time for the faces was the same (five seconds; Ellis et al. did not mention the interstimulus interval), the same 'YES-NO' response option was used and the number of faces was approximately the same (18 vs. 20).

The present study used 20 subjects in each group; similarly Ellis et al. (1978) used 20 subjects in study II and 27 subjects in study III. Thus, the statistical power in the present study and that of Ellis et al. was approximately the same.

The viewing angle and distance from the screen was controlled across the three conditions by presenting the stimuli to subjects in groups of three to six. However, it is not clear whether Ellis et al. (1978) were careful in controlling viewing angle and distance. This is an important point because long viewing distances may have detrimental effects on recognition, especially for lined faces. If, for example, subjects in the recognition phase sat further away from the screen than in the study phase, then the change in viewing distance may have affected performance.

Failure to replicate the results of Ellis et al. (1978) meant that the original intention of this investigation had to be changed. It will be recalled that if lines on the faces were shown to affect recognition accuracy then it would be of great interest to ascertain whether these lines affected either the encoding stage or the retrieval stage of the memory process. The failure to replicate the study of Ellis et al. made the proposed study rather meaningless.

Nonetheless, the Ellis et al. (1978) investigation is a most important one, not only for its theoretical implications with regard to how lines on faces affect the encoding or storage of faces, but also for methodological reasons. Their studies suggest that composite (photofit) faces should not be used in laboratory studies in place of ordinary photographs because recognition rates are lower for the composite faces. Thus, it is important to show that the results can stand up to replication. Therefore, it was decided to attempt a second replication of their study.

EXPERIMENT II

Every attempt was made in Experiment I to keep conditions the same as in Ellis et al. (1978) to allow replication a maximum chance of success. However, due to ceiling (100% hits) and floor (zero false alarms) effects, the key recognition accuracy index, d', could not be computed for a number of subjects. Thus, it was decided to increase the difficulty of the task confronting the subjects for the second replication attempt. Podd (in press) showed that this could be readily achieved by increasing the number of stimulus faces shown in both the study and recognition phases. Accordingly, the number of stimulus faces was increased at study from 20 to 35 and at recognition phase from 40 to 70. The resulting lower hit rates and higher false alarm rates allowed d' to be calculated for all subjects.

The recognition index, A', based on one pair of hit and false alarm rates (Grier, 1971) was used in Experiment I. Once again, it was decided to obtain data to calculate the area under the ROC curve for the second study. This was to ensure that a second measure of recognition accuracy was available should d' be unobtainable for some subjects as before. However, subjects completed a rating scale task (very sure old , fairly sure old, fairly sure new, very sure new) for Experiment II rather than the YES/NO task used in Experiment I. This allowed the area under the ROC curve, A_g (see McNicol, 1972), to be calculated on the basis of several pairs of hits and false alarms. Podd (in press) appears to be the only researcher to have used A_g as a measure of recognition accuracy in face studies. This is rather surprising considering that A_g is a measure that enables average recognition rates to be calculated irrespective of the obtained hit and false alarm rates. A_g can range from 50% (chance performance) to 100% (perfect recognition) whereas d' values range from zero to infinity. The finite range of A_g renders it a very useful measure for obtaining group averages when subjects are likely to produce perfect hit rates or zero false alarm rates. Podd showed that A_g produced essentially the same results as d' for both the effects of memory load and delay.

Ellis et al. (1978) obtained their results by testing almost immediately after the study phase. However, the effects of delay (the interval between the study and recognition phases) on lined faces have not been reported in the literature. In fact, there has been little systematic research on the effects of delay on facial recognition in general, leading Goldstein and Chance (1981) to lament the lack of information on the effects of different delay intervals on facial memory. One aim of the second experiment was to ascertain whether reasonably long delays interfered with the recognition of lined faces more than unlined faces. It is not the author's intention to review all studies on delay. However, some of the major studies are reviewed to provide a background for the second study and to help determine the delay interval required to produce a reliable effect on recognition accuracy.

In general, short delays of up to two days have no statistically significant effect on facial recognition rates. For example, Yarmey (1971) showed 80 pictures of upright vs. rotated pictures of well known faces, unknown faces, canine faces and architecture. Recognition performance was superior for upright and familiar faces after immediate testing. There was no decrease in recognition accuracy after a 20-minute retention interval.

Brown, Deffenbacher, and Sturgil (1977) showed their subjects 50 photographs of faces in slide form in two different rooms. There was a two hour break after the first 25 photographs before the subjects viewed the remaining 25. For the recognition phase, they were tested in a third room with a 100 photographs shown in pairs (one new, one old) after a two day delay. Correct identification of faces was above 90% (but recall of rooms was poor) again showing little forgetting.

Chance et al. (1975) had 144 Caucasian subjects view 14 target faces in slide form of either White, Black or Japanese faces. Half were tested immediately after study, and the other half after two days. The 14 target faces were mixed with 70 distractor faces (White, Black and Japanese). Recognition was best for white faces followed by black faces. There was no difference in performance after a two day delay.

Goldstein and Chance (1970) showed their subjects 14 targets including inkblots, snow crystals and faces in slide form for two to three seconds with five to eight seconds between slides. For the recognition phase, 70 distractors of each stimulus were mixed with the 14 targets. There was no difference in performance for faces between immediate testing and a two day retention interval.

Krouse (1981) investigated the effects of different poses on facial recognition performance. In her study subjects saw slides of eight full face pose and eight three-quarter face pose photographs. They were shown for four seconds each, with a one second interstimulus interval. Some photographs were shown in the same pose or a different pose immediately after study and after a two and three day delay. Performance was better when the pose was the same at test and at study with recognition performance being best when the stimuli were three-quarter pose rather than full face pose. Recognition performance decreased after the two and three day retention interval compared to immediate testing.

Overall, most face recognition studies with up to about a two day delay have little or no effect on recognition rates. However, the Krouse (1981) study shows that there are some exceptions.

Interestingly, even longer periods of delay do not by any means consistently produce a drop in recognition accuracy. Laughery et al. (1974) had their subjects view a film portraying a commercial exchange in a shop. Later, they were asked to recognise the target in slide form interspersed amongst 149 decoys from four minutes to one week later. There was no effect on memory performance for any of the delay periods.

Shepherd and Ellis (1973) studied the effect of attractiveness on recognition accuracy. Their subjects viewed 27 slides of female faces (independently rated for attractiveness) at study. Each slide was seen for three seconds with a one second interstimulus interval. The subjects were tested immediately after study, and after one day, six days and 35 days delay. There was no decrease in recognition performance for high and low attractive faces for all three retention intervals. However, there was a decline in performance for the average looking faces after a delay of 35 days. Shepherd (1981) points out that there was no deterioration at one day and six days perhaps because the subjects only had three faces in each group to remember. The lack of difficulty may have been responsible for the effects of delay occurring only at 35 days. Also, Shepherd states that the decrease in performance seen after 35 days may have been due to a 'ceiling effect' caused by the small number of faces used in each group.

Deffenbacher et al. (1981) showed their subjects slides of faces, pictures and words for five seconds with a one second interstimulus interval at study. There was a decrease in recognition after an interval of two weeks for pictures and words but not for faces.

In the previously mentioned study of Podd (in press), memory load and delay were varied systematically to study their effects on facial recognition. Podd showed his subjects 20, 35, and 50 slides of faces at study and 40, 70, and 100 at recognition. They were tested at retention intervals of 10 minutes, one week and two weeks. There was a small, marginally significant effect ($p < 0.08$) on d' with delay.

Other research has found modest decreases in facial recognition with delay. For example, Yarmey (1979) had 126 male and female subjects view 30 male and female slides of faces independently rated for their physical attractiveness, distinctiveness, and likeability. Each slide was shown for two seconds. The subjects had to memorise the faces for later recognition and to make a decision on their level of attractiveness, distinctiveness and likeability. The subjects were tested in groups of 42 immediately after study, one week later and one month later. Identification performance was superior for faces rated high and low for attractiveness, liking and distinctiveness immediately after study compared to the average rated faces on the three measures. There was a marked decrease in identification performance on all three measures after the two delay periods with the average rated faces affected the most.

Courtois and Mueller (1981) studied the effect of target and distractor typicality on facial recognition. Subjects viewed five typical and five atypical slides of faces (as assessed by independent judges) at study for 15 seconds each. At study, a typical target was paired with three typical distractors and three atypical distractors. Similarly, an atypical target was paired with three atypical distractors and three typical distractors. The subjects were tested immediately after study, after two days and after 28 days. Recognition performance deteriorated with increased delay. Performance was better with an atypical target-atypical distractor pairing compared to the typical target-typical distractor pairing at all three retention intervals.

A study by Chance and Goldstein(1987) compared the effects of three delay intervals on recognition accuracy for Caucasian and Japanese faces. In addition to the usual recognition measures, the researchers measured the response times of the subjects. Chance and Goldstein argued that response times would be more sensitive to any decrement in facial memory over time than hits and false alarms. The subjects were tested immediately after study, two days and seven days later. There was no significant difference in the number of correct identifications (hits) across the three retention intervals. However, there was a small increase in false alarms. Response times to make correct identifications after two days and seven days were longer compared to immediate testing. Recognition accuracy (increased hits and decreased false alarms) was better for Caucasian than for Japanese faces. Likewise, response times were longer for Japanese faces than Caucasian faces.

Taken together, studies of retention interval suggest that facial memory does not automatically deteriorate with delay. The discrepancy in results across the studies may be due to differences in encoding time of faces at study and recognition, differences in instructions to subjects, uniqueness of targets and distractors, number of targets and distractors and type of stimuli used (slides, photographs, or live line-ups).

Overall, studies which have varied the retention interval up to two weeks have generally shown only small effects on recognition accuracy. However, a meta-analysis of over 190 studies conducted by Shapiro and Penrod (1986) show that delay (averaging 4.5 weeks) has a highly significant effect on recognition accuracy. Similarly, Deffenbacher (1986) found that delay had a definite effect across 33 studies in another meta-analysis study. Thus, delay has a strong effect when several studies are taken together. However, the present review seem to show that delay does not always produce statistically reliable effects with individual studies, even with delays of up to two weeks.

Ellis et al. (1978) demonstrated that lines cause a deterioration in facial memory and speculated that this may be due to fragmenting the facial image thereby causing disruption in the encoding, storage or retrieval stage. Therefore, it is possible that delay will cause a bigger decrease in recognition rates for lined than unlined faces. That is, the effect of fragmentation may be to cause higher rates of forgetting for the lined faces. If this is so, then one would predict an interaction between delay and facial lines.

The studies reviewed on the effects of delay suggest that a reliable effect should occur after about three weeks. To be certain of producing a delay effect, it would have been better to use a retention interval of four to five weeks, but the practical problems of getting the subjects to return after long delays made it necessary to choose three weeks as the delay period for the present study.

In summary, the major aim of Experiment II was to attempt a second replication of the work of Ellis et al. (1978). The number of stimulus faces was increased in the study and recognition phases to negate ceiling and floor effects, and a rating scale task was used in place of a YES-NO task so that A_g could be calculated. A subsidiary aim was to study the effects of delay on recognition rates for lined faces. If the lines fragment the image of the face, as suggested by Ellis et al., it may be that more forgetting occurs with delay on lined faces compared to unlined faces.

METHOD

SUBJECTS

Ninety subjects (54 females and 36 males), aged 18-30 years (mean 19.8 years) were drawn from the same pool of volunteers used for Experiment I. These subjects were allocated at random to one of six groups with the restriction that each group consisted of the same ratio of sexes (nine females and six males).

MATERIALS AND DESIGN

Thirty new stimulus faces were randomly selected from the library of faces. Three sets of black and white prints were made of each colour print, and lines (photofit and random) were drawn on two sets of the extra prints as described for Experiment I. A third set was left free of lines. Slides were then produced from these extra photographs and added to those used in Experiment I. Therefore, in total, there were three sets of 70 stimulus faces consisting of no lines, photofit lines, and random lines.

The stimulus faces were randomly divided into 35 target and 35 distractor faces. The same procedure as for Experiment I was used in the construction of the sequences of slides shown in the study and recognition phases to the groups in the different conditions (with the restraint that not more than three targets or distractors occurred in sequence in the recognition phase).

Taking into consideration the 3-week delay, the design was a 3 (no lines, photofit lines, and random lines) x 2 (zero and three weeks delay) between-groups factorial design with an equal number of subjects (15) assigned to each group.

PROCEDURE

In general, the procedure was the same as for Experiment I; the main exception was that a rating task was used in place of the YES-NO task. Therefore, the instructions given in the recognition phase was modified as follows:

> "This part of the experiment will test how well you can recognise the faces you saw in the first part of the experiment. Faces you have already seen are called OLD FACES. These are mixed in with faces you have not previously seen. These are called NEW FACES. Your task is to rate how certain you are that each face is old (previously seen) or new (not previously seen). Try to use all the rating categories available- very sure old (1), fairly sure old (2), fairly sure new (3), and very sure new (4). On each trial the presentation of an old or new face is equally likely; that is 50/50. Each face will be presented for five seconds followed by a three second interval during which you can record your response. The first five trials are for practice purposes only. These trials with blank slides will give you an idea of how long each slide will be presented and the time you have to respond."

The response sheets consisted of a set of 70 four-point rating scales (20 per page) with the verbal descriptors for each of the rating values at the top and bottom of each page (see Appendix 3).

In all other respects the procedure was identical to that used in the first experiment.

RESULTS

Hits, defined as the number of old responses (ratings 1 and 2 combined) for old faces, and false alarms, defined as the number of old responses (ratings 1 and 2 combined) for new faces were calculated across all 70 trials for each subject. Hits minus false alarms, (H-FA), the recognition measure used by Ellis et al. (1978), was also calculated for each subject along with the index d' (defined as the z-score of the false alarm rate minus the z-score of the hit rate). A_g was ascertained for each subject by plotting the cumulative hit rate values as a function of the cumulative false rate values obtained from the rating scale data (e.g., see McNicol, 1972).

Means and standard deviations for all measures for each condition are presented in Table 4. (Individual data for each subject for all measures can be found in Appendix 4a to f).

As may be seen from Table 4, for immediate testing there was a 6% drop in hits from no lines to photofit lines and a 7% decrease from no lines to random lines. The addition of lines on the photographs had no observable effect on false alarms across all three conditions. As a result of the change in hit rate, there was a 6% drop in (H-FA) from no lines to random lines. Also, there was a small reduction in d' from unlined to lined faces. Similarly, A_g showed equivalent though small decreases for the different stimuli.

Table 4

Means and standard deviations for hits, false alarms, (H-FA), d' and A_g tested immediately and following three weeks delay.

Condition	Hits	False Alarms	H-FA	d'	A_g
Immediate testing					
No lines					
M	0.80	0.18	0.62	1.88	0.87
SD	0.08	0.11	0.12	0.47	0.06
Photofit lines					
M	0.74	0.17	0.56	1.65	0.83
SD	0.11	0.07	0.12	0.43	0.06
Random Lines					
M	0.73	0.18	0.55	1.65	0.82
SD	0.13	0.10	0.17	0.65	0.08
3 week delay					
No lines					
M	0.69	0.24	0.45	1.28	0.79
SD	0.13	0.11	0.13	0.40	0.07
Photofit Lines					
M	0.68	0.25	0.44	1.22	0.77
SD	0.08	0.11	0.14	0.47	0.07
Random lines					
M	0.65	0.22	0.44	1.29	0.76
SD	016	0.11	0.18	0.60	0.09

Note: Due to rounding errors, (H-FA) may not equal the average hits minus false alarm values given in the table.

Following the three week retention interval, the results in Table 4 suggest that the addition of lines to the faces had only very small effects on hits and A_g.

A 2 (delay) x 3 (lines) analysis of variance was performed on each of the measures to determine the statistical significance of the mean differences across the three conditions shown in Table 4. These results are shown in Table 5.

The addition of lines on the photographs had no significant effect on hits $F(2, 89) = 1.54$, $p = 0.22$ (see Table 5A). Similarly, there were no significant effects on false alarms: $F(2,89) = 0.16$, $p = 0.86$ (Table 5b), (H-FA): $F(2,89) = 0.67$, $p = 0.52$ (Table 5c), d': $F(2,89) = 0.66$, $p = 0.52$ (Table 5d) or $A_g = F(2,89) = 1.91$, $p = 0.16$ (Table 5d).

Thus, while hit rate and the recognition measures (H-FA), d', and A_g in the immediate testing conditions (no delay) showed similar trends to the results obtained by Ellis et al. (1978), the addition of both photofit and random lines to the photographs produced no statistically significant effects on recognition accuracy.

<u>DELAY</u>

Table 4 shows that delay had a marked effect on hit and false alarm rates and on the recognition accuracy measures. Collapsed across the three different face stimuli, hits decreased by 9% and false alarms increased by 6%. The recognition indices, d' and A_g, decreased over delay from 1.73 to 1.26 and 0.84 to 0.77 respectively.

Separate two-way analyses of variance conducted on each of the measures shown in Table 4 were performed to assess the statistical significance of the mean differences across delay. The results of these are shown in Table 5.

Delay had a significant effect on hits, $F(1,89) = 9.81$, $p = 0.002$ (see Table 5a) and false alarms, $F(1,89) = 6.73$, $p = 0.01$ (Table 5b). Similarly, delay had highly significant effects on (H-FA): $F(1,89) = 19.16$, $p < 0.0001$ (Table 5c), d': $F(1,89) = 18.15$, $p < 0.0001$ (Table 5d) or A_g: $F(1,89) = 19.16$, $p < 0.0001$ (Table 5e).

There was no interaction between lines and delay on any of the five measures (all F values < 1; see Tables 5a to e).

Table 5

Summary tables of two-way ANOVAS for (A) Hits;
(B) false alarms; (C) (H-FA); (D) d', and (E) A_g

A

Source	SS	df	MS	F	F-Prob
A (lines)	0.04335	2	0.02167	1.54	0.22
B (delay)	0.13845	1	0.13845	9.81	0.002
AB	0.01042	2	0.00521	0.37	0.69
S/AB (Error)	1.18521	84	0.01411		
Total	1.37743	89	0.01548		

Table 5 continued

B

Source	SS	df	MS	F	F-Prob
A (lines)	0.00322	2	0.00161	0.16	0.86
B (delay)	0.06944	1	0.06944	6.73	0.01
AB	0.00610	2	0.00305	0.30	0.75
S/AB (Error)	0.86669	84	0.01032		
Total	0.94545	89	0.01062		

C

Source	SS	df	MS	F	F-Prob
A (lines)	0.02820	2	0.01410	0.67	0.52
B (delay)	0.40401	1	0.40401	19.16	0.00001
AB	0.01233	2	0.00616	0.29	0.75
S/AB (Error)	1.77133	84	0.02109		
Total	2.21587	89	0.02490		

Table 5 continued

D

Source	SS	df	MS	F	F-Prob
A (lines)	0.34300	2	0.17150	0.66	0.52
B (delay)	4.70596	1	4.70596	18.15	0.0001
AB	0.23505	2	0.11752	0.45	0.64
S/AB (Error)	21.78051	84	0.25929		
Total	27.06452	89			

E

Source	SS	df	MS	F	F-Prob
A (lines)	0.02056	2	0.01028	1.91	0.16
B (delay)	0.10336	1	0.10336	19.16	0.00001
AB	0.00198	2	0.00099	0.18	0.83
S/AB (Error)	0.45324	57	0.00540		
Total	0.57914	89	0.00651		

DISCUSSION

The main purpose of Experiment II was another replication attempt of the results of Ellis et al. (1978). Once again, the attempt was unsuccessful. Although some trends in the results were similar to those obtained by Ellis et al. (who showed that recognition memory was adversely affected by lined faces), the differences were not sufficient to produce statistically significant effects. The addition of photofit and random lines to facial stimuli appear not to cause a deterioration in recognition accuracy.

As mentioned previously for Experiment I, response bias seemed to have some influence on the results, but because of the rating procedure used, it was not possible to ascertain whether it played a part in the results of Experiment II. When using a rating scale, subjects are assumed to be holding multiple decision criteria, one for each rating category. However, a rough estimate of the effect of response bias can be made by collapsing ratings 1 and 2 (very sure old, fairly sure old) and estimated B", as for Experiment I. This procedure yielded mean B" values of 0.09 at both zero delay and three week delay for the no lined condition, 0.12 at zero delay and 0.11 at three weeks delay for the photofit-type condition, and 0.14 at zero delay and 0.15 at three weeks delay for the random lined condition. Thus, as for the first experiment, the trend (although not statistically significant: $F < 1$) was for a slightly stricter decision criterion (more "new" responses) for lined faces. Once again, the possibility of a small change in response bias needs to be taken into account when evaluating the changes in hits and (H-FA) in the Ellis et al. (1978) study. The effects of changes of decision criteria will be further discussed in the general discussion.

DELAY

Delay produced a strong and highly significant effect on all the recognition indices used. In particular, both d' and A_g showed that a retention interval of three weeks produces a marked decrement in performance for black and white photographs. However, it is interesting to note that there was no interaction between the period of delay and the lines on the stimulus faces. Thus, to the extent that lines do disrupt recognition memory (and the present results suggest that the effect may be very small to nonexistent), it would seem that the passage of time does not play any special role in the process. All that can be said is that a retention interval of three weeks affects recognition accuracy fairly equally for lined and unlined faces. It is certainly the case that the possible weak effects of the addition of lines seen at zero delay is not strengthened after a three week delay. These results strongly suggest that the period of time for which the faces are stored in memory plays little or no part in any such process.

In summary, like Experiment I, Experiment II failed to replicate the Ellis et al. (1978) effect. The addition of lines had no statistically significant effect, though it should be noted that trends in some of the recognition accuracy measures paralleled those of Ellis et al.

A subsidiary aim was to find out if delay had any differential effect on lined and unlined faces. There were no such effects, although delay itself produced a reduction in recognition accuracy. Therefore, it seems that fragmenting the face with photofit or random lines does not cause any deterioration in recognition memory over time.

GENERAL DISCUSSION AND CONCLUSIONS

The results of the two experiments reported here suggest that the addition of photofit-type lines or random lines to full frontal, black and white photographs has little or no effect on recognition memory. Strictly interpreted, therefore, the results fail to support the findings of Ellis et al. (1978).

However, the present results pose an interesting problem of interpretation. On the one hand, a rigid interpretation using a significant level of 5% shows that not a single recognition measure produced a reliable result. Thus, the most parsimonious conclusion is that no evidence could be found to support the results of Ellis et al. (1978). On the other hand, one cannot ignore the fact that some of the results obtained in both Experiment I and II showed trends similar to the results obtained by Ellis et al. Taken as a whole and looking at the results descriptively, Experiment II and III of Ellis et al. and the present two studies suggest that adding lines to stimulus faces may have an effect on recognition accuracy, albeit a small one. Therefore, although no significant results were obtained in the present investigations the trends obtained, taken together with the Ellis et al. results suggest that the most prudent course of action might be to avoid using faces fragmented by lines, or composite faces (e.g., photofit) in research on recognition memory until the issues are further clarified. It would seem most unwise at the present time to conclude that adding lines to faces has no effect. Indeed, even if only the small effects noted in the present studies occur, they may add to the effects of the independent variables in face studies using photofit to produce spurious results.

In Experiment II the period of delay between the study and test phases was varied in order to find out whether the retention interval may have differentially affected recognition accuracy for lined and unlined faces. Delay itself produced the expected effect on recognition accuracy, but there was no evidence at all that delay differentially affected memory for lined and unlined faces. The conclusion to be drawn from the effect of delay is that if lines do tend to fragment the face and produce a small decrement in recognition accuracy, then storage time in memory seems not to be an important factor. Thus, it seems more likely that any effects are occurring either at the encoding stage or during the retrieval of faces from memory.

In the introduction, it was argued that the (H-FA) measure used by Ellis et al. (1978) was suspect and it was implied that this may have affected the results they obtained. Presumably they used this unorthodox measure because they were unable to calculate d' for all subjects due to ceiling and floor effects for hits and false alarms respectively. It was shown in the introduction that for a fixed hit rate, d' plotted as a function of false alarms produced an accelerating function while (H-FA) produced a linear function. The differences in the rate of change for the two measures can become quite large for small false alarm rates (see Figure 1a and 1b). Moreover, (H-FA) is not a pure recognition index because it can be affected by response bias. It is usual to assume normal-normal, equal variance underlying "old" and "new" distribution of faces (see Banks (1970). With such an assumption, any change in response bias can be shown to produce a larger change in hit rate than in false alarm rate (or vice versa). Thus, (H-FA) may change partially as a function of bias since hits and false alarms are differentially affected. Theoretically at least, the d' and A_g indices are not so affected.

However, Ellis et al. 1978 show that the addition of lines to the face causes not only a decrease in hits but an increase in false alarms also. Thus, even though their measure, (H-FA), may have been affected by bias, it seems likely that changes in recognition accuracy also occurred.

Nevertheless, the results of the present study show that the addition of lines to stimulus faces may cause a small change in response bias in the direction of the subjects being less likely to report a face as a target. This is an interesting finding and any further research on the effects of fragmenting facial stimuli should seriously consider estimating the effects of response bias as well as recognition accuracy.

Overall, a strict interpretation of the present results in terms of the conventional statistical decision model shows that none of the measures used (hits, false alarms, d', A_g, A', and B") produced reliable changes in performance. However, it seems that caution is required in using such a strict interpretation because the trends in the present two studies follow the results obtained by Ellis et al. (1978). It therefore seems premature to state categorically that the addition of lines to stimulus faces has no effect on recognition accuracy. However, if such effects do exist then the evidence obtained here, and by Ellis et al., suggest they are very small. Further research is required, using different sets of faces, to further ascertain whether the effect has theoretical implications for how faces are stored in memory. Given the size of the effect obtained in the present study, and in that of Ellis et al., it seems unlikely that the results have any serious implications for the use of the photofit-kit in practical situations, and as an investigative tool for the police. The seeming inability to construct accurate likenesses from photofit (e.g., see Davies, 1981) poses a far more serious problem for the kit in every day use.

REFERENCES

Baddeley, A., and Woodhead, M. (1982). Depth of processing context, and face recognition. *Canadian Journal of Psychology*, **36**, 148-164.

Baddeley, A., and Woodhead, M. (1983). Improving face recognition ability. In S. M. A. Lloyd-Bostock and B. R. Clifford (Eds.), *Evaluating witness evidence* (pp. 125-136). Chichester: Wiley.

Banks, W. P.(1970). Signal detection theory and human memory. *Psychological Bulletin*, **74**, 81-89.

Benton, A. L. (1980). The neuropsychology of face recognition. *American Psychologist*, **35**, 176-186.

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

Brigham, J. C., Maass. A., Snyder, L. D., and Spaulding, K. (1982). Accuracy of eyewitness identification in a field setting. *Journal of Personality and Social Psychology*, **42**, 673-681.

Brown, E. L., Deffenbacher, K. A., and Sturgil, W. (1977). Memory for faces and the circumstances of encounter. *Journal of Applied Psychology*, **62**, 311-318.

Bruce, V. (1982). Changing faces: Visual and non-visual coding processes in face recognition. *British Journal of Psychology*, **73**, 105-116.

Bruce, V. (1989). *Recognising faces*. Hillsdale, U.S.A: Laurence Earlbaum Associates.

Bruyer, R. (1989). Disorders of face processing. In H. D. Ellis and A. W. Young (Eds.), *Handbook of research on face processing* (pp. 437-473). Netherlands: Elsevier.

Carey, S. (1981). The development of face perception. In G. Davies, H. Ellis and J. Shepherd (Eds,), *Perceiving and remembering faces* (pp. 9-38). London: Academic Press.

Chance, J. E., and Goldstein, A. G. (1987). Retention interval and face recognition: Response latency measures. *Bulletin of the Psychonomic Society*, **6**, 415-418.

Chance, J. E., Goldstein, A. G., and McBride, L.(1975). Differential experience and recognition memory for faces. *Journal of Social Psychology*, **97**, 243-253.

Christie, D. F. M., and Ellis, H. D. (1981). Photofit construction versus verbal description of faces. *Journal of Applied Psychology*, **66**, 358-363.

Cohen, M. E., and Carr, W. J. (1975). Facial recognition and the Von Restorff effect. *Bulletin of the Psychonomic Society*, **6**, 383-384.

Courtois, M. R., and Mueller, J. R. (1981). Target and distractor typicality in facial recognition. *Journal of Applied Psychology*, **66**, 639-645.

Craik, F. I. M., and Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, **11**, 671-684.

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

Davies, G. M. (1983). Forensic face recall: the role of visual and verbal information. In S. M. A. Lloyd-Bostock and B. R. Clifford (Eds.), *Evaluating witness evidence* (pp. 103-123). London: Wiley.

Davies, G. (1986) The recall and reconstruction of faces. Implications for theory and practice. In H. D. Ellis, M. A. Jeeves, F. Newcombe and A. Young (Eds.), *Aspects of face processing* (pp. 388-397). Dordrecht: Martinus Nijhoff.

Davies, G., Ellis, H. D., and Shepherd, J. W. (1978a). Face identification: The influence of delay upon accuracy of photofit construction. *Journal of Police Science and Administration*, **6**, 35-42.

Davies, G., Ellis, H. D. and Shepherd, J. W. (1978b). Face recognition accuracy as a function of mode of representation. *Journal of Applied Psychology*, **63**, 180-187.

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

Deffenbacher, K. A. (1986). On the memorability of the human face. In H. D. Ellis, M. A. Jeeves, F. Newcombe and A. Young (Eds.), *Aspects of face processing* (pp. 61-70). Dordrecht: Martinus Nijhoff.

Deffenbacher, K. A., Carr, T. H., and Leu, J. R. (1981). Memory for words, pictures, and faces: retroactive interference, forgetting, and reminiscence. *Journal of Experimental Pschology: Human Learning and Memory*, **7**, 299-305.

Egan, D., Pittner, M., and Goldstein, A. G. (1977). Eyewitness identification: Photographs vs. live models. *Law and Human Behavior*, **1**, 199-206.

Ellis, H. D.(1981). Theoretical aspects of face recognition. In G. Davies, H. Ellis and J. Shepherd (Eds.), *Perceiving and remembering faces* (pp. 171-197). London: Academic Press.

Ellis, H. D., Davies, G. M., and Shepherd, J. W. (1978). Remembering pictures of real and 'unreal' faces: Some practical and theoretical considerations. *British Journal of Psychology*, **69**, 467-474.

Ellis, H. D., Shepherd, J. W., and Davies, G. M. (1975). An investigation of the use of the photofit technique for recalling faces. *British Journal of Psychology*, **66**, 29-37.

Gehring, R. E., Toglia, M. P., and Kimble, G. A. (1976). Recognition memory for words and pictures at short and long retention intervals. *Memory and Cognition*, **4**, 256-260.

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

Goldstein, A. G., and Chance, J. E. (1970). Visual recognition memory for complex configurations. *Perception and Psychophysics*, **9**, 237-241.

Goldstein, A. G., and Chance. J. E. (1981). Laboratory studies of face recognition. In G. Davies, H. Ellis and J. Shepherd (Eds.), *Perceiving and remembering faces* (pp. 81-104). London: Academic Press.

Green, D. M., and Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Grier, B. J. (1971). Non-parametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin*, 75, 424-429.

Hecaen, H. (1981). The neuropsychology of face recognition. In G. Davies, H. Ellis and J. Shepherd (Eds.), *Perceiving and remembering faces* (pp.39-54). London: Academic Press.

Krouse, F. L. (1981). Effects of pose, pose change, and delay on face recognition performance. *Journal of Applied Psychology*, **66**, 651-654.

Laughery, K. R., Alexander, J. F., and Lane, A. B. (1971). Recognition of human faces: Effects of target exposure time, target position, pose position, and type of photograph. *Journal of Applied Psychology*, **51**, 477-483.

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

Laughery, K. R., Rhodes, B., and Batten, G. (1981). Computer-guided recognition and retrieval of facial images. In G. Davies, H. Ellis and J. Shepherd (Eds.), *Perceiving and remembering faces* (pp. 251-269). London: Academic Press.

Light, L. L., Kayra-Stuart, F., and Hollander, S. (1979). Recognition memory for typical and unusual faces. *Journal of Experimental Psychology: Human Learning and Memory*, **5**, 212-228.

Loftus, E. F. (1979). *Eyewitness testimony*. Cambridge, Mass: Harvard University Press.

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

McNicol, D. (1972). *A primer of signal detection theory*. London: George Allen and Unwin.

Patterson, K. E., and Baddeley, A. D. (1977). When face recognition fails. *Journal of Experimental Psychology: Human Learning and Memory*, **3**, 406-417.

Podd, J. (in press). The effects of memory load and delay on facial recognition. *Applied Cognitive Psychology*.

Salzen, E. A. (1981). Perception and emotion in faces. In G. Davies, H. Ellis, and J. Shepherd (Eds.), *Perceiving and remembering faces* (pp. 133-169). London: Academic Press.

Sergent, J. (1984). An investigation into component and configural processes underlying face perception. *British Journal of Psychology*, **75**, 221-242.

Shapiro, P. N., and Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin*, **100**, 139-156.

Shepherd, J. (1981). Social factors in facial recognition. In G. Davies, H. Ellis and J. Shepherd (Eds.), *Perceiving and remembering faces* (pp.55-79). London: Academic Press.

Shepherd, J. W. (1983). Identification after long delays. In S. M. A. Lloyd-Bostock and B. R. Clifford (Eds.), *Evaluating witness evidence* (pp. 173-181). London: Wiley.

Shepherd, J. W., and Ellis, H. D. (1973). The effect of attractiveness on recognition memory for faces. *American Journal of Psychology*, **86**, 627-633.

Shepherd, J. W., Ellis, H. D., and Davies, G. M. (1982). *Identification evidence: A psychological evaluation.* Aberdeen University Press.

Venner, B. R. H. (1969). Facial identification techniques. *Police Research Bulletin*, No. **13**, 17-20.

Winograd, E. (1981). Elaboration and distinctiveness in memory for faces. *Journal of Experimental Psychology: Human Learning and Memory*, **7**, 181-190.

Woodhead, M. M., Baddeley, A. D., and Simmonds, D. C. V. (1979). On training people to recognise faces. *Ergonomics*, **22**, 333-343.

Yarmey, A. D. (1971). Recognition memory for familiar 'public' faces: Effects of orientation and delay. *Psychonomic Science*, **24**, 286-288.

Yarmey, A. D. (1979). The effects of attractiveness, feature saliency, and liking on memory for faces. In M. Cook and G. Wilson (Eds.), *Love and attraction* (pp. 51-53). New York: Pergamon Press.

APPENDICES

APPENDIX 1: Response sheet used by subjects in Experiment I.

PLEASE MARK THE RESPONSE SHEET WITH Y FOR "YES" AND N FOR "NO"

1.	26.
2.	27.
3.	28.
4.	29.
5.	30.
6.	31.
7.	32.
8.	33.
9.	34.
10.	35.
11.	36.
12.	37.
13.	38.
14.	39.
15.	40.
16.	41.
17.	42.
18.	43.
19.	44.
20.	45.
21.	46.
22.	47.
23.	48.
24.	49.
25.	50.

APPENDIX 2a:

Individual data for all measures
for subjects who saw unlined
faces in Experiment I.

	Hits	False alarms	H-FA	A'	B"
	0.85	0.10	0.75	0.93	0.17
	0.80	0.15	0.65	0.89	0.11
	0.75	0.10	0.65	0.89	0.35
	0.80	0.05	0.75	0.83	0.54
	0.85	0.30	0.55	0.86	-0.24
	0.85	0.30	0.55	0.86	-0.24
	0.70	0.10	0.60	0.88	0.40
	0.80	0.20	0.60	0.88	0.00
	1.00	0.00	1.00	0.99	0.00
	0.90	0.00	0.90	0.98	1.00
	0.80	0.15	0.65	0.89	0.11
	0.55	0.05	0.50	0.86	0.68
	0.70	0.05	0.65	0.90	0.63
	0.95	0.15	0.80	0.95	-0.46
	0.85	0.10	0.75	0.93	0.17
	0.80	0.05	0.75	0.93	0.54
	0.85	0.05	0.80	0.95	0.46
	0.95	0.05	0.90	0.97	0.00
	0.90	0.30	0.60	0.88	-0.40
	0.85	0.40	0.45	0.82	-0.31
M	0.83	0.13	0.69	0.90	0.18
SD	0.10	0.11	0.14	0.05	0.40

APPENDIX 2b:

Individual data for all measures
for subjects who saw photofit-type
faces in Experiment I

	Hits	False Alarms	H-FA	A’	B"
	0.70	0.00	0.70	0.93	1.00
	0.70	0.20	0.50	0.83	0.14
	0.75	0.10	0.65	0.90	0.35
	0.85	0.05	0.80	0.95	0.46
	0.70	0.00	0.70	0.93	1.00
	0.90	0.25	0.65	0.90	-0.35
	0.75	0.15	0.60	0.88	0.19
	0.80	0.10	0.70	0.91	0.28
	0.80	0.05	0.75	0.93	0.54
	0.75	0.20	0.55	0.86	0.08
	0.75	0.00	0.75	0.94	1.00
	0.65	0.10	0.55	0.86	0.43
	0.90	0.10	0.80	0.94	0.00
	0.85	0.15	0.70	0.91	0.00
	0.55	0.05	0.50	0.86	0.68
	1.00	0.05	0.95	0.99	-1.00
	0.55	0.10	0.45	0.83	0.47
	0.85	0.10	0.75	0.93	0.17
	0.75	0.05	0.70	0.92	0.60
	0.95	0.15	0.80	0.95	-0.46
M	0.78	0.10	0.68	0.91	0.28
SD	0.12	0.07	0.12	0.04	0.50

APPENDIX 2c:

Individual data for all measures
for subjects who saw randomly
lined faces in Experiment I

	Hits	False Alarms	H-FA	A'	B"
	0.85	0.05	0.80	0.95	0.46
	0.80	0.00	0.80	0.95	1.00
	0.85	0.05	0.80	0.95	0.46
	0.75	0.05	0.70	0.92	0.60
	0.85	0.00	0.85	0.96	1.00
	0.70	0.05	0.65	0.90	0.63
	0.80	0.35	0.45	0.81	-0.17
	0.65	0.20	0.45	0.81	0.17
	0.75	0.15	0.60	0.88	0.19
	0.75	0.05	0.70	0.92	0.60
	0.75	0.15	0.60	0.88	0.19
	0.80	0.20	0.60	0.88	0.00
	0.80	0.05	0.75	0.93	0.54
	0.90	0.05	0.85	0.96	0.31
	0.60	0.00	0.60	0.90	1.00
	0.85	0.35	0.50	0.84	-0.28
	0.75	0.25	0.50	0.83	0.00
	0.85	0.15	0.70	0.91	0.00
	0.55	0.15	0.40	0.80	0.32
	0.70	0.10	0.60	0.88	0.40
M	0.77	0.12	0.65	0.89	0.37
SD	0.09	0.11	0.14	0.05	0.37

APPENDIX 3:

Response sheets used by subjects in Experiment II

Trial	**Very Sure Old**	**Fairly Sure Old**	**Fairly Sure New**	**Very Sure New**
1	1	2	3	4
2	1	2	3	4
3	1	2	3	4
4	1	2	3	4
5	1	2	3	4
6	1	2	3	4
7	1	2	3	4
8	1	2	3	4
9	1	2	3	4
10	1	2	3	4
11	1	2	3	4
12	1	2	3	4
13	1	2	3	4
14	1	2	3	4
15	1	2	3	4
16	1	2	3	4
17	1	2	3	4
18	1	2	3	4
19	1	2	3	4
20	1	2	3	4
	Very Sure Old	**Fairly Sure Old**	**Fairly Sure New**	**Very Sure New**

Trial	Very Sure Old	Fairly Sure Old	Fairly Sure New	Very Sure New
21	1	2	3	4
22	1	2	3	4
23	1	2	3	4
24	1	2	3	4
25	1	2	3	4
26	1	2	3	4
27	1	2	3	4
28	1	2	3	4
29	1	2	3	4
30	1	2	3	4
31	1	2	3	4
32	1	2	3	4
33	1	2	3	4
34	1	2	3	4
35	1	2	3	4
36	1	2	3	4
37	1	2	3	4
38	1	2	3	4
39	1	2	3	4
40	1	2	3	4
	Very Sure Old	Fairly Sure Old	Fairly Sure New	Very Sure New

Trial	Very Sure Old	Fairly Sure Old	Fairly Sure New	Very Sure New
41	1	2	3	4
42	1	2	3	4
43	1	2	3	4
44	1	2	3	4
45	1	2	3	4
46	1	2	3	4
47	1	2	3	4
48	1	2	3	4
49	1	2	3	4
50	1	2	3	4
51	1	2	3	4
52	1	2	3	4
53	1	2	3	4
54	1	2	3	4
55	1	2	3	4
56	1	2	3	4
57	1	2	3	4
58	1	2	3	4
59	1	2	3	4
60	1	2	3	4
	Very Sure Old	Fairly Sure Old	Fairly Sure New	Very Sure New

Trial	Very Sure Old	Fairly Sure Old	Fairly Sure New	Very Sure New
61	1	2	3	4
62	1	2	3	4
63	1	2	3	4
64	1	2	3	4
65	1	2	3	4
66	1	2	3	4
67	1	2	3	4
68	1	2	3	4
69	1	2	3	4
70	1	2	3	4
71	1	2	3	4
72	1	2	3	4
73	1	2	3	4
74	1	2	3	4
75	1	2	3	4
76	1	2	3	4
77	1	2	3	4
78	1	2	3	4
79	1	2	3	4
80	1	2	3	4
	Very Sure Old	Fairly Sure Old	Fairly Sure New	Very Sure New

APPENDIX 4a:

Individual data for all measures for subjects in the immediate testing condition who saw unlined faces in Experiment II

	Hits	False Alarms	H-FA	d'	A_g	B"
	0.83	0.14	0.69	2.03	0.91	0.08
	0.69	0.03	0.66	2.38	0.89	0.76
	0.83	0.03	0.80	2.83	0.93	0.66
	0.94	0.26	0.68	2.19	0.92	-0.55
	0.83	0.09	0.74	2.29	0.92	0.27
	0.74	0.43	0.31	0.82	0.70	-0.12
	0.86	0.31	0.55	1.58	0.85	-0.28
	0.74	0.17	0.57	1.59	0.81	0.15
	0.77	0.17	0.60	1.69	0.87	0.11
	0.83	0.20	0.63	1.79	0.87	-0.06
	0.80	0.17	0.63	1.79	0.89	0.06
	0.60	0.06	0.54	1.80	0.85	0.62
	0.80	0.29	0.51	1.40	0.79	-0.13
	0.83	0.17	0.66	1.90	0.91	0.00
	0.89	0.20	0.69	2.07	0.89	-0.24
M	0.80	0.18	0.62	1.88	0.87	0.09
SD	0.08	0.11	0.12	0.47	0.06	0.37

APPENDIX 4b:

Individual data for all measures for subjects in the immediate testing condition who saw photofit-type faces in Experiment II

	Hits	False Alarms	H-FA	d'	A_g	B"
	0.83	0.09	0.74	2.29	0.92	0.27
	0.60	0.11	0.49	1.48	0.81	0.42
	0.77	0.23	0.54	1.48	0.85	0.00
	0.74	0.14	0.60	1.72	0.84	0.23
	0.94	0.14	0.80	2.63	0.94	-0.36
	0.83	0.26	0.57	1.59	0.87	-0.15
	0.89	0.20	0.69	2.07	0.91	-0.24
	0.74	0.11	0.63	1.87	0.89	0.33
	0.69	0.23	0.46	1.24	0.77	0.09
	0.74	0.17	0.57	1.59	0.81	0.15
	0.51	0.11	0.40	1.26	0.74	0.44
	0.71	0.17	0.54	1.50	0.84	0.19
	0.74	0.34	0.40	1.05	0.77	-0.08
	0.69	0.23	0.46	1.24	0.76	0.09
	0.63	0.09	0.54	1.67	0.79	0.48
M	0.74	0.17	0.56	1.65	0.83	0.12
SD	0.11	0.07	0.12	0.43	0.06	0.25

APPENDIX 4c:

Individual data for all measures for subjects in the immediate testing condition who saw randomly lined faces in Experiment II.

	Hits	False Alarms	H-FA	d'	A_g	B"
	0.71	0.17	0.54	1.50	0.86	0.19
	0.54	0.11	0.43	1.33	0.75	0.43
	0.60	0.26	0.34	0.90	0.76	0.11
	0.91	0.11	0.80	2.57	0.95	-0.09
	0.71	0.26	0.45	1.20	0.76	0.03
	0.83	0.03	0.80	2.83	0.93	0.66
	0.83	0.11	0.72	2.18	0.89	0.18
	0.46	0.11	0.35	1.13	0.68	0.43
	0.69	0.26	0.43	1.14	0.74	0.05
	0.69	0.34	0.35	0.92	0.76	-0.02
	0.86	0.06	0.80	2.63	0.91	0.36
	0.63	0.20	0.43	1.17	0.76	0.19
	0.86	0.17	0.69	2.03	0.89	-0.08
	0.83	0.20	0.63	1.79	0.84	-0.06
	0.83	0.34	0.49	1.36	0.81	-0.23
M	0.73	0.18	0.55	1.65	0.82	0.14
SD	0.13	0.10	0.17	0.65	0.08	0.24

APPENDIX 4d:

Individual data for all measures for subjects who saw unlined faces in the three week delay condition in Experiment II.

	Hits	False Alarms	H-FA	d'	A_g	B"
	0.69	0.17	0.52	1.46	0.85	0.21
	0.83	0.26	0.57	1.59	0.86	-0.15
	0.34	0.09	0.25	0.93	0.67	0.47
	0.60	0.31	0.29	0.76	0.69	0.06
	0.69	0.20	0.49	1.34	0.79	0.14
	0.80	0.40	0.40	1.10	0.82	-0.20
	0.71	0.34	0.37	0.96	0.72	-0.04
	0.77	0.31	0.46	1.24	0.79	-0.09
	0.54	0.29	0.25	0.66	0.72	0.09
	0.74	0.09	0.65	1.98	0.84	0.40
	0.66	0.29	0.37	0.96	0.73	0.04
	0.77	0.14	0.63	1.82	0.84	0.19
	0.63	0.14	0.49	1.41	0.82	0.32
	0.77	0.14	0.63	1.82	0.88	0.19
	0.86	0.46	0.40	1.18	0.77	-0.35
M	0.69	0.24	0.45	1.28	0.79	0.09
SD	0.13	0.11	0.13	0.40	0.07	0.23

APPENDIX 4e:

Individual data for all measures for subjects who saw photofit-typed faces in the three week delay condition for Experiment II

	Hits	False Alarms	H-FA	d'	A_g	B"
	0.69	0.06	0.63	2.06	0.87	0.58
	0.71	0.23	0.48	1.30	0.80	0.08
	0.51	0.29	0.22	0.58	0.66	0.10
	0.71	0.23	0.48	1.30	0.84	0.08
	0.77	0.37	0.40	1.07	0.77	-0.14
	0.71	0.23	0.48	1.30	0.78	0.08
	0.71	0.26	0.45	1.20	0.77	0.03
	0.69	0.11	0.58	1.74	0.85	0.37
	0.71	0.40	0.31	0.81	0.70	-0.08
	0.66	0.11	0.55	1.64	0.79	0.39
	0.71	0.34	0.37	0.96	0.73	-0.04
	0.51	0.34	0.17	0.44	0.66	0.05
	0.66	0.37	0.29	0.74	0.70	-0.02
	0.71	0.20	0.51	1.40	0.78	0.13
	0.80	0.17	0.63	1.79	0.85	0.06
M	0.68	0.25	0.44	1.22	0.77	0.11
SD	0.08	0.11	0.14	0.47	0.07	0.19

APPENDIX 4f:

Individual data for all measures for
subjects who saw randomly lined
faces in the three week delay condition

	Hits	False alarms	H-FA	d'	A_g	B"
	0.80	0.34	0.46	1.25	0.75	-0.17
	0.37	0.09	0.28	1.01	0.71	0.48
	0.71	0.26	0.45	1.20	0.77	0.03
	0.69	0.06	0.63	2.06	0.88	0.58
	0.51	0.29	0.22	0.58	0.60	0.10
	0.69	0.20	0.49	1.34	0.78	0.14
	0.77	0.31	0.46	1.24	0.78	-0.09
	0.34	0.14	0.20	0.67	0.61	0.30
	0.63	0.06	0.57	1.88	0.81	0.61
	0.54	0.34	0.20	0.51	0.66	0.05
	0.71	0.31	0.40	1.06	0.77	-0.02
	0.86	0.20	0.66	1.92	0.86	-0.14
	0.63	0.31	0.32	0.84	0.70	0.04
	0.86	0.06	0.80	2.63	0.94	0.36
	0.71	0.26	0.45	1.20	0.79	0.03
M	0.65	0.22	0.44	1.29	0.76	0.15
SD	0.16	0.11	0.18	0.60	0.09	0.25