1.2.1 The uniqueness of faces

We have a remarkable ability to discriminate faces. When we know a face well, we can discriminate it from the face of anyone we meet. Familiar faces appear to be unique. Pliny (23-79 A.D.) in Historia Naturalis wrote, "The human features and countenance, although composed of but some ten parts or little more, are so fashioned that among so many thousands of men there are no two in existence who cannot be distinguished from one another." Thomas Browne (1643, p. 57) attempted to explain "how among so many millions of faces, there should be none alike" by an analogy with the vast number of words which have, more or less carelessly, been assembled out of the letters of the alphabet. The analogy is a bad one as there are plenty of homographs to be found, but probably no true facial 'doubles'. However Browne's point is that as long as there are enough features to vary, the uniqueness of faces is not surprising. But how many features do there have to be? Or, to update the question, how much information must we hold about a face for it to be apparently unique?

It is possible to put an approximate lower limit on the information needed to recognise a familiar face. First it is necessary to make a tentative statement about face recognition by the average person: a representative person has a number of close acquaintances whose faces would not be confused with anyone else in the country. Let us say the population of the country is n people. The features of one acquaintance are memorised using a certain amount of information. What is the total number of different faces that could be encoded with this same amount of information? Suppose i discrete faces could be encoded at this level of information. Now each of the n people in the country is going to have one of these i faces. If we assume that everyone has an equal probability of having any of the i faces, this will be the most favourable distribution with regard to the information that has to be stored. As we are only concerned with finding a lower limit, the assumption of a random distribution is justified.

Clearly i must be larger than n if faces are to appear unique, but how much larger? The random distribution can be described by a Poisson distribution where any of the i faces may belong to nobody, or to just one person, or to two people, or to three people, and so on. The mean number of people who possess any particular face will be n/i. According to the Poisson distribution the probability that any face will belong to one person and one person only, is given by,

$$\frac{n\,e^{-n/i}}{i}$$

And so the actual number of faces that belong to only one person will be,

$$i \cdot \frac{n\,e^{-n/i}}{i} = n\,e^{-n/i}$$

Some consideration will show that this must also be equal to the number of people in the country who have a face that no one else possesses. The fraction of the country's population with unique faces is given by,

$$\frac{n\,e^{-n/i}}{n} = e^{-n/i}$$

This argument is in terms of one representative person's perception. It is not, of course, possible to talk about discriminating faces without reference to a perceptual system. The argument is limited to the case of recognising the face of a single acquaintance - other faces are considered solely as possible sources of confusion. Two rather arbitrary estimates are needed to turn this algebra into arithmetic. First let us assume that the original statement (that the face of a close acquaintance will not be confused with any other face in the country) has at least a .95 probability of being true, and second, that the country's population is 50 million. We can now write,

$$e^{-5\times10^{7}/i} \gg .95$$

$$\frac{-5\times10^{7}}{i} \gg \log_e .95$$

$$i \gg \frac{5\times10^{7}}{.051} \simeq 10^{9}$$

At least 10^9 discriminable faces could be stored at this level of information. Therefore the information for one face must be at least $\log_2 10^9 \simeq 30$ bits.

This probably represents a realistic lower limit on the information that must be stored about a face that is to be discriminated with reasonable accuracy from a large number of other people. Thirty bits is not remarkable in itself: the information is about the same as in a typical S.T.D. telephone number (e.g. 0273-681021, where the only redundancy is in the initial O). But faces are such similar looking things, it is necessary to explain how useful information in excess of thirty bits can be so easily extracted from them. Almost certainly, part of the process of learning faces must involve much more than a static process of storing cues. There must be a considerable amount of sampling of the distribution of facial characteristics in the general population. A face becomes familiar over some period of time. Perhaps this face is unconsciously compared with the faces of many other people in order to select the most useful cues for its discrimination. As this process goes on, the familiarity of the face grows gradually. On the other hand the sampling process may not occur in this way at all, but it could simply be done on the basis of a massive sampling of the distribution of features that is continually being updated.

In conclusion, what is surprising is not that faces appear to be unique, nor the amount of information we need to store them. Telephone numbers are also unique and also require a substantial amount of information to remember them. But what is most surprising, and most difficult to explain, is the encoding of this information from a class of such apparently similar objects as faces.