Face Perception

Pawan Sinha[1], Yuri Ostrovsky[1] and Richard Russell[2]

[1]Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology

[1]Department of Psychology
Harvard University

Faces constitute one of the most pervasive and important classes of objects in our visual environment. The human visual system is remarkably proficient at various face perception tasks, including identification, emotion assessment and gaze estimation, besides gender, race and age classification. Much research effort has been directed towards elucidating the facial cues that underlie these different assessments, with identification attracting the lion's share. Here we provide an overview of four key aspects of human face perception performance: 1. Tolerance for degradations, 2. Relative contribution of geometric and photometric cues, 3. Development of face perception skills, and 4. Neural correlates of face-perception.

1. What are the limits of human face recognition skills?

The high resolution of foveal vision allows the human visual system, in principle, to discriminate between individuals on the basis of fine differences in their facial features. Analogously, progressive improvements in camera resolution make it possible to use increasing amounts of detail in face representations in machine vision systems. The advent of iris based biometric systems, which use the fine details of iris patterns as identifying signatures, is a case in point. However, the problem that such details-based schemes often have to contend with is that high-resolution images are not always available. This is particularly true in situations where individuals have to be recognized at a distance. When seen from a distance, the image of a face occupies a small area of the retina, and thus is sampled by very few photoreceptors. In effect, only the coarse low-resolution structure of distant faces is available for further processing. Precisely how does face identification performance change as a function of image resolution? This question is relevant not only for characterizing human face recognition performance, but also for understanding the fundamental face representation strategies used by the visual system.

Working with block averaged images of familiar faces of the kind shown in figure 1a, researchers have found that observers can recognize faces well above chance levels with image resolutions of merely 7x10 pixels, and reach ceiling levels by resolutions of 25x30 pixels. The remarkable tolerance of the human visual system to resolution reduction demonstrates that fine featural details are not necessary to obtain good face recognition performance. Furthermore, given the indistinctness of the individual features at low resolutions, it appears likely that diagnosticity resides in their overall configuration. However, precisely which aspects of this configuration are important, and how we can computationally encode them, are largely open questions. Tentative clues to the answer come from experiments probing the relative saliencies of different facial features. Figure 1b, for instance, demonstrates the perceptual importance of overall head configuration for face recognition. The internal features on their own and even their mutual configuration appear to be insufficient to account for the impressive recognition performance of subjects with full face images.

Figure 1*. (a) Images such as the one shown here have been used by several researchers to assess the limits of human face identification processes.(After Harmon and Julesz, 1973). (b) Although this image appears to be a fairly run-of-the-mill picture of Bill Clinton and Al Gore, a closer inspection reveals that both men have been digitally given identical inner face features and their mutual configuration. Only the external features are different. It appears, therefore, that the human visual system makes strong use of the overall head shape in order to determine facial identity. (From Sinha and Poggio, 1996)*

2. What cues do humans use for face recognition?

For a cue to be useful for recognition, it must differ across faces and be consistent for a given face. Such a cue would be an 'intrinsic' attribute of the face, such as its two and three-dimensional shape, or pigmentation. A cue that differs indiscriminately, irrespective of facial identity, would be expected to be ignored by the human visual system. This would be an 'extrinsic' cue, such as illumination. How perceptually important are these two sources of facial image variability?

Illumination greatly affects the image level appearance of a face. Surprisingly, humans appear to be unable to completely ignore illumination induced image changes while performing recognition tasks. When observers are asked to match different images of the same face, performance is worse when the two images of a face to be matched are illuminated differently. However, this confound is less pronounced for familiar faces, and naming performance for such faces is largely invariant across different illumination conditions, save for some very unusual circumstances, such as lighting from below. Taken together, these results suggest that illumination induced image changes are included in facial representations, but with increasing familiarity with a face (and, presumably, viewing it across many different lighting conditions), the contribution of illumination to recognition performance becomes less pronounced.

Besides lighting, two additional sources of variability for faces are their shape, and their pigmentation. Given that we can recognize individuals even in unpainted stone busts, it is clear that shape cues on their own can support recognition. These anecdotal observations are supported by results from systematic studies with 3D laser-scanned faces that similarly lack variation in pigmentation. However, the ability to recognize a face in the absence of pigmentation cues does not mean that pigmentation cues are not used when available. Indeed, there is reason to believe that pigmentation may also be important for face recognition. Unlike other objects, faces are much more difficult to recognize from a line drawing than from a photograph suggesting that the missing pigmentation cues may well be important.

Recently, some researchers have attempted to directly compare the relative importance of shape and pigmentation cues for face recognition. The basic approach is to create a set of face representations with the shape of a particular face and the average pigmentation of a large group of faces, and a separate set of face representations with the pigmentation of an actual face and the average shape of many faces. The rationale here is that to distinguish among the faces from the set that all have the same pigmentation, subjects must use shape cues, and vice versa for the set of faces with the same shape. Subjects are trained with one or the other set of face images, and are subsequently tested for memory. Recognition performance is found to be about equal with both the shape and pigmentation sets, suggesting that both cues are important for recognition.

To summarize the findings on the question of what cues are used in face recognition, there is evidence that all sources of facial image variability impact the internal representations. When considering extrinsic cues such as illumination, this is somewhat of a drawback, since a factor that is unrelated to facial identity ends up influencing performance. However, for the intrinsic cues of shape and pigmentation, the visual system behaves in an appropriately opportunistic fashion, making use of all cues that vary across exemplars of this class to achieve its impressive face recognition skills.

3. What is the timeline of development of human face recognition skills?

Basic face perception skills develop surprisingly rapidly in the course of an infant's development. As early as a few days after birth, infants appear to be able to distinguish their mother's face from other faces. Yet, after this initial phase of rapid progression in face recognition abilities, the full maturing of children's face processing abilities is a long drawn out process, estimated to take over a decade. This section outlines the trajectory of development of face processing.

Bootstrapping Face Processing: Evidence from Neonates

As is the case for most visual skills, face processing must be bootstrapped with some primitive mechanism from which more advanced processes can be learned. A key question is whether infants are born with some innate abilities to process faces or are those abilities a consequence of general visual learning mechanisms? To examine this issue, neonates (newborns) are assessed for various abilities as soon as is practical after birth. Three major findings have historically been taken as evidence for innate facial processing abilities: (1) The initial preference for faces over non-faces (2) The ability to distinguish the mother from strangers, and (3) imitation of facial gestures. We look at each of these in turn.

Innate facial preference

Are infants pre-wired with a face detection algorithm? If infants knew how to locate the faces in an image, it would be a valuable first step in the subsequent learning of face recognition processes. John Morton and Mark Johnson formalized this idea into a theory called *CONSPEC and CONLERN*. CONSPEC is the structural information which guides newborn preferences for face-like patterns. CONLERN is a learning process which extracts further visual characteristics from patterns localized based on CONSPEC.

Some supporting evidence for this theory comes from the fact that newborns do indeed preferentially orient their gaze to face-like patterns. When presented with a rudimentary drawing of a face, upright or inverted (figure 2), newborns are reported to gaze longer at the upright pattern. Since the two patterns are largely identical except for their "faceness" and the infants have had very little visual experience prior to the presentation, this finding is interpreted as a demonstration of newborns' innate preference for faces.

Figure 2. *Newborns preferentially orient their gaze to the face-like pattern on the left, rather than the one shown on the right, suggesting some innately specified representation for faces. (After Johnson and Morton, 1991).*

Distinguishing the mother from strangers

Further evidence for innate mechanisms for facial processing comes from the remarkable ability of newborn infants within less than 50 hours of birth to discriminate their mothers from strangers. Newborns suck more vigorously when viewing their mother's face on a videotaped image. They also are capable of habituating to a mother's image, eventually preferring a novel image of a stranger, showing a classic novelty preference. External facial features, especially hair, are believed to underlie much of this discriminative performance of newborns.

Imitation of facial expressions

It has been reported that newborns are capable of imitating their caregivers' facial expressions. This would entail the infant's recognizing the expression, then mapping it to its own motor functions. Many studies have shown evidence of imitation. However, there have been some failures to replicate this result. Moreover, consistent positive results have been found primarily for one expression, namely, tongue protrusion. This action might be an innate releasing mechanism, perhaps to a surprising stimulus, or an attempt at oral exploration. Thus, the action might be a response to the stimulus, not a recognition of it followed by imitation.

Developmental Progression Beyond Infancy

Although infant face recognition proficiency develops rapidly within the first few months of life, performance continues to improve up to the age of 10 years or even later. The best-studied progression is the use of featural versus configural facial cues.

Adults match upright faces more easily than inverted faces (the so-called inversion effect, (Yin, 1969)). This is believed to be a consequence of the disruption in configural processing in inverted faces. Interestingly, four month old infants do not show this decrement in performance if the images to be matched are otherwise identical. However, the inversion effect does appear in 4-month olds if, for instance, pose is manipulated at each presentation of the face. Thus, there is evidence that configural cues start to play some role in face processing at this early age, although the processing of these cues is still rudimentary. Processing based on features appears to play the primary role in infant facial recognition. Given the early onset of the usage of configural cues in child face recognition, rudimentary though it may be, one would expect that full maturation of such a fundamental system would ensue rapidly. However, numerous studies have found that, although face recognition based on features reaches near-adult levels by the age of 6 years, configural processing still lags behind until at least 10 years of age, with a gradual progression of greater accuracy and dependence on configural cues.

4. What are the neural underpinnings of human face-perception?

Whether or not faces constitute a "special" class of visual stimuli has been the subject of much debate for many years. Since the first demonstrations of the "inversion effect" described above, it has been suspected that unique cognitive and neural mechanisms may exist for face processing in the human visual system.

Indeed, there is a great deal of evidence that the primary locus for human face processing may be found on the fusiform gyrus of the extra-striate visual cortex. This region shows an intriguing pattern of selectivity (schematic faces do not give rise to much activity) and generality (animal faces do elicit a good response), suggesting a strong domain-specific response for faces. In keeping with behavioral results, the "fusiform face area" (FFA) also appears to exhibit an "inversion effect". Overall, the characterization of the FFA as a dedicated face processing module appears very strong. Developmental studies indicate a progressive strengthening of FFA activity with age. In a paradigm comparing the passive viewing of faces versus houses, younger children (8-10 years) showed weaker activation in the fusiform gyrus, relative to older (12-14 years) ones. Thus, although the neuroimaging data are very preliminary at the present, they do appear to be broadly consistent with the behavioral data showing continuing development of face processing well into adolescence.

It must be noted that the debate over faces being "special" is far from over. It has been suggested that rather than being a true "face module," the FFA may be responsible for performing either subordinate or "expert-level" categorization of generic objects. For instance, in dog experts, this cortical patch might be involved in discriminating one dog exemplar from another, while in a car expert, it might help distinguish different cars. The reason this area exhibits responsivity to faces might be that most observers have attained expertise in face discrimination. There are results from both behavioral studies and neuroimaging studies that lend some support to this "perceptual expertise" account.

There is neurophysiological evidence that at least some face recognition tasks may be carried out over a surprisingly short period of time. Neurons in primate inferotemporal (IT) cortex can exhibit selectivity to stimuli that are more complicated than the simple gratings and bars that elicit responses from cells in early visual areas. In particular, it has been noted that there are some cells in IT cortex that are selective for faces. Moreover, the latency of response in these cells is in the neighborhood of 80-160ms. The computational relevance of these results is that recognition as it is performed up to the level of IT cortex probably requires only one feed-forward pass through the visual system. Feedback and iterative processing are likely not major factors in the responses recorded in these studies.

References

Carey, S., & Diamond, R. (1977). From Piecemeal to Configurational Representation of Faces. *Science,* **195**(4275), 312-314.

Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *The Journal of Neuroscience,* **4**(8), 2051-2062.

Harmon, L. D. (1973). The recognition of faces. *Scientific American*, **229**(5), 70-83.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17*(11), 4302-4311.

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of Facial and Manual Gestures by Human Neonates. *Science,* **198**(4312), 75-78.

Morton, J., & Johnson, M. H. (1991). Conspec and Conlern - a 2-Process Theory of Infant Face Recognition. *Psychological Review,* **98**(2), 164-181.

Sinha, P. and Poggio, T. (1996) I think I know that face…, *Nature*, **384**, 404.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology,* **81**, 141-145.