PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Modelling Face Recognition [and Discussion]

Vicki Bruce, A. Mike Burton, Ian Craw, T. Valentine, E. T. Rolls and H. D. Ellis

Phil. Trans. R. Soc. Lond. B 1992 **335**, 121-128
doi: 10.1098/rstb.1992.0015

Modelling face recognition

VICKI BRUCE[1], A. MIKE BURTON[1] AND IAN CRAW[2]

[1]*Department of Psychology, University of Nottingham, Nottingham NG7 2RD, U.K.*
[2]*Department of Mathematical Sciences, University of Aberdeen, Aberdeen AB9 2UB, U.K.*

SUMMARY

Much early work in the psychology of face processing was hampered by a failure to think carefully about task demands. Recently our understanding of the processes involved in the recognition of familiar faces has been both encapsulated in, and guided by, functional models of the processes involved in processing and recognizing faces. The specification and predictive power of such theory has been increased with the development of an implemented model, based upon an 'interactive activation and competition' architecture. However, a major deficiency in most accounts of face processing is their failure to spell out the perceptual primitives that form the basis of our representations for faces. Possible representational schemes are discussed, and the potential role of three-dimensional representations of the face is emphasized.

1. EARLY RESEARCH ON FACE PROCESSING

It is perhaps understandable that early research into the recognition of faces was conducted in the absence of any guiding theoretical framework (H. D. Ellis 1975; Bruce 1979). Much of this research was aimed at accounting for factors that affected the ease or difficulty of remembering previously unfamiliar faces, tasks which were of obvious applied (forensic) relevance. However, the kinds of theories of learning and memory which developed with the 'cognitive' revival in the 1960s and early 1970s were poorly equipped to deal with explaining the retention of non-verbalizable materials. Models of memory of the 'modal' kind (see, for example, Atkinson & Shiffrin (1968)) emphasized the retention of verbal information with an intrinsically sequential structure. It was difficult to adapt such models to materials whose basic structure is spatial, and where tasks such as serial recall make little ecological sense.

Arguably, however, the kind of theoretical framework more appropriate for face recognition is not a model of episodic memory at all, any more than theories of object or word recognition are built upon the study of remembering particular occurrences of objects or words. Most everyday tasks of face recognition involve us recognizing, or failing to recognize, the faces of friends, relatives or celebrities. Successful recognition involves a 'complete' identification of the person. By complete identification we mean a knowledge of why the person is familiar, from what contexts and in what role they are known, and retrieval of the name (if known) when introducing the person to another. Moreover, when recognizing our friends in such ways we are able to comment on their appearance (they look tired, or old, or particularly glamorous with a new haircut). We can usually recognize them in these unusual or novel views. The typical recognition memory experiment, in contrast, requires subjects to discriminate 'old' (previously presented) from 'new' (distractor) items on the basis of mere familiarity. Usually the old faces at test are shown in the same views as they were at study, allowing performance to be based on the matching of 'pictorial' codes (literal descriptions of a particular picture) rather than 'structural' codes (abstract descriptions of a particular face) (Bruce 1982, 1983; Bruce & Young 1986). If novel viewpoints are presented at test, recognition accuracy is reduced considerably (Bruce 1982). Although semantic codes may be involved in recognition memory for unfamiliar faces ('that is the face I thought particularly handsome'; 'that's the one that looks like Uncle Joe'), these are not the same kind of semantic codes that specify individual identity in terms of, for example, occupation or place of abode. A face may look like your Uncle Joe, but in actuality belong to a villanous murderer, and however much you know about Uncle Joe will not help you behave appropriately towards the person who resembles him. Bruce & Young (1986) distinguished visually derived semantic codes (meanings based on the literal appearance of the face, for example, looks kind, looks happy, looks like Joe) from identity-specific semantic codes, which specify who a person really is rather than what type of person they look like.

Our everyday task of face recognition involves the retrieval of identity-specific semantic codes from faces that vary from moment to moment (as lighting or expressions change), from day to day (as health, hair or cosmetics change) or from year to year (as age changes). Studying the task of remembering or matching identical pictures of faces whose identities are unknown can reveal rather little of these processes. Even the usual task faced by the eye witness is to determine from an identification parade or photograph file which, if any, of the faces is that of the person seen committing the crime. The feeling of mere familiarity in such forensic contexts is dangerous, and may account for occasional cases where witnesses have identified innocent bystanders, or national celebrities, from line-ups.

2. FUNCTIONAL MODELS OF FACE PROCESSING

Young (this symposium) and A. W. Ellis (this symposium) describe how psychologists have recently attempted to understand face recognition by using functional models (see, for example, Hay & Young (1982); Bruce & Young (1986)). Such models have made a number of important contributions. First, their development has forced us to be explicit about the different uses made of facial information (for deciphering emotions, for lipreading, for identification), and their inter-relationships. On the basis of neuropsychological and experimental investigations, it has been shown that the logical independence of certain categorizations made to faces is reflected in a functional independence of one processing route from another. The clearest example of this comes with the distinction between the processing of expression and identification, where there is now considerable converging evidence for the functional independence of expression and identity processing (see Young & Bruce (1991) for an extended discussion).

Second, empirical research has allowed us to be explicit within these models about the nature and sequence of stages that underly the full identification of the face. In particular, considerable evidence (Young *et al.* 1986*a*, 1988; McWeeny *et al.* 1987) suggested that the retrieval of a name for a seen face involves a stage of processing additional to the retrieval of semantic information about a person's identity. Particularly compelling is the observation that decisions requiring access to names take longer than decisions requiring access to semantic information, even using a small number of well-learned faces as stimuli (Young *et al.* 1988; Johnston & Bruce 1990). It is as though we must get to names via a preceding stage of retrieving identity-specific semantic codes.

Third, such functional models make it easier to make explicit the similarities and differences between the processes underlying the recognition of faces and those underlying the recognition of other visual objects and written words (see Bruce & Young (1986) for a discussion). A number of demonstrations of priming (A. W. Ellis, this symposium) suggest that the basic mechanisms whereby recognition can be facilitated by prior presentation of the same or related items are shared by faces, other objects and words. Studies of the effects of interference between faces and names suggest some asymmetry in the processing stages or routes available for faces compared with names, but the nature of these asymmetries is the same as that found for objects and words (see, for example, Young *et al.* (1986*b*, 1987)).

However, models of this 'box and arrow' type, although useful and potentially falsifiable (for example, see Brennen *et al.* (1990)) can allow us to remain vague about detailed mechanisms. The Bruce & Young (1986) model postulated three stages within the person identification route (cf. Hay & Young 1982). The first stage involves the perceptual classification of faces by the face recognition units (FRUS), logogen-like (Morton 1969, 1979) entities which were proposed to respond to any recognizable view of a known person's face. The second stage involved retrieving semantic information about the person via the person identity nodes (PINS), which were assumed also to be involved when a person was recognized via their name or voice. Finally there was a stage of name retrieval, which involved the access of a particular identifying label for the person. It was assumed that a feeling of familiarity of a face arose somehow as a result of sufficient activation at the level of the face recognition units, yet no clear mechanism for making familiarity judgements was proposed by Bruce & Young (1986). It was also left unclear whether semantic information about personal identity was held at the PINS or accessed via the PINS. Finally, much of the burden of explanation was placed at the level of the 'cognitive system' which, as in other similar models of visual cognition, was left completely unconstrained.

A further level of theoretical rigour is imposed when an attempt is made to implement such functional models in a working computer model. Burton *et al.* (1990) describe how the central features of the Bruce & Young (1986) framework can be expressed within a connectionist implementation making use of the 'interactive activation and competition' architecture described by McClelland (1981; also McClelland & Rumelhart 1981) (see figure 1). In the Burton *et al.* model, familiarity decisions are taken at the level of the PINS as a result of activity passing there from the FRUS (or name input units). This means that sensory evidence from the face is combined with that from other routes to achieve identification of the person

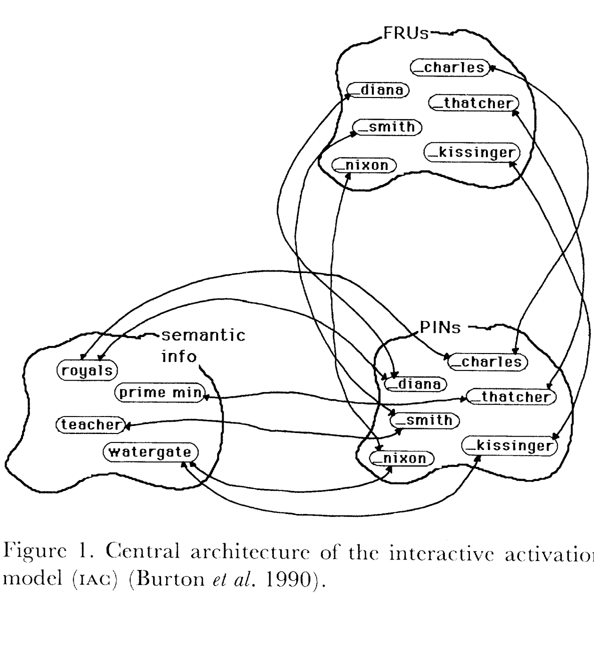

Figure 1. Central architecture of the interactive activation model (IAC) (Burton *et al.* 1990).

rather than of their face in isolation. In providing this mechanism for making familiarity judgements, the Burton *et al.* model expressly separates the PINS from the semantic information that they access. The PINS act as multi-modal nodes that are the entry point to the semantic system, whereas the FRUS act as uni-modal nodes pooling information over a set of more primitive feature analysers.

The central properties of this architecture arise from excitation between corresponding nodes in different pools of units, combined with inhibition between nodes within a pool. Burton *et al.* (1990) describe how such a model can exhibit semantic priming from prior presentation of a related item (Bruce & Valentine 1986). This occurs because activity levels for related PINS can be raised by their shared semantic information despite the within-pool competition. In this architecture, with familiarity decisions taken at the PINS, semantic priming crosses domains (e.g. faces prime names as much as faces prime faces) and is obliterated by an item intervening between the 'prime' and 'target' face. Both these properties are consistent with what is known empirically (A. W. Ellis, this symposium). However, we have yet to establish whether cross-domain semantic priming is identical to within-domain, and at present it is not clear whether semantic priming is abolished by the mere passage of time or only by the presence of an intervening face. Further experiments will allow us to establish whether the detailed predictions from this model are upheld.

The effects of repetition priming (see Bruce & Valentine 1985; A. W. Ellis *et al.* 1987) are explained within this model as arising from the strengthening of links between the recognition units and the PINS. The mechanism of repetition priming is thus separated from that of semantic priming which is compatible with the very different timecourse of the two types of priming, and the domain-specificity of repetition priming where names will not prime subsequent face familiarity judgements (A. W. Ellis, this symposium). The mechanism of repetition priming in this model is also consistent with the observation that priming is only found when the test phase involves identification, and is not exhibited by mere repetition of faces in non-identification tasks such as identifying sex or expressions (A. W. Ellis *et al.* 1990), because there is no reason to suppose that strengthening connections into the identification system would have any influence on the speed of other tasks which can be conducted without reference to identity (Bruce 1986; Bruce *et al.* 1987).

The advantages of this implemented functional model is that it allows us to account explicitly for certain patterns of effect, and has the additional bonus that findings which appeared difficult for the 'box and arrow' version of the model may actually occur quite naturally within this implemented version. For example, Burton *et al.* (1991) describe how the model could account for 'covert' recognition of faces by prosopagnosic patients (see Young, this symposium). On this model, prosopagnosic patients who show 'covert' recognition have reduced connection strengths between intact FRUS and PINS, so that threshold activation of the PINS is impossible when familiar faces are presented. Nevertheless, sub-threshold activation of the PINS can still produce sufficient activation of related PINS to produce semantic priming of related names, and interference from unrecognized faces in classifying names according to occupational category (see Burton *et al.* (1991) for detailed simulations).

Burton & Bruce (1991) discuss various possible schemes for the addition of a stage of name retrieval to this basic model. They propose that names are simply semantic information units (SIUS) whose special properties arise as a by-product of their pattern of interconnections with other units. Most SIUS will have connections from many PINS (e.g. the SIU for 'politician' will have links from all PINS corresponding to known politicians, similarly for the SIU for 'dead', 'American', etc.). Assuming that the SIU corresponding to a name encodes a name in its entirety (i.e. 'John Kennedy', rather than 'John') there will usually only be a single PIN connecting to each SIU which represents a name. Burton & Bruce (1991) show how SIUS with unique connections will always become activated more slowly, and reach lower levels of activation than those with more than one connection, and discuss how this could account for apparent name-specific deficits arising from brain damage. This model therefore makes the prediction that unique pieces of semantic information (e.g. an address) should behave like names, i.e. be relatively slower, and harder, to retrieve. (Note that equating names with SIUS in person identification is not to deny the necessity both for input systems for recognizing spoken and heard names, nor output systems for articulation as a result of activation at SIU level.)

However, the Burton *et al.* model, like the Bruce & Young (1986) framework preceding it, still has deficiencies. The first problem is that the model represents a steady state of an adult cognitive system. The model has no way of learning beyond the strengthening of connections of units corresponding to already-learned faces and hence cannot acquire units for new individuals. However, the human person recognition system is immensely flexible. We acquire innumerable new acquaintances during the course of a lifetime as new celebrities become famous, or as new students appear in class. The second, and related, problem is that the model lacks a perceptual interface: it contains no very convincing account of the representation of faces in terms of some set of perceptual primitives. Indeed, the Burton *et al.* (1990) model may be better thought of as a model of the portion of semantic memory concerned with person identification (cf. McClelland 1981) than as a model of face recognition at all. This problem of the lack of a perceptual interface is the more fundamental one as its solution is necessary if we are to solve one aspect of the perceptual learning problem: how we acquire stable visual representations of individuals' faces from varying exemplars (see Bruce *et al.* 1991).

3. PERCEPTUAL CODING OF FACES

Once we begin to search for a way of implementing a 'front end' for a face recognition system, with a capacity to learn new patterns as well as respond to old ones, it becomes natural to examine the literature on machine recognition of faces. Bruce & Burton (1989) review this literature and find remarkably few serious attempts to build automatic face recognition devices, given their obvious commercial potential. The attempts that have been made seem to vary a great deal in how much they assume the face image has been 'understood' during the learning phase. At one extreme are those such as WISARD (e.g. Stonham 1986) where the face is stored without any explicit preprocessing, and the raw pixel intensities are used directly to control recognition units. In contrast, a number of attempts have been made to understand the face image well enough to locate individual face features (for example, see Kanade (1977); Craw *et al.* (1987)). Explicit measurements can then be taken and used to define parameters, such as the distance between the eyes, or the ratio of nose length to head height, on which to base recognition. These systems have the advantage that the required measurements can be checked manually, but need the recognition algorithms to function perfectly before the required parameters are obtained.

A more recent strategy is to do some preprocessing without necessarily obtaining the type of measurement that can be taken (and checked) by hand. Starting with Kohonen *et al.* (1981) there has been interest in applying Principal Component Analysis (PCA) to the task (Millward & O'Toole 1986; O'Toole *et al.* 1991). The individual pixel values are regarded as a set of correlated measurements, with each image described by, for example, a 16384-tuple of intensities. Given an ensemble of different faces, PCA is done to extract significant axes to describe a 'face subspace' and hence capture the essence of a 'face'. Kirby & Sirovich (1990) used this technique (also Sirovich & Kirby 1987), studying both the dimension of the 'face space' and its generality. Working with an ensemble of 100 face images, from which they extract a central cameo, they find that the first 50 axes or eigenfaces, account for 95% of the total variance of the ensemble. Having obtained axes – essentially a collection of 'useful' mixes of the ensemble faces – these are used to construct, as accurately as possible, all other faces. The representation has the advantage that it is extremely compact (typically 50 bytes) and can be obtained very rapidly by projecting the new face onto the 'face subspace'.

There are objections to this methodology. Because the representation is in effect averaging among existing face images, very little of substance should be expected if they have different scales, orientations or even locations within the whole image. Thus the method implicitly assumes that segmentation, e.g. into figure and ground, has already been done, and in particular that scale and position invariance has been achieved. In fact very careful normalizations have been done in published work, ensuring that as far as possible corresponding features in the different faces coincide, to overcome these problems.

One way to do this is to choose a number of control points on each face, and then use 'cartooning' (Benson & Perrett 1991) on these points to map the face to a standard position. Reconstruction can then be done with much greater accuracy as long as the initial distortion is recorded as part of the coding (Craw & Cameron 1991). For this reason a plausible face representation consists of the eigenface coordinates together with this distortion vector.

Because of the need for prior normalization, PCA methods should be regarded as requiring more knowledge of the face image than is at first apparent. We refer to this as the 'bootstrap' problem, and it is common to most recognition schemes. Starting with the raw face image, one first step is to seek 'blobs' to give the initial head location. Turk & Pentland (1991) adopt this strategy, apparently for purely pragmatic reasons, whereas Watt (1988) suggests there are good neurophysiological arguments to do so. Turk & Pentland then use the eigenface representation to detect heads among 'blobs' found by a tracking system, using the ability of the eigenface representation to describe 'faceness'. They also describe the use of eigenface coordinates for recognition of a small set of faces in a limited context, within which the images appear to be carefully normalized.

An alternative approach moves from blobs to a plausible face outline (Bennett & Craw 1991). At this point more detailed understanding of the facial features can be attempted with a view to mapping the image to a standard position, from which reliable eigenface coordinates can be obtained in the same way as Turk & Pentland. Such a scheme has been studied by Shackleton & Welsh (1991) solely for eyes, which they are able to locate with some success using geometric template matching following Yuille *et al.* (1988).

Despite the wide variety of methods, at present systems for automatic face description and recognition are only successful at recognizing a relatively small set of faces tested under certain, rather limited conditions. No systems have yet been produced that can recognize faces despite changes in size, position, background, expression, viewpoint and lighting, though some can tolerate variability within a range upon which they have been trained. Although human recognition memory for once-viewed faces is impaired after a change of viewpoint or expression (Bruce 1982), we do not know the detailed effects of systematic variations of these and other factors. It would now be timely to explore systematically the sensitivity of human face memory to deviations between presented and test views in order to establish the tolerance of the human recognition system to different kinds of variation, and to explore possible mechanisms for achieving this tolerance. Note that we here suggest a return to the study of episodic memory for faces, now motivated by clearer theoretical issues.

4. FACES AS THREE-DIMENSIONAL SURFACES

In reviewing the poverty of ideas about the basic perceptual primitives which allow facial representations to be constructed, Bruce (1988) has suggested that progress has been hindered by a lack of proper understanding of the nature of the input. Most psychological and computer models have assumed that the front end of face processing can be understood by an appeal to pattern processing algorithms. However, faces are not flat patterns, but three-dimensional surfaces, an observation that has been acknowledged by artists (e.g. Bridgeman 1924), orthodontists (Enlow 1982), animators (Waters & Terzopoulos, this symposium) and telecommunications engineers (Pearson, this symposium). In recent perceptual research, Bruce & Burton have developed a psychophysical approach to face perception based on the facial surface rather than on pixels or two-dimensional features (see, for examples, Bruce *et al.* (1989, 1991); Bruce (1990)). Such an approach allows us to assess the sensitivity of the visual system to changes made to the shape of the face rather than to displacements made to an image of a face.

In addition to such methodological developments, understanding that faces are surfaces rather than patterns could have different types of theoretical implication for the encoding of faces by people or by machines. The first possibility would be that a visual system might explicitly represent the surface structure of the face when building representations. At University College, London, A. M. Coombes and colleagues (see, for example, Coombes *et al.* 1990, 1991) have applied a description of the basic surface types to faces for the purposes of planning reconstructive facial surgery. On their scheme, a face can be described as an 'array' of different surface types such as the peak of the nose, or the pit at the corner of the eye, at different spatial locations. In collaboration with this group, we are currently investigating whether the human visual system likewise represents surface types in its encoding of faces.

A second possibility is that a visual system does not explicitly represent the surface structure of faces it encodes, but that an implicit knowledge of the structure of faces is used to solve problems of normalization and figure-ground segmentation, as well as general problems such as view and lighting invariance. If the face image were stored in a limited number of views, then recognition could be performed by rotating the observed image to find a match. Such rotations can be achieved on a computer by using a three-dimensional model of a generic head plus an appropriate face image (Aitchison & Craw 1991). Such a method also offers the possibility of building view-point invariance into the eigenface representation described above, by using control points in three dimensions.

A third possibility is that the structure of surfaces in general is used to drive the development of image-processing algorithms that operate essentially in two dimensions. A good example here is Pearson & Robinson's (1985) algorithm for the automatic production of line-drawings of faces (see Pearson, this symposium), which draws lines at locations where the surface shifts sharply away from the viewer's line of sight. Given front- or top-lighting, such places will be characterized by luminance valleys in the image.

At present it is not clear which if any of these strategies the human visual system adopts when deciphering faces, nor is it clear which strategy will lead to most rapid progress in the automatic recognition of faces. Attempts to side-step the issue of perceptual primitives altogether by the employment of PDP approaches to pattern encoding (e.g. Cottrell & Fleming 1990) are of interest; however, such methods usually work best when the input is tuned carefully, and this may involve the type of preprocessing already described for the eigenface representation. In other words, it may still be necessary to solve the bootstrapping problem efficiently, and it is still not clear what methods will prove most suitable for this.

What all these approaches have in common, however, is that the encoding of faces must be achieved by making use of some explicit or implicit knowledge of the general structure of faces, whether in two or in three dimensions. In human vision, such knowledge could be hard-wired or learned through exposure to faces during development. There is considerable evidence that accummulated knowledge of faces acquired throughout life affects the encoding of novel faces and the recognition of familiar ones (see, for example, Valentine (1991)), and there is also evidence which suggests that human infants may enter the world equipped with a very general face 'schema' to mediate selective attention to faces (see, for example, Morton & Johnson (1991)). The possible relation(s) between innate knowledge, acquired knowledge. and the encoding of faces by the human visual system merits further attention in future research aimed at understanding the perceptual basis of our extraordinary abilities to identify individuals from their faces.

5. CONCLUDING REMARKS

Cognitive and computer sciences equip us with powerful new tools with which we can develop and extend our theories of how the human visual system recognizes faces. At present we have successfully implemented a model of person identification, but have yet to build a plausible perceptual interface, with a capacity to learn. The modelling tools and issues reviewed in the second half of this paper offer some promising directions for extending this model in the future.

Work reviewed in this paper has been funded by an ESRC programme (XC15250001 to Vicki Bruce at Nottingham University; XC15250002 to Ian Craw at Aberdeen University; XC15250003 to Hadyn Ellis at University of Wales College of Cardiff; XC15250004 to Andy Ellis and Andy Young at Lancaster University; XC15250005 to David Perrett at St Andrews University) and SERC projects to Vicki Bruce, Mike Burton and Alf Linney (GR/F 33698 and 72178).

REFERENCES

Aitchison, A. & Craw, I. 1991 Synthetic images of faces – an approach to model-based face recognition. *Proceedings of the British Machine Vision Conference BMVC'91*. Turing Institute Press and Springer Verlag. (In the press.)

Atkinson, R.C. & Shiffrin, R.M. 1968 Human memory: a proposed system and its control processes. In *The psychology of learning and motivation*, Vol. 2 (ed. K. W. Spence & J. T. Spence) pp. 89–195. New York: Academic Press.

Bennett, A. & Craw, I. 1991 Finding image features using deformable templates and detailed prior statistical knowledge. *Proceedings of the British Machine Vision Conference BMVC'91*. Turing Institute Press and Springer Verlag. (In the press.)

Benson, P.J. & Perrett, D.I.P. 1991 Perception and recognition of photographic quality facial caricatures: implications for the recognition of natural images. *Eur. J. Cogn. Psychol.* **3**, 105–135.

Brennen, T., Baguley, T., Bright, J. & Bruce, V. 1990 Resolving semantically induced tip-of-the-tongue states for proper nouns. *Mem. Cogn.* **18**, 339–347.

Bridgman, G.B. 1924 *Bridgman's life drawings*. (1971 edition.) New York: Dover.

Bruce, V. 1979 Searching for politicians: An information processing approach to face recognition. *Q. Jl exp. Psychol.* **31**, 373–395.

Bruce, V. 1982 Changing faces: visual and nonvisual coding processes in face recognition. *Br. J. Psychol.* **73**, 105–116.

Bruce, V. 1983 Recognising faces. *Phil. Trans. R. Soc. Lond.* B **302**, 423–436.

Bruce, V. 1986 Influences of familiarity on the processing of faces. *Perception* **15**, 387–397.

Bruce, V. 1988 *Recognising faces*. London: Lawrence Erlbaum.

Bruce, V. 1990 Perceiving and recognising faces. *Mind Lang.* **5**, 342–364.

Bruce, V. & Burton, M. 1989 Computer recognition of faces. In *Handbook of research on face processing* (ed. A. W. Young & H. D. Ellis), pp. 487–506. Amsterdam: North Holland.

Bruce, V., Burton, M. & Doyle, T. 1991 Faces as surfaces. In *Processing images of faces* (ed. V. Bruce & M. Burton). Norwood, New Jersey: Ablex (In the press.)

Bruce, V., Burton, M., Doyle, T. & Dench, N. 1989 Further experiments on the perception of growth in three dimensions. *Percept. Psychophys.* **46**, 528–536.

Bruce, V., Doyle, A., Dench, N. & Burton, M. 1991 Remembering facial configurations. *Cognition* **38**, 109–144.

Bruce, V., Ellis, H., Gibling, F. & Young, A. 1987 Parallel processing of the sex and familiarity of faces. *Can. J. Psychol.* **41**, 510–520.

Bruce, V., Healey, P., Burton, M., Doyle, T., Coombes, A. & Linney, A. 1991 Recognising facial surfaces. *Perception*. (In the press.)

Bruce, V. & Valentine, T. 1985 Identity priming in the recognition of familiar faces. *Br. J. Psychol.* **76**, 363–383.

Bruce, V. & Valentine, T. 1986 Semantic priming of familiar faces. *Q. Jl exp. Psychol.* **38A**, 125–150.

Bruce, V. & Young, A. 1986 Understanding face recognition. *Br. J. Psychol.* **77**, 305–327.

Burton, A.M. & Bruce, V. 1991 I recognise your face but I can't remember your name: a simple explanation? *Br. J. Psychol.* (In the press.)

Burton, A.M., Bruce, V. & Johnston, R.A. 1990 Understanding face recognition with an interactive activation model. *Br. J. Psychol.* **81**, 361–380.

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

Coombes, A.M., Linney, A.D., Richards, R. & Moss, J.P. 1990 A method for the analysis of the 3D shape of the face and changes in the shape brought about by facial surgery. *Proceedings SPIE Conference on Biostereometrics & Applications*. Boston, Massachusetts.

Coombes, A.M., Moss, J.P., Linney, A.D., Richards, R. & James, D.R. 1991 A mathematical method for the comparison of three-dimensional changes in the facial surface. *Eur. J. Orthodont.* **13**, 95–110.

Cottrell, G.W. & Fleming, M. 1990 Face recognition using unsupervised feature extraction. *International Neural Network Conference*, pp. 322–325. Dordrecht: Kluver.

Craw, I. & Cameron, P. 1991 Parameterising images for recognition and reconstruction. To appear in the *Proceedings of the British Machine Vision Conference BMVC'91*. Turing Institute Press and Springer Verlag. (In the press.)

Craw, I.G., Ellis, H.D. & Lishman, J.R. 1987 Automatic extraction of face features. *Pattern Recogn. Lett.* **5**, 183–187.

Ellis, A.W., Young, A.W. & Flude, B.M. 1990 Repetition priming and face processing: priming occurs within the system that responds to the identity of a face. *Q. Jl exp. Psychol.* **39A**, 193–210.

Ellis, A.W., Young, A.W., Flude, B.M. & Hay, D.C. 1987 Repetition priming of face recognition. *Q. Jl exp. Psychol.* **39A**, 193–210.

Ellis, H.D. 1975 Recognizing faces. *Br. J. Psychol.* **66**, 409–426.

Enlow, D.H. 1982 *Handbook of facial growth*. Philadelphia: W. B. Saunders.

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

Johnston, R.A. & Bruce, V. 1990 Lost properties? Retrieval differences between name codes and semantic codes for familiar people. *Psychol. Res.* **52**, 62–67.

Kanade, T. 1977 *Computer recognition of human faces*. Basel: Birkhauser Verlag.

Kirby, M. & Sirovich, L. 1990 Application of the Karhunen-Loève 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 *Parallel models of associative memory* (ed. G. Hinton & J. A. Anderson), pp. 105–143. Hillsdale, New Jersey: Lawrence Erlbaum Associates Inc.

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. 1981 An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychol. Rev.* **88**, 375–407.

McWeeny, K.H., Young, A.W., Hay, D.C. & Ellis, A.W. 1987 Putting names to faces. *Br. J. Psychol.* **78**, 143–151.

Millward, R. & O'Toole, A. 1986 Recognition memory transfer between spatial-frequency analysed faces. In *Aspects of face processing* (ed. H. D. Ellis, M. A. Jeeves, F. Newcombe & A. Young), pp. 34–44. Dordrecht: Martinus Nijhoff.

Morton, J. 1969 Interaction of information in word recognition. *Psychol. Rev.* **76**, 165–178.

Morton, J. 1979 Facilitation in word recognition: experiments causing change in the logogen model. In *Processing of visible language, 1* (ed. P. A. Kolers, M. Wrolstad & H. Bouma), pp. 259–268. New York: Plenum.

Morton, J. & Johnson, J.H. 1991 CONSPEC and CONLERN: A two-process theory of infant face recognition. *Psychol. Rev.* **98,** 164–181.

O'Toole, A.J., Abdi, H., Deffenbacher, K.A. & Bartlett, J.C.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

1991 Classifying faces by race and sex using an autoassociative memory trained for recognition. In *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society* (ed. K. J. Hammond & D. Gentner). Hillsdale, New Jersey: Erlbaum.

Pearson, D.E. & Robinson, J.A. 1985 Visual communication at very low data rates. *Proceedings of the IEEE* **73**, 795–812.

Shackleton, M.A. & Welsh, W.J. 1991 Classification of facial features for recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (CVIP-91), 1991, pp. 573–579.

Sirovich, L. & Kirby, M. 1987 Low-dimensional procedure for the characterization of human faces. *J. opt. Soc. Am.* **4**, 519–524.

Turk, M. & Pentland, A. 1991 Eigenfaces for recognition. *J. Cogn. Neurosci.* **3**, 71–86.

Stonham, J. 1986 Practical face recognition and verification with WISARD. In *Aspects of face processing* (ed. H. D. Ellis, M. A. Jeeves, F. Newcombe & A. Young), pp. 426–441. Dordrecht: Martinus Nijhoff.

Valentine, T. 1991 A unified account of the effects of distinctiveness, inversion and race in face recognition. *Q. Jl exp. Psychol.* **43A**, 161–204.

Watt, R.J. 1988 *Visual processing*. London: Erlbaum.

Young, A.W. & Bruce, V. 1991 Perceptual categories and the computation of grandmother. *Eur. J. Cogn. Psychol.* **3**, 5–49.

Young, A.W., Ellis, A.W. & Flude, B.M. 1988 Accessing stored information about familiar people. *Psychol. Res.* **50**, 111–115.

Young, A.W., Ellis, A.W., Flude, B.M., McWeeny, K.H. & Hay, D.C. 1986*b* Face-name interference. *J. exp. Psychol. (Hum. Percept. Perform.)* **12**, 466–475.

Young, A.W., Flude, B.M., Ellis, A.W. & Hay, D.C. 1987 Interference with face naming. *Acta Psychol.* **64**, 93–100.

Young, A.W., McWeeny, K.H., Ellis, A.W. & Hay, D.C. 1986*a* Naming and categorising faces and written names. *Q. Jl exp. Psychol.* **38A**, 297–318.

Yuille, A., Cohen, D. & Hallinan, P. 1988 Facial feature extraction by deformable templates. Technical Report CICS-P-124, Center for Intelligent Control Systems, Cambridge, Massachusetts.

Discussion

T. Valentine (*Department of Psychology, University of Manchester, U.K.*). Can Professor Bruce be more specific about the nature of the nodes representing names that are included in the pool of semantic information units in the Burton & Bruce IAC model of face naming? Do they represent the first stage of lexical access for the name (e.g. to the semantic lexicon)? If not, what is the functional significance of a set of nodes that have a one-to-one mapping to the person identity nodes?

V. Bruce. The semantic information units (SIUs) corresponding to names in the Burton & Bruce (1991) account of naming should not be confused with units needed to recognize input names (written or spoken) for which we assume separate input recognition devices. The new idea is that there are semantic units with content like 'is a politician', others with content like 'born in Grantham', others with content like 'lived at 10 Downing Street' and others with content like 'is called Margaret Thatcher'. If you know only one Margaret Thatcher, then this SIU will indeed have a one-to-one correspondence with the PIN for Margaret Thatcher. Nevertheless the PINS and the SIUS play very different functional roles. The PINS are a point of multimodal convergence: they allow us to 'add' activation coming from an input face, voice, clothing (presumably) and so forth to decide whether the person is known. They then allow access of the semantic information associated with that person, but do not themselves have semantic content. Of course it is possible to have PINS with no associated 'names' at all (e.g. the PIN corresponding to the person you see on the bus to work each day).

E. T. Rolls (*Department of Experimental Psychology, University of Oxford, U.K.*). An interactive activation (IAC) model with the architecture the authors describe does not seem to be the most natural network to implement a hierarchical processing scheme. For example, how is it consistent with the fact that the latency for naming is longer than that for semantic access (of occupation); or with other evidence for hierarchical organization, showing for example that patients who are unable to access names but able to access semantic data are found; whereas patients unable to access semantic data but able to access names are not found.

V. Bruce. The first thing to note is that the IAC model is presented as by Burton *et al.* (1990, 1991) and Burton & Bruce (1991) is hierarchical: units for material-specific perceptual classification activate units for multi-modal person recognition which in turn activate units deemed to have semantic content. Common phenomena such as knowing that a face is familiar, but not knowing why, are explained in terms of some temporary failure of the route to SIU from PINS. With respect to the specific questions Dr Rolls raises about the relative difficulty of naming, Burton & Bruce (1991) present the results of simulations which show that it is not necessary to distinguish names from other semantic information units to produce the patterns Dr Rolls describes. Given their different pattern of interconnection, SIUS with unique links from PINS become active more slowly, and reach lower eventual levels of activation than those with multiple links. If one assumes that one form of brain damage may weaken the links between PINS and SIUS, it is straightforward to show that such weakening will still allow SIUS of the kind corresponding to occupations to become active while those corresponding to names will not. There is no pattern of supposed 'damage' that could preserve activation of 'unique' SIUS while impairing activation of those with multiple links. Of course this hypothesis about why names are hard to retrieve predicts that certain other kinds of semantic information (such as addresses perhaps) should behave like names. This is an empirical question. Finally, note that all SIUS will in turn have to activate further levels of units responsible for output phonology: our model says nothing about the processes leading to articulation.

H. D. Ellis (*School of Psychology, University of Wales College of Cardiff, U.K.*). I should like to ask two related questions regarding the authors' decision to move familiarity judgements from the FRU stage (Bruce & Young 1986) to the PIN stage in the IAC model. It seems to me that this poses at least two problems.

1. The cognitive system seems to have no way of knowing whether familiarity has been signalled following face, voice or name.

2. More importantly, Dylan Jones and I have been recently experimenting with people's inability to provide information on voices beyond saying they are familiar. We find that in some 25% of cases this tip-of-the-tongue state can be easily induced: an incidence far higher than for faces.

According to the IAC model once a multimodal PIN has been activated the likelihood of any subsequent retrieval of semantic information should then be independent of the original modality.

V. Bruce. In answer to the first question, the problem here is not really any different from that in other models: e.g. in the Bruce & Young (1986) model, how does the subject know that the semantic information accessed applies to a face, voice or name? Of course there must be other categorical decisions occurring in perception which will enable the person to know what it is they are responding to; for example the object recognition system (not addressed by our model at present) must classify an incoming pattern as a face as opposed to some other object. We propose that it is information provided by other perceptual or cognitive processes that will allow the disambiguation of any apparently disembodied familiarity signals.

I'm afraid it is difficult for me to answer the second question of Professor Ellis without seeing the data to which he refers. First, any comparison of rates of tip-of-the-tongue (TOT) states between materials requires the use of the same items. Second, it would be important for comparison that the voices did not say anything that would give clues to identity from the content of their speech, as this would complicate the supposed routes to identity. Third, his question seems to imply that you are using a feeling of familiarity as synonymous with a TOT state. According to our model, a TOT state involves maximum activation at the PIN combined with a block from PIN to SIUS. We agree that if the rates of genuine TOT states differed radically for faces and voices of the same people then this would pose some problems for our model. However, 'feelings of familiarity', as opposed to genuine TOT states, can be of different strengths (this is true for faces also) and it may be that subjects use different criteria for reporting these depending on input material. These different criteria might arise for a number of reasons (one speculative example: it is rude not to recognize a friend from their face, but not so rude if you don't recognize their voice on the telephone, therefore we may need to operate a different criterion to seek additional information when the input is a voice rather than a face).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES