TR－H－095

Exemplar-based and Norm-based Models of Face Recognition

Robert A. Johnston

1994. 8. 30

ATR人間情報通信研究所

〒619-02 京都府相楽郡精華町光台2-2　☎07749-5-1011

ATR Human Information Processing Research Laboratories

2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02 Japan

Telephone: +81-7749-5-1011
Facsimile: +81-7749-5-1008

Exemplar-based and Norm-based Models of Face Recognition

Robert A. Johnston

ATR Human Information Processing Research Laboratories

2-2 Hikari-dai, Seiko-cho, Soraku-gun

Kyoto 619-02 JAPAN[1]

This paper gives an overview of recent attempts to explore a heuristic for the representation of faces in memory. Valentine (1991a, 1991b) has described a framework whereby face representations may be thought of as locations within a multidimensional space. This framework has proven very efficient at accommodating and accounting for many of the phenomena observed in face recognition research. Valentine has proposed that there are at least two possible models which can be based on this framework: an exemplar-based model and a norm-based model. This discussion examines recent attempts to determine between these two accounts.

1 Introduction

Valentine has suggested that a useful heuristic for understanding this how this is achieved is to view the adult face space as a multidimensional space (Valentine, 1991a; 1991b). Facial representations in memory can be viewed as locations within this multidimensional space. So far it has not been possible to specify the dimensions of this space but it would not be unreasonable to assume that they will be based on those that

[1] Currently at the School of Psychology, University of Wales College of Cardiff, Cardiff, Uk, CF1 3YG. email sapraj@thor.cf.ac.uk

would best serve to discriminate among faces. Indeed there are many reasonable candidates which include such feature dimensions as face shape, hair length, hair colour, or perceived age derived from multidimensional scaling studies (e.g., see Shepherd, Ellis and Davies, 1977).

The origin of the multi-dimensional face space will be the central tendency of the dimensions and it is assumed that the feature dimensions of faces experienced will vary normally around this point. Typical faces, by definition, are more often experienced than distinctive faces and so the density of points throughout this space will not be uniform. There will be a higher density of face representations around the central tendency (i.e., the region where representations of typical faces are located). The putative framework, which Valentine has suggested allows us neatly to account for many effects described in the face recognition literature (e.g.,distinctiveness vs. typicality effects, inversion phenomena and face classification effects).

Using this theoretical framework, Valentine has identified two specific models based on the multidimensional nature of face space. One version he describes as being a norm-based model (however, this label is intended to cover a variety of similar theoretical constructs). Specifically, the theoretical approaches included are the prototype hypothesis (Valentine and Bruce, 1986a, 1986b), the norm-based coding model (Rhodes, Brennan and Carey, 1987); and the schema theory (Goldstein and Chance, 1980). All these accounts assume that storing representations of faces in memory entails the abstraction of something that can be called a face norm, prototype or schema. The norm-based model proposes that each individual face is stored in memory according to its deviation from a single, general face norm or prototype. This would be located at the origin of the face space. For an n-dimensional face-space, an n-dimensional vector from the origin to a point representing the dimension values of a face would uniquely specify that face. The process of recognising a face involves encoding the stimulus face as an n-dimensional vector and deciding if the resultant stimulus matches the stored vector of a face already encountered.

This is not the only way in which a multi-dimensional space could be described of course, alternative formulations, for example may not require a prototype face: instead these could be based on inter-stimulus similarity. In fact Valentine (1991a) has described such versions which he groups under the heading of exemplar-based models. It is appropriate in these models to think of faces as being encoded as points rather than vectors. Here, the origin of the space plays no role in encoding stimuli, it is simply the area of greatest exemplar density.

Valentine (1991a, 1991b) has shown how each of these models (norm-based and exemplar-based) can predict a particularly robust effect in the face recognition literature. Research employing faces in the recognition memory experiments has shown that subjects are better at recognising distinctive faces. A typical experiment would involve showing subjects a set of unfamiliar faces and subsequently asking them to identify these stimuli later when given a larger selection of unfamiliar faces. It has been demonstrated that subjects perform at a superior level with distinctive faces compared with typical faces. This differential performance can be demonstrated in a number of different ways. These include a higher recognition hit rate for previously seen faces if they are distinctive; and also faster recognition latencies. The distinctiveness advantage is also exhibited by the occurrence of fewer false positive decisions to distinctive distracter faces. An overall advantage for the recognition of distinctive as opposed to typical faces can often be demonstrated using a measure of sensitivity such as d' or A' (Going and Read, 1974; Cohen and Carr, 1975; Light, Kayra-Stuart, & Hollander, 1979; Winograd, 1981; Bartlett, Hurry, Thorley, 1984; Valentine, 1991a, 1991b; Shepherd, Gibling and Ellis, 1991). Moreover, Ellis, Shepherd, Shepherd, Flin and Davies (1989) demonstrated that, although when subjects searched for a target face using FRAME (a computerised mug-shot system) they were as good at retrieving distinctive faces as typical faces, when other subjects used a traditional mugshot album to recognise faces there was an enormous advantage for distinctive faces, particularly when they occurred later in a series of 1000 mugshots.

There is some variation in how the researchers mentioned above, refer to the faces denoted above as either distinctive or typical. Some share this nomenclature, while others instead use the labels 'memorableness', 'uniqueness' or 'unusualness' instead. Vokey and Read (1992) even approach the dimension from the other direction and talk of typicality of faces where distinctive faces are considered atypical. Nevertheless, all researchers mean this to describe the range of variation present in *ordinary* faces. We do not intend the appellation 'distinctive' to conjure an image of a face that is deformed or has one eye or a huge scar.

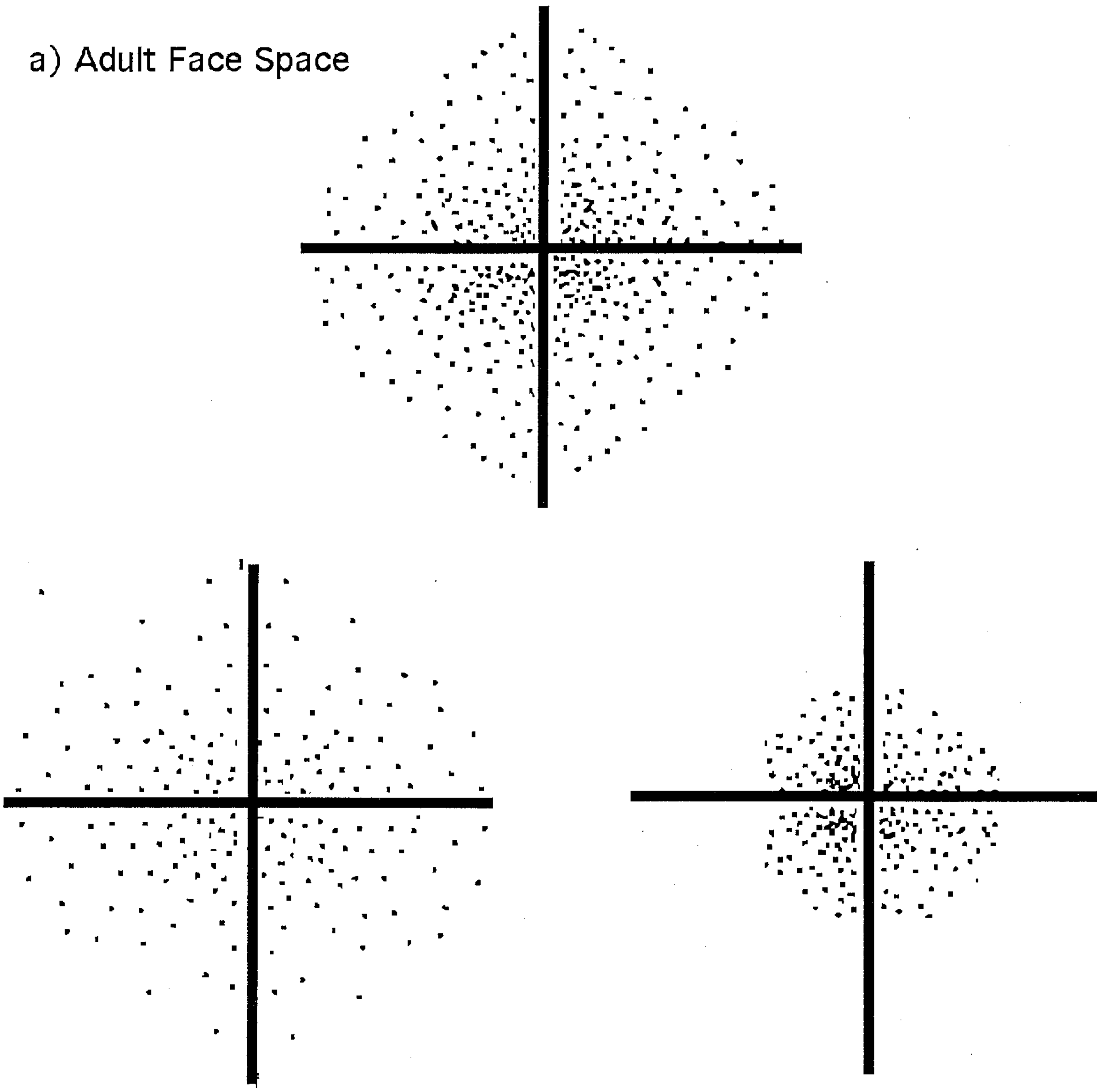
As mentioned earlier in this discussion, various workers have shown that the ability to discriminate novel faces from ones already encountered is a skill which improves steadily with age (e.g., Goldstein and Chance, 1964; Flin, 1980). However, compared to the abundance of work which has looked at this task with adult subjects when stimuli are controlled for distinctiveness, there is little research available on the performance of children. Ellis (1992) described some preliminary work on school-age children which suggested that the characteristic adult advantage for distinctive faces was absent in children aged around 6 years of age. He suggested that young children either fail to encode those aspects which make faces distinctive or that they store both typical and distinctive faces in the same manner. Ellis also demonstrated that even by the age of 13 years, subjects were not able to discount distinctive distracter faces more easily than typical distracters. In research employing adult subjects this advantage is a robust effect (e.g., Bartlett et al., 1984; Valentine and Endo 1992), which suggests that the adult level of performance occurs after puberty.

Valentine's multi-dimensional face space can readily accommodate the distinctiveness advantage shown by adults in recognition memory experiments. According to either norm-based or exemplar-based models, typical faces are located in the areas of highest exemplar density: so, when a typical face is presented for test it is more likely to resemble another face, which produces a situation with greater uncertainty and hence more opportunity for error: it will be reflected both in longer response latencies to typical faces and in more false positives. Is it possible that this framework can also account for effects observed with child subjects?

In order to answer this question it is necessary to speculate on how the way faces are represented in children's memory may differ from the adult arrangement. It is scarcely controversial to suggest that, in some way, a child's face space would be smaller than an adult's. This claim can be sensibly made on the basis that children have simply seen, and hence represented, fewer faces than adults. What is more contentious, however, is how this 'smallness' might be manifested.

One version is that the face space has the same general framework and parameters of the adult space but that it is less densely populated (Figure 1b) than the adult space (Figure 1a) and so, consequently, the density gradient of the space may be attenuated. Alternatively, the space might be based simply on a smaller volume than the adult space (Figure 1c). In this arrangement relative difference in the density of face space enjoyed by typical and distinctive faces are preserved and, as the child experiences more faces, the volume of the face space increases to accommodate the additions. (The face-spaces displayed in Figure 1 are shown in two dimensions for the purposes of illustration only.)

Figure 1: Possible Models of Face Space



a) Adult Face Space

b)Possible child face space

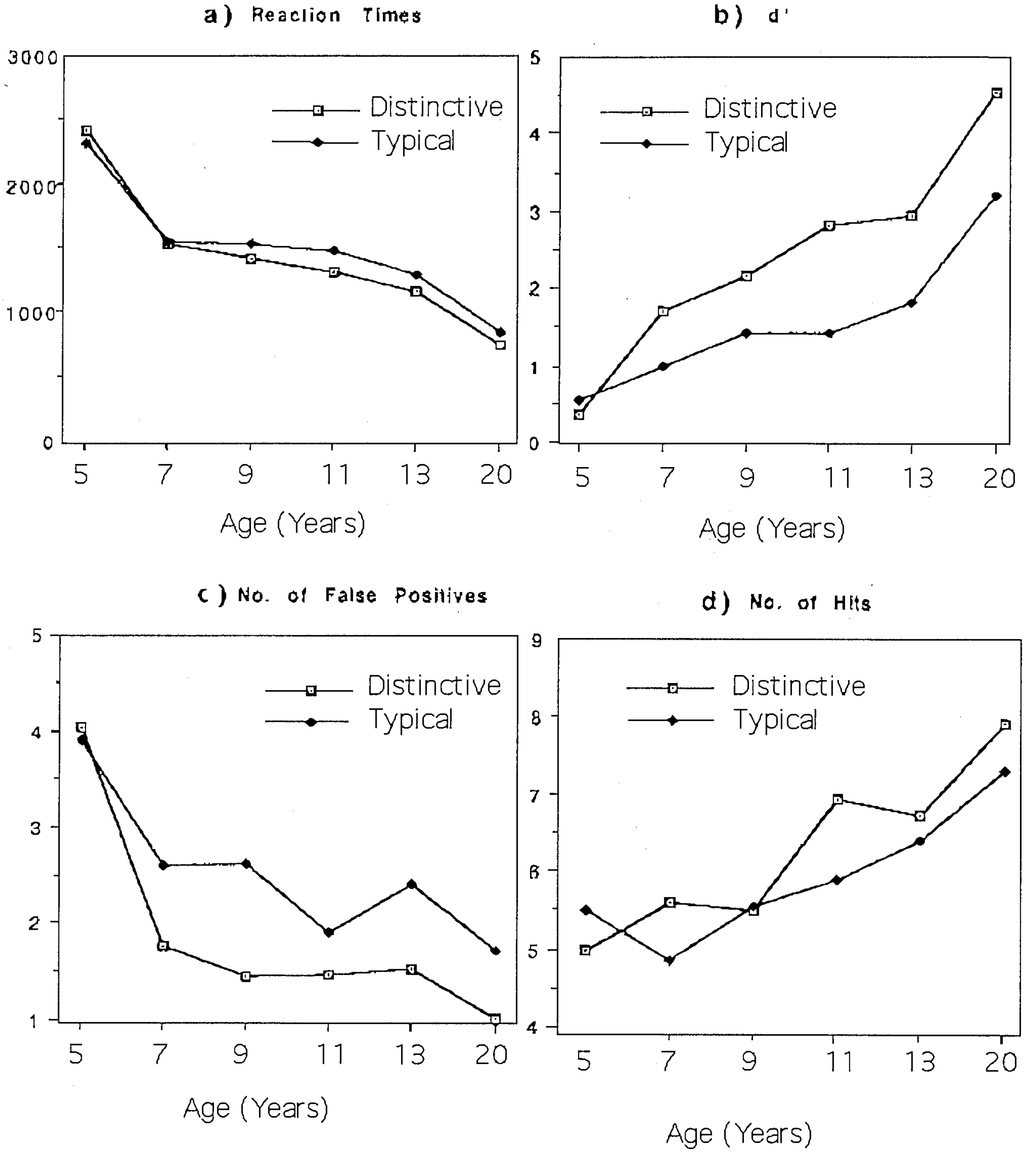
c)Possible child face space

It is not immediately apparent how additional dimensions can be added to the face space. We might propose that faces that are encountered early in childhood are not encoded along as many dimensions by the child's face-processing system. This could occur because s/he is not sensitive to some parameters or because there is no requirement to use them. As increasingly more faces are experienced greater complexity in encoding will be necessary to distinguish among them and so additional dimensions are invoked that take account of second-order and higher-order features (Carey, 1992). This need not depend on any lack of 'power' or completeness of the child's face-processing capacity and is surely the means by which adult specialists become attuned to discriminating among exemplars that for non-experts would appear undifferentiated (e.g. sheep for shepherds; McNeil and Warrington, 1993; or cows for farmers Bruyer et al., 1983; dogs for canine show judges, Diamond and Carey, 1986; Bewick swans for ornithologists, Bateson, 1977). The change in the person's power of discrimination arises through frequent and varied exposure to the exemplars. Valentine (1991a) proposes that it is precisely the implicit knowledge acquired from one's lifetime exposure to faces which leads to the normal distribution of faces in the adult face space.

2 Studies using children

In a recently completed study, Johnston and Ellis (In press), conducted experiments to extend Ellis' (1992) earlier findings and examine differential processing by children of distinctive and typical faces. Subjects were recruited across an age range from 5 to 20 years. Several performance measures were used to ascertain how subjects responded to typical and distinctive faces. These included the number of hits, the number of false positives and latencies for correct recognition. Hits and false positives were also combined to yield d' prime scores to permit an overall measure of accuracy. For three out

Figure 2: Performance on Face Recognition Memory Task



of four performance measures (response latencies, d' scores and number of false positives) it was demonstrated that the characteristic advantage shown by adults when processing distinctive faces only emerges completely at 9 years of age and was not present at all in 5-year-old subjects. A graphical representation of these findings can be seen in Figure 2.

3 Models of face space

It is now appropriate to consider how these findings can be reconciled with the alternative architectures outlined earlier for putative child face spaces (Figures 1b and 1c). In order to do this effectively it is necessary to consider the other possible refinement which Valentine has ascribed to the multidimensional face space. The two architectures suggested for the child face space can be implemented as either a norm-based or exemplar-based model. Each of these alternatives can accommodate the findings reported above, but they would do so using different explanations.

The competing hypotheses are as follows:

N1. The Figure 1b architecture and a norm-based model suggests that the distinctiveness advantage is not present because typical faces are no closer to their neighbours than distinctive faces. This model suggests that the density gradient in the child face-space will be less steep than in the adult version.

N2. The Figure 1c architecture, together with a norm-based model suggests that the distinctiveness advantage is not present because distinctive faces now have a neighbour which is as close as the neighbour of a typical face. This model suggests the density gradient in the child face-space can be effectively the same as in the adult version.

E1. The Figure 1b architecture combined with an exemplar-based model implies that the distinctiveness advantage is not present because typical faces are no closer to their neighbours than distinctive faces. This model suggests that the density gradient in the child face-space will be less steep than in the adult version.

E2. The Figure 1c architecture and an exemplar-based model suggests that the distinctiveness advantage is not present because distinctive faces now have a neighbour which is as close as the neighbour of a typical face.

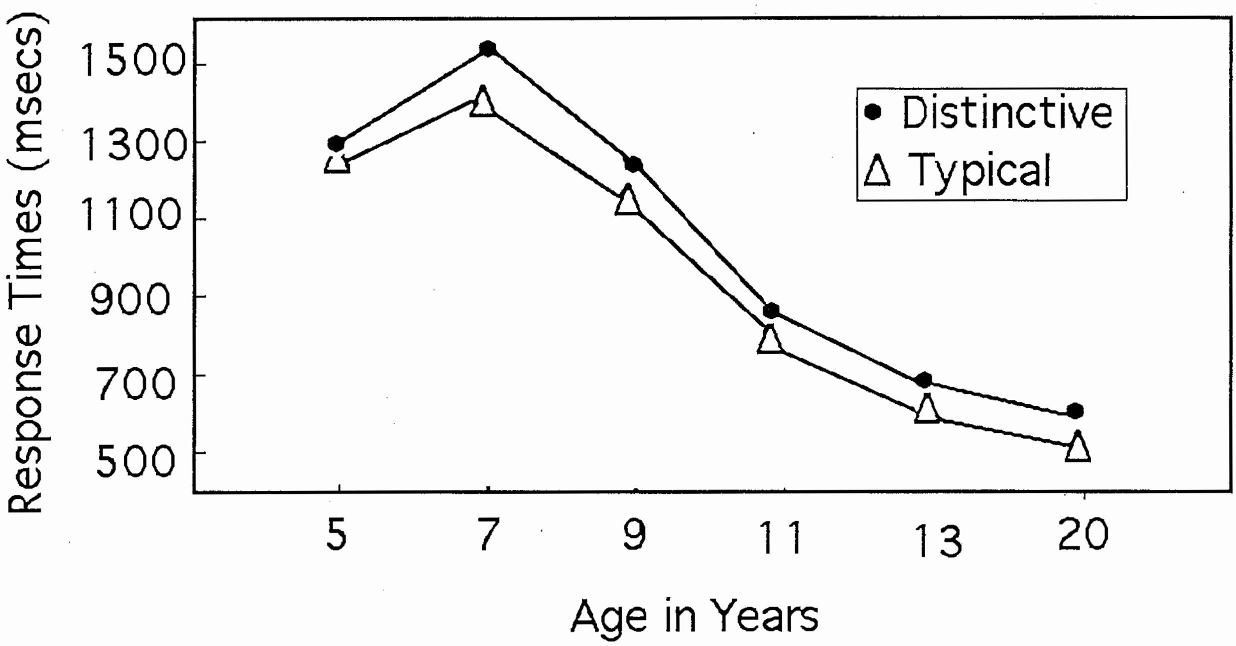
While all of the above hypotheses provides an adequate explanation for the distinctiveness effect in the recognition memory experiment paradigm, there may be ways in which distinctiveness effects can be manifested with which they have more difficulty. In all the above cases the differential processing of distinctive and typical faces is based on differences in the distance/angle between a target face and its nearest neighbour. In the young child's face-space it appears that the distance between a typical face and its neighbour is little different from the distance between a distinctive face and its neighbour. There is another way in which distinctiveness effects are manifested with respect to face processing by adults, however, and this phenomenon provides a means to discriminate further between these four options.

Some years ago, Valentine and Bruce (1986a) devised a paradigm which they called the face classification task in which subjects are shown either intact or jumbled faces and are asked to make a judgement of whether the target stimulus is a face or not. Valentine and Bruce demonstrated that "facedness" decisions could be made faster to typical rather than distinctive faces. Valentine (1991b) has explained how both the exemplar-based model and the norm-based model can permit this effect. Fortuitously, the mechanisms which Valentine outlines are quite different for norm-based and exemplar-based architectures. The exemplar-based model offers an explanation which is based on exemplar density. To classify the presented stimulus as a face it is necessary to judge how closely that stimulus resembles the central tendency of the population of all known faces. Consequently, the high exemplar density which impedes recognition of a previously-encountered face can be advantageous when the task is simply to determine that a face is a face. The norm-based model predicts this same effect but via a different mechanism. Classifying a face is considered to depend on how closely the stimulus resembles the prototypical or norm face. Therefore the time needed to respond will be related to the length of the vector derived from the stimulus face, i.e. the distance between the face and the norm. Coincidentally, the length of this vector will usually be negatively correlated with exemplar density, but the effect is independent of this feature.

Given these observations it should be possible to exploit this difference in order to reduce further the number of possible arrangements for describing a child's face space and indeed Johnston and Ellis (In press) were able to do just that . They managed to show experimentally that children of the same age groups tested in their original experiment on recognition memory (5, 7, 9, 11 and 13 years) were able to classify typical faces as 'faces' more rapidly than they were able to respond to distinctive faces. Subjects of all age groups exhibited the characteristic adult advantage for classifying typical faces as faces (see Figure 3).

We are thus in a position to explore how the four accounts (N1, N2, E1 and E2) which were described above are capable of accommodating two different manifestations of the distinctiveness effect in the youngest proposed face space - that of the 5 year old subjects. In other words, which combinations of the architectures derived from Ellis (1992) and Valentine's (1991a) exemplar-based or norm-based alternatives will provide the most efficient explanations?

Figure 3: Time to classify correctly stimuli as faces.



N1: As children of 5 years of age show the advantage for typical faces in the facedness task but not for distinctive faces in a recognition memory task it is preferable to explain each of these phenomena by a different mechanism. This account of the young child's face-space provides just such an opportunity. The density gradient in the child face-space is not as steep as in the adult version, but the relative distance from the norm of distinctive and typical faces is the same since the child's space shares the parameters of the adult space. As a consequence of this, neighbouring faces will be generally no closer to target typical faces than they are to target distinctive faces and so there is no advantage for recognition of the latter. However, distinctive faces will usually be farther from the norm compared with typical faces and so distinctive faces will require more time to be classified as a face.

N2: This explanation accounts for the lack of a recognition advantage for distinctive faces by suggesting that the face-space is compressed on to fewer dimensions and, consequently, that the distinctive faces are now as disadvantaged (in terms of closeness of neighbour) as typical faces. Unfortunately, this arrangement does not fit easily with the finding that 5 year olds do show a faster classification time for typical faces compared with distinctive ones. If the time to make a facedness decision is dependent on the distance between a representation and the norm, we should not expect the distinctive-typical difference in the 5 year old group to be present or at least as strong as in the adult group. In the 5 year olds' face-space the difference in relative distance from the norm is very much reduced. Not only is it unexpected to find this effect at all, but an examination of Johnston and Ellis' data revealed that the difference between response times to typical and distinctive faces for 5 year olds and adults was 13% and 3.5%, respectively, of the mean time required for a facedness decision. Account N2 would predict that if the typical face advantage were to occur at all then this difference should lie in the opposite direction.

E1: This account relies on the dissociation of exemplar density from the distance between a target face and its nearest neighbour. As in N1, it is assumed that in the 5 year old face space neither distinctive nor typical target faces have nearest neighbours close enough to interfere in the recognition task. The advantage for typical faces in the facedness task is explained by the fact that the distance from a face to its nearest neighbour can be independent of the exemplar density of the region in which it is situated. For example, a face may only suffer interference in the recognition task if its nearest neighbour is closer than κ units away. The closest neighbour to typical faces in Figure 1b may be at 2κ units and the closest neighbour to distinctive faces may also be at 2ʜ units (or further) and so there would not be an observed advantage for distinctive faces in a recognition memory experiment. For typical faces, however, there may be many more neighbouring faces at a distance of 2ʜ compared with the situation for distinctive faces. This means that typical faces still have an advantage in the face classification task. But this account does not cope so effectively with the facedness results. While this model allows a differential density gradient for the areas surrounding distinctive and typical faces it should not be as steep as in the adult version and certainly cannot be steeper. Once again this account suggests an absence, or at least dissipation, of the typical advantage in the facedness task which does not fit with what was found.

<u>E2</u>: The last solution to be considered also involves a compression of face representations owing to faces being coded on fewer dimensions. However, while there are fewer faces, there are also likely to be a smaller number of dimensions and so we expect that the distribution of faces through this space will reflect the same arrangement as in an adult space: (i.e., a steep density gradient with typical faces in the area of greatest exemplar density). This explains the facedness result in a parsimonious way; as in the adult space, there are simply more faces around a typical face which help in making a facedness decision. Using this explanation to account for findings in the recognition memory task needs more detailed consideration. Essentially, this is the converse of E1. Once more it is dependent on a disentanglement of the distance from a target face to its nearest neighbour and the local exemplar density of the target face. Here it is assumed that if the recognition of a face is impaired when the nearest neighbouring face is less than ʜ units away then all faces - typical and distinctive - have neighbouring faces within that radius. This is a more persuasive account for the lack of a distinctiveness advantage in the recognition task. It also suggests, that for the purposes of the recognition task, the child face-space contains face representations which are all effectively typical.

I will now try to weigh up these various solutions. First we can dispense fairly rapidly with N2. This explanation cannot properly handle the data from either task. The account admits differential numbers of neighbours for distinctive and typical faces and yet there is no recognition advantage for distinctive faces. In addition it suggests the typical advantage in the facedness task should be absent or at least weakened in the youngest group and yet this is as strong as in adults.

Solution E1 suggests that in a child's face space all faces are effectively distinctive. This is not an unreasonable proposition as we could imagine that a face which may seem very typical to an adult because s/he has already seen many similar ones could still be very distinctive to a child with a less extensive experience of faces. A consequence of this, however, should surely be that a child's accuracy in a recognition task should be as good if not better than that of an adult. The data which was collected did not suggest that this was the case. The children were revealed as less accurate than the adult subjects (see Figure 2). This is also at odds with data from elsewhere that suggest children are troubled more than adults by similarities between faces (e.g., Ellis et al, in prep; Flin, 1983). If children find faces so distinctive why are they so liable to category-inclusive errors? Also the typicality advantage in the facedness task fits uncomfortably with this model. Assuming an exemplar density gradient which is not as steep as the one present in the adult suggests that the typicality advantage should be reduced or eliminated in the child face-space but instead it appears as strong or even stronger than in the adult space.

On the other hand, the explanation offered in E2 to account for the lack of a distinctiveness effect appears more likely. If all faces are coded as representations which makes them 'typical' (as defined by target faces having a nearest neighbour close enough to impair recognition); then the poorer rates of accuracy for younger children would be expected.

Even more evidence is available from Valentine (1991b), he considers that the representations of other-race faces in an adult face-space may be compressed because 'the dimensions of the space might be inappropriate' (1991b, p.120). Johnston and Ellis (In press) proposed that while the dimensions of the child's face-space displayed in Figure 1c are not inappropriate, but they may be less appropriate - particularly if we take into account the assumption that the adult face-space employs further dimensions to facilitate storage. It is reasonable to suppose that a similar compression contributes to the phenomenon of all faces being typical.

The proposed explanation of N1 is able to fit very easily with the data described here. This version explains the differential processing of typical and distinctive faces in the recognition memory task and the facedness task through different mechanisms. Therefore we do not need any further elaborated rationalisations for the performance of 5 years olds in the two different tasks.

Consequently Johnston and Ellis (In press) suggested that there are two viable alternatives for the nature of the child face space; N1 or E2. Either children have a face-space based on the adult parameters but which is less densely populated with faces (Figure 1b) and is implemented as a norm-based account. Alternatively, children have a face-space which is compressed because faces are coded on fewer dimensions but where there is a normal distribution of face representations (Figure 1c) and this is implemented as an exemplar-based account.

Given that both the accounts outlined in N1 and E2 could be good descriptions for this data, would it be possible to decide between them? Certainly at a general level the preference would be for an exemplar-based account. Indeed, Valentine and Endo (1992) have pointed out that an exemplar-based account is the more parsimonious since it does not require the storage and abstraction of a norm and yet can still reconcile all the empirical recognition phenomena. Nevertheless, is it possible to find stronger empirical reasons on which to base a decision?

It has been suggested that one route to this may lie in taking a further look at how children are able to discriminate among faces. Moving on from the findings of Ellis et al. that young children have more difficulty than adults in rejecting similar alternatives to a familiar face, workers have tried to examine how they fare with unfamiliar faces. In order to do this it was first necessary to establish some firm baselines with adult subjects with which to compare children's performance.

4 Multidimensional scaling studies

It is acknowledged that Valentine's multidimensional model is a powerful way of understanding a large amount of existing face recognition data relating to the processing of distinctive tasks. Up to this point, however, I have only referred to two ways in which a distinctiveness effect is often seen (in recognition memory tasks and with facedness decisions). Nevertheless, there are other experimental paradigms where this dimension has an influence (e.g., familiar distinctive faces can be recognised as familiar faster than familiar typical faces, Valentine and Bruce, 1986; distinctive faces are learned more easily than typical faces, Ellis, Shepherd, Gibling and Shepherd, 1988). It is interesting to reflect, however, that researchers have moved to a multi-dimensional model for faces on the basis of experiments using faces rated along a unidimensional scale. A scale where the labels "typical" and "distinctive" represent the ends of a continuum. In a usual rating exercise subjects would be shown faces and required to rate them along a scale of typicality/distinctiveness. To give a particular example, subjects could be asked to imagine they are looking for this person at a railway station and should try to imagine how easy/difficult this would be to do; the easier they think it would be to spot the face, the higher the distinctiveness rating they should give to it (from Bruce and Valentine, 1986). Consequently, faces are attributed a single value dependent on their overall distinctiveness and two very 'different' faces could have an identical score. (This could also apply to faces which are rated as relatively typical, but of course in terms of the Valentine framework this would be less of a problem.) It is also worth passing mention that, although what constitutes distinctiveness is hard to put into words, subjects do not find classifying faces in this way an unreasonable or ridiculous request. Moreover, there is a good degree of reliability across subjects.

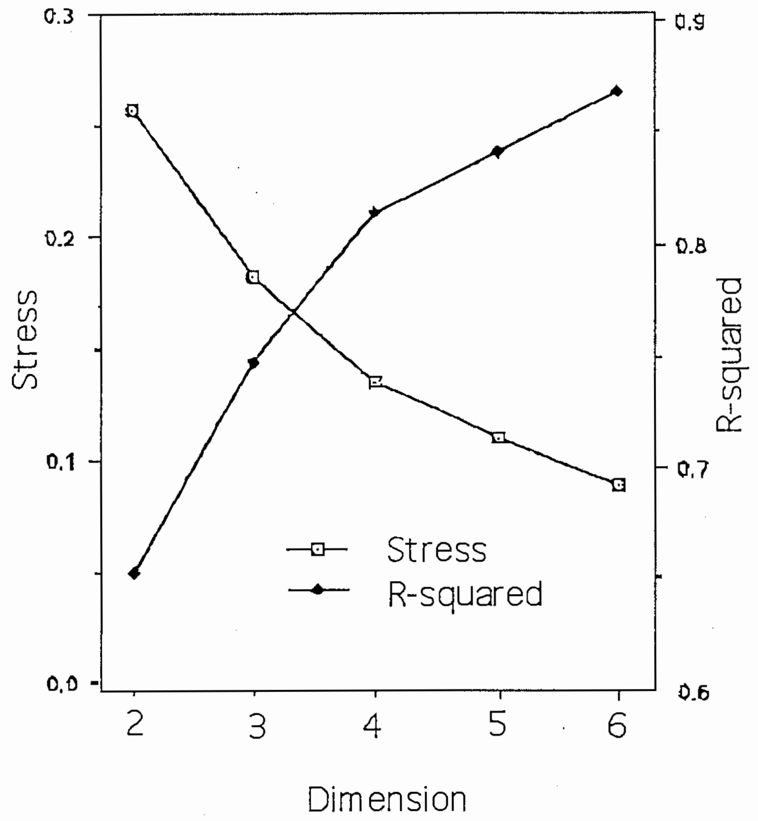
In order to explore this 'curiosity' Johnston, Milne, Williams and Hosie (submitted) took a set of faces already rated for distinctiveness and asked subjects to make similarity judgements to pairs of sequentially presented faces. Their intention was to try to explore the face space using MDS techniques. Subjects were asked to make similarity judgements but they were not instructed as to what aspect or features of the faces to use. It was predicted, however, that it would be possible to determine the attributes that were considered important from the subsequent analysis. MDS also permitted an estimation of what is an appropriate number of dimensions employed to encode faces.

Ratings were collected for every possible combination of 36 faces (18 distinctive and 18 typical), ratings were made for each possible pair in both directions (Face A followed by Face B and also vice versa).The ratings given to each possible pairing of different faces were collapsed across subjects and order of presentation to produce a set of 630 similarity ratings. The resulting matrix was submitted to the SPSS-X procedure ALSCAL

(Takane, Young, and de Leeuw, 1977). Five analyses were performed, generating solutions in two to six dimensions. The goodness-of-fit measures S-Stress and r^2 are shown in Figure 4. As can be seen from this graph, the improvement in r^2 levels off beyond dimension 4 indicating that the 4 dimensional solution provides an adequate representation of the data.

Correlation coefficients were computed between the linear distinctiveness rating of each face and the distance it lay from the origin of the space for all dimensional solutions. For all five spaces there was a high positive correlation between distance from the centre of the space and the distinctiveness rating. This supports Valentine's proposal that distinctive faces will tend to be located on the periphery of the face space, while typical ones will be located more towards the centre. Analyses of the higher-order constructs indicate that the arrangement of typical and distinctive faces is similarly consistent with Valentine's proposals.

Figure 4: Stress and R-square as a function of dimension of ALSCAL solution.



The difference in spatial position between the groups of faces which had been originally designated as typical or distinctive were investigated using t-tests and in every dimensional space it was found that the mean distance of the distinctive group of faces from the origin is greater than the mean distance of the typical group of faces. This provides neat support for the architecture of the face space which has been proposed by Valentine (1991a,b).

While the above findings are clearly coherent with a multidimensional face space, they cannot directly inform us about whether than space should be thought of as norm-based or exemplar-based. Each of these alternatives could underlie the arrangement of face representations. To make use of these data in a way which can help solve this puzzle it is necessary to have an independent measure of the ease/difficulty with which faces are discriminated. This was obtained by recording response latencies to decide if pairs of sequentially-presented faces were the same or different. Interobject visual similarity has been measured by the time required to say that two objects are not identical (Podgorny and Garner, 1979). The more similarity between two items the longer it will require to decide that they are different.

A series of experiments has been conducted which at first produced single comparison latencies for all possible combinations of faces, and then produced latencies for several presentations of the same pairs. In the latter case the faces chosen were a subset of the original 36 faces which included the eight faces closest to the origin and the eight faces most distant from it (Johnston, Milne, Williams and Ellis, in prep).

Taking the face spaces built up from similarity ratings it is possible to designate the distance between any pair of faces in two ways. First, it could be assumed that a norm-based account is appropriate and thus we should measure the distance between two faces by determining the angle between the vectors which represent them. Alternatively, it could be assumed that an exemplar-based account is appropriate and view the face representations as simple points in space. In the latter case the distance between faces should be computed as their straightforward linear separation. To some extent, these two solutions will be correlated with one another: faces separated by a large angle may also be separated by a large distance. However, this will not exclusively be the case. It is entirely plausible that one of these solutions will generate a better fit than the other for the reaction time data.

There are two measures available from the response latency data to determine how appropriate it is, but only one of these will have any direct bearing on the norm-based versus exemplar-based discussion. Each model posits that distinctive faces are further away from the origin of the space and located in an area of reduced exemplar density than typical faces. Whatever constitutes 'distinctiveness' it should be easier to assert that two copies of a distinctive face are identical because a distinctive face will have fewer neighbours to make this decision more confusing. In line with this we have found significant negative correlations between the distance a face lies from the origin of the face space and the time needed to decide that a pair of identical faces are the same (i.e., the farther a face is from the origin the faster subjects are able to respond on identical trials). In both of these studies this effect has proven significant.

The second measure that is available relates directly to the issue of norm-based versus exemplar-based models. By correlating the measure of response times for each pair of different faces with the separation of that pair, based either on the angle between two vectors or the distance between two locations in space it should be possible to determine which model provides the better fit for the data. In the first experiment, where single decisions to every pair were available, very small negative correlations were found between the separation of pairs of faces and the time required to make a 'different' judgement. Nevertheless, these were significant and there was a better correlation between response times and separation based on linear distance between faces i.e. the measure appropriate to an exemplar based space. The exemplar-based account is at an advantage over the norm-based model for all solutions from 2 to 6 dimensions.

When the study was repeated using a subset of the original faces and response latencies were collected for multiple presentations of each pair the same result was found. The farther the faces are apart on the exemplar-based solutions the faster subjects are to say that faces are different. In the latter experiment the correlation coefficients are much larger for the exemplar-based space (r reaches values above -0.4 for 120 comparison pairs a highly significant result $p<0.001$). Although it must be recognised that these correlation coefficients only account for a relatively small portion of variance in the response times, they do provide a good basis for choosing one model over the other.

To return to the original objective, of trying to determine how spaces were arranged in children experimenters returned to asking younger subjects to make similarity judgements to pairs of faces (Johnston, Ellis, and Williams, in prep). In the original adult study (Johnston, Milne, Williams and Hosie, submitted) subjects were asked to rate pairs of faces along a seven point scale where a rating of 1 meant that the pair was of two identical faces and a rating of 7 meant that they were very different. This task was considered too hard for young subjects. Also, constraints due to limited access to subjects and the attention span of young children dictated that it would not be possible to have subjects rate every pair of faces. In order to overcome these difficulties Johnston, Ellis, and Williams decided to ask subjects to rate pairs of a smaller subset of faces. Four distinctive faces and four typical faces were selected; the choice was based on the distance of these faces from the origin of face spaces that had already been constructed. In this case subjects were asked to rate pairs of faces along a three point scale. No identical pairs of faces were shown in a further attempt to minimise confusion for younger subjects. A rating of 1 indicated that two faces were very similar and a rating of 3 meant that two faces were very different. This arrangement allowed for twenty eight comparisons to be made in total. This study was conducted across a range of ages from 5 to adult (5, 7, 9, 11, 13 and 20 years of age).

Mean ratings of the similarity of all twenty eight comparisons were obtained for each age group. These ratings were then correlated with the separation of these faces in the original models according to either the norm-based or exemplar-based metric. Here we find significant positive correlations for judgements of dissimilarity and distance apart in an exemplar-based space. Correlations were smaller and non-significant for the ratings with the norm-based model separations. If faces were far apart in terms of an exemplar-based face they were judged as being more dissimilar across all the age groups. This is particularly impressive in that the original formulation was made on a seven point scale and still hold up when made on a three point scale by new subjects.

5 Conclusion

In conclusion this suggests that the most appropriate way to understand the developing face space is as an exemplar-based model. Earlier I suggested that either solutions N1 or E2 were the better way to understand how the face space in young children was organised, I am now able to conclude tentatively that E2 is the preferable one of this pair. The child face space should be considered to have the architecture illustrated in Figure 1c (i.e., an arrangement with fewer dimensions that the adult space) with an underlying exemplar-based implementation. Of course this leaves open questions of what the 'missing' dimensions will be, whether they are implemented gradually or whether there is a sudden shift from a child space to an adult space, and indeed how they added at all, but by a judicious combination of experimental techniques it should be possible to begin to tackle what hitherto seemed impossibly difficult theoretical questions concerning the way children learn to process faces.

Acknowledgements

I am extremely grateful to Dr S. Akamatsu and Dr T. Kato for their invitation to work at ATR. Their kindness in putting half a world between myself and my students has provided me with a valuable opportunity to think more deeply about the work described in this paper.

References

Bartlett, J. C., Hurry, S. and Thorley, W. (1984). Typicality and familiarity of faces. Memory and Cognition, 12, 219-228.

Bateson, P.R.G. (1977). Testing an observer's ability to identify individual animals. Animal Behaviour, 25, 247-248.

Bruyer, R., Laterre, C., Seron, X., Feyereisen, P., Strypstein, E., Pierrard, E., & Rectem, D. (1983). A case of prosopagnosia with some preserved covert remembrance of familiar faces. Brain and Cognition, 2, 257-284.

Carey, S. (1992). Becoming a face expert. Philosophical Transactions of the Royal Society of London. B, 335, 95-103.

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

Diamond, R. and Carey, S. (1986). Why faces are and are not special: an effect of expertise. Journal of Experimental Psychology: General, 14, 107-117.

Ellis, H.D. (1981). Theoretical aspects of face recognition. In G. Davies, H. Ellis, & J. Shepherd (Eds). Perceiving and Remembering Faces. London: Academic Press.

Ellis, H.D. (1990). Developmental trends in face recognition. The Psychologist, 3, 114-119.

Ellis, H.D. (1991). The development of face processing skills. Final report to the ESRC on Grant XC15250003.

Ellis, H.D. (1992). The development of face processing skills. Philosophical Transactions of the Royal Society of London. B, 335, 105-111.

Ellis, H.D., Ellis, D. M., and Hosie, J. (In prep). Children's processing of similar and dissimilar faces.

Ellis, H.D., Shepherd, J.W., Gibling, F. & Shepherd, J. (1988). Stimulus factors in face learning. In M.M. Gruneberg, P.E. Morris, & R.N. Sykes (Eds.), Practical aspects of memory: Current research and issues: Vol. 1. Memory in everyday life (pp 136-143). Chichester, U.K.: Wiley.

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

Flin, R.H. (1980). Age effects in children's memory for unfamilar faces. Developmental Psychology, 16, 373-374.

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

Goldstein, A. G. and Chance, J. E. (1964). Recognition of children's faces. Child Development, 35, 129 -136.

Goldstein, A. G. and Chance, J. E. (1980) Memory for faces and schema theory. Journal of Psychology, 105, 47-59.

Johnston, R.A. and Ellis, H.D. (In press) "Age effects in the processing of typical and distinctive faces.

Johnston, R.A., Milne, A.B., Williams, C.L. and Hosie, J. (submitted). Do distinctive faces come from outer space: An investigation of the status of a multi-dimensional face space.

Johnston, R.A., Milne, A.B., Williams, C.L. and Ellis, H.D. (In prep). Determining between two models of face space.

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

McNeil, J.E. & Warrington, E.K. (1993). Prosopagnosia: A Face-specific disorder. Quarterly Journal of Experimental Psychology, 46A, (1), 1-10.

Mann, V.A., Diamond, R. and Carey, S. (1979). Development of face voice recognition: Parallels with face recognition. Journal of Experimental Child Psychology, 27, 153-165.

Podgorny, P. and Garner, W.R. (1979). Reaction time as a measure of inter- and intraobject visual similarity: Letters of the alphabet. Perception and Psychophysics, 26, (1), 37-52.

Rhodes, G. (1988). Looking at faces: First order and second order features as determinants of facial appearance. Perception, 17, 43-63.

Shepherd, J.W., Gibling, F. and Ellis, H.D. (1991). The effects of distinctiveness, presentation time and delay on face recognition. European Journal of Cognitive Psychology, 3, 173-146.

Valentine, T. and Bruce, V. (1986a). The effects of distinctiveness in recognising and classifying faces. Perception, 15, 525-535.

Valentine, T. and Bruce, V. (1986b). Recognising familiar faces: The role of distinctiveness and familiarity. Canadian Journal of Psychology, 40, 300-305.

Valentine, T. (1991a). A unified account of the effects of distinctiveness, inversion and race on face recognition. Quarterly Journal of Experimental Psychology, 43A, 161-204.

Valentine, T. (1991b). Representation and process in face recognition. In: Watt, R. (ed.) Vision and visual dysfunction. Vol. 14: Pattern recognition in Man and Machine. (series editor J. Cronley-Dillan). London: MacMillan.

Valentine, T. and Endo, M. (1992). Towards an exemplar model of face processing: The effects of race and distinctiveness. Quarterly Journal of Experimental Psychology,

Vokey, J. R. and Read, J.D. (1992). Familiarity, memorability, and the effect of typicality on the recognition of faces. Memory and Cognition, 20, (3), 291-302.

Young, A.W., Hay, D. and Ellis, A.W. (1985). The faces that launched a thousand slips: Everyday difficulties and errors in recognising people. British Journal of Psychology, 76, 495-523.