From pixels to people: a model of familiar face recognition

A. Mike Burton[1], Vicki Bruce[2] & P.J.B. Hancock[2]

1. University of Glasgow, Scotland
2. University of Stirling, Scotland

Address correspondence to

Mike Burton
Department of Psychology
University of Glasgow
Glasgow G12 8QQ
Scotland, UK

mike@psy.gla.ac.uk

This research was supported in part by an ESRC grant to Vicki Bruce and Mike Burton (ESRC GR 00023 4573) and in part by a SERC grant to Mike Burton, Vicki Bruce & Ian Craw (GRH 93828).

Abstract

Research in face recognition has largely been divided between those projects concerned with front-end image processing and those projects concerned with memory for familiar people. These *perceptual* and *cognitive* programmes of research have proceeded in parallel, with only limited mutual influence. In this paper we present a model of human face recognition which combines both a perceptual and a cognitive component. The perceptual front-end is based on principal components analysis of images, and the cognitive back-end is based on a simple interactive activation and competition architecture. We demonstrate that this model has a much wider predictive range than either perceptual or cognitive models alone, and we show that this type of combination is necessary in order to analyse some important effects in human face recognition. In sum, the model takes varying images of "known" faces and delivers information about these people.

Introduction

Face recognition attracts interest from a very broad range of scientists. The issues surrounding this ability have been studied by neurophysiologists (e.g. Perrett, Hietanen, Oram & Benson, 1992; Gross, 1992; Sergent, Ohta, MacDonald & Zuck, 1994), cognitive psychologists (e.g. Bruce & Young, 1986; Rhodes, Brake & Atkinson, 1993; Ellis, 1992), social psychologists (e.g. Ekman & Friesen, 1976; Shepherd, 1989) and computer scientists (e.g. Kohonen, Oja & Lehtio, 1981; Pentland, Moghaddam & Starner, 1994; Lades, Vorbruggen, Buhmann, Lange, von der Malsburg, Wurtz & Konen, 1993; for a review see Chellappa, Wilson & Sirohey, 1995). These various disciplinary groups have brought converging evidence to the problem of how faces are recognised. In this article we attempt explicitly to bring together work from two previously rather disparate fields.

The two areas of research on which we will concentrate will broadly be called perceptual and cognitive aspects of face recognition. We use the term *perceptual* as a short-hand to denote those processes which allow mapping of a visual image onto a given representation or label. In these terms, the problem of face recognition is how to individuate a particular image, or how to label it the same as other images of that person. This is normally captured in artificial systems by postulating some core canonical representation for each known face, against which input patterns are matched in some way. The cognitive aspects of the system are those which follow analysis of the image. Once the face is individuated, we need to ask how information about that particular person is retrieved. To answer this question various cognitive models have been proposed, and we will discuss some of these below.

These two (perceptual and cognitive) research programmes have proceeded largely in parallel. For many researchers concerned with the perceptual aspects of the system, the problem is solved once an image has been given the appropriate label. In contrast, many researchers in the cognitive aspects of face recognition simply assume some initial processing to take place, and construct models of person recognition which are agnostic with respect to early processing of images. In this paper we will put forward a new model of face recognition which attempts to marry perceptual and cognitive aspects of the ability. We do this by attaching a "front-end" to an existing cognitive model of the process. We show that the resultant model provides the facility to examine phenomena outside the range of either perceptual or cognitive models which have preceded it.

The plan of the article is as follows. First we will briefly review the IAC model of face recognition. This is a model of the cognitive aspects of the processes which the authors have been developing over a number of years. We then briefly review the available candidates for providing a front-end, image-processing capability for this model. We describe in detail the chosen front-end architecture, which is based on Principal Components Analysis of images. We next describe the construction of the complete model. We then test the model against some human data on face recognition.

Finally, we answer some frequently asked questions about this model, and attempt to set out just what is and is not captured by the model.

1. Review of the IAC model

The IAC model of person recognition (Burton, Bruce & Johnston, 1990; Burton, Young, Bruce, Johnston & Ellis, 1991; Burton & Bruce, 1993; Bruce, Burton & Craw, 1992) is shown in Figure 1. The architecture is interactive activation and competition (McClelland, 1981). This is a very simple form of connectionist architecture comprising pools of simple processing units. Within pools all units inhibit one another (these links are not shown in Figure 1), and there are excitatory links connecting individual units across pools (these links are shown). Activation passes between units along these links, and in accordance with a standard unit update function (see Appendix). Following other IAC models, all links are initially of equal strength, and are bi-directional. There is also a global decay operation on units, which drives activation towards a standard resting state. The effect is to eliminate unit activation (over time) in the absence of input, and to stabilise unit activation in the presence of input.

We have used this architecture to extend previous functional accounts of face recognition, and in particular, that of Bruce & Young (1986). Following these early models, we propose a pool of units corresponding to classification of a face, these are called Face Recognition Units (FRUs). There is one FRU for each known face, and the notion is that these units are *view-independent*, meaning that any recognisable view of the face will cause activation in the appropriate FRU. The next level of classification occurs at the Person Identity Nodes, or PINs. This is classification of the *person* rather than the face, and once again there is one unit for each known person. At this level all domains for recognition converge. Figure 1, taken from Burton & Bruce (1993) shows convergence of face and name recognition. We would also expect other domains (e.g. voice recognition) to converge at the PIN.

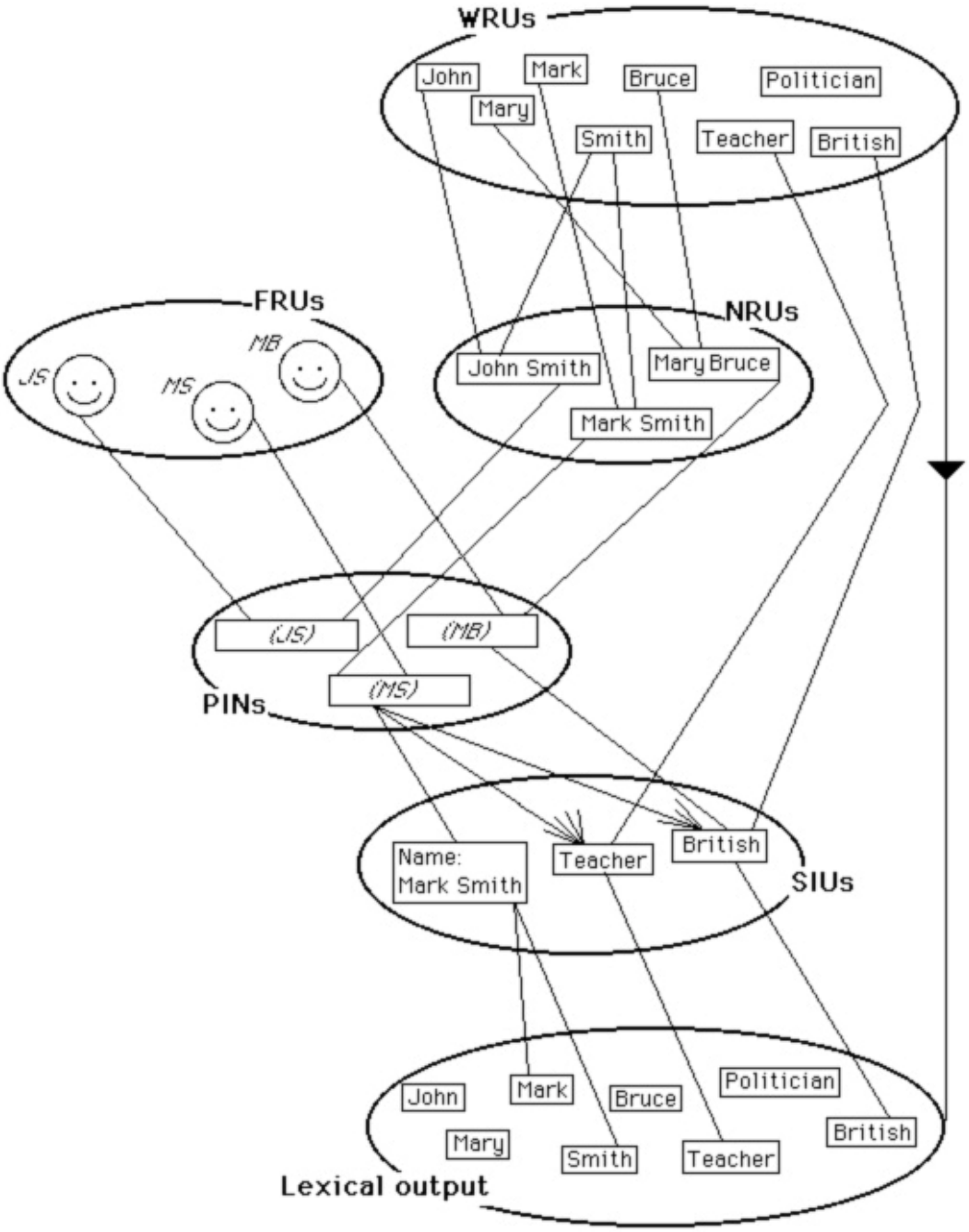

Figure 1: The IAC model of face and name recognition. From Burton & Bruce (1993).

A great deal of research in face recognition has examined the processes which allow people to decide that a face is familiar. Many experiments use the face familiarity decision task (Bruce, 1983) in which subjects make speeded familiar/unfamiliar judgements to a succession of faces. Similarly, research on naturalistic and laboratory-based breakdown has studied circumstances in which people can only state that they recognise a person as familiar, but cannot recall any further information (Young, Hay &

Ellis, 1985; Hay, Young & Ellis, 1991). We propose that the locus for familiarity decisions is the PINs. When any PIN reaches a common activation threshold, familiarity is signalled. This has the implication that the same decision mechanism is used for all person familiarity judgements, regardless of whether they are made to faces, names or other kinds of information. Note that the threshold is merely a device for signalling familiarity. There are no thresholds for passing activation within the model. Instead activation is passed in a cascade fashion throughout.

Following PINs, there is a pool labelled Semantic Information Units (SIUs) which code information about known individuals. Information about a person is coded in the form of a link between the person's PIN and the relevant SIU. Note that many SIUs will be shared (e.g. there may be many people represented with occupation "actor" or with nationality "British"). The notion is that activation of any of these units to a common threshold allows retrieval of that piece of information. Finally, there is a pool of units labelled "lexical output" which are intended to capture the first stage of processes involved in speech and other output modalities.

This model also includes a recognition route for domains other than faces, and the architecture for doing so represents an adaptation and rationalisation of an architecture put forward by Valentine, Brédart, Lawson & Ward (1991). There is an input lexicon, labelled WRUs (Word Recognition Units). Those words which code names (both forenames and surnames) have links directly to a pool of Name Recognition Units, or NRUs. These NRUs are linked to PINs in the same way as FRUs are linked to PINs. The WRUs which do not correspond to names are linked to SIUs. This is intended to capture the idea that words such as "Peter" will be processed as names, while words such as "British" will have access directly to their meanings. Finally, all WRUs are connected directly to the lexical output units. In this way, the model contains the elements of a "dual route" model of reading.

The architecture of this model is very simple and it clearly operates at a very coarse scale. For example, we have not implemented a detailed model of reading. Instead, we have intended to capture gross aspects of the architecture of person recognition which allow comparison and integration of recognition through different domains. Despite its simplicity, this model is able to capture a large number of previously enigmatic effects in the face recognition literature. For example, it offers explanations of semantic and repetition priming (Burton et al, 1990), covert recognition in prosopagnosia (Burton et al, 1991), name recall (Burton & Bruce, 1992), name recognition (Burton & Bruce, 1993), and (with some additions which we will discuss later) learning of new faces (Burton, 1994).

Readers are referred to original sources for details of these various accounts. However, we will give a very brief overview of two of these phenomena in order to provide a flavour of theorising in the model. Consider the phenomena of priming in face recognition. Like word recognition, face recognition gives rise to two types of priming, semantic (or associative) priming and repetition priming. Associative priming is most often demonstrated using the face familiarity decision task. Many researchers have

shown that a face is recognised faster if immediately preceded with the face of an associated person (e.g. Bruce & Valentine, 1986). For example, Stan Laurel's face is recognised faster if immediately preceded by Oliver Hardy's face. The IAC account of this phenomenon is as follows. First Stan Laurel's face is perceived, and the FRU corresponding to that face becomes active. This FRU activation causes activation in the relevant PIN, and then in the SIUs related to that PIN. Now, many of the SIUs related to Laurel will also be related to Hardy. As links are bi-directional, some activation passes back from these SIUs to Hardy's PIN. If we subsequently activate Hardy's FRU, his PIN (which now has some above-rest activation) will rise to the recognition threshold faster than would be the case had the unit started at rest. This is the simple account of semantic priming: recognition of a person causes sub-threshold, but above-resting, activation of the PIN of an associated person, and this is exploited by subsequent presentation of the primed person's face.

This account of semantic priming has two attractive features. First, it predicts that priming will cross domains. As the effect relies on activation in the PINs, at the point where domains converge, recognition of a name should prime subsequent recognition of both a name and a face. Indeed, this is pattern found in empirical studies (Young, Hellawell & de Haan, 1988; Young, Flude, Hellawell & Ellis, 1994). Further, we would expect the effect to be short-lived. Transient activation of a PIN is obliterated by subsequent recognition of other people. The within-pool competition ensures that this is the case. We therefore predict that priming will be attenuated by an intervening stimulus item, and will be short-lived. Again, this is the pattern found in the literature (Bruce, 1986).

We now consider repetition priming. This refers to the fact that a face is recognised faster if that same face has been seen previously (Bruce & Valentine, 1985; Ellis, Young, Flude & Hay, 1987a). This effect is comparatively long-lasting (experiments standardly use at least a 20 minute gap between prime and test phases), and is strongest when the prime and target are the same image, but still present when different images of the same person are used as prime and target (Bruce & Valentine, 1985; Ellis et al, 1987a). We have proposed that this phenomenon can be captured by global Hebbian strengthening in the model. When a face is recognised as familiar, the person's FRU and PIN will both become active. If we allow simple Hebbian updates, then the link between the FRU and PIN will become strengthened. This means that the next time that person's face is seen (i.e. the person's FRU is activated), it will take a shorter time for the person's PIN to become active. Note that this account does not predict cross domain priming: strengthening an FRU-PIN link will not aid subsequent recognition of a name (which uses the NRU-PIN route). This pattern has been demonstrated many times: at the intervals used in this type of research, repetition priming does not cross domains (Ellis et al, 1987a, Ellis, 1992; Ellis, Flude, Young & Burton, in press).

This brief account of priming is intended to provide a flavour of the level of theorising available in this model. Full details of simulations demonstrating these two priming effects can be found in Burton et al (1990). In summary, the model, simple as it is, provides a coherent account of a number of phenomena. It has also been predictive.

Researchers have used it to derive hypotheses about work in cognitive neuropsychology (de Haan, Young & Newcombe, 1991), priming between part faces (Ellis, Burton, Young & Flude, submitted), self-semantic priming (Young et al, 1994) and the development of face recognition (Scanlan & Johnston, in press).

Finally, we should also note that some peripheral aspects of the model remain the subject of debate. We have used a variant of the model, shown here, to provide an account of the difficulty people experience in retrieving names from faces (Burton & Bruce, 1992). Other researchers have challenged our account of this phenomenon (Brédart, Valentine, Calder & Gassi, 1995; Hanley, 1995) and have proposed that name retrieval mechanisms would be better located elsewhere in the model. This debate continues, and we will not rehearse it here. The important point to note is that the debate centres on the appropriate location of units which all agree are down-stream of the integration of perceptual and cognitive processes. None of the effects we describe below would be affected by differences between any of the currently available models of name retrieval. The integration we describe is therefore independent of the resolution of this debate, and we will not pursue it further here.

2. A front-end to the IAC model

The model described so far has been useful in exploring cognitive effects in familiar face recognition. However, its scope is clearly limited. In particular, the model makes no statements about processing of visual images. The scope of this model is consistent with a number of other functional approaches to the problem (e.g. Bruce & Young, 1986; Hay & Young, 1982; Ellis, Young & Hay, 1987). However, there are many phenomena in face recognition which require an understanding of the image processing aspects of the problem, and we will describe some phenomena below which require understanding across both perceptual and cognitive domains in combination. We are therefore faced with the problem: how do the FRUs become active in the first place? How might we implement a system in which FRUs act as localised *view-independent* units for individual faces?

In previous work (Burton, 1994; Ellis et al, submitted) we have suggested that faces are parameterised in some way. We proposed that the set of faces is represented by some combination of a set of known elements. However, we have not made any commitment to the nature of these elements. The simplest way to think about this is to imagine a photofit tool of the type developed for police work. In such systems, individuals' faces are represented as a combination of a set of elements. By combination in this way, it is possible to represent a very large number of faces with a relatively small number of constituent elements. The problem which remains is specifying the nature of the elements: what are the primitives of face recognition?

We have argued elsewhere that certain candidates for the primitives of face recognition are unlikely to be true. In particular, we have argued (Burton, Bruce & Dench, 1993; Bruce, Burton & Dench, 1994; Bruce, 1994) that descriptions based upon simple 2D measures in the picture plane are unlikely to form the basis for human face recognition. 2D picture plane measures remain constant over transformations which

render face recognition almost impossible for humans. For example, recognition of people represented in photographic negative is extremely poor (usually at floor) in studies of human perception (Galper, 1970; Phillips, 1972; Bruce & Langton, 1994). Moreover, recognition of faces from line drawings made by careful tracing of the outlines of face features such as eyes, brows and hairline are extremely difficult to recognise (Davies, Ellis & Shepherd, 1978; Rhodes, Brennan & Carey, 1987) even though such drawings should preserve all the measurements that might form the basis of facial descriptions. Line drawings of faces become recognisable when information about the pattern of light and dark from the original is added (Bruce, Hannah, Dench, Healy & Burton, 1992). This observation, together with effects of negation and dramatic effects of lighting direction (e.g. Hill & Bruce, in press) suggest that face descriptions are based upon image features rather than edge features. In what follows we explore one such description scheme based on image features.

3. Principal Components Analysis

The use of PCA on images has developed as a technique in engineering image processing. The advantage of the technique is that it delivers a radical data compression, and hence allows narrow-bandwidth communications channels to carry additional information. In this next section we give an introductory overview of the technique.

The aim of PCA is to deliver a new basis to a set of multidimensional data. The most commonly used form of PCA in psychology is the Factor Analysis technique developed for multivariate statistics. In factor analysis one has typically taken a number of measures on a set of cases. One then needs to know whether the variability in the measures can be coded on fewer dimensions. This is the same procedure as is used in PCA on images. Imagine an image represented in a computer. Such an image is simply an array of pixels, for example 10000 pixels for a 100x100 image. For each of these pixels a single number is stored in memory, representing the grey-scale value of that pixel (or its intensity). We can therefore think of any image of this size as a point in 10000-dimensional space. If we take many of these images, we can then perform a PCA treating each image as a case. The aim is to establish whether there is a smaller number of dimensions (smaller than 10000) on which the set can be described. PCA delivers a new set of axes, each of which can be displayed in an image of the same size as the originals. These new axes are called "eigenfaces" (Kirby & Sirovich, 1990). The original cases can be reconstructed by a weighted sum of these new axes (eigenfaces). The coefficients for this weighted sum are the new representation for that image, and the goodness of the reconstruction can easily be compared with the original image, i.e. it is possible to measure how well the new dimensions code the faces.

This technique was introduced into face recognition by Kirby & Sirovich (1990) and by Turk & Pentland (1991). They showed that faces could be represented in very few dimensions, and in many subsequent studies researchers have used as few as 50 eigenfaces. This represents a radical data compression, reducing the storage requirement per face from 10000 numbers (say) to 50 numbers. A good introduction and review of the technique can be found in Valentin, Abdi & O'Toole (1994).

Recently, researchers have begun to ask whether the PCA approach to face recognition might have some correspondence with human face perception. O'Toole, Deffenbacher, Valentin & Abdi (1994) have demonstrated that it provides a natural account of the other race effect. This refers to the fact that people show more errors in differentiating between members of another race than for their own race. The PCA account of O'Toole et al suggests that one's eigenfaces, generated to capture variation in the population of faces from one's own experience (and hence race) do not code faces from another race well. In short, one's eigenfaces will reflect the dimensions of variation in the faces one encounters. However, these dimensions of variation may not reflect the dimensions of variation of faces from another race. This makes these faces confusable. In our own laboratory, we have examined the psychological phenomenon of *distinctiveness* in relation to an eigenface coding. Like O'Toole et al (1994) we found that there is a large and reliable correlation between these measures: faces which humans find distinctive also tend to have extreme eigenface values (Hancock, Burton & Bruce, 1996).

PCA, like any image-based recognition technique, is prone to influence by spurious image factors such as the size and position of a face within the image. Most researchers address this issue by performing some standardisation of images before subjecting them to PCA. This typically takes the form of standardising eye-position for all faces. However, Craw (1995; Craw & Cameron, 1991) has provided a more effective standardisation. Craw's technique is to standardise the shape of the face before PCA. This is achieved by overlaying each face with a standard grid, with key points at the eyes, nose mouth and round the shape of the face. The faces are all then morphed to a standard shape, typically the average of all the images used. The resultant images, called "shape-free faces" by Craw (1995) are then subject to PCA. This means that the eigenfaces are completely independent of background - as all faces have the same shape, only those pixels within the shape are analysed. Secondly, it means that gross features of each face (e.g. mouth and nose) are in the same position for each face. An example of a manipulation to a shape-free face is shown in Figure 2. Third, the resulting eigenfaces can be combined in linear form and will give rise to face-like objects. To see this, imagine taking the image average of two faces with different shapes. The average will be a ghostly image with no clear boundaries. However, if the shape-free manipulation is performed first, the "average" of the two faces will itself be a face, with a standard boundary. These considerations have led various researchers interested in image-based face recognition to develop techniques for treating the shape of a face separately from the intensity information in the face, usually called its "texture" (Beymer, 1995; Vetter & Troje, 1995; Troje & Bülthoff, in press).

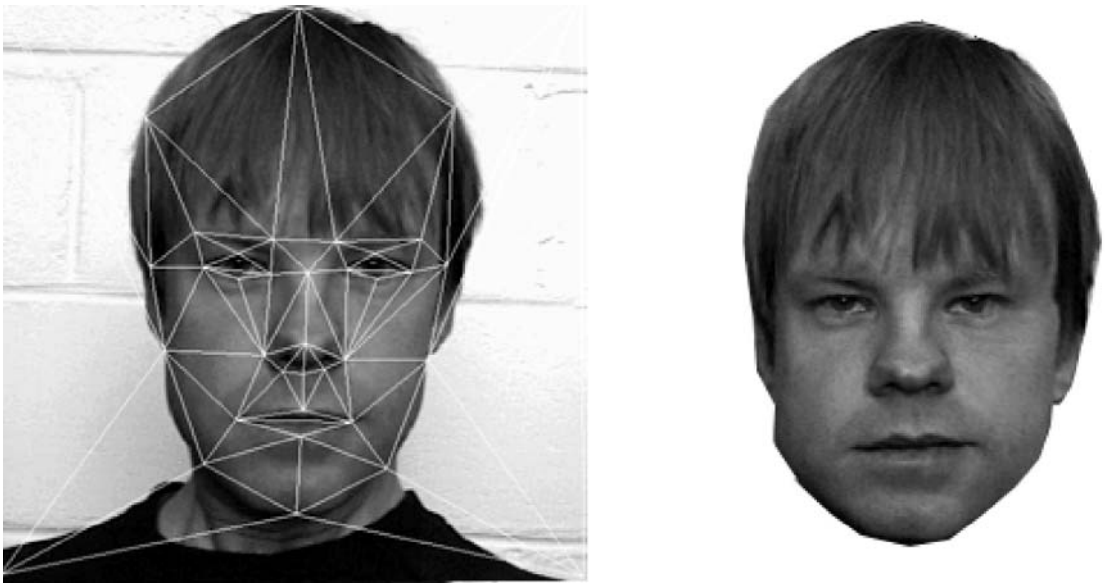

Figure 2a: An original face, showing the set of control points and triangulations used in the transformation to shape-free representation. 2b: The same face morphed to the average shape.

As with other techniques, the shape-free pre-processing manipulation allows one to examine the shape and the texture separately. It is possible to code a face in terms of shape-free eigenfaces, *in conjunction with* some representation of the original shape, e.g. the co-ordinates of the grid used in the original transformation. In the work described below, we will examine a system based on both shape-free information and shape, and compare it with a system based only on the shape-free information. It is worth noting that in preliminary studies Costen, Craw & Akamatsu (1995) have found that the shape-free information alone gives a good representation, and statistical recognition systems based on this information alone perform well. In previous work, we have found that shape-free faces give a good account of rated distinctiveness, and of some tests of memory for faces (Hancock et al, 1996). In fact, the term "shape-free" is perhaps unfortunate, as it suggests representations which are independent of a face's shape. In fact, this is not the case. The shape-free transformation (morphing) will produce images which nonetheless have residual information arising from the face's original shape. For example, the shading pattern arising from a big chin (say) will be different to the shading pattern arising from a small chin. When the two chins are morphed to the same shape, their patterns of shading will remain different. In this way, the influence of shape remains in the shape-free faces, and so it is perhaps not so surprising that the shape-free faces appear to be useful in themselves.

The shape-free manipulation has effects beyond those of simple outline shape. The morph transforms an original image (or polygon set) into a standard shape (or polygon configuration). This means that some information concerning expression is removed in the shape-free process. Furthermore, the manipulation eliminates small differences between different viewpoints. These manipulations seem to us to be desirable in a system based on image properties. In what follows we compare the efficiency of

different systems based on (i) raw images of faces; (ii) shape-free images of faces only; and (iii) independent contributions from shape-free images plus a representation of shape.

4. Description of the model

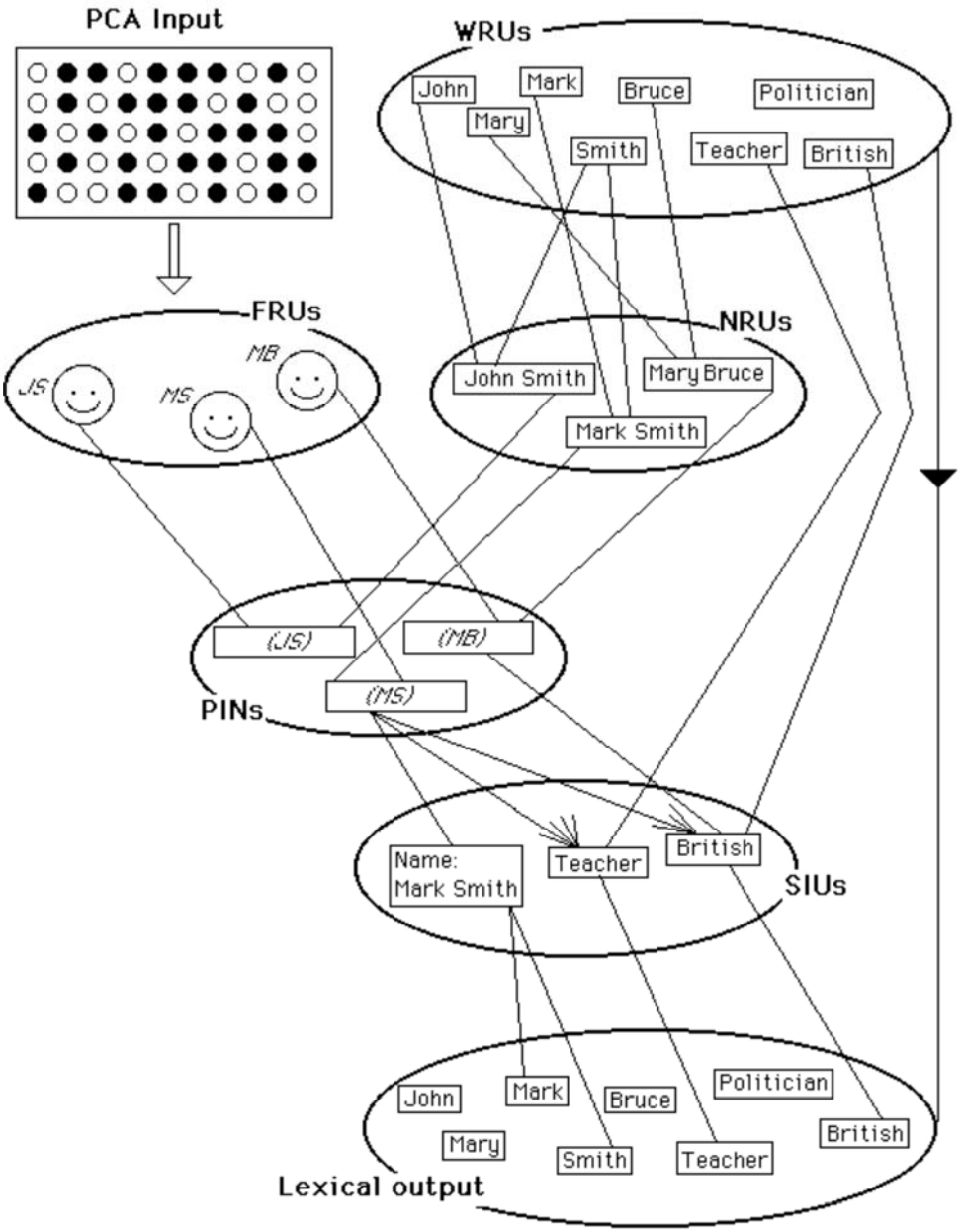

Figure 3: The IAC model with new PCA units

The model combining perceptual and cognitive aspects of face recognition is described in this section. The simulation was written in C using the Rochester Connectionist Simulator (Goddard, Lynne, Mintz & Bukys, 1989). Implementation details, and a list of global parameters are given in the Appendix. Figure 3 shows the

outline of the model. The IAC component, representing cognitive aspects of face recognition, is much the same as previous instantiations, and we will describe it in detail below. We will first describe the PCA front-end.

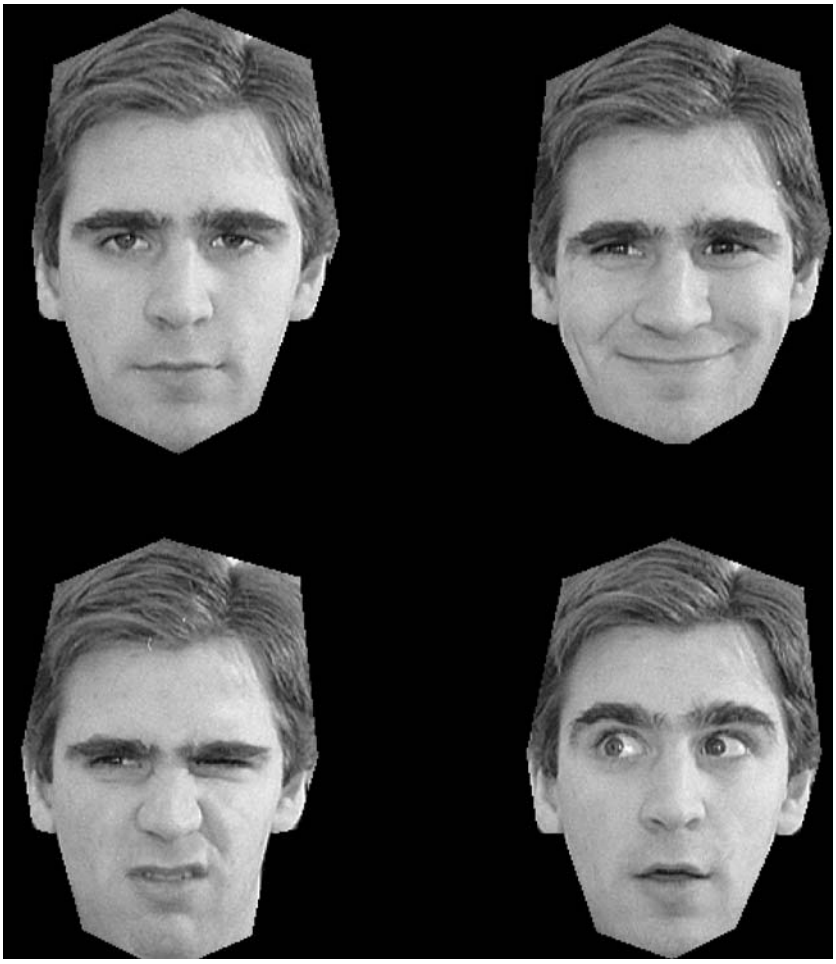

Figure 4: One of the faces used in the recognition set. 4a shows the neutral face. The remaining faces show variations typical of the set.

The model was constructed to "know" 50 people. To provide real face data for these representations, 50 young men were photographed. Each person was photographed in full-face view with a neutral expression, and a further twice or three times in full-face view with another expression. Figure 4 shows an example of one of the faces. Figure 4a is the neutral face, while the remaining are labelled "expressive". Figure 4 shows the amount of variation typically present in the faces photographed. By this procedure, we collected 50 neutral faces, and a further 136 expressing faces (comprising two or three extra photographs from these 50 people).

All photographs were captured onto grey-level (8 bit) computer images at resolution 280x240 pixels. Shape-free versions of all the images were generated by specifying the co-ordinates of 31 points on each face by hand. These co-ordinates were triangulated and used to morph all the images to the average shape of the 50 neutral

faces. The triangulation shown in Figure 2 shows the grid used for the morphing. Note that this process can be automated by feature finding (Craw, Tock & Bennet, 1992) or by one of the new generation of optic flow applications to extract shape (Troje & Bülthoff, in press). However, in this paper we are concerned explicitly with the interface between perceptual and cognitive models, and we therefore present the model free from any errors due to an automatic shape extraction procedure.

PCA was performed on three different types of data, representing the three different types of model we wish to compare. In all three cases, only the neutral faces were subject to PCA. First, the "raw image" data was subject to PCA. All images were scaled in 2d and aligned such that the eyes of each face were coincident. These images were then reduced to 50x66 pixels and the neutral images were subject to PCA. The first 50 components (eigenfaces) were extracted. The reconstruction coefficient for each component (the value of each eigenface dimension for the face) was stored for each face and we call this set of 50 numbers the signature of a face. In essence, we code each face as a set of 50 numbers, such that these can be used as coefficients in a weighted sum of eigenfaces which will reconstruct the original. Note that once the eigenfaces have been extracted, any new image can be coded in terms of these eigenfaces. We are not guaranteed that this coding will give a *good* representation of the new image, but it is important to note that images which were not used to generate the basis (eigenfaces) may nonetheless be coded by them. Of particular interest to researchers in this field is the robustness of the representation. Will it be the case that two views of the same face will have a similar signature?

The same PCA procedure was used for shape-free faces. The neutral faces were morphed to a common shape, representing the mean x,y co-ordinates for each point in the grid. The resulting shape-free faces were reduced to 50x66 pixels and subject to PCA, generating a shape-free signature for each face from the first 50 components generated. Similarly, a shape-free signature can be generated for each expressing face (i.e. those not used to generate the basis of eigenfaces).

Finally, we analysed the shape itself. This was done by performing PCA on the individual x and y co-ordinates for each of the 31 points in the grid. These co-ordinates were entered separately into a PCA, giving 62 points for each of the 50 neutral faces. The first 20 components of this analysis were used to code the "shape signature" of each face.

In order to code these faces into the various simulations, we store for each face only the coarsest of information about its signature. In the simulations described here we store only the sign of each component coefficient for each face. The PCA procedure delivers coefficients of mean zero across all faces for each component (eigenface) and so there is large discriminability in this set. For example, in the version of the model which codes 50 shape-free components, there are a possible 2^{50} patterns. This has the advantage of further data reduction, since the original faces are now stored as 50 bits.

The front-end system is tied to the IAC model through the FRUs. We tested three versions of this model. One in which the FRUs code the 50 component signature for the

raw image data, one in which the FRUs code the 50 component signature for the *shape-free* faces, and one in which the FRUs code the 50 component values for the *shape-free* signature plus an additional 20 component values representing the *shape* signature from that face. The implementations contain a new set of units, labelled PCA input units. So, in the *raw image* and *shape-free* models there were 50 PCA input units, while there were 70 in the *shape-free plus shape* model.

PCA input units are maximally connected to FRUs according to the signature of that face. If a face has positive value on component 1, and negative on component 2, a link of strength +1 is constructed between PCA unit 1 and that FRU, and a link of strength -1 is constructed between PCA unit 2 and that FRU. Input to the model is through the PCA units. A face is parameterised in the new PCA basis (into its signature) and converted into 50-bits or 70-bits of information (according to the model under test). This signature is presented to the PCA units, and activation propagates through to the FRUs.

We now describe the cognitive aspects of the model. The IAC component, also shown in Figure 1, is the same implemented model as we have used in the past (for example Burton & Bruce, 1993). The implementation codes knowledge about 50 people, and so there are 50 FRUs, NRUs and PINs. Each known person is connected to 6 SIUs, one coding that person's name, and a further 5 chosen at random from the available SIUs. There are 120 SIUs, with 50 of these coding individuals' names, and the remaining 70 coding other information (for example occupations and nationalities). This link arrangement means that some of the SIUs will be quite common, and others quite rare, due to the fact that the 70 general SIUs are used to select 5 random semantic links for each person. There are a further two pools of units: one coding word recognition units, and one coding lexical output units. There are 110 units in each of these pools, 10 coding forenames, 30 coding surnames and 70 coding general information. Name WRUs are connected to NRUs and non-name WRUs are connected to SIUs. There are also direct, unidirectional connections between each WRU and its corresponding lexical output unit. The name units are chosen as described in Burton & Bruce (1993): all surnames are equally common, but the frequency of forenames is variable. We will not use name frequency in any of the simulations described here.

There are some aspects of this model which are clearly artificial, and which need to be mentioned explicitly. First, this model "knows" everybody equally-well. So, the same number of facts is known about each person. This is clearly implausible. Second, all excitatory links and all inhibitory links within the model have equal weight. This means that all facts are represented as being equally well-known. Again this is implausible, for example the fact that Prince Charles is a royal is likely to be a stronger association for most people than the fact that Prince Charles studied in Cambridge. The decision to hold all these factors constant in simulations was taken to allow examination of the architecture *per se*. We have discovered in the past that judicious manipulation of local parameter values can result in quite differing behaviours. We are not interested to demonstrate the power of a particular implementation of this model, with particular parameter values. Instead, we would like to demonstrate the architecture in general. We

have no theoretical reasons for manipulating parameters locally, and so we continue with a generic system. For this reason, the results of our simulations are qualitative only. This aspect of the model is discussed in the final section, frequently asked questions.

5. Testing the model

5.1 Face recognition

One of the primary aims of the model is to implement a form of face recognition which is generalisable over some (certainly limited) range of viewing conditions. In the first test we present the model with the PC values (the 50-bit or 70-bit signature) for the various face images, and observe behaviour elsewhere in the system. Recall that the PIN level is the locus for familiarity decision. We describe as a *hit* the situation in which the correct PIN becomes most active on presentation of an image.

Table 1 shows the results of tests with the three different versions of the implemented model. First note that the implemented system recognises the neutral faces perfectly in each case. This is an unsurprising result (though not in fact necessary), because the models were hand-wired to code these neutral-expression faces. Of more interest are the expressing faces. The model based on raw face images, performs reasonably well, though worst of all the three versions. Removing shape from the images gives a substantial increase in the hit rate, to 95%. We next consider the model which codes both shape-free faces and information about the shape itself. In fact, this performs only very slightly better than the model coding the shape-free faces. Only a further two faces are identified. This is a very small gain for an extra 20 bits of input.

Table 1: Correct hits (correct maximally active PINs) for neutral and expressing faces in the three different versions of the model.

	Neutral faces (/50)	Expressing faces (/136)
raw image (50 bit)	50	113 (83%)
shape-free (50 bit)	50	129 (95%)
shape-free plus shape (70 bit)	50	131 (96%)

These results suggest that the shape-free representation is the most efficient tested, and this is the version of the model which we will examine in further simulations. Note, again, that the efficiency of shape-free representations does not mean that shape is unimportant.

We next examine the details of the performance of the shape-free version of the model. This model made 7 errors. Examination of these errors showed that no two errors were made to any individual person. So, for every individual known, the model was able to recognise at least one and usually two views which had not been explicitly coded (i.e. expressing faces). In two of the 7 mis-identified faces, the correct FRU became most active, but complex interactions between units led to the incorrect PINs becoming most active. In a further two, the correct PIN achieved second-most active status among the

PINs. On the basis of this performance, we progressed to different tests with the shape-free (50-bit) version of the model.

5.2 Multimodal input and cueing

It is possible to cue recognition of faces in this model through simultaneous presentation of a face and another piece of information. For example, if we activate a name unit (i.e. a forename or a surname in the WRU pool), activation will accumulate in the PINs (via the NRUs) of people having that name. Similarly, if we activate the WRU corresponding to a semantic fact (say "footballer") some PINs will gain some activation (via the SIUs). To test this facility in the model, we examined the seven face images mis-identified in the previous section. These faces each presented simultaneously with their correct forenames, or with a single WRU corresponding to a correct fact known about that person. In each of the seven cases, this single extra piece of information was sufficient to resolve the mis-identification. In each case, the correct PIN became most highly active.

In order to check for the possibility that WRUs might be having an overpowering effect on recognition, we presented 10 expressing images which were correctly recognised in the original test of face recognition. We examined recognition of each of these faces in two ways: first by presenting it simultaneously with an incorrect forename, and second by presenting it with an incorrect fact (i.e. a name WRU or a semantic WRU). The forenames and facts were chosen at random, and were all units which coded information about other people. In none of these simulations was the misleading forename or fact sufficient to suppress access to the correct PIN. In each case, the incorrect cue failed to prevent the PIN corresponding to the face from becoming most active. So, a small amount of information appears to be sufficient to resolve a difficult recognition problem, whereas a correspondingly small amount of information is not sufficient to destroy intact recognition. Of course, there may be situations in which an incorrect cue results in incorrect identification. However, this is not a general consequence of multi-modal input.

5.3 Distinctiveness

Distinctiveness effects manifest themselves in face recognition in two ways. For unfamiliar faces, those rated as distinctive are subsequently recognised with higher accuracy than those rated as typical. For familiar faces, those rated distinctive are recognised faster than those rated typical (Valentine & Bruce, 1986a,b). In order to test the model for typicality, we had the neutral faces rated. Subjects were shown each of the 50 neutral faces in turn and asked to rate each on a 15-point scale (where 1 is "very typical" and 15 is "very distinctive"). These same faces have been used in previous experimental work (Bruce, Burton & Dench, 1994) and these ratings were gathered for other purposes. In fact, each face had been rated twice, once (by 10 subjects) in their entirety, and once (by a further 30 subjects) with the hair concealed. Interestingly, these two ratings correlate significantly, but not very highly ($r=0.33$, $p < 0.05$).

In previous work we have demonstrated a link between human rated distinctiveness and PCA values. Hancock, Burton & Bruce (1996) showed that a face's signature could be used to predict distinctiveness, and to predict hits and false positives in a memory test using the same images. O'Toole et al (1994) were the first to show this effect on memorability with previously unfamiliar faces. However, in the model presented here we have, for the first time, the opportunity to relate PCA signature to distinctiveness effects with *familiar* faces. Recall that distinctive familiar faces are recognised faster than typical familiar faces. As the model presented here is intended to capture recognition of familiar faces, this effect should be within its range.

To test the performance of the model, we set a standard level of activation for a PIN which would act as a threshold for recognition of faces as familiar. We can then take the number of processing cycles needed to achieve this threshold as a measure of recognition latency. There is no theoretical reason to choose any particular threshold, and the absolute values of unit activations would change according to global factors such as the total number of units, or the overall strengths of links. We chose a value of 0.45 as the recognition threshold simply because it provided a reasonable spread across PINs and did not lead to floor or ceiling effects. Using this common threshold, we presented the model with the 50 known neutral faces, all of which are recognised, and noted the number of processing cycles required for the appropriate PIN to reach the recognition threshold level. These latency values were correlated with the distinctiveness ratings allocated to these faces by human raters.

The value of the product-moment correlation between number of cycles to reach threshold and the "without hair" human distinctiveness ratings is -0.31, giving a significant negative correlation ($p < 0.05$). The correlation between cycles to threshold and the "with hair" ratings is -.0.22, which fails to reach significance. The significant correlation with one of these measures strikes us as remarkable. The model was not constructed to analyse distinctiveness. Indeed, one plausible locus for distinctiveness has been eliminated from the input representations. Faces are coded as binary values, so effects of extreme values on particular eigenfaces will not show up. Nonetheless, faces which humans had rated as distinctive were recognised faster by the model, exactly as one finds with human subjects. To interpret this effect one needs to postulate that there are distinctive *patterns* of binary values across the PCA inputs, and the distinctiveness of these patterns is related to human ratings of distinctiveness. The non-significant correlation between the "with hair" ratings and performance shows a trend in the predicted direction. These rating data are less reliable as they come from a smaller number of subjects, and we note that the two ratings measures themselves are rather weakly correlated.

5.4 Semantic Priming

The model shows the normal effects of semantic priming. As described above, we have previously suggested that semantic priming occurs due to transient activation at a PIN. On presentation of a prime face, the relevant PIN and SIUs will become active. The two-way nature of the PIN-SIU links ensures that related PINs (those sharing SIUs) also

become slightly active (i.e. above rest, but below the threshold for recognition). This residual activation can be exploited in subsequent recognition of a target. If the target person's PIN is already at above resting levels, then it will take fewer cycles to reach threshold than would be the case if an unrelated person had preceded the target. As with human data, semantic priming crosses input domains: faces prime names and names prime faces. Note also that the model predicts a short time-course for semantic priming: any intervening stimulus item tends to force down residual unit activations. This prediction has been confirmed in recent experimental work (Calder & Young, in press).

To demonstrate this effect with the model developed here we carried out the following experiment. Ten neutral-expression faces were chosen as target faces, these were simply faces 1 to 10 from the population of 50. The images of these faces were presented to the model under two conditions: (i) following the face of an unrelated person; and (ii) following the face of a related person. For these purposes we defined an unrelated person as someone who shares no SIUs with the target. For each target, the first unrelated person meeting this criterion was chosen from the remaining population (people 11 to 50). We defined a related person as someone who shares two SIUs with the target. For each target, the first person meeting this criterion was chosen from the remaining population (people 11 to 50). In fact, because of the random allocation of semantic units, there were four target faces which did not share two SIUs with any other person. For these people we chose as "related" stimuli the first person sharing a single SIU with the target. These four pairs are therefore related less strongly.

The procedure was as follows. The prime face was presented to the system, and the model was allowed to cycle until it settled (we used 100 cycles here). The prime face was then removed, and there followed an inter-stimulus interval of 20 cycles, causing some decay in unit activation. The target face was then presented. We used the same recognition criteria as chosen in Section 5.3 to signal familiarity. The dependent variable was therefore the number of cycles required for the target PIN to reach threshold (0.45 here).

Table 2: Mean cycles to threshold for faces primed by related and unrelated faces and names.

	Unrelated prime	Related prime
Face prime	65	38
Name prime	63	41

The results of this experiment are shown in the first line of Table 2. The mean number of cycles to reach threshold is lower in the *related* condition than in the *unrelated* condition. Related-means t-test showed this difference to be reliable ($t(9) = 6.2$, $p < 0.01$). In fact the inferential statistics are redundant here, as the faces primed by related faces are *always* recognised faster than faces primed by unrelated faces.

The bottom line of Table 2 shows the results of a replication of this experiment, using the same target faces. However, in this case, faces are primed by names. An exactly similar procedure was used. At prime stage, both the forename and surname units of the semantically prime person were activated fully. The model was allowed to cycle until it stabilised (100 cycles was used). The same 20 cycle ISI as above was used, and the target presentation followed the same procedure as with face primes. Table 2 shows that priming occurs across domains, and this difference is reliable ($t(9) = 4.0$, $p < 0.01$)

These data replicate those found in human subjects. Readers are referred to Burton et al (1990, 1991) for a more detailed description of semantic priming. The important point is that the model presented here is capable of demonstrating cross-domain effects in priming. It is consistent with all our previously demonstrated effects of semantic priming, though in the past these were demonstrated in the absence of real face input. Only a model comprising both perceptual and cognitive components is capable of this demonstration.

5.5 Repetition priming

We have described our previous work on repetition priming above. Within the IAC model, the proposal is that repetition priming occurs as a result of strengthening links between FRUs and PINs. A simple Hebb-like procedure allows this to be performed in an unsupervised way. We have developed this theme in simulations of a number of different phenomena including face learning (Burton, 1994), part-face priming (Ellis, Burton, Young & Flude, submitted), priming different parts of the system (Ellis, Young, Flude & Burton, in press), learning new facts about people (Young & Burton, submitted) and imagery priming (Cabeza, Burton, Kelly & Akamatsu, submitted). However, in all previous work, we have had to make some rather arbitrary assumptions about the nature of the face input to the system. The lack of a developed front-end has meant that some fundamental aspects of repetition priming could not be simulated. In particular, it has been shown on very many occasions that repetition priming for faces is strongest when the same image is used as prime and target, and weaker, though still present, when different images of the same person are used as prime and target (Bruce & Valentine, 1986; Ellis et al, 1987a). This is an example of a phenomenon which requires *both* perceptual and cognitive models: we have proposed a cognitive account, but the human data show that the effect is moderated by front-end factors (image change). In this section we show that the combined model simulates this effect naturally.

To demonstrate repetition priming in the model, we presented it with a set of faces in three conditions (i) unprimed, (ii) primed by the same image, and (iii) primed by a different image of the same face. In this simulation only *expressing* images were used, i.e. not the neutral faces hard-wired into the model. The first 10 faces from the population were chosen for the experiment, and the first expressing face was used as target in each case, with the second expressing face being used as the "different picture" prime. Repetition priming is a long-term effect, and we have proposed that it is not due to transitory unit activations, but rather to link-strengthening. We therefore presented all faces in the context of having just seen an unrelated filler face. The same "filler" face was

used for all trials, as this shared no SIUs with any of the targets. (In fact, the same pattern of data occur if the model is tested entirely from rest each time.)

The procedure was as follows. To generate an unprimed "response time", the filler face was presented and the simulation allowed to cycle until it settled (100 cycles). There was then an ISI of 20 cycles with no presentation following which the target face was presented. We used the same recognition criteria as in previous experiments to signal familiarity. The dependent variable was therefore the number of cycles required for the target PIN to reach threshold (0.45 here). To generate primed responses the following procedure was used. First the filler face was presented and the system cycled, there then followed an ISI of 20 cycles and then the prime face was presented and the system was allowed to settle (100 cycles). At this stage, a Hebb-like operation was applied to all FRU-PIN links. This is the same link-update function that has been used in all previous work with the model (see Appendix). The system was then reset to rest to eliminate all unit activations. The target face was then presented in exactly the same manner as for unprimed (i.e. filler face, ISI, target face). As we are interested in demonstrating the effect individually for different faces, the model's links were reset to their original strength before each trial. The results are shown in Table 3.

Table 3: Mean cycles to threshold for faces primed by the same different pictures.

Unprimed	Primed with same image	Primed with different image
78.6	60.1	64.8

One-way analysis of variance showed a significant effect of condition ($F(2,18) = 12.6$, $p < 0.01$). Comparison between conditions showed significant differences between all conditions, by sign test. For all 10 target faces, unprimed > primed with different image > primed by same image. Consistent with human data, the model shows priming for same and different images, but the larger effect for the same image.

It is worth considering for a moment why the model behaves in this way. Recall that only *expressing* faces were used in this simulation, i.e. no face is coded perfectly. When the face is presented to the system various FRUs gain some activation, though the "correct" FRU gains most (in all recognised faces). Similarly, the correct PIN gains most activation, though others may gain a little too. Different images of the same face cause slightly different patterns of activation in the FRU pool. This means that when the global Hebb-like update occurs between FRUs and PINs, the system can be said to be "learning" a particular pattern. So, a face primed by the same image will be using strengthened links corresponding to exactly the pattern which was reinforced on a previous occasion. In contrast, a face primed by a different image of the same person will be using links which were strengthened only in as much as the two images are coded similarly. This means that a localist connectionist system of the IAC-type can code not only a central key representation (the neutral faces here) but can also show evidence of picture memory.

Once again, this seems to us to be an attractive property of the model, and one which has not been available in our previous simulations that lack an image-based front-end.

5.6 Summary and range of the simulation

We have now demonstrated that this model seems to capture various multi-modal effects in person recognition. Our argument has been that the combination of a perceptual and cognitive model can provide accounts of phenomena outside the range of models which exclusively concentrate on cognitive or perceptual aspects of the system. In particular, the effects of cross-modal cueing and distinctiveness seem to reflect interactions between perceptual and cognitive processes, and are therefore simply unavailable to previous models of face recognition. We have also demonstrated that cross-domain effects such as semantic priming and image effects as seen in repetition priming are properties of the model. In fact, we have previously offered accounts of these last two effects (e.g. Burton et al, 1990) though always in a model which simply assumes some perceptual input without specifying its nature. We have now shown that these effects are consistent with a front-end based on PCA of images.

We should also note that there are various other effects of face recognition which have previously yielded to explanations in terms of the IAC (cognitive) part of the model. In particular, it has been useful in capturing effects from neuropsychology, and especially covert recognition. Some prosopagnosic patients demonstrate an apparent recognition of faces when tested covertly, but have no experience of familiarity (e.g. Bauer, 1984; Bruyer, 1991; Farah, O'Reilly & Vecera, 1993; Young, 1994). So, for example, patient PH (Young et al, 1988) has been shown to demonstrate normal patterns of semantic priming from faces to associated name recognition, despite the fact that he has no experience of recognising the face. Indeed, PH scores at chance in a forced-choice test in which he is asked to sort faces into known and unknown, despite recognising all the "known" people from their names. We have previously argued that covert recognition can be captured by an IAC model in which the FRU-PIN links are attenuated (though not deleted). This attenuation means that presentation of a face leads to activation in the PIN which is below the recognition threshold. However, some activation is present, and this can be passed to associated SIUs, and then back to related PINs. Once again, sub-threshold, but super-resting activations in PINs can subsequently be exploited by input from another modality, e.g. names.

We will not present simulations of these effects, as they have been described in detail in the original papers. The important point to note is that they remain unaffected by the combination of perceptual and cognitive models. This combination not only extends the range of effects in person recognition which can captured, but it represents incremental progress, in the sense that previous explanations are not damaged by the extension.

6. Frequently asked questions

Over the time that we have been constructing this model, we have often been asked the same questions of it. We take this to mean that we have not been sufficiently

clear on some points about the architecture of the system. In this section we attempt to remedy this by listing the answers to some frequently asked questions. The aim here is to be as explicit as possible about what we are and are not claiming for this model.

6.1. Surely semantic units should not inhibit one another: does this not lead to absurdities such as "British" inhibiting "actor"?

The use of inhibition in models of this kind is sometimes confusing. It is important to realise that semantic units (SIUs) are connected not only directly, with inhibitory links, but also indirectly through PINs which share them. So, "actor" and "British" will be properties of very many people, and so will be linked, excitatorily, and bi-directionally, to many PINs. This means that semantic units which actually are correlated (say "actor" and "comedian") may have a net connectivity which is positive, despite the fact that there is a single within-pool negative link connecting them directly. It is easy to demonstrate this in the model. If a WRU coding, say, "actor" is activated, this activation flows to the SIUs, and then on to the PINs. Inspection of the SIU pool, following a number of cycles, shows that other related SIUs (i.e. those which tend to share people in common with "actor") will have also gained some small activation. This mechanism is exactly like that of semantic priming for face and name recognition: despite direct inhibitory links, associated units within the same pool can become active during the same presentation cycle.

We have shown that within-pool inhibitory links are not disadvantageous. In fact, they are very advantageous, for two reasons. First, these links allow the model to be dynamic, by avoiding hysteresis (Grossberg, 1978; McClelland & Rumelhart, 1988), the phenomenon in which unit activations resist decay. Imagine a comparable model in which SIU units were connected excitatorily, reflecting semantic structure (e.g. "actor" and "comedian" would have a positive link connecting them). Such networks are subject to a paralysing hysteresis. This is best illustrated by an example. Imagine we present the face of Bob Hope. This might lead to activation of that person's FRU, PIN, and SIUs representing, say, "actor", "comedian" and "American". We now stop presenting the face, and present another, say that of Prince Charles. In the normal course of events, the PIN of Bob Hope will decay, as the PIN of Charles rises. However, in the SIU pool the two units "actor" and "comedian" are in a circular relation, each receiving activation along excitatory links between them. There is therefore great resistance to decay as units within the system can bolster each other's activation in the absence of external input. For this reason (although it is rarely stated explicitly) models relying on this semantic network notion are actually reset to rest by the experimenter between input trials. There can therefore be no examination of transitory effects between items due to residual activation. This problem is true of traditional semantic network models (Burke, Mackay, Worthley & Wade, 1991) but also true of some more modern connectionist models (Farah et al, 1993) and is eliminated by within-pool inhibitions. An IAC-like system which contains within-pool inhibitions as well as excitatory connections via other pools does not suffer from this problem: one can present stimuli sequentially, and hence observe dynamic behaviour, rather than resetting the whole model before each stimulus presentation.

The second advantage for within-pool inhibition is that it provides the facility to represent exclusives. For example, though many people are both actors and comedians, many people are not. A semantic network representation would require excitatory links between actor and comedian, which would result in some activation for "comedian" on presentation of Richard Burton's face. This seems inappropriate as this actor was not a comedian in any sense. This is a more prosaic reason to prefer the architecture we have chosen, but again seems to us to represent a significant advantage over a representational scheme in which semantic properties are connected independently of the people who instantiate the properties. We must note that this is a model of person recognition, not of semantics per se. It is clear that the system must somehow be able to represent semantic relations in the absence of personal information (for example to code abstract relations). However, it seems to us that a model of person recognition is best served by a person-based semantic system.

6.2. Is the model not inefficient? It seems that there is some duplication of structure between the FRUs, NRUs and PINs.

In the model we have presented, there are one-to-one relations between the FRUs and PINs and between the NRUs and PINs. However, these pools are intended to represent different levels of classification: the face, name and person respectively. In fact, we know many people by only one of these routes. The person who serves in a shop we often use, or the person who stands next to us in bus queues each morning, may be recognised by face without our needing to know the person's name. Similarly, we know many people by name only. For example many of us would be unable to recognise the face of Charles Dickens or Crick & Watson, despite knowing a great deal about these people. It is therefore necessary to separate out these recognition routes in the model. The fact that we have given everyone represented by the model a name and a face is simply to ease modelling (and to avoid behaviour arising out of arbitrary architectural decisions).

6.3 Is number of cycles to threshold a good analogue of RT data?

The dependent variable used in the simulations described here, and in previous work, is the number of processing cycles needed for particular units to reach some threshold. There is clearly some arbitrariness in this. First, the threshold chosen plays a part. In fact, the nature of these systems is such that in almost all cases (and in all cases we present here) units which rise fastest, also rise highest. The choice of threshold therefore does not affect the direction of predictions, only their size. Second, the nature of the relation between RT and cycles is important. We take the relation between RT and processing cycles to be positive and monotonic. This means that the model is restricted in the effects it can simulate. Ordinal predictions of the form "the unit will reach threshold faster in condition X than condition Y" can be captured. However, predictions which rely on differences in *size* between effects are not available for modelling. Without a more detailed assertion of the relation between RT and cycles, effects such as non-crossover interactions cannot be captured in the model. This is, of course, true of any simulation which does not spell out the exact nature of the relation of the simulated DV and the DV from human studies, though this fact is sometimes ignored.

6.4 The model is static, is this type of architecture suitable for learning?

We have argued at length elsewhere (Young & Burton, submitted) that learning is an attractive feature of models only when that learning is consistent with human learning. So, for example, connectionist models which require repeated presentation of the entire to-be-learned set of stimuli do not capture human learning. Burton (1994) presents a learning mechanism for IAC models. This mechanism has some of the features which we take to be true of humans learning faces: it is incremental (i.e. learning a new face does not affect representations of previously-learned faces); it allows different levels of learning (i.e. representations can be more or less "known"); it is unsupervised and automatic. The learning mechanism is based on exactly the same principles as is repetition priming: ubiquitous Hebb-like updates throughout the model. However, the learning mechanism was developed in the absence of a front-end perceptual system.

In the model presented we have tried to capture a snap-shot of the person recognition system, and we have not built in a learning mechanism. We have not wished to confound results of perceptual-cognitive interactions with any effects of learning. We note that the system is consistent with the learning mechanism described by Burton (1994), but integration of this mechanism with analysis of real image-based input must be the subject of further development. Such development will require human experimentation, as there is surprisingly little known about the processes by which faces become familiar.

6.5 Surely humans do not do PCA on pixel-like properties of images

The claim in the model described here is that input from images forms a suitable front-end for a model of face recognition. This apparently uncontentious statement in fact represents a theoretical position which, though growing in popularity, is not universally held. In the past, researchers have often looked for representations of a more abstractive type. So, for example, many researchers have assumed that faces are represented in memory in terms of some set of distance measures in the picture plane. In choosing PCA we are claiming that faces are represented in terms of the patterns of light across the whole image, rather than in some abstract form. Principal components analysis is one way to capture the regularities in this *image-based* approach, however, it is clearly not the only way. For example, von der Malsburg and colleagues have provided an alternative image-based face recognition system based on a deformable template and Gabor wavelets (e.g. Würtz, Vorbrüggen, von der Malsburg & Lange, 1992; Konen, Maurer & von der Malsburg, 1994; Wiskott & von der Malsburg, 1995). Moreover, the use of pixels as input to out PCA-based system is clearly an over-simplification and we have recently been experimenting with PCA-based upon the outputs of images filtered in ways resembling early visual processing mechanisms (e.g. Hancock, Burton & Bruce, 1995). A complete account would probably combine outputs of filters with different spatial scales. What is impressive to date is that even based upon "mere" pixel level representation of the image, our PCA-based description fares so well at accounting for human face memory and perception. In summary, our claim is not that the details of the PCA account are correct. Rather it is that a linearised compact coding of human face images represents a promising hypothesis for human representation of faces. It is worth noting additionally that such an approach, while able to deal with variation such as the expression changes described here, will not readily deal with changes in viewpoint

without representing discrete viewpoints separately. This view-specific property also seems to be function of human face recognition (Bruce, 1994).

7. Conclusion

We have presented a model of the complete process of face recognition. The model takes images of faces, and processes them through a recognition system which simulates the process of familiarity, retrieval of personal information and naming. The model is the result of combining approaches from the perceptual literature on face recognition with a model of the cognitive processes.

We hope to have demonstrated that the combined model represents an advance over either component on its own. The combination considerably extends the range of findings which the model can simulate, and allows one to examine the interaction between perceptual and cognitive processes. In particular, it allows one to examine top-down effects (such as cued recognition) and effects which seem to be dependent on both perceptual and representational processes (such as distinctiveness).

Although this model represents an attempt to capture the complete process of face recognition, it is far from complete itself. In particular, a satisfactory account must be able to integrate learning into the system, and so far this has not been achieved. On a related issue, a complete model must be able to capture decisions which we can make with unfamiliar faces. This model has nothing to say on the issue of unfamiliar faces: it cannot decide about the sex or age of a face or decide whether it looks attractive, grumpy or distinguished. It cannot decide that a face looks like Ronald Regan but certainly isn't, and it cannot decide that a face would look better without the beard. However, we believe that many of these effects will be captured only in a model which contains both perceptual and cognitive processes. The purpose of this paper is to propose just such a model.

References

Bauer, R.M. (1984). Autonomic recognition of names and faces in prosopagnosia: a neuropsychological application of the guilty knowledge test. Neuropsychologia, 22, 457-469.

Beymer (1995). Vectorizing face images by interleaving shape and texture computation. AI Memo 1573, MIT Center for Biological and Computational Learning.

Brédart, S., Valentine, T., Calder, A. & Gassi, L. (1995). An interactive activation model of face naming. Quarterly Journal of Experimental Psychology, 48A, 466-486.

Bruce, V. (1983). Recognizing faces. Philosophical Transactions of the Royal Society, London, B302, 423-436.

Bruce, V. (1994). Stability from variation: The case of face recognition. The M.D. Vernon Memorial Lecture. Quarterly Journal of Experimental Psychology, 47, 5-28.

Bruce, V. (1986). Recognising familiar faces. In H.D. Ellis, M.A. Jeeves, F. Newcombe & A.W. Young (Eds.) Aspects of Face Processing. Dordrecht: Martinus Nijhof.

Bruce, V., Burton, A.M. & Dench, N. (1994). What's distinctive about a distinctive face? Quarterly Journal of Experimental Psychology, 47, 119-141.

Bruce, V., Hannah, E., Dench, N., Healy, P. & Burton, M. (1992). The importance of "mass" in line drawings of faces. Applied Cognitive Psychology, 6, 619-628.

Bruce, V. & Langton, S. (1994). The use of pigmentation and shading information in recognising the sex and identities of faces. Perception, 23, 803-822.

Bruce, V. & Valentine, T. (1985). Identity priming in the recognition of familiar faces. British Journal of Psychology, 76, 363-383.

Bruce, V. & Valentine, T. (1986). Semantic priming of familiar faces. Quarterly Journal of Experimental Psychology, 38A, 125-150.

Bruce, V. & Young, A. (1986). Understanding face recognition. British Journal of Psychology, 77, 305-327.

Bruyer, R. (1991). Covert face recognition in prosopagnosia: a review. Brain and Cognition, 15, 223-235.

Burke, D.M., Mackay, D.G., Worthley, J.S. & Wade, E. (1991). On the tip-of-the-tongue: What causes word finding failures in young and older adults. Journal of Memory and Language, 30, 542-579.

Burton, A.M. (1994). Learning new faces in an interactive activation and competition model. Visual Cognition, 1, 313-348.

Burton, A.M. & Bruce, V. (1992). I recognise your face but I can't remember your name: a simple explanation? British Journal of Psychology, 83, 45-60.

Burton, A.M. & Bruce, V. (1993). Naming faces and naming names: exploring an interactive activation model of person recognition. Memory, 1, 457-480.

Burton, A.M., Bruce, V. & Dench, N. (1993). What's the difference between men and women? Evidence from facial measurement. Perception, 22, 153-176.

Burton, A.M., Bruce, V. & Johnston, R.A. (1990). Understanding face recognition with an interactive activation model. British Journal of Psychology, 81, 361-380.

Burton, A.M., Young, A.W., Bruce, V., Johnston, R. & Ellis, A.W. (1991). Understanding covert recognition. Cognition, 39, 129-166.

Cabeza, R., Burton, A.M., Kelly, S. & Akamatsu, S. (submitted). Investigating the relation between imagery and perception: Evidence from face priming.

Calder, A. & Young, A.W. (in press). Self priming. Quarterly Journal of Experimental Psychology, A.

Chellappa, R., Wilson, C.L. & Sirohey, S. (1995). Human and machine recognition of faces: a survey. Proceedings of the IEEE, 83, 705-740.

Costen, N., Craw, I. & Akamatsu, S. (1995). Automatic face recognition: what representation? Paper presented to the British Machine Vision Association (Scottish chapter) meeting, Glasgow, November 1995.

Craw, I. (1995). A manifold model of face and object recognition. In T. Valentine (Ed.) Cognitive and computational aspects of face recognition. London: Routledge.

Craw, I. & Cameron, P. (1991). Parameterising images for recognition and reconstruction. In P. Mowforth (Ed.) Proceedings of the British Machine Vision Conference, 1991. Berlin: Springer Verlag.

Craw, I., Tock, D. & Bennet, A. (1992). Finding face features. Proceedings of the 2nd European conference on computational vision. 92-96.

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

de Haan, E.H.F., Young, A.W. & Newcombe, F. (1991b). A dissociation between the sense of familiarity and access to semantic information concerning familiar people. European Journal of Cognitive Psychology, 3, 51-67.

Ekman, P. & Friesen, W.V. (1976). Pictures of facial affect. Palo Alto, Ca: Consulting Psychologists Press.

Ellis, A.W. (1992). Cognitive mechanisms of face processing. Philosophical Transactions of the Royal Society, London, B335, 113-119.

Ellis, A.W., Burton, A.M., Young, A.W. & Flude, B.M. (submitted). Repetition priming between parts and wholes: tests of a computational model of familiar face recognition.

Ellis, A.W., Flude, B.M., Young, A.W. & Burton, A.M. (in press). Two loci of repetition priming in the recognition of familiar faces. Journal of Experimental Psychology: Learning, Memory, and Cognition.

Ellis, A.W., Young, A.W., Flude, B.M. & Hay, D.C. (1987a). Repetition priming of face recognition. Quarterly Journal of Experimental Psychology, 39A, 193-210.

Ellis, A.W., Young, A.W. & Hay, D.C. (1987b). Modelling the recognition of faces and words. In P.E. Morris (Ed.), Modelling cognition (pp. 269-297). Chichester: Wiley.

Farah, M.J., O'Reilly, R.C. & Vecera, S.P. (1993). Dissociated overt and covert recognition as an emergent property of a lesioned neural network. Psychological Review, 100, 571-588.

Galper, R.E. (1970). Recognition of faces in photographic negative. Psychonomic Science, 19, 207-208.

Goddard, N.H., Lynne, K.J., Mintz, T. & Bukys, L. (1989). Rochester connectionist simulator Technical Report 233 (Revised). Department of Computer Science, University of Rochester, New York.

Gross, C.G. (1992). Representation of visual stimuli in inferior temporal cortex. Philosophical Transactions of the Royal Society of London, 335, 3-10.

Grossberg, S. (1978). A theory of visual coding, memory and development. In E.L.J. Leeuwenberg & H.F.J.M. Buffart (Eds.), Formal theories of visual perception. New York: Wiley.

Hancock, P.J.B., Burton, A.M. and Bruce, V. (1995). Pre-processing images of faces: correlations with human perceptions of distinctiveness and familiarity. In Proceedings of IEE Fifth International Conference on Image Processing and its Applications, Edinburgh, July 1995.

Hancock, P.J.B., Burton, A.M. & Bruce, V. (1996). Face processing: human perception and principal components analysis. Memory and Cognition, 24, 26-40.

Hanley, J.R. (1995). Are names difficult to recall because they are unique? A case study of a patient with anomia. Quarterly Journal of Experimental Psychology, 48A, 487-506.

Hay. D.C. & Young, A.W. (1982). The human face. In A.W. Ellis (Ed.) Normality and pathology in cognitive functions. 173-202. London: Academic Press.

Hay, D.C., Young, A.W. & Ellis, A.W. (1991). Routes through the face recognition system. Quarterly Journal of Experimental Psychology, 43A, 761-791.

Hill, H. & Bruce, V. (in press) Effects of lighting on the perception of facial surfaces. Journal of Experimental Psychology: Human Perception and Performance.

Kirby, M & Sirovich, L. (1990). Applications of the Karhunen-Loeve procedure for the characterisation of human faces. IEEE: Transactions on Pattern Analysis and Machine Intelligence, 12,103-108.

Kohonen, T., Oja, E. & Lehtio, P. (1981). Storage and processing of information in distributed associative memory systems. In G. Hinton & J. Anderson (Eds.) Parallel Models of Associative Memory, 105-143. Hillsdale, NJ: Lawrence Erlbaum.

Konen, W., Maurer, T. & von der Malsburg, C. (1994). A Fast Dynamic Link Matching Algorithm for Invariant Pattern Recognition. Neural Networks, 7, 1019--1030.

Lades, M., Vorbruggen, J.C., Buhman, J., Lange, J., von der Malsburg, C., Wurtz, R.P. & Konen, W. (1993). Distortion invariant object recognition in the dynamic link architecture. IEEE Transactions on Computation, 3, 300-311.

McClelland, J.L. (1981). Retrieving general and specific information from stored knowledge of specifics. Proceedings of the Third Annual Meeting of the Cognitive Science Society, 170-172.

McClelland, J.L. & Rumelhart, D.E. (1988). Explorations in parallel distributed processing. Cambridge, Massachusetts: Bradford Books.

O'Toole, A.J., Deffenbacher, K.A., Valentin, D. & Abdi, H. (1994). Structural aspects of face recognition and the other race effect. Memory & Cognition, 22, 208-224.

Pentland, A., Moghaddam, B. & Starner, T. (1994). View-based and modular eigenspace for face recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 84-91.

Perrett, D I, Hietanen, J K, Oram, M W and Benson, P J (1992) Organisation and functions of cells responsive to faces in the temporal cortex. Philosophical Transactions of the Royal Society of London, 335, 23-30.

Phillips, R.J. (1972). Why are faces hard to recognise in photographic negative? Perception and Psychophysics, 12, 425-426.

Rhodes, G., Brake, S. & Atkinson, A. (1993). What's lost in inverted faces? Cognition, 47, 25-57.

Rhodes, G., Brennan, S. & Carey, S. (1987). Identification and rating of caricatures: Implications for mental representations of faces. Cognitive Psychology, 19, 473-497.

Scanlan, L.C. & Johnston, R.A. (in press). I recognize your face but I can't remember your name: a grown-up explanation. Quarterly Journal of Experimental Psychology.

Sergent, J., Ohta, S., MacDonald, B & Zuck, E. (1994). Segregated processing of facial identity and emotion in the human brain: a PET study. Visual Cognition, 1, 349-369.

Shepherd, J. (1989). The face and social attribution. In A.W. Young & H.D. Ellis Handbook of Research on Face Processing. 289-320. Amsterdam: North Holland.

Troje, N. and Bülthoff, H. (in press) Face recognition under varying pose: The role of texture and shape. Vision Research.

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

Valentine, T., Brédart, S., Lawson, R. & Ward, G. (1991). What's in a name? Access to information from people's names. European Journal of Cognitive Psychology, 3, 147-176.

Valentin, D., Abdi, H. & O'Toole, A. (1994). Categorization and identification of human face images by neural networks: A review of the linear autoassociative and principal component approaches. Journal of Biological Systems, 2, 413-423.

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

Valentine, T. & Bruce, V. (1986b). The effects of distinctiveness in recognising and classifying faces. Perception, 15, 525-536.

Vetter, T. and Troje, N. (1995) Separation of texture and two-dimensional shape in images of human faces. In: Sagerer, G., Posch, S. and Kummert F., Mustererkennung 1995, Reihe Informatik aktuell, pp. 118-125, Springer Verlag.

Wiskott, L. & von der Malsburg, C. (1995). Recognizing faces by dynamic link matching. ICANN'95: International Conference on Artificial Neural Networks, Paris, France: EC2 & Cie.

Würtz, R.P., Vorbrüggen, J.C., von der Malsburg, C. & Lange. J. (1992). Recognition of human faces by a neuronal graph matching process. In H. G. Schuster, (ED.), Applications of Neural Networks, 181-200. VCH: Weinheim.

Young, A.W. (1994). Conscious and nonconscious recognition of familiar faces. In C. Umiltà & M. Moscovitch (Eds.), Attention and Performance, XV: conscious and nonconscious information processing (pp. 153-178). Cambridge, Massachusetts: MIT Press/Bradford Books.

Young, A.W., Flude, B.M., Hellawell, D.J. & Ellis, A.W. (1994). The nature of semantic priming effects in the recognition of familiar people. British Journal of Psychology, 85, 393-411.

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

Young, A.W., Hellawell, D. & de Haan, E.H.F. (1988). Cross-domain semantic priming in normal subjects and a prosopagnosic patient. Quarterly Journal of Experimental Psychology, 40A, 561-580.

Appendix: Technical details.

The IAC simulations reported here were run using the Rochester Connectionist Simulator (Goddard, Lynne, Mintz & Bukys, 1989). The unit update function was the standard IAC update; see McClelland and Rumelhart (1988) p.13 for equations.

The Hebb-like rule used for repetition priming was taken from Burton (1994) and is as follows:

If $a_i a_j > 0$, $\Delta w_{ij} = \lambda a_i a_j (1 - w_{ij})$

Otherwise $\Delta w_{ij} = \lambda a_i a_j (1 + w_{ij})$

where λ is a global learning rate parameter.

Global parameters were set as follows for all simulations:

Maximum unit activation = 1.0
Minimum unit activation = –0.2
Rest = –0.1
Decay rate = 0.1
External strength = 0.4
Alpha = 0.1
$\lambda = 0.25$

In each of the simulations all excitatory and inhibitory connections had strength 1.0 and -0.8 respectively.