NIH Public Access
Author Manuscript
Annu Rev Neurosci. Author manuscript; available in PMC 2009 January 21.

Published in final edited form as:
Annu Rev Neurosci. 2008 ; 31: 411–437. doi:10.1146/annurev.neuro.30.051606.094238.

Mechanisms of face perception

Doris Y. Tsao and
Centers for Advanced Imaging & Cognitive Sciences Bremen University D-28334 Bremen, FR Germany and Athinoula A. Martinos Center for Biomedical Imaging Charlestown MA 02129, USA email: doris@nmr.mgh.harvard.edu

Margaret S. Livingstone
Department of Neurobiology Harvard Medical School Boston, MA 02115, USA email: mlivingstone@hms.harvard.edu

Abstract

Faces are among the most informative stimuli we ever perceive: Even a split-second glimpse of a person's face tells us their identity, sex, mood, age, race, and direction of attention. The specialness of face processing is acknowledged in the artificial vision community, where contests for face recognition algorithms abound. Neurological evidence strongly implicates a dedicated machinery for face processing in the human brain, to explain the double dissociability of face and object recognition deficits. Furthermore, it has recently become clear that macaques too have specialized neural machinery for processing faces. Here we propose a unifying hypothesis, deduced from computational, neurological, fMRI, and single-unit experiments: that what makes face processing special is that it is gated by an obligatory detection process. We will clarify this idea in concrete algorithmic terms, and show how it can explain a variety of phenomena associated with face processing.

Keywords

Face processing; Face cells; Holistic processing; Face recognition; Face detection; Temporal lobe

1. Introduction

The central challenge of visual recognition is the same for both faces and objects: we must distinguish among often similar visual forms despite substantial changes in appearance arising from changes in position, illumination, occlusion, etc. Although face identification is often singled out as demanding particular sensitivity to differences between objects sharing a common basic configuration, in fact such differences must be represented in the brain for both faces and non-face objects. It is true that most humans can easily identify hundreds of faces (Diamond & Carey 1986), but, even if one cannot recognize a hundred different bottles by name, one can certainly distinguish them in pairwise discrimination tasks. Further, most of us can recognize tens of thousands of words at a glance, not letter by letter, a feat requiring expert detection of configural patterns of non-face stimuli. Thus, face perception is in many ways a microcosm of object recognition; and the solution to the particular problem of face recognition will undoubtedly yield insights into the general problem of object recognition.

The system of face-selective regions in the human and macaque brain can be defined precisely using fMRI, so we can now approach this system hierarchically and physiologically to ask

Correspondence to: Margaret S. Livingstone.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

mechanistic questions about face processing at a level of detail previously unimaginable. Here we review what is known about face processing at each of Marr's levels: computational theory, algorithm, and neural implementation.

Computer vision algorithms for face perception divide the process into three distinct steps. First, the presence of a face in a scene must be detected. Then the face must be measured, to identify its distinguishing characteristics. Finally, these measurements must be used to categorize the face in terms of identity, gender, age, race, and expression.

Detection

The most basic aspect of face perception is simply detecting the presence of a face, which requires the extraction of features that it has in common with other faces. The effectiveness and ubiquity of the simple T-shaped schematic face (eye, eye, nose, mouth) suggest that face detection may be accomplished by a simple template-like process. Face detection and identification have opposing demands: The identification of individuals requires a fine-grained analysis to extract the ways in which each face differs from the others despite the fact that all faces share the same basic T-shaped configuration, whereas detection requires extracting what is common to all faces. A good detector should be poor at individual recognition, and vice versa.

Another reason why detection and identification should be separate processes is that detection can act as a domain-specific filter, ensuring that precious resources for face recognition (e.g., privileged access to eye movement centers (Johnson et al 1991)) are used only if the stimulus passes the threshold of being a face. Such domain-specific gating may be one reason for the anatomical segregation of face processing in primates (it is easier to gate cells that are grouped together). A further important benefit of preceding identification by detection is that detection automatically accomplishes face segmentation, i.e., isolates the face from the background clutter, and can aid in aligning the face to a standard template. Many face recognition algorithms require prior segmentation and alignment and will fail with non-uniform backgrounds or varying face sizes.

Measurement & categorization

After a face has been detected, it must be measured in a way that allows for accurate, efficient identification. The measurement process must not be so coarse as to miss the subtle features that distinguish one face from another. On the other hand, it must output a set of values that can be efficiently compared to stored templates for identification. There is a zero sum game between measurement and categorization: the more efficient the measurement, the easier the classification; conversely, less efficient measurement (e.g., a brute force tabulation of pixel gray values), makes the classification process more laborious.

2. Computer Vision Algorithms

A comprehensive review of computer algorithms for face recognition can be found in Zhao et al. (2003) and Shakhnarovich and Moghaddam (2004). Our goal here is to discuss algorithms that offer special insights into possible biological mechanisms.

Detection

How can a system determine if there is a face in an image, regardless of whose it is? An obvious approach is to perform template matching (e.g., search for a region containing 2 eyes, a mouth, and a nose, inside an oval). In many artificial face-detection systems a template is swept across the image at multiple scales, and any part of the image that matches the template is scored as a face. This approach works, but it is slow.

IPR2023-01014
Nielsen Company EX2012 Page 2

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

To overcome this limitation, Viola and Jones (2004) introduced the use of a cascade of increasingly complex filters, or feature detectors. Their reasoning was that the presence of a face can be ruled out most of the time with a very simple filter, thus avoiding the computational effort of doing fine-scale filtering on uninformative parts of the image. The first filter in their cascade consists of only 2 simple features, each composed of a few rectangular light and dark regions (Figure 1a). Subsequent stages of filtering are performed only on regions scoring positive at any preceding stage. This cascade approach proved just as accurate, but 10 times faster, than single-step face detector algorithms.

Sinha's face detection algorithm (Sinha 2002a) is based on the observation that qualitative contrast relationships between different parts of a face are highly conserved, even under different lighting conditions (Figure 1b). Even though any single contrast relationship between two facial regions would be inadequate to detect a face, a set of such relationships could be (because probabilities multiply). A subset of Sinha's directed contrasts ([r2, r3] and [r4, r5]) are equivalent to the first stage of the Viola Jones face detector.

Effective primitives for face detection can also be computed using an information theory approach by identifying “fragments” (subwindows) of face images that are maximally informative about the presence or absence of a face (Ullman et al 2002). The resulting fragments consist of medium-resolution face parts, e.g., an eye, rather than the whole face, so in this algorithm face detection is triggered by detection of a threshold number of such fragments.

All three algorithms discussed above use basic feature detectors much simpler than a whole face (rectangle features in the Viola Jones algorithm, qualitative contrast ratios between pairs of face regions in the Sinha algorithm, and face parts in the Ullman algorithm). Yet, all three algorithms perform holistic detection, that is, they obligatorily detect faces as correctly arranged wholes. This is because all three algorithms detect overlapping constellations of elemental features that cover the whole face. The feature overlaps implicitly enforce the correct overall arrangement of features.

Measurement

Once a face has been detected, it may need to be identified or classified. Algorithms for the identification of individual faces generally are either feature-based or holistic. In feature-based methods, fiducial points (e.g., eyes, mouth, nose) are identified and used to compute various geometric ratios. As long as the features can be detected, this approach is robust to position and scale variations. In holistic methods, the entire face is matched to memory templates without isolating specific features or parts. One advantage of holistic methods is that all parts of the face are used, and no information is discarded.

The simplest holistic recognition algorithm is to directly correlate a presented image to a bank of remembered templates, but having templates for every face is expensive in time and storage space. Turk and Pentland (1991) developed the Eigenface algorithm to overcome these limitations. The Eigenface algorithm exploits the fact that all faces share a common basic structure (round, smooth, symmetric, two eyes, a nose, and a mouth). Thus the pixel arrays defining various faces are highly correlated, and the distinguishing characteristics of a face can be expressed more efficiently if these correlations are removed using principal components analysis (PCA). When PCA is performed on a large set of faces, the eigenvectors with largest eigenvalues all look like faces, and hence are called “Eigenfaces” (Figure 2a). An arbitrary face can be projected onto a set of Eigenfaces to yield a highly compressed representation; good face reconstructions can typically be obtained with just 50 Eigenfaces and passable ones with just 25. In other words, something as ineffable as an identity can be reduced to 25 numbers (Figure 2b).

IPR2023-01014
Nielsen Company EX2012 Page 3

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

PCA on sets of faces varying in both expression and identity generates some principal components that are useful for only expression or only identity discrimination, and others that are useful for both (Calder et al 2001). This partial independence of PC's can successfully model the independent perception of expression and identity (Cottrell et al 2002).

The Eigenface algorithm does not perform well if the sample face is not accurately aligned in scale and position to the template Eigenfaces. Human face perception, on the other hand, is tolerant to changes in both scale (Figure 5d) and position. Moreover, if a face is transformed further along the morph line representing the deviation of that face from the average face, the transformed face is easily recognized as the same individual (e.g. Figure 5c); this is the basis of caricature (Leopold et al 2001). The process of morphing one individual into another (Wolberg 1996) involves both an intensity transform (which Eigenfaces model very well) and a simultaneous geometric transform. Because Eigenfaces represent axes of intensity values on a fixed spatial basis, the Eigenface approach does not interpret caricature transformations as the same individual.

Jebara (2003) proposed a clever way to get around the spatial rigidity of the original Eigenface approach: instead of performing PCA on the intensity values, the size of the representation is tripled, so each pixel conveys not only the image intensity value but also the intensity value's (x,y) location. Then PCA can be done on the triple-sized image containing a concatenation of (x,y,I) values (Figure 3). The power of this approach is that spatial coordinates are treated just like intensity coordinates and can also be represented in the resultant Eigenfaces. The fact that this “bags of pixels” approach performs three orders of magnitude better than standard Eigenface analysis on face sets with changes in pose, illumination, and expression is computational proof of the importance of representing geometric variations in addition to intensity variations.

Categorization

Turk and Pentland used a simple Euclidean distance metric on face Eigen-coordinates to perform recognition. More powerful classifiers that have been applied to the problem of face recognition include Fisher linear discriminants (Belhumeur et al 1997), Bayesian Estimation (Moghaddam et al 2000), and Support Vector Machines (Shakhnarovich & Moghaddam 2004). These classification techniques can be regarded as second-tier add-ons to the basic Eigenface measurement system. Measurement yields analog descriptions, while classification is non-linear and yields discrete boundaries between descriptions.

Separating the process of measurement from the process of classification gives a computational system maximum flexibility, because different categorizations (e.g., emotion, identity, gender) can all operate on the same set of basic eigenvector projections. Gender determination can be based on large eigenvalue eigenvectors, while identification of individuals relies on lower value eigenvectors (O'Toole et al 1993). Furthermore, because classifications are necessarily nonlinear, if mechanisms for classification are indeed independent from the mechanisms for measurement, this would be very exciting from an experimental point of view, because the templates used for measurement could be linear and therefore their detailed structure could be mapped. We will return to the idea of linear measurement when we discuss tuning properties of face cells.

Invariance

Developing position and scale invariant recognition is a huge challenge for artificial face recognition systems. Initial attempts to compute a meaningful set of Eigen-coordinates for a face required that the face be accurately aligned in scale, position, and rotation angle to the template Eigenfaces. However, if, as we propose, face detection precedes measurement, the

IPR2023-01014
Nielsen Company EX2012 Page 4

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

detector can determine the location, size, and rotation angle of the eyes and face outline, and then use these to normalize the input to face measurement units.

Summary

The main lesson we can extract from artificial systems for face processing is that detection and recognition are distinct processes, with distinct goals, primitives (coarse contrast relationships versus detailed holistic templates), and computational architectures (filter cascade versus parallel measurements). By preceding recognition, detection can not only act as a domain-specific filter to gate subsequent processing, but it can include alignment and segmentation, preparing faces for subsequent measurement. The effectiveness of the Eigenface algorithm for face recognition shows that faces can be represented by their deviation from the average in a compressed subspace. To most effectively characterize faces, this subspace needs to include spatial variations as well as intensity variations.

Some machine vision models of recognition use common “meta-algorithms” to learn the primitives for both detection and recognition of faces (Riesenhuber & Poggio 2000, Ullman 2007) thus the two processes may share core computational principles. Whether biological systems use discrete steps of detection, measurement, and classification to recognize faces is a question that can only be resolved empirically.

3. Human Behavior and functional imaging

The extensive behavioral literature on face perception provides a rich source of clues about the nature of the computations performed in processing faces. One of the hallmarks of face processing is that recognition performance drops substantially when faces are presented upside down (Figure 5a) or in negative contrast, and both effects are much smaller for objects (Kemp et al 1990, Yin 1969). We propose that both these properties can be explained if only upright, positive-contrast faces gain access to the face processing system, i.e., if an upright, positive-contrast template is used for face detection. This template may be innate in humans, as evidenced by the tendency for newborns to track normal schematic faces longer than scrambled schematic faces (Johnson et al 1991, Simion et al 1998).

Detection

As argued above, it is computationally efficient to separate detection and recognition, and to have detection precede recognition because detection can act as a domain-specific filter to make the recognition process more efficient (by focusing recognition on regions actually containing faces). That there are also separate detection and recognition stages in human face processing fits with one of the most striking findings from the neuropsychology literature: Patient CK, who is severely impaired at object recognition, including many basic midlevel visual processes, is nonetheless 100% normal at face recognition (Moscovitch et al 1997). His pattern of deficits indicates that face processing is not simply a final stage tacked onto the end of the non-face object recognition pathway, but rather a completely different pathway that branches away from object recognition early in the visual hierarchy, and it is this branching off that we propose to equate with the detection process. CK's dissociation is illustrated by his perception of the face made up of vegetables by Arcimbaldo—CK sees the face, but not the constituent vegetables.

CK's ability to recognize famous or familiar faces is at least as good as normal controls, until the faces are shown upside down, and then his performance becomes much worse than that of controls. Conversely prosopagnosics perform better than controls in recognizing inverted faces (Farah et al 1995). This double dissociation of the inversion effect is consistent with the existence of a face-specific processing system that can be accessed only by upright faces,

IPR2023-01014
Nielsen Company EX2012 Page 5

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

present in CK and absent in prosopagnosics. Presumably CK can only process objects using the face-specific system; prosopagnosics have a general object recognition system but not the face-specific system, and normal subjects have both systems. The general non-face object system is not as good at processing faces as the face-specific system (hence the inversion effect in normal subjects), is missing in CK (hence his disproportionate deficit for inverted faces), and is the only way prosopagnosics can process any face (hence their relatively better behavior than normals with inverted faces, because of extra practice).

"Holistic" Processing of Faces

Face processing is said to be distinct from non-face object processing in that it is more "holistic"; that is, faces are represented as non-decomposed wholes, rather than as a combination of independently-represented component parts (eyes, nose, mouth), and the relations between them (Farah et al 1998). Evidence for holistic processing of faces comes from a number of behavioral paradigms, of which the two most cited are the part-whole effect (Tanaka & Farah 1993) and the composite effect (Young et al 1987). In the part-whole effect, subjects are better able to distinguish two face parts when the parts are presented in the context of a whole face than in isolation. In the composite effect, subjects are slower to identify half of a chimeric face if it is aligned with an inconsistent other half-face than if the two half-faces are misaligned (Young et al 1987). As with the part-whole effect, the composite effect indicates that even when subjects attempt to process only part of the face, they suffer interference from the other parts of the face, suggesting a lack of access to parts of the face and mandatory processing of the whole face.

One interpretation of uniqueness of face processing is that it uses special neural machinery not shared by other kinds of objects, an idea that is consistent with functional imaging studies, as described below. Another interpretation is that holistic processing is characteristic of any kind of object that must be distinguished on a subordinate level, especially objects with which the subject is highly trained or familiar (Diamond & Carey 1986). It is not yet clear what the perceptual phenomenology of holistic processing implies either mechanistically or computationally. We suggest that holistic face processing can be explained by an obligatory detection stage that uses a coarse upright template to detect whole faces (Figure 4). An aligned chimera would be detected as a whole face and therefore would be processed as a unit by subsequent measurement and classification stages.

However, we cannot rule out alternatives, such as "one-stage" models in which both face detection and identification are carried out by the same set of face-selective cells. In this case, to explain holistic properties of face processing, we would have to postulate that individual face cells unlike non-face cells are selective not just for local features but for whole faces, or that the readout of face information must comprise all or most of the population code. Either or both of these models would produce the behavioral holistic effects, even without an earlier detection gate. The key evidence favoring our early detection gating hypothesis over a single-stage system comes from the identification of a series of face-selective areas in the macaque (Pinsk et al 2005, Tsao et al 2003) and the finding that an area early in this hierarchy already consists entirely of face-selective cells (Tsao et al 2006); both of these results will be discussed more extensively below.

Although faces are unique in the degree to which they are processed holistically, other non-face objects can also show holistic effects, especially well-learned categories; for review see (Gauthier & Tarr 2002). Words may approach faces in the degree to which they are processed holistically: Coltheart et al. (1993) found that some acquired dyslexics can read whole words and understand their meanings, but cannot distinguish individual letters making up the words. And Anstis (2005) showed that words can show the composite effect, in that observers cannot tell whether two words have same or different top halves.

IPR2023-01014
Nielsen Company EX2012 Page 6

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Norm-based coding

Caricatures are remarkably powerful in evoking recognition (Figure 5c): caricatured faces are often more identifiable than veridical photographs (Lee et al 2000). This finding has led to the proposal that faces are coded in terms of their deviation from the norm, or average, face (Leopold et al 2001, Rhodes et al 1987). Further, the existence of face aftereffects (Figure 5b) shows that the face norm is adaptable (Webster & MacLin 1999). Because such face aftereffects transfer across retinal positions (Leopold et al 2001), and image sizes (Jeffery et al 2006) they apparently do not reflect adaptation to specific low-level image features, but instead indicate adaptation of higher-level representations. This "face identity aftereffect" was interpreted as indicating that adaptation to a given face shifts the norm or average face in the direction of the adapting face, making faces on the opposite side of the norm more distinctive (i.e., more different from the norm). To explain these results Rhodes et al (2006) propose that face identity is coded by pairs of neural populations that are adaptively tuned to above-average and below-average values along each dimension of face space.

Opposite adaptation can occur simultaneously for upright and inverted faces, consistent with the idea that two distinct neural pathways underlie the coding (and adaptation to) upright and inverted faces (Rhodes & Jeffery 2006). Finally, although norm-based coding can only work for classes of stimuli that have similar enough first-order shape that a "norm" can be defined, this situation may not be unique to faces. Rhodes & Mclean (1990) showed evidence for norm-based coding for images of birds, and adaptation effects can also be observed for simple shapes such as taper and overall curvature (Suzuki & Cavanagh 1998). Thus it is possible that adaptive norm-based coding is a general feature of high-level form coding processes.

Human functional imaging

PET studies initially showed activation of the fusiform gyrus in a variety of face perception tasks (Haxby et al 1991, Sergent et al 1992), and subsequently fMRI revealed more of the specificity of these cortical regions for faces, with demonstrations of fusiform regions that responded more strongly to faces than to letter strings and textures (Puce et al 1996), flowers (McCarthy et al 1997), everyday objects, houses, and hands (Kanwisher et al 1997). Although face-specific fMRI activation can also be seen in the superior temporal sulcus (fSTS) and in part of the occipital lobe (the "occipital face area", OFA), the most robust face-selective activation is consistently found on the lateral side of the right mid-fusiform gyrus, the "fusiform face area" or FFA (Kanwisher et al 1997) (Figure 6). The fact that this part of the brain is activated selectively in response to faces indicates that activity in this region must arise at or subsequent to a detection stage.

A number of studies support the idea that the FFA is activated specifically by faces, and not by the low-level stimulus features usually present in faces, that is, activity in the FFA indicates that stimuli have been detected as faces: The FFA shows increased blood flow in response to a wide variety of face stimuli: front and profile photographs of faces (Tong et al 2000), line drawings of faces (Spiridon & Kanwisher 2002), and animal faces (Tong et al 2000). Furthermore, the FFA BOLD signal to upright "Mooney faces" is almost twice as strong as to inverted Mooney stimuli (which have similar low-level features but do not look like faces) (Kanwisher et al 1998). Finally, for bistable stimuli such as the illusory face-vase, or for binocularly rivalrous stimuli in which a face is presented to one eye and a nonface is presented to the other eye, the FFA responds more strongly when subjects perceive a face than when they do not, even though the retinal stimulation is unchanged (Andrews et al 2002, Hasson et al 2001).

Although the FFA shows the strongest increase in blood flow in response to faces, it does also respond to non-face objects. Therefore two alternative hypotheses have been proposed to the

IPR2023-01014
Nielsen Company EX2012 Page 7

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

idea that activity in the FFA represents face-specific processing: The Expertise Hypothesis. According to this hypothesis the FFA is engaged not in processing faces per se, but rather in processing any sets of stimuli that share a common shape, and for which the subject has gained substantial expertise (Tarr & Gauthier 2000). Distributed coding: In an important challenge to a more modular view of face and object processing, Haxby et al. (2001) argued that objects and faces are coded via the distributed profile of neuronal activity across much of the ventral visual pathway. Central to this view is the suggestion that "nonpreferred" responses, for example to objects in the FFA, may form an important part of the neural code for those objects. The functional significance of the smaller but still significant response of the FFA to non-face objects will hopefully be unraveled by the combined assaults of higher resolution imaging in humans and single-unit recordings in non-human primates.

Measurement & Categorization

Does the human brain use separate systems for face measurement and face classification? Some fMRI evidence suggests that it does. For example, in a study of morphing between Marilyn Monroe and Margaret Thatcher, adaptation strength in the OFA followed the amount of physical similarity along the morph line, while in the FFA it followed the perceived identity (Rotshtein et al 2005), suggesting that the OFA performs measurement and the FFA performs classification. However, another study indicates that release from adaptation occurs in the FFA when there are physical differences unaccompanied by changes in perceived identity (Yue et al 2006).

According to Bruce and Young (1986) the processing of facial expression (one form of categorization) and facial identity (another form of categorization) take separate routes. A possible neural basis for this model has been proposed by Haxby and colleagues (2000). According to this proposal, the inferior occipital gyri are involved in early perception of facial features (i.e., measurement). The pathway then diverges, with one branch going to the superior temporal sulcus, which is proposed to be responsible for processing changeable aspects of faces including direction of eye gaze, view angle, emotional expression, and lip movement. The other projection is to the lateral fusiform gyrus, which is responsible for processing identity. A recent review has challenged the Bruce and Young model, arguing that changeable aspects and invariant identity may instead be processed together and rely on partially overlapping visual representations (Calder & Young 2005).

Invariance

Several studies have used fMRI-adaptation for face identity in the FFA, and found invariance to position (Grill-Spector et al 1999), image size (Andrews 2004, Grill-Spector et al 1999), and spatial scale (Eger et al 2004). Thus representations in the FFA are not tied to low-level image properties, but instead show at least some invariance to simple image transformations, though not to viewpoint (Pourtois et al 2005).

Summary

Behavioral studies complement computational approaches in indicating that specialized machinery may be used to process faces and that a face-detection stage gates the flow of information into this domain-specific module. Also reminiscent of successful computational approaches, the gating or detection step may use coarse, simple filters to screen out non-face images. These filters, or templates, require an upright, positive contrast face, with the usual arrangement of features, and images that do not fit the template are analyzed only by the general object recognition system. Even images that pass into the face-specific module are probably processed in parallel by the general system. The face module appears to process images differently from the general object system: Face processing is holistic, in the sense that we cannot process individual face parts without being influenced by the whole face. We suggest

IPR2023-01014
Nielsen Company EX2012 Page 8

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

that this difference arises early in the face module. The face-detection stage may, in addition to gating access, obligatorily segment faces as a whole for further processing by the face module. Finally, substantial recent evidence suggests that face identity is coded in an adaptive norm-based fashion.

Human imaging studies converge on the conclusion that faces are processed in specific locations in the temporal lobe, but the degree of specialization for faces within these locations is debated. The modular interpretation is consistent with neurological findings, and, as described below, with single-unit recordings in macaques. The role of experience in determining both the localization of face processing and its holistic characteristics is also debated. And the relationship, if any, between modular organization and holistic processing is completely unexplored. Only a few visual object categories show functional localization in fMRI: faces, body parts, places, and words (for review see (Cohen & Dehaene 2004, Grill-Spector & Malach 2004)). Faces, bodies, and places are all biologically significant, and their neural machinery could be genetically determined, but the use of writing arose too recently in human history for word processing to be genetically determined, therefore at least one kind of anatomical compartmentalization must be due to extensive experience. We have suggested that the existence of discrete regions of the brain dedicated to face processing implies an obligatory detection stage and that an obligatory detection stage then causes holistic processing. What we know about word processing suggests that it too displays holistic properties, and it is localized, interestingly, in the left hemisphere in an almost mirror symmetric location to the position of the FFA in the right hemisphere (Cohen & Dehaene 2004, Hasson et al 2002).

Monkey fMRI & single unit physiology

Detection

The seminal finding by Gross and his colleagues (1969, 1972) that there exist cells in inferotemporal cortex driven optimally by complex biologically relevant stimuli, such as hands or faces, was novel and initially not well accepted, despite the fact that Konorski (1967) had predicted the existence of face selective cells, or “gnostic” units and that they would be found in IT. Although IT cells do not generally appear to be “detectors” for complex objects, there are consistently observed populations of cells selectively responsive to faces, bodies, and hands, suggesting that faces, bodies, and hands are treated differently from other types of complex patterns, consistent with their also being among the only object category besides faces and words that show localization in human fMRI. But the strong possibility remained that these cells were not really tuned to environmentally relevant objects, but rather to some more abstract basis set, in which all possible shapes are represented by different cells, and some cells were tuned to particular parameters that happened to fit the face or hand stimuli better than any of the other objects tested. Foldiak et al. (2004) recently provided evidence that face selectivity is not just an incidental property of cells tuned to an exhaustive set of image features: they presented 600-1200 stimuli randomly chosen from several image archives to cells recorded from both the upper and lower bank of the STS and found that the distribution of tuning to these images showed bimodality, i.e. cells were either predominantly face selective or not face selective. It is not unprecedented to have specialized neural systems for socially important functions: birds have evolved specialized structures for the perception and generation of song, and in humans there are specialized parts of the auditory and motor systems devoted specifically to language.

Direct evidence that some face cells are used for face detection comes from a microstimulation study by Afraz et al. (2006). Monkeys were trained to discriminate between noisy pictures of faces and non-face objects. Through systematic sampling, Afraz et al. identified cortical locations where clusters of face-selective cells could be reliably recorded. When they

IPR2023-01014
Nielsen Company EX2012 Page 9

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

microstimulated these regions and observed the monkeys' perceptual choices, they found a shift in the psychometric curve favoring detection of a face.

Holistic processing of faces

In general face cells require an intact face and are not just selective for individual features (Bruce et al 1981, Desimone et al 1984, Kobatake & Tanaka 1994., Leonard et al 1985, Oram & Perrett 1992, Perrett et al 1982, Perrett et al 1984, Scalaidhe et al 1999, Tsao et al 2006). Figure 7 shows nonlinear combinatorial response properties of a face selective cell recorded in IT by Kobatake and Tanaka (1994). Out of a large number of 3D objects, this cell responded best to the face of a toy monkey (a), and by testing various simplified 2D paper stimuli, they determined that the cell would also respond to a configuration of two black dots over a horizontal line within a disk (b), but not in the absence of either the spots or the line (c and d) or the circular outline (e). The contrast between the inside and the outside of the circle was not critical (g), but the spots and the bar had to be darker than the disk (h). In other words, the cell only responded when the stimulus looked like a face, no matter how simplified.

The response selectivity of face cells indicates that they must not only combine features nonlinearly but also require them to be in a particular spatial configuration, however such spatial-configuration selective responses and nonlinear combination of features is not restricted to face cells, as such behavior has been reported for other kinds of complex-object selective cells in the temporal lobe (Baker et al 2002, Kobatake & Tanaka 1994, Tanaka et al 1991). Even earlier in the temporal pathway, nonlinear combinatorial selectivity can be seen (Brincat & Connor 2004).

Anatomical specialization of face cells

Most studies on face cells reported face-selective cells scattered throughout the temporal lobe, though they tended to be found in clusters (Perrett et al 1984). Because other kinds of shape selectivities also tend to be clustered (Desimone et al 1984, Fujita et al 1992, Tanaka et al 1991, Wang et al 1996), it was assumed that within the temporal lobe there was a columnar organization for shape, in which face columns represented just one of many shape-specific types of columns. However this view was inconsistent with emerging evidence from human neurology and functional imaging that human face processing was localized to specific, reproducible regions of the temporal lobe. The apparent discrepancy was resolved by two recent studies by Tsao et al. (2003, 2006) who found that in monkeys, as in humans, face processing, as revealed by functional imaging, is localized to discrete regions of the temporal lobe, and they further showed that even at the single unit level, face processing is highly localized (Figure 8, note also Figure 7, top).

Tsao et al. used functional imaging to localize regions in the macaque temporal lobe that were selectively activated by faces, compared to non-face objects, and then they recorded almost 500 single units within the largest of these face-selective regions in two monkeys. They found a remarkable degree of face selectivity within this region, with 97% of the cells being face selective, on average showing almost 20 fold larger responses to faces than to non-face objects. The region where they recorded was quite posterior in the temporal lobe (6 mm anterior to the interaural canal, corresponding to posterior TE/anterior TEO). The fact that an area consisting almost entirely of face-selective cells exists so early in the ventral stream provides strong support for the hypothesis that the face pathway is gated by an obligatory detection stage.

In light of the clear large-scale organization of face processing in macaques revealed by Tsao et al. and recently by Pinsk et al. (2005) we re-examined all previous physiological studies that mapped out locations of face-selective cells, and by remapping their face-cell localizations onto a common map, we found that, taken en masse, these studies do show a higher

IPR2023-01014
Nielsen Company EX2012 Page 10

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

concentration of face selectivity in two major regions of the temporal lobe, regions that correspond to the middle and anterior face patches described by Tsao and colleagues using functional imaging (Figure 8d).

The functional significance of the ananatomical localization of face processing

The cerebral cortex is functionally parcellated: neurons concerned with similar things are organized into areas and columns, each having extensive interconnections and common inputs and outputs (Mountcastle 1997). It is not surprising that face processing, being an identifiable and discrete form of object recognition, is also organized into anatomically discrete processing centers. Individual neurons connect with only a small fraction of the rest of the neurons in the brain, usually to nearby cells, because longer axons delay neural transmission, are energetically expensive, and take up space. Barlow (1986) has pointed out that facilitatory interactions within a functional area or column could underlie Gestalt linking processes—clustering cells concerned with color or motion might facilitate interactions between parts of the visual field having common color or motion. However enriched local inhibitory interactions might be an even more important function of co-localization, since inhibitory neurons are always local, and long-range intracortical connections are invariably excitatory (Somogyi et al 1998). Wang et al. (2000) recorded responses in anterior IT to a set of complex stimuli before during and after applying the GABA receptor antagonist bicuculline near the recording electrode. In many cases, for both face selective and non-face selective cells, blocking local inhibition revealed responses to previously non-activating stimuli, which were often activating stimuli for neighboring cells. This suggests that neighboring cells refine each other's response selectivity by mutual inhibition.

Timecourse of feature-combination responses

Although a large fraction of the information about which face stimulus was shown is carried by the earliest 50 ms of the response of face-selective cells (Tovee et al 1993), several studies have shown that the information carried by the early part of the response is different from the information carried by later spikes. In particular, the earliest spikes in a response are sufficient for distinguishing faces from other object categories, but information about individual facial identity does not develop until about 50 ms later (Sugase et al 1999, Tsao et al 2006).

Similarly, responses in IT to non-face stimuli also become more selective, or sparser, over time (Tamura & Tanaka 2001, Tanaka et al 1991). Similar temporal dynamics indicative of early detection activity followed by later individual identification activity have been observed for face-selective MEG responses in human occipitotemporal cortex (Liu et al 2002). The observation that global information precedes finer information is consistent with a role for local inhibition in sharpening tuning within a local cluster of cells having similar response properties. Such response dynamics suggest a feedback or competitive process, whereby cells that respond best to a given stimulus inhibit nearby cells, resulting in a winner-take-all situation.

Norm-based coding

Recently an idea has emerged for both face processing and general object coding in the temporal lobe—that firing rate represents the magnitude of deviation from a template or norm for that property. Cells in V4 can be tuned to curvature, but the optimal values for curvature are most often found at either extreme or zero curvature, with few cells tuned for intermediate curvature (Pasupathy & Connor 2001). Kayaert and colleages (2005a) found norm based tuning for shapes in IT; neurons tuned for different shapes tended to show monotonic tuning, with maximum responses for extreme values of those shapes. Lastly, Leopold et al. (2006) recorded from face-responsive cells in anterior IT, and found that most cells were tuned around an identity ambiguous average human face, showing maximum firing to faces farthest from an average face (i.e., tuning was V-shaped around the average). Freiwald et al., (2005) on the

IPR2023-01014
Nielsen Company EX2012 Page 11

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

other hand, reported that many cells in the macaque middle face patch showed monotonic turning curves to different feature dimensions in a large cartoon face space, with the maximum response at one extreme and the minimum response at the opposite extreme (Figure 9). This ramp-shaped tuning is consistent with the model proposed by Rhodes et al. (2004) for explaining the face adaptation illusion (Figure 5b)—that each face feature axis is coded by two opponent cell populations. Interestingly for both faces and for non-face objects many cells show tuning to several feature dimensions, and the tuning is separable, or independent, for the different tuning axes (Freiwald et al 2005, Kayaert et al 2005b).

Invariance

Face selective cells in the temporal lobe are usually position and scale invariant in their ability to detect and distinguish faces, but they are seldom view and angle invariant (Desimone et al 1984, Perrett et al 1989, Perrett et al 1991, Perrett et al 1984, Perrett et al 1985, Rolls & Baylis 1986, Tanaka et al 1991, Tovee et al 1994, Tsao et al 2006). The marked view selectivity of some IT cells may reflect a role in interpreting social gestures (who is looking at whom) (Argyle & Cook 1976, Bertrand 1969). De Souza et al. (2005) recently found a striking pattern of view selectivity in rostral vs caudal anterior STS. In caudal anterior STS, they found mirror-symmetric view-tuned cells, but in rostral anterior STS, view tuning was not mirror-symmetric, and furthermore, view angle and gaze direction interacted, with neurons selective for a particular combination of face view and direction of gaze and often strongly modulated by eye contact.

Recordings from the medial temporal lobe of human epilepsy patients have revealed the existence of cells that respond to familiar individuals in a highly invariant manner (Quiroga et al 2005), as expected of a “grandmother cell”. For example, some cells each responded to multiple pictures of a well-known individual as well as to a letter string of their name, but were unresponsive to all other images. Such individual-specific cells have not been found in the lateral inferior temporal lobe, where most face cells in monkeys have been recorded, though as a population cells in the anterior inferior temporal gyrus of the macaque can represent view-invariant identification (Eifuku et al 2004).

Summary

The correlation between fMRI localization of face processing in macaques and the strong clustering of physiologically identified face-selective cells supports the idea of domain specificity, suggested earlier by neurological findings and fMRI studies in humans. The strength and predominance of face selectivity within the middle face patch are not consistent with either the expertise hypothesis or the distributed coding model. The existence of neurons located at an early stage of form processing in the macaque brain that respond selectively to faces supports the idea that face processing begins with a detection stage, and the response properties of face cells indicate that this stage is highly nonlinear.

However face cells seem to measure different face variables independently and linearly, so how does this reconcile with perceptual evidence that face perception in humans is `holistic'; i.e. how can we explain the composite effect and the part-whole” effect neurally? We suggest that both these apparently non-linear perceptual effects can be consistent with a linear neural measuring stage if the preceding detection stage is holistic. One surprising result from physiological studies on face processing is the preponderance of view-selective units, but what role they play in face processing is still unclear.

IPR2023-01014
Nielsen Company EX2012 Page 12

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Future Research Directions

1. Is face processing special? We don't yet understand the details of how either faces or non-face objects are represented in the brain—perceptual studies have shown major differences in the ways that faces and objects are recognized, but there are nevertheless similarities in the response properties between face-selective cells and object-selective cells in IT. Both face and object selective cells in IT show tuning characteristic of a norm-based code. A variety of evidence suggests that our perception of faces is holistic, but processing of some non-face objects, like words, also shows important context effects. One fact is clear: the basic computational challenges to face processing are common to all object recognition (namely: detection, measurement, and classification). What is a face template in computational and neural terms, and how does it differ from a chair template? A truly satisfying answer to this question will come only when we understand the precise neural mechanism underlying both face and non-face object recognition.

2. Is face processing modular? Perhaps the most striking result to come from the neurobiological research on face perception in the past decade is that specialized machinery is used for processing faces. There is a fundamental specialization both at the gross anatomical level and at the level of single cells. It will be exciting to move forward along this pathway, to understand how these face cells are used for different high level percepts and behaviors; e.g., conveying invariant identity, expression, direction of attention, social dominance, etc. But we believe that equally important new insights will come from looking back, asking how these cells acquire their face selectivity—undertaking a systematic study of the process of face detection.

3. What makes face processing special? We have proposed that what is special about face processing is that it is gated by an obligatory detection process. Such a design would be computationally elegant (by allowing for fast domain-specific filtering, segmentation, and alignment prior to fine-grained identification), and could explain the existence of face cells, face areas, prosopagnosia, and holistic processing. This detection-gating hypothesis naturally leads to the idea that there may be two distinct classes of face cells: face recognition cells, which encode precise face templates, and face detector cells, which (contrary to their name) could perform the triple function of detection, segmentation, and alignment. However, it is also possible that detection and discrimination are carried out by the same cells (either simultaneously or sequentially). Either way, we should at least be able to find out the answer. Because we know that face-selective cells are coding faces, we can distinguish detection-related activity from discrimination-related activity, impossible for a cell whose form specialization is unknown. Perhaps what is truly special about face processing is that it is now amenable to being understood. We have a beautiful hierarchy, a gift from nature, and we should exploit it, in both directions.

Mini glossary

BOLD signal, Hemodynamic signal measured in fMRI experiments. Active neurons consume oxygen, leading to a delayed blood flow increase 1-5 seconds later.; Caricature, An artistic technique to enhance the recognizability of a face by exaggerating features distinguishing that face from the average face.; Distributed coding, Representation scheme using distributed activity of coarsely-tuned units. A key challenge is specifying how distributed codes can be read out.; Eigenface, An eigenvector of the covariance matrix defined by a set of faces that allows a compressed representation.; expertise hypothesis, That face-processing mechanisms are used to process any stimuli sharing a common shape and visual expertise.; fMR adaptation, Controversial technique for deducing tuning properties of single cells using fMRI, which

IPR2023-01014
Nielsen Company EX2012 Page 13

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

averages activity of tens of thousands of cells.; gnostic unit, Or “grandmother cell”. A hypothetical cell responding exclusively to a single high-level percept in a highly invariant manner.; inferotemporal cortex, Ventral temporal lobe, including the lower bank of the STS and outer convexity, specialized for visual object recognition.; inversion effect, The ability to recognize an object better when it is upright than inverted, especially strong for faces and words.; Prosopagnosia, A highly specific inability to recognize faces, due to either congenital brain miswiring (“developmental prosopagnosia”) or focal brain lesions (“acquired prosopagnosia”)..

Acronyms

BOLD signal, blood-oxygen-level dependent signal; STS, superior temporal sulcus; PCA, principal components analysis.

LITERATURE CITED

Afraz SR, Kiani R, Esteky H. Microstimulation of inferotemporal cortex influences face categorization. Nature 2006;442:692–5. [PubMed: 16878143]

Andrews TJ, Ewbank MP. Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. Neuroimage 2004;23:905–13. [PubMed: 15528090]

Andrews TJ, Schluppeck D, Homfray D, Matthews P, Blakemore C. Activity in the fusiform gyrus predicts conscious perception of Rubin's vase-face illusion. Neuroimage 2002;17:890–901. [PubMed: 12377163]

Anstis S. Last but not least. Perception 2005;34:237–40. [PubMed: 15832573]

Argyle, M.; Cook, M. Gaze and mutual gaze. Cambridge University Press; Cambridge, Eng. ; New York: 1976. p. xip. 210

Baker CI, Behrmann M, Olson CR. Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. Nat Neurosci 2002;5:1210–6. [PubMed: 12379864]

Barlow HB. Why have multiple cortical areas? Vision Res 1986;26:81–90. [PubMed: 3716216]

Belhumeur PN, Hespanha JP, Kriegnam DJ. Eigenfaces vs Fisherfaces: recognition using class specific linear projection. IEEE Trans. Patt. Anal. Mach. Intell 1997;19:711–20.

Bertrand, M. The behavioural repertoire of the stumptail macaque: a descriptive and comparative study. Karger; Basel: 1969.

Brincat SL, Connor CE. Underlying principles of visual shape selectivity in posterior inferotemporal cortex. Nat Neurosci 2004;7:880–6. [PubMed: 15235606]

Bruce C, Desimone R, Gross CG. Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. J Neurophysiol 1981;46:369–84. [PubMed: 6267219]

Bruce V, Young A. Understanding face recognition. Br J Psychol 1986;77(Pt 3):305–27. [PubMed: 3756376]

Calder AJ, Burton AM, Miller P, Young AW, Akamatsu S. A principal component analysis of facial expressions. Vision Res 2001;41:1179–208. [PubMed: 11292507]

Calder AJ, Young AW. Understanding the recognition of facial identity and facial expression. Nat Rev Neurosci 2005;6:641–51. [PubMed: 16062171]

Cohen L, Dehaene S. Specialization within the ventral stream: the case for the visual word form area. Neuroimage 2004;22:466–76. [PubMed: 15110040]

Coltheart M, Curtis B, Atkins P, Heller M. Models of reading aloud: Dual-route and parallel-distributed-processing approaches. Psychological Review 1993;100:589–608.

Cottrell, GW.; Branson, KM.; Calder, AJ. Presented at 24th Annual Meeting of the Cognitive Science Society, Fairfax; Virginia. 2002.

De Souza WC, Eifuku S, Tamura R, Nishijo H, Ono T. Differential characteristics of face neuron responses within the anterior superior temporal sulcus of macaques. J Neurophysiol 2005;94:1252–66. [PubMed: 15857968]

IPR2023-01014
Nielsen Company EX2012 Page 14

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Desimone R, Albright TD, Gross CG, Bruce C. Stimulus-selective properties of inferior temporal neurons in the macaque. J Neurosci 1984;4:2051–62. [PubMed: 6470767]

Diamond R, Carey S. Why faces are and are not special: an effect of expertise. J Exp Psychol Gen 1986;115:107–17. [PubMed: 2940312]

Eger E, Schyns PG, Kleinschmidt A. Scale invariant adaptation in fusiform face-responsive regions. Neuroimage 2004;22:232–42. [PubMed: 15110013]

Eifuku S, De Souza WC, Tamura R, Nishijo H, Ono T. Neuronal correlates of face identification in the monkey anterior temporal cortical areas. J Neurophysiol 2004;91:358–71. [PubMed: 14715721]

Farah MJ, Wilson KD, Drain HM, Tanaka JR. The inverted face inversion effect in prosopagnosia: evidence for mandatory, face-specific perceptual mechanisms. Vision Res 1995;35:2089–93. [PubMed: 7660612]

Farah MJ, Wilson KD, Drain M, Tanaka JN. What is “special” about face perception? Psychol Rev 1998;105:482–98. [PubMed: 9697428]

Foldiak P, Xiao D, Keysers C, Edwards R, Perrett DI. Rapid serial visual presentation for the determination of neural selectivity in area STSa. Prog Brain Res 2004;144:107–16. [PubMed: 14650843]

Freiwald WA, Tsao D, Tootell RB, Livingstone MS. Single-unit recording in an fMRI-identified macaque face patch. II. Codin along multiple feature axes. Soc. Neurosci. Abstr 2005;362:6.

Fujita I, Tanaka K, Ito M, Cheng K. Columns for visual features of objects in monkey inferotemporal cortex. Nature 1992;360:343–6. [PubMed: 1448150]

Gauthier I, Tarr MJ. Unraveling mechanisms for expert object recognition: Bridging brain activity and behavior. J Exp Psychol Hum Percept Perform 2002;28:431–46. [PubMed: 11999864]

Grill-Spector K. The neural basis of object perception. Curr Opin Neurobiol 2003;13:159–66. [PubMed: 12744968]

Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzchak Y, Malach R. Differential processing of objects under various viewing conditions in the human lateral occipital complex. Neuron 1999;24:187–203. [PubMed: 10677037]

Grill-Spector K, Malach R. The human visual cortex. Annu Rev Neurosci 2004;27:649–77. [PubMed: 15217346]

Gross CG, Bender DB, Rocha-Miranda CE. Visual receptive fields of neurons in inferotemporal cortex of the monkey. Science 1969;166:1303–6. [PubMed: 4982685]

Gross CG, Rocha-Miranda CE, Bender DB. Visual properties of neurons in inferotemporal cortex of the Macaque. J Neurophysiol 1972;35:96–111. [PubMed: 4621506]

Hasson U, Hendler T, Ben Bashat D, Malach R. Vase or face? A neural correlate of shape-selective grouping processes in the human brain. J Cogn Neurosci 2001;13:744–53. [PubMed: 11564319]

Hasson U, Levy I, Behrmann M, Hendler T, Malach R. Eccentricity bias as an organizing principle for human high-order object areas. Neuron 2002;34:479–90. [PubMed: 11988177]

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 2001;293:2425–30. [PubMed: 11577229]

Haxby JV, Grady CL, Horwitz B, Ungerleider LG, Mishkin M, et al. Dissociation of object and spatial visual processing pathways in human extrastriate cortex. Proc Natl Acad Sci U S A 1991;88:1621–5. [PubMed: 2000370]

Haxby JV, Hoffman EA, Gobbini MI. The distributed human neural system for face perception. Trends Cogn Sci 2000;4:223–33. [PubMed: 10827445]

Jebara, T. Images as Bags of Pixels. Presented at Ninth IEEE International Conference on Computer Vision (ICCV'03); 2003.

Jeffery L, Rhodes G, Busey T. View-specific coding of face shape. Psychol Sci 2006;17:501–5. [PubMed: 16771800]

Johnson MH, Dziurawiec S, Ellis H, Morton J. Newborns' preferential tracking of face-like stimuli and its subsequent decline. Cognition 1991;40:1–19. [PubMed: 1786670]

Kanwisher N, Tong F, Nakayama K. The effect of face inversion on the human fusiform face area. Cognition 1998;68:B1–B11. [PubMed: 9775518]

IPR2023-01014
Nielsen Company EX2012 Page 15

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Kanwisher NG, McDermott J, Chun MM. The fusiform face area: A module in human extrastriate cortex specialized for face perception. Journal of Neuroscience 1997;17:4302–11. [PubMed: 9151747]

Kayaert G, Biederman I, Op de Beeck H, Vogels R. Tuning for shape dimensions in macaque inferior temporal cortex. European Journal of Neuroscience 2005a;22:212–24. [PubMed: 16029211]

Kayaert G, Biederman I, Vogels R. Representation of regular and irregular shapes in macaque inferotemporal cortex. Cereb Cortex 2005b;15:1308–21. [PubMed: 15616128]

Kemp R, McManus C, Pigott T. Sensitivity to the displacement of facial features in negative and inverted images. Perception 1990;19:531–543. [PubMed: 2096371]

Kobatake E, Tanaka K. Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. J Neurophysiol 1994;71:856–67. [PubMed: 8201425]

Konorski, J. Integrative activity of the brain; an interdisciplinary approach. University of Chicago Press; Chicago: 1967. p. xiip. 531

Lee K, Byatt G, Rhodes G. Caricature effects, distinctiveness, and identification: testing the face-space framework. Psychol Sci 2000;11:379–85. [PubMed: 11228908]

Leonard CM, Rolls ET, Wilson FA, Baylis GC. Neurons in the amygdala of the monkey with responses selective for faces. Behav Brain Res 1985;15:159–76. [PubMed: 3994832]

Leopold DA, Bondar IV, Giese MA. Norm-based face encoding by single neurons in the monkey inferotemporal cortex. Nature 2006;442:572–5. [PubMed: 16862123]

Leopold DA, O'Toole AJ, Vetter T, Blanz V. Prototype-referenced shape encoding revealed by high-level aftereffects. Nat Neurosci 2001;4:89–94. [PubMed: 11135650]

Liu J, Harris A, Kanwisher N. Stages of processing in face perception: an MEG study. Nat Neurosci 2002;5:910–6. [PubMed: 12195430]

McCarthy G, Luby M, Gore J, Goldman-Rakic P. Infrequent events transiently activate human prefrontal and parietal cortex as measured by functional MRI. Journal of Neurophysiology 1997;77:1630–4. [PubMed: 9084626]

Moghaddam B, Jebara T, Pentland A. Bayesian face recognition. Pattern Recognition 2000;33:1771–82.

Moscovitch M, Winocur G, Behrmann M. What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. Journal of Cognitive Neuroscience 1997;9:555–604.

Mountcastle VB. The columnar organization of the neocortex. Brain 1997;120(Pt 4):701–22. [PubMed: 9153131]

O'Toole A, Abdi H, Deffenbacher K, Valentin D. Low dimensional representation of faces in high dimensions of the space. Journal of the Optical Society of America A 1993;10:405–10.

Oram MW, Perrett DI. Time course of neural responses discriminating different views of the face and head. J Neurophysiol 1992;68:70–84. [PubMed: 1517829]

Pasupathy A, Connor CE. Shape representation in area V4: position-specific tuning for boundary conformation. J Neurophysiol 2001;86:2505–19. [PubMed: 11698538]

Perrett DI, Harries MH, Bevan R, Thomas S, Benson PJ, et al. Frameworks of analysis for the neural representation of animate objects and actions. J Exp Biol 1989;146:87–113. [PubMed: 2689570]

Perrett DI, Oram MW, Harries MH, Bevan R, Hietanen JK, et al. Viewer-centred and object-centred coding of heads in the macaque temporal cortex. Exp Brain Res 1991;86:159–73. [PubMed: 1756786]

Perrett DI, Rolls ET, Caan W. Visual neurones responsive to faces in the monkey temporal cortex. Exp Brain Res 1982;47:329–42. [PubMed: 7128705]

Perrett DI, Smith PA, Potter DD, Mistlin AJ, Head AS, et al. Neurones responsive to faces in the temporal cortex: studies of functional organization, sensitivity to identity and relation to perception. Hum Neurobiol 1984;3:197–208. [PubMed: 6526706]

Perrett DI, Smith PA, Potter DD, Mistlin AJ, Head AS, et al. Visual cells in the temporal cortex sensitive to face view and gaze direction. Proc R Soc Lond B Biol Sci 1985;223:293–317. [PubMed: 2858100]

Pinsk MA, Desimone K, Moore T, Gross CG, Kastner S. Representations of faces and body parts in macaque temporal cortex: A functional MRI study. Proc Natl Acad Sci U S A 2005;102:6996–7001. [PubMed: 15860578]

IPR2023-01014
Nielsen Company EX2012 Page 16

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Pourtois G, Schwartz S, Seghier ML, Lazeyras F, Vuilleumier P. View-independent coding of face identity in frontal and temporal cortices is modulated by familarity: an event-related fMRI study. Neuroimage 2005;24:1214–24. [PubMed: 15670699]

Puce A, Allison T, Asgari M, Gore JC, McCarthy G. Differential sensitivity of human visual cortex to faces, letterstrings, and textures: A functional magnetic resonance imaging study. Journal of Neuroscience 1996;16:5205–15. [PubMed: 8756449]

Quiroga RQ, Reddy L, Kreiman G, Koch C, Fried I. Invariant visual representation by single neurons in the human brain. Nature 2005;435:1102–7. [PubMed: 15973409]

Rhodes G, Brennan S, Carey S. Identification and ratings of caricatures: Implications for mental representations of faces. Cognitive Psychology 1987;19:473–97. [PubMed: 3677584]

Rhodes G, Jeffery L. Adaptive norm-based coding of facial identity. Vision Res 2006;46:2977–87. [PubMed: 16647736]

Rhodes G, Jeffery L, Watson TL, Jaquet E, Winkler C, Clifford CW. Orientation-contingent face aftereffects and implications for face-coding mechanisms. Curr Biol 2004;14:2119–23. [PubMed: 15589154]

Rhodes G, McLean IG. Distinctiveness and expertise effects with homogeneous stimuli: Towards a model of configural coding. Perception 1990;19:773–94. [PubMed: 2130375]

Riesenhuber M, Poggio T. Models of object recognition. Nat Neurosci 2000;3:1199–204. [PubMed: 11127838]

Rolls ET, Baylis GC. Size and contrast have only small effects on the responses to faces of neurons in the cortex of the superior temporal sulcus of the monkey. Exp Brain Res 1986;65:38–48. [PubMed: 3803509]

Rotshtein P, Henson RN, Treves A, Driver J, Dolan RJ. Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. Nat Neurosci 2005;8:107–13. [PubMed: 15592463]

Scalaidhe SP, Wilson FA, Goldman-Rakic PS. Face-selective neurons during passive viewing and working memory performance of rhesus monkeys: evidence for intrinsic specialization of neuronal coding. Cereb Cortex 1999;9:459–75. [PubMed: 10450891]

Sergent J, Ohta S, MacDonald B. Functional neuroanatomy of face and object processing. A Positron emission tomography study. Brain 1992;Part 1:15–36. [PubMed: 1559150]

Shakhnarovich, G.; Moghaddam, B. Face recognition in subspaces. In: Li, SZ.; Jain, AK., editors. Handbook of Face Recognition. Springer-Verlag; 2004.

Simion F, Valenza E, Umilta C, Dalla Barba B. Preferential orienting to faces in newborns: a temporal-nasal asymmetry. J Exp Psychol Hum Percept Perform 1998;24:1399–405. [PubMed: 9778830]

Sinha, P. Lecture Notes in Computer Science. Springer-Verlag; 2002a. Qualitative representations for recognition; p. 249-62.

Sinha P. Recognizing complex patterns. Nat Neurosci 2002b;5:1093–7. [PubMed: 12403994]

Sinha, P.; Balas, BJ.; Ostrovsky, Y.; Russell, R. Face Recognition by Humans: Nineteen Results All Computer Vision Researchers Should Know About. Presented at Proceedings of the IEEE; 2006.

Sinha P, Poggio T. I think I know that face. Nature 1996;384:404. [PubMed: 8945457]

Somogyi P, Tamas G, Lujan R, Buhl EH. Salient features of synaptic organisation in the cerebral cortex. Brain Res Brain Res Rev 1998;26:113–35. [PubMed: 9651498]

Spiridon M, Kanwisher N. How distributed is visual category information in human occipito-temporal cortex? An fMRI study. Neuron 2002;35:1157–65. [PubMed: 12354404]

Sugase Y, Yamane S, Ueno S, Kawano K. Global and fine information coded by single neurons in the temporal visual cortex. Nature 1999;400:869–73. [PubMed: 10476965]

Suzuki S, Cavanagh P. A shape-contrast effect for briefly presented stimuli. J Exp Psychol Hum Percept Perform 1998;24:1315–41. [PubMed: 9778826]

Tamura H, Tanaka K. Visual response properties of cells in the ventral and dorsal parts of the macaque inferotemporal cortex. Cereb Cortex 2001;11:384–99. [PubMed: 11313291]

Tanaka JW, Farah MJ. Parts and wholes in face recognition. Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology 1993;46A:225–45.

IPR2023-01014
Nielsen Company EX2012 Page 17

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Tanaka K, Saito H, Fukada Y, Moriya M. Coding visual images of objects in the inferotemporal cortex of the macaque monkey. J Neurophysiol 1991;66:170–89. [PubMed: 1919665]

Tarr MJ, Gauthier I. FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. Nat Neurosci 2000;3:764–9. [PubMed: 10903568]

Thompson P. Margaret Thatcher: a new illusion. Perception 1980;9:483–4. [PubMed: 6999452]

Tong F, Nakayama K, Moscovitch M, Weinrib O, Kanwisher N. Response properties of the human fusiform face area. Cognitive Neuropsychology 2000;17:257–79.

Tovee MJ, Rolls ET, Azzopardi P. Translation invariance in the responses to faces of single neurons in the temporal visual cortical areas of the alert macaque. J Neurophysiol 1994;72:1049–60. [PubMed: 7807195]

Tovee MJ, Rolls ET, Treves A, Bellis RP. Information encoding and the responses of single neurons in the primate temporal visual cortex. J Neurophysiol 1993;70:640–54. [PubMed: 8410164]

Tsao DY, Freiwald WA, Knutsen TA, Mandeville JB, Tootell RB. Faces and objects in macaque cerebral cortex. Nat Neurosci 2003;6:989–95. [PubMed: 12925854]

Tsao DY, Freiwald WA, Tootell RB, Livingstone MS. A cortical region consisting entirely of face-selective cells. Science 2006;311:670–4. [PubMed: 16456083]

Turk M, Pentland A. Eigenfaces for recognition. J. Cognitive Neurosci 1991;3:71–86.

Ullman, s. Object recognition and segmentation by a fragment-based hierarchy. Trends Cogn Sci 2007;11:58–64. [PubMed: 17188555]

Ullman S, Vidal-Naquet M, Sali E. Visual features of intermediate complexity and their use in classification. Nat Neurosci 2002;5:682–7. [PubMed: 12055634]

Viola P, Jones M. Robust Real-Time Face Detection. Intl. J. Computer Vision 2004;57:137–54.

Wang G, Tanaka K, Tanifuji M. Optical imaging of functional organization in the monkey inferotemporal cortex. Science 1996;272:1665–8. [PubMed: 8658144]

Wang Y, Fujita I, Murayama Y. Neuronal mechanisms of selectivity for object features revealed by blocking inhibition in inferotemporal cortex. Nat Neurosci 2000;3:807–13. [PubMed: 10903574]

Webster MA, MacLin OH. Figural aftereffects in the perception of faces. Psychon Bull Rev 1999;6:647–53. [PubMed: 10682208]

Wolberg, G. Recent Advances in Image Morphing. Presented at Computer Graphics International; Pohang, Korea. 1996.

Yin R. Looking at upside-down faces. J. Exp. Psychol 1969;81:141–5.

Young AW, Hellawell D, Hay DC. Configurational information in face perception. Perception 1987;16:747–59. [PubMed: 3454432]

Yue X, Tjan BS, Biederman I. What makes faces special? Vision Res. 2006doi:10.1016/j.visres.2006.06.017

Zhao W, Chellappa R, Phillips PJ, Rosenfeld A. Face recognition: a literature survey. ACM computing Surveys 2003;35:399–458.

IPR2023-01014
Nielsen Company EX2012 Page 18

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

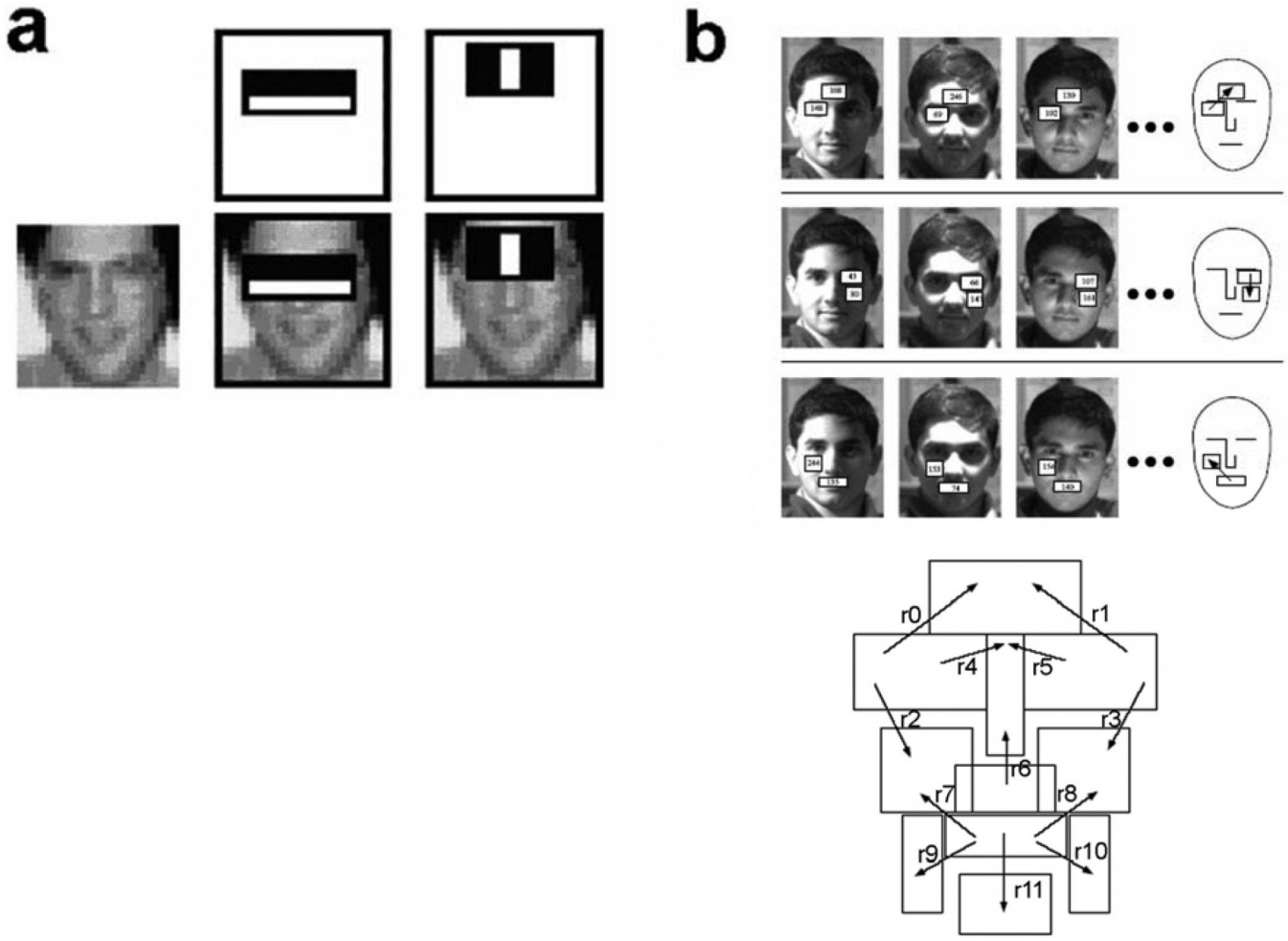

Figure 1.
(a) The two most diagnostic features defining a face comprise the first level of the detection cascade in the Viola-Jones algorithm for face detection. From (Viola & Jones 2004). (b) The Sinha algorithm for face detection, showing the "ratio-templates" defining a face. From (Sinha 2002a).

IPR2023-01014
Nielsen Company EX2012 Page 19

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

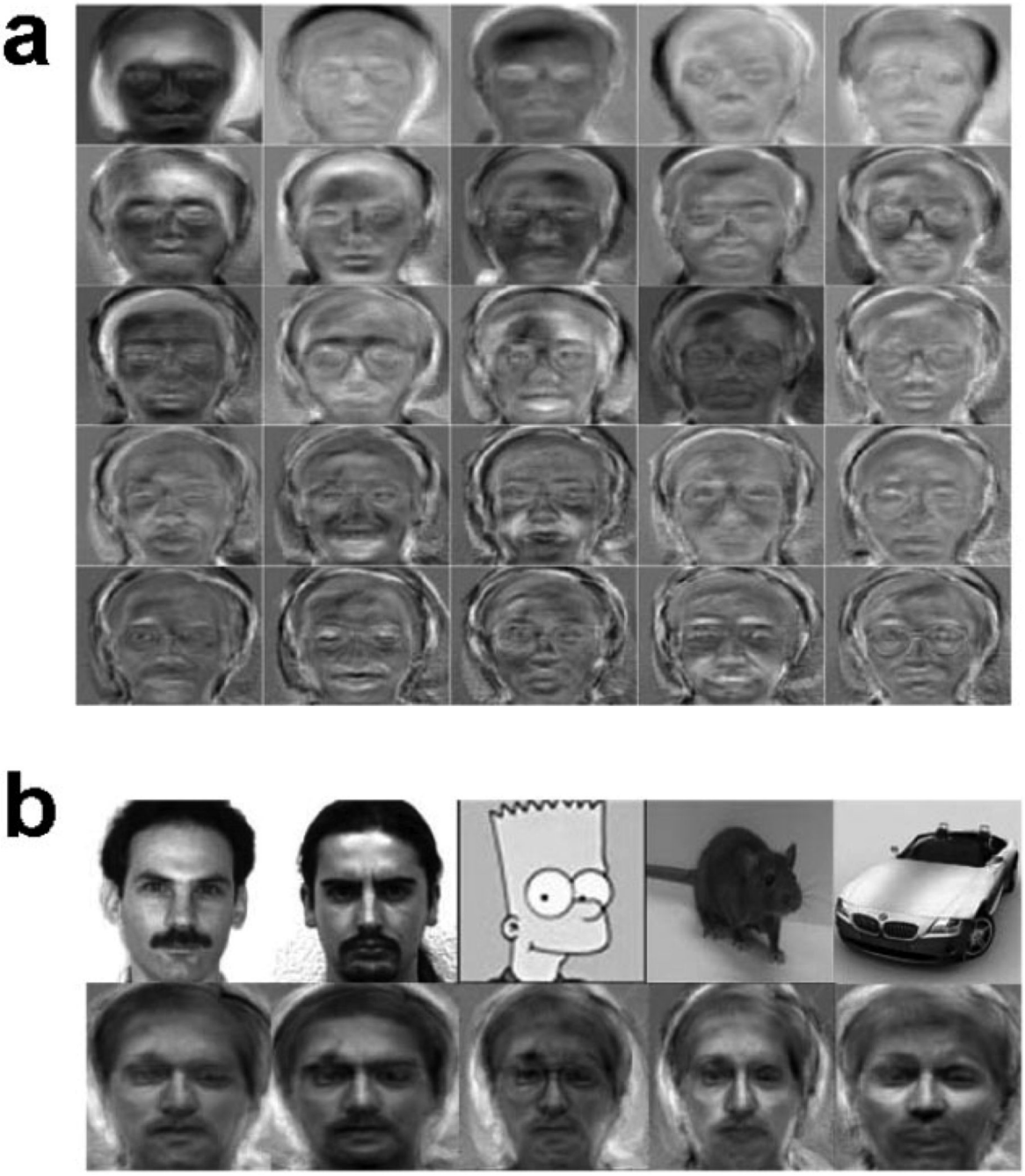

Figure 2.
The Eigenface algorithm for face recognition. (a) The first 25 eigenvectors computed from the Yale face database (a collection of 165 face images). (b) Eigenface reconstructions of 5 different images, using the 25 Eigenfaces shown in (a). Note that nonface images can have nontrivial projections onto Eigenfaces. Courtesy of Christopher DeCoro.

IPR2023-01014
Nielsen Company EX2012 Page 20

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

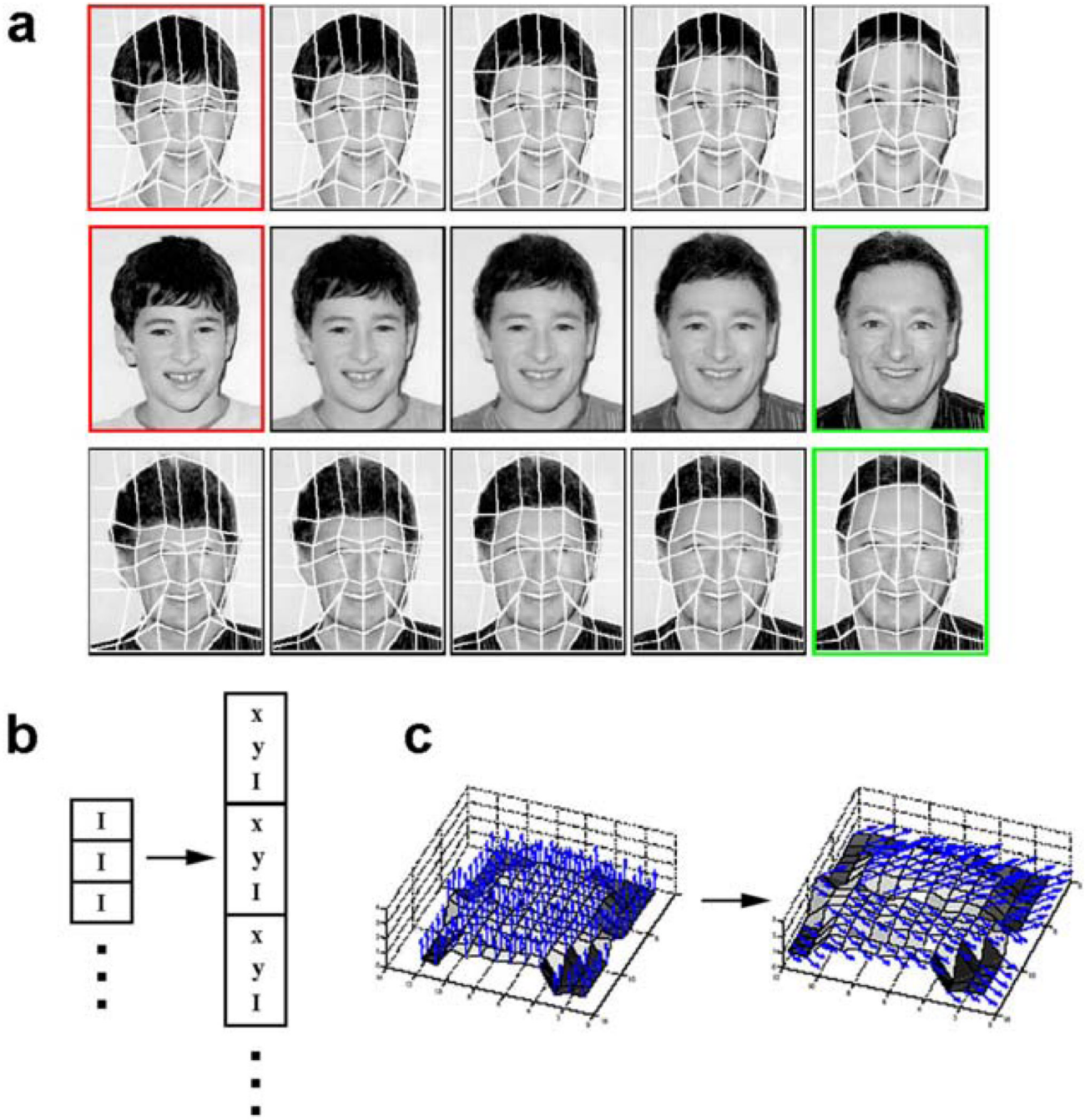

Figure 3.

A computational approach that can represent both spatial and intensity variations. (a) The computer graphics technique of morphing, in which the identity of one individual can be continuously transformed into that of another, provides insight about the nature of the face template. In the middle row, the individual outlined in red is continuously morphed into the individual outlined in green, which requires both a geometric transform and an intensity transform. The top and bottom rows show pure geometric transforms (morphing of the mesh) of the same 2 faces (the top rows shows the geometric distortion of the red face into the shape of the green face, while the bottom row shows the distortion of the green face into the shape of the red face). The middle row shows a weighted intensity average of the aligned meshes

IPR2023-01014
Nielsen Company EX2012 Page 21

from the top and bottom rows. From (Wolberg 1996). (b) “Bags of Pixels” variant on the Eigenface algorithm. The (x,y) coordinate of each pixel is elevated to the same status as the intensity value. (c) Adding or subtracting traditional Eigenfaces to an average face produces only intensity variations at each pixel. Adding or subtracting Eigenfaces computed using Bags of Pixels, however, can produce geometric variations in addition to intensity variations. From (Jebara 2003).

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

IPR2023-01014
Nielsen Company EX2012 Page 22

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

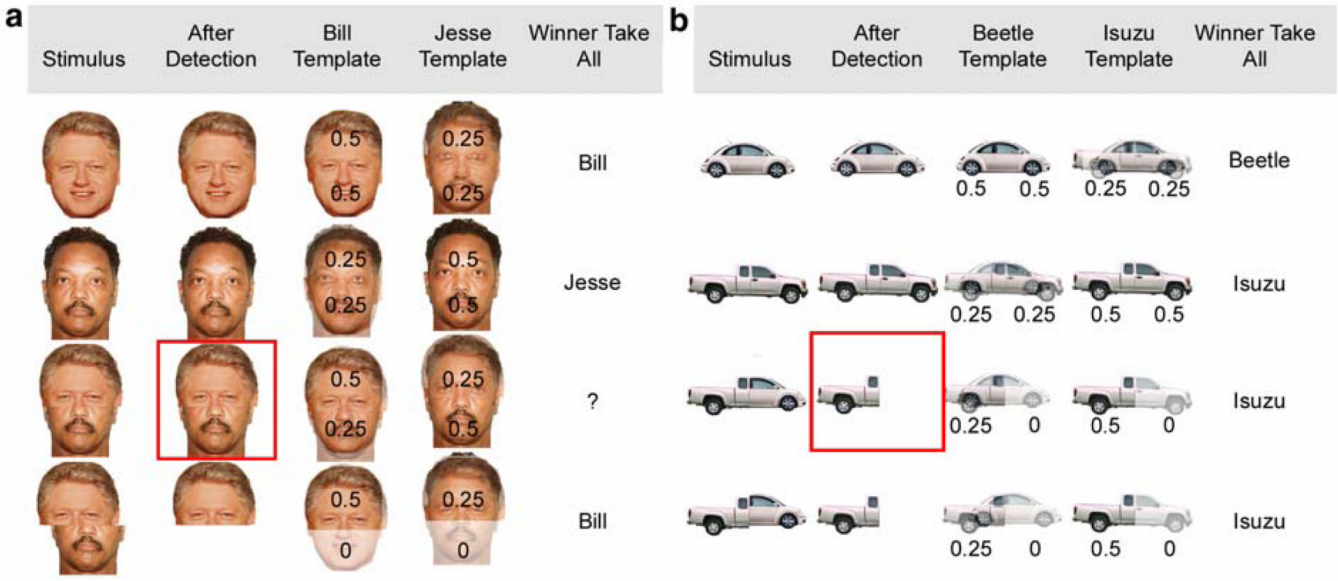

Figure 4.

Holistic (composite) effects of face processing can be explained by a detection stage that obligatorily segments faces as a whole. Subjects are asked to identify the top (faces) or left (car) part of each chimera (third and fourth rows), or to simply identify the object (first and second rows). Four face (a) and car (b) stimuli are detected, projected onto holistic templates, and then identified through a winner-take-all mechanism. The numbers in the third and fourth columns indicate the result of projecting each stimulus, after detection, onto the respective templates. Aligned faces are obligatorily detected as a whole, but misaligned faces and cars are not, and therefore their attended parts can be processed independently. According to this hypothesis, the essential difference between face (a) and non-face (b) processing occurs at the detection stage (red boxes). Subsequent measurement and classification could use similar mechanisms.

IPR2023-01014
Nielsen Company EX2012 Page 23

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

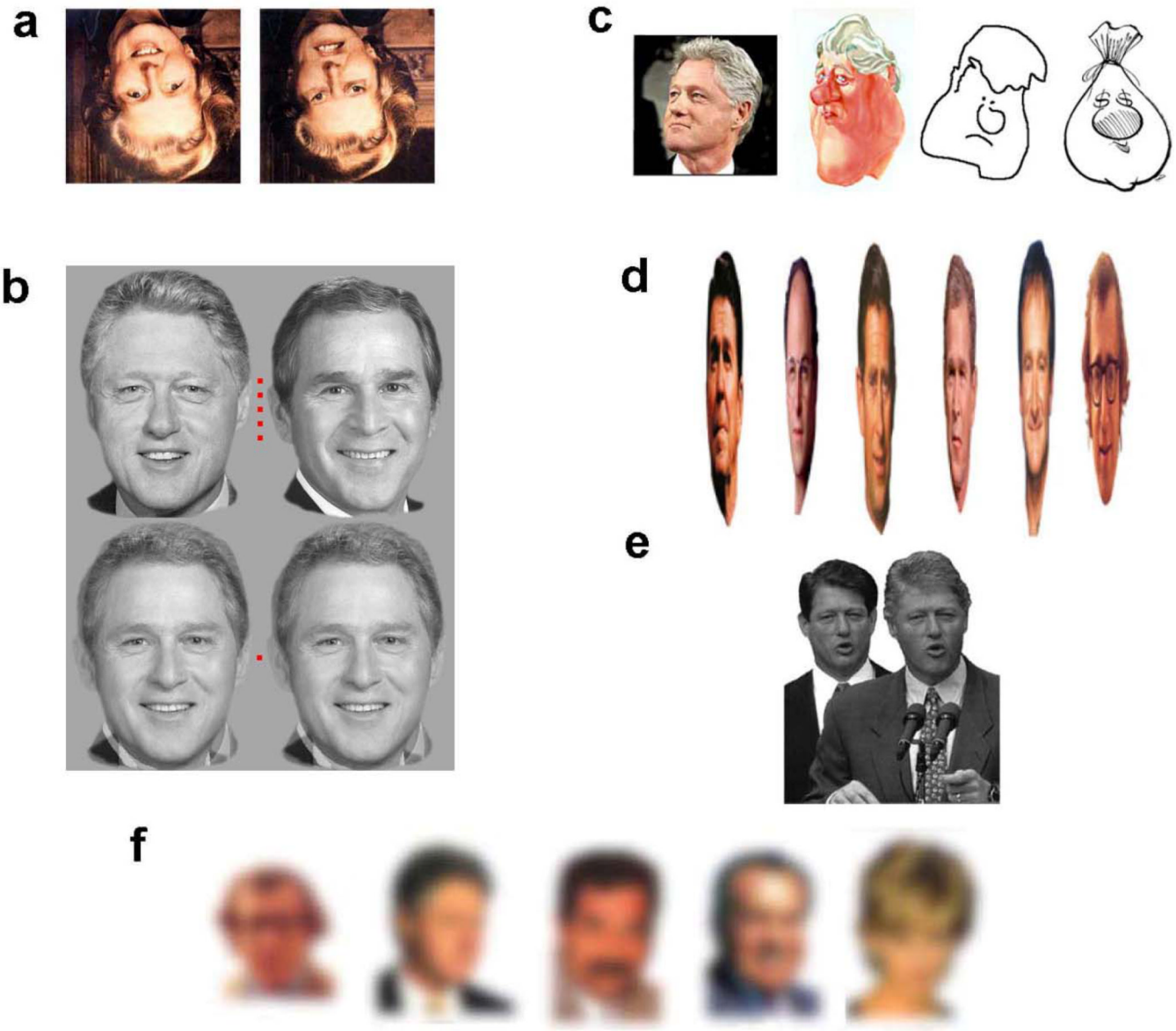

Figure 5.
Behavioral experiments probing the nature of human face processing. (a) Flip the page upside down. The Thatcher Illusion shows that faces are obligatorily processed as wholes (an identical pair of features such as the upright and inverted mouth can appear similar or dramatically different depending on the surrounding context). From (Thompson 1980). (b) Adaptation: run your eyes along the 5 red dots for a minute, and then shift your gaze to the single red dot. Courtesy of Seyed-Reza Afraz. (c) Robustness of face identification to caricature. (d) Robustness to compression. From (Sinha & Poggio 1996). (e) The importance of external features. From (Sinha et al 2006). (f) Robustness to low resolution. From (Sinha 2002b).

IPR2023-01014
Nielsen Company EX2012 Page 24

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

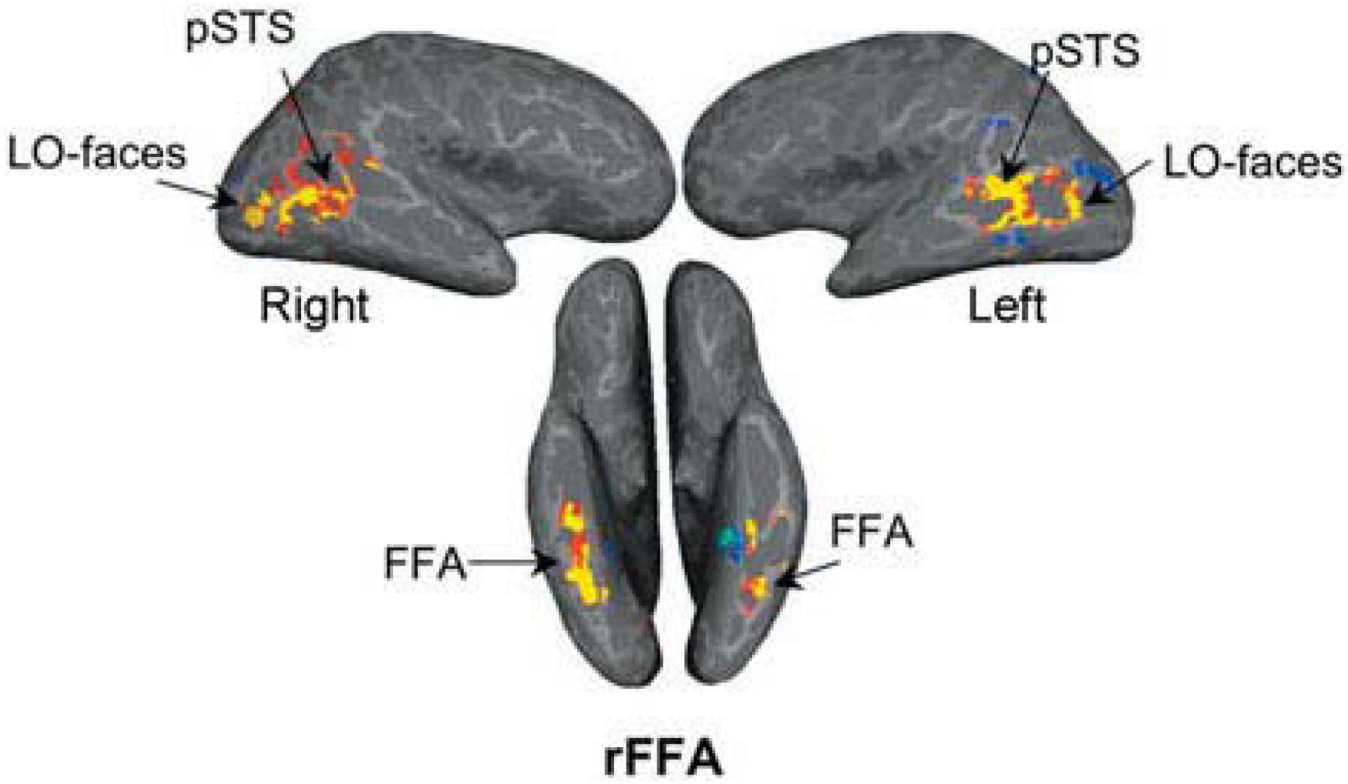

Figure 6.

Face-selective regions in one representative subject. Face-selective regions (yellow) were defined as regions that respond more strongly to faces than houses, cars and novel objects ($P < 10^{-4}$) From (Grill-Spector 2003).

IPR2023-01014
Nielsen Company EX2012 Page 25

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

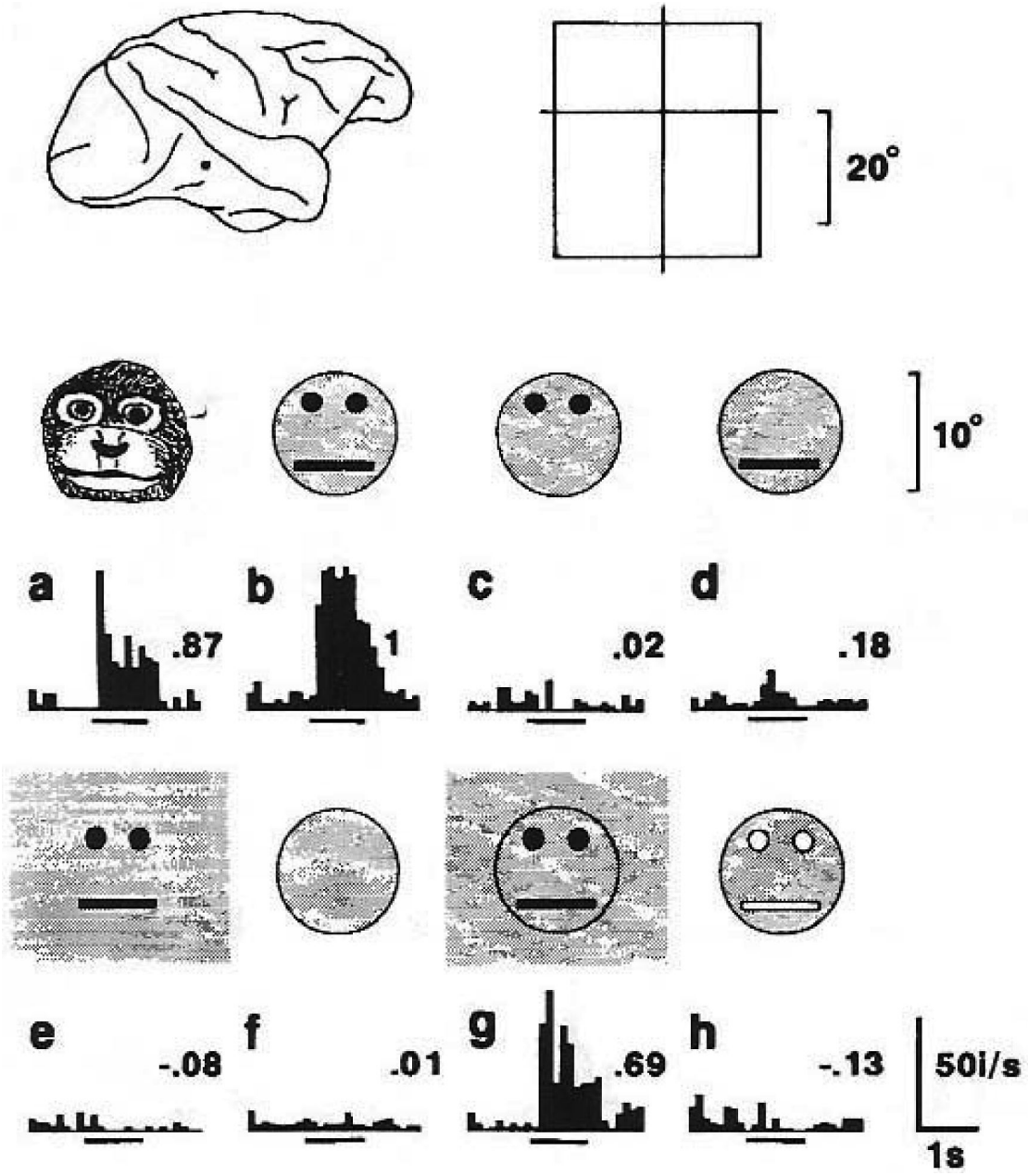

Figure 7.
Holistic face detection. (top) recording location and receptive-field location of a face cell. (a-h) Response selectivity. From Kobatake and Tanaka (1994).

IPR2023-01014
Nielsen Company EX2012 Page 26

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

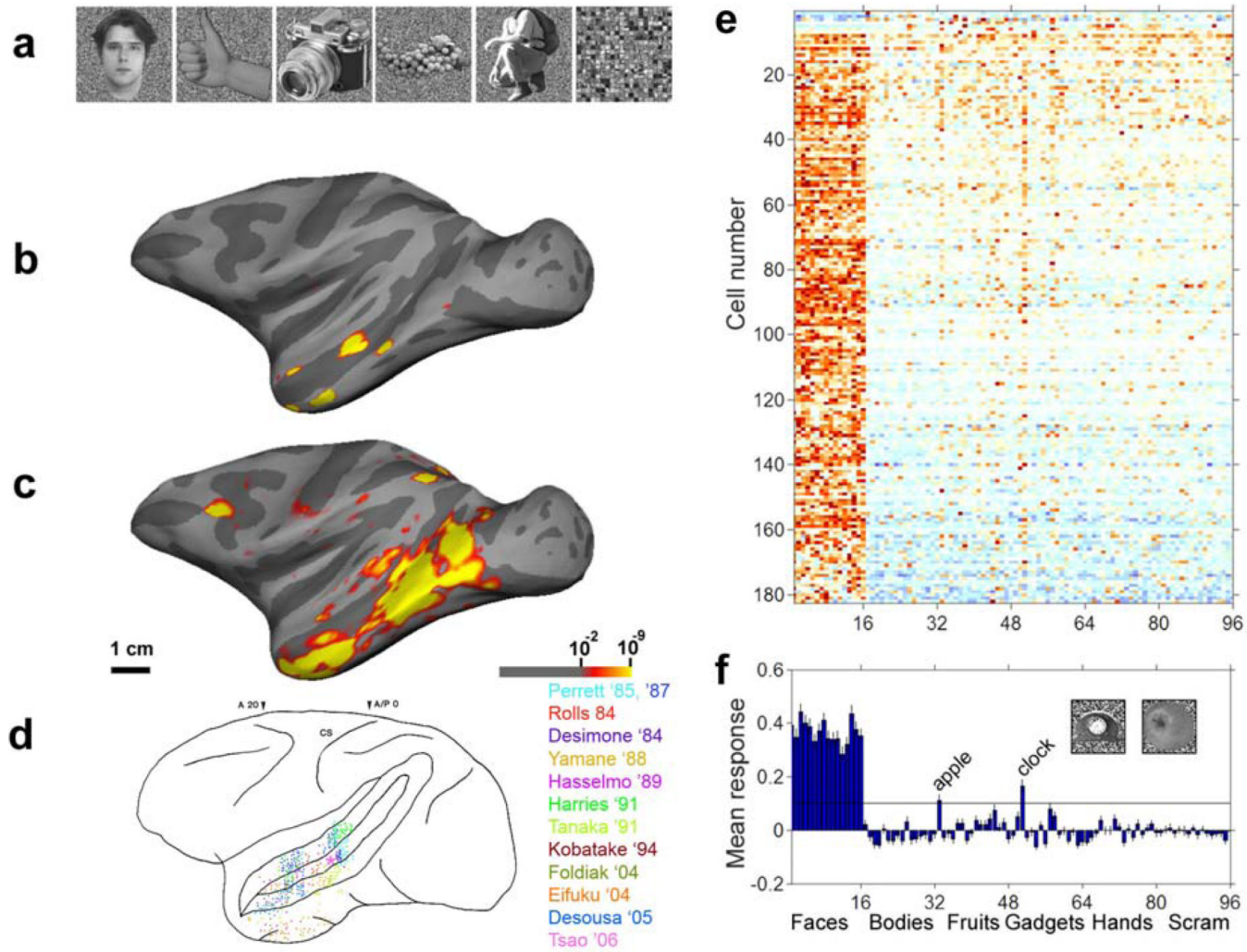

Figure 8.

Mapping face and object selectivity in the monkey brain. (a) Five stimulus categories included faces, four non-face object categories (hands, gadgets, fruits, and bodies), and grid scrambled patterns. (b) Map of faces > objects. (c) Map of objects > scrambled. (d) Meta-analysis showing the location of physiologically identified face-selective cells; studies identified by first author and date. 500 face-selective cells were recorded by Tsao et al., 2006 at the location indicated by the pink asterisk. (e) Responses of 182 neurons from M1's middle face patch to 96 images of faces and non-face objects. (f) Average normalized population response to each image. (a,b, c, e &f) From Tsao et al., 2006.

IPR2023-01014
Nielsen Company EX2012 Page 27

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

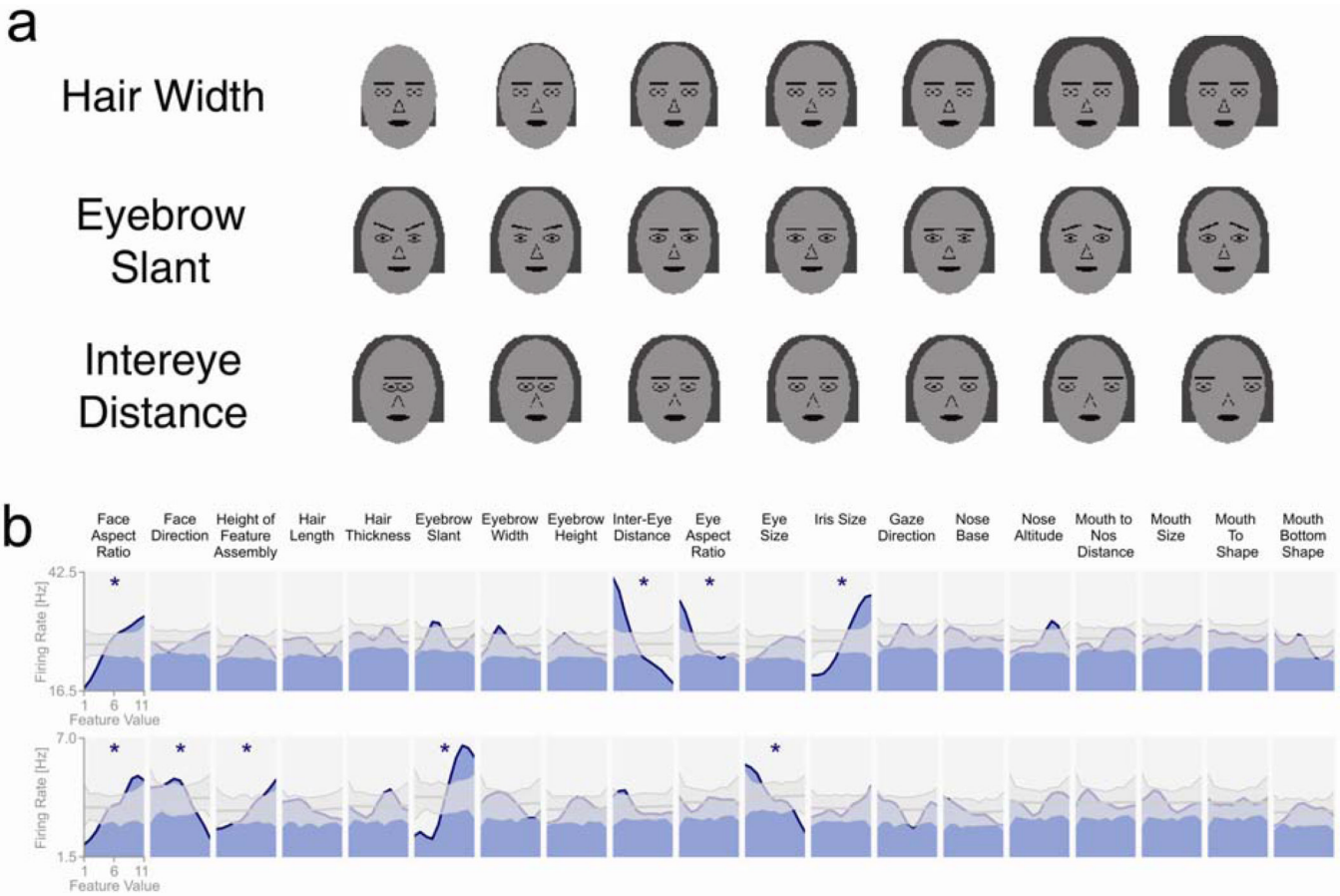

Figure 9.

Tuning of face cells to a cartoon face space. (A) Three example dimensions of the 19-dimensional cartoon space. Each row shows example values for one parameter, with all other parameters fixed at their mean. (B) Tuning curves of two example cells to each of the 19 feature dimensions. In grey maximal, minimal and mean values from shift predictor are shown. Stars mark significant modulation. (From Freiwald et al., 2007)

IPR2023-01014
Nielsen Company EX2012 Page 28