

What's so special about the average face?

Doris Y. Tsao^{1,2} and Winrich A. Freiwald^{1,2}

¹ Brain Research Institute, Center for Advanced Imaging, University of Bremen D28334, Bremen, Germany

² Department of Neurobiology, Harvard Medical School, Boston, MA 02115, USA

The past few years have seen a marked increase in our understanding of how faces are represented in the brain, with the discovery of new anatomical structures and new algorithms for representing faces. Still, the basic computational mechanism used by the primate visual system to identify faces remains a topic of intense debate. Are faces represented by matching to a set of stored exemplars, or by measuring the distance from a standard prototype along a set of different axes? A recent article by Rhodes and Jeffery provides compelling psychophysical evidence in favor of the latter 'axis' model.

Primates can distinguish thousands of different individuals of their own species effortlessly, despite changes in illumination, size and viewpoint, or occlusion – far better than any machine. Functional MRI and neurophysiology show that faces are represented in several specialized face-selective areas of the temporal lobe [1,2], at least one of which consists almost exclusively of face-selective neurons [3]. Ultimately, we want to understand how these neural representations translate into our perception of faces. In their recent article, Rhodes and Jeffery [4] use psychophysical methods to probe the neural organization of 'face space', and they find evidence for the 'axis' model of face representation.

Exemplars versus axes

The space of faces includes faces of many different identities, views, genders and ethnicities, as well as cartoon faces, and so on. How does the brain make sense of this vast space? Intuitively, one can think of two models for representing face space. In the exemplar model, face space is tiled by cells with bell-shaped tuning functions whose peaks exhaustively sample the entire space (Figure 1a). One difficulty for the exemplar model is to explain the caricature effect: it is often easier to recognize a caricature of a face (a face in which certain distinctive features are grossly exaggerated) than the original face [5]. According to the exemplar model, cells tuned to an individual face should show decreased firing to a caricature of the same face (because the caricature would fall on the edge of the bell-shaped tuning curve). This has led to an alternative model for representing face space: the axis model [6] (Figure 1b). In this model, different directions in face space represent different identities, and magnitude of firing is proportional to identity strength (thus extreme caricatures would elicit maximal responses).

Two key features distinguish the axis model from the exemplar model: (i) in the axis model, face space is not isotropic (equivalent at all locations). Instead, the *average* face is special in that all faces are represented by their deviation from it. (ii) Because in the axis model different cells encode different *trajectories* in face space (rather than specific faces), each face is paired with an 'anti-face', a face that lies on the same identity trajectory but on the opposite side of the average (e.g. if 'Jim' has larger eyes than average, then 'anti-Jim' would have smaller eyes than average). Mutually opponent cell populations encode the face and its anti-face.

Evidence for axes

A face adaptation illusion

Several years ago, Webster and MacLin discovered a new illusion: after adaptation to a distorted face (either compressed or expanded), a normal face appears distorted in the opposite direction [7]. Leopold *et al.* [8] asked whether such aftereffects could occur even to natural configural changes responsible for distinguishing different individuals in the real world, and if so, whether the aftereffects are different for different identities. In other words, can adaptation be used to probe the neural representation of face identity? Their results showed that identity-based aftereffects do occur: after adaptation to an anti-face, an average face is identified as the face opposite to the anti-face (the 'target' face) with above chance probability. This is surprising because appearance-wise, a face/anti-face pair looks simply like two different faces – yet the illusion shows that these two particular faces are intimately linked in the brain.

What the illusion specifically suggests is that face/anti-face pairs are represented by opponent poles of a single axis (Figure 1b,c). By analogy with color, the average face represents the white point, and different face identities represent different colors (Figure 2a). In color vision, when one adapts to a red spot, then a subsequently presented white spot will look cyan, because red and cyan mutually inhibit each other (and white is defined by a balance between the two).

Does face adaptation affect an axis or a region of face space?

Careful examination shows that the face adaptation illusion could also be accommodated by the exemplar model: suppose that the percept of a particular identity is determined by a weighted average of *all* exemplars, and adaptation to a face causes the percept of all nearby face stimuli to be biased away from the adapting face. If this were the

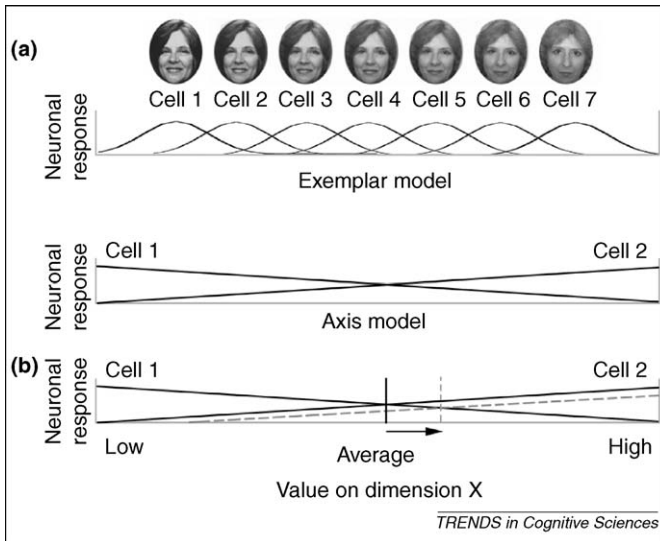


Figure 1. Two opposing models of face representation. (a) In the exemplar model (top), each different face is represented by a different cell. In the axis model (bottom), face space is organized parametrically into opponent axes, such that different faces along an axis can be coded by the relative firing of just two cells. (b) According to the axis model, adaptation to an extreme face should cause the average face to appear shifted in the opposite direction.

case, then following adaptation to an anti-face, the percept of the average face would also shift away from the anti-face towards the target face. The critical difference between the two models, then, is that according to the axis model, repulsion should not act isotropically in all directions surrounding a particular face, but rather, should be *specific to a particular axis*, which always passes through the center of face space.

Leopold and colleagues tried to demonstrate this last property by showing that adaptation to a non-opposite anti-face had no effect on target discrimination. However, in their set of face stimuli, it seems that the distance from each face to its matching anti-face was greater than the distance to the three non-matching anti-faces (for example, in Figure 2c, Face 1 looks a bit like anti-Face 2). Adaptation effects are known to be sensitive to the precise distance between adapting and test stimuli. A control experiment demonstrating that adaptation to an *average* face did not change target identification thresholds goes further towards dispelling the possibility of generalized repulsion [7,8].

Rhodes and Jeffery have now removed this possibility by making a clever addition: they tested face triplets

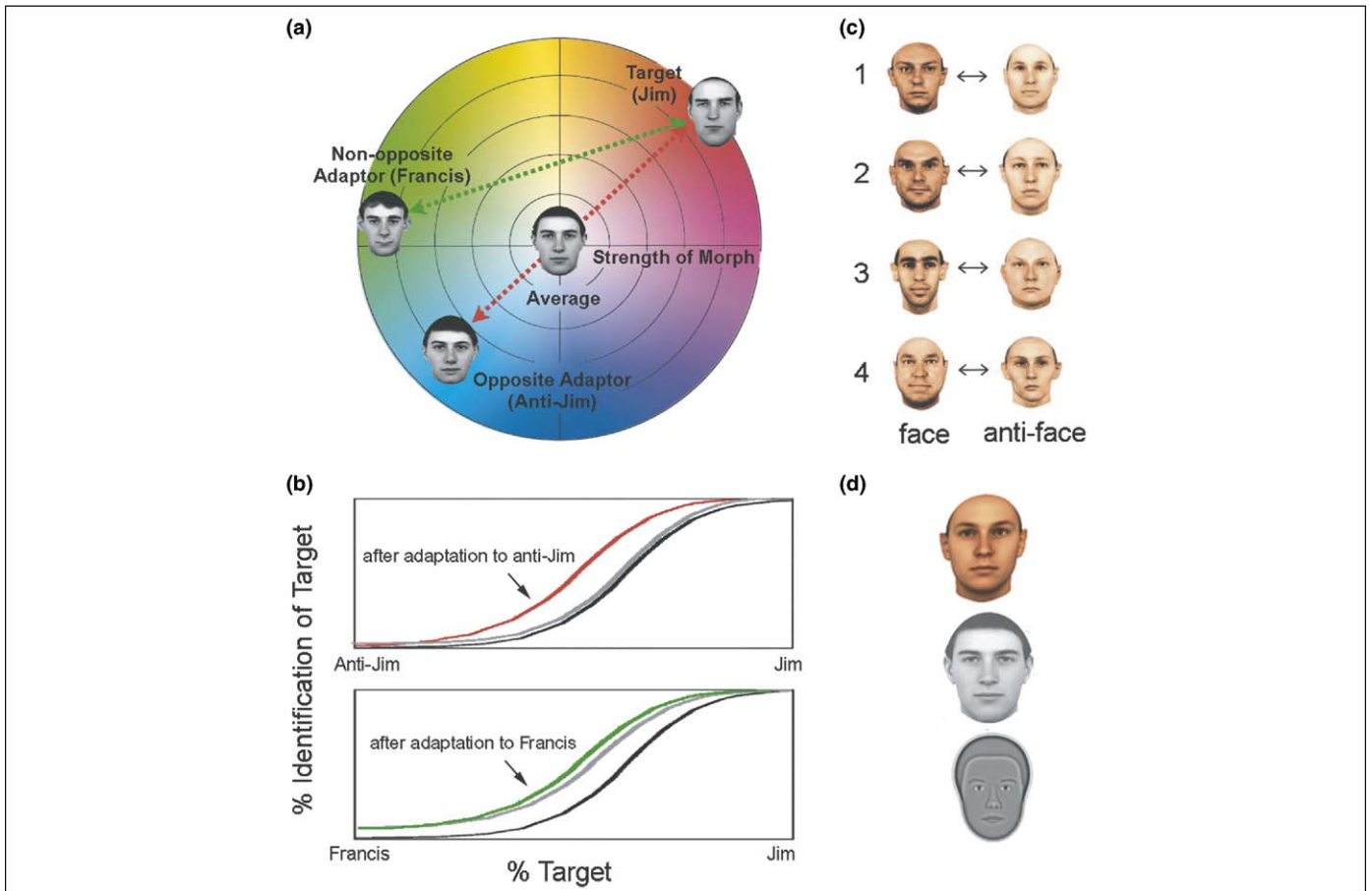


Figure 2. Adaptation to a face biases face perception towards the opposite direction in face space. (a) In the experiment by Rhodes and Jeffery [4], non-opposite adapting faces were matched in perceptual similarity to opposite adapting faces, and target identification thresholds were measured along morph trajectories connecting adaptor and target for both types of adaptor. (Adapted with permission from [4]). (b) A schematic of the identification curves obtained for opposite (top) and non-opposite (bottom) adaptation. Black curves: baseline before adaptation trials. Gray curves: baseline after adaptation trials. (c) Faces and anti-faces used in the experiment by Leopold *et al.* [8] The non-opposite adapting faces were more similar to the test faces than were the opposite adapting faces, creating uncertainty as to the specificity of the face adaptation illusion elicited by these anti-faces. (d) The average face in the different experiments (top [8], middle [4], bottom [12]). (Adapted with permission from [4,8,12]).

Box 1. Questions for future research

- What specific face variables do different axes in the axis model represent? There are multiple ways to decompose a face, e.g. along a set of feature dimensions or along eigenspace dimensions computed through PCA analysis of face space [13]. Recent electrophysiological experiments in the macaque middle face area suggest face cells in this region are tuned to simple feature dimensions [14].
- How does the average face change with learning and task requirements? The 'average face' used in three studies of the face adaptation illusion [4,8,12] look quite different from each other (see Figure 2d in main text), yet all three studies argue for a special role for their average face.
- Are faces special or can the geometry of other objects be measured using similar mechanisms? The prevalence of tuning to extreme values of curvature [15] in IT cells suggests that axis-based representation may be a general feature of IT coding, allowing efficient discrimination of shapes sharing the same basic parts and configuration. There could be even more general computational reasons for using axis-based coding, potentially linking a wide variety of disparate phenomena including color vision, face perception, and even social cognition (e.g. the formation of social stereotypes).

consisting of a face/anti-face pair (e.g. Jim/anti-Jim), plus a third face (e.g. Francis) whose perceptual distance to Jim was equal to that between anti-Jim and Jim. Subjects were tested for their ability to recognize the target face (Jim) along Jim/anti-Jim morph trajectories, and along Jim/Francis morph trajectories (Figure 2a).

Following adaptation to anti-Jim, the threshold for recognizing Jim was significantly lowered compared with the baseline threshold (measured both before and after adaptation trials) (Figure 2b). On the other hand, following adaptation to Francis, the threshold for recognizing Jim was only slightly lower than the baseline threshold (measured after adaptation trials). This supports the axis model: adaptation is specific to morph trajectories that pass through the average face.

Surprisingly, Rhodes and Jeffery found a second kind of plasticity in face perception: the baseline threshold measured after the Francis adaptation trials was much lower than the baseline threshold measured before Francis adaptation trials (Figure 2b bottom, compare gray and black curves). Apparently, during the adaptation trials, subjects somehow learned to associate Francis and Jim. Subjects never did this for anti-Jim and Jim (Figure 2b top, compare gray and black curves), again highlighting the special status of the average face. This learning effect might be tapping into basic mechanisms for rapid visual association of faces. To speculate, perhaps given a set of identities together with possible exemplars (including strong and weak versions), one learns to associate all the exemplars

in a connected region of face space as long as this region does not intersect a region occupied by another identity. Faces and anti-faces are never associated because they are connected via the average face, which by definition intersects every identity region. Such learning could be what allows us to recognize faces irrespective of image degradation and occlusion.

Future prospects

Abraham Lincoln once remarked, 'The Lord prefers common looking people. That is the reason he made so many of them.' The results of Rhodes and Jeffery make an important addition to a mounting body of evidence that the average face has a special status [8–12], serving as the norm against which incoming faces are measured. Clarifying the detailed mechanism by which this norm is represented in the brain will be an exciting topic for further research (Box 1).

References

- 1 Kanwisher, N. *et al.* (1997) The fusiform face area: A module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17, 4302–4311
- 2 Tsao, D.Y. *et al.* (2003) Faces and objects in macaque cerebral cortex. *Nat. Neurosci.* 6, 989–995
- 3 Tsao, D.Y. *et al.* (2006) A cortical region consisting entirely of face-selective cells. *Science* 311, 670–674
- 4 Rhodes, G. and Jeffery, L. Adaptive Norm-Based Coding of Facial Identity. *Vision Res.* 46, 2977–2987.
- 5 Rhodes, G. *et al.* (1987) Identification and ratings of caricatures: Implications for mental representations of faces. *Cognit. Psychol.* 19, 473–497
- 6 Valentine, T. (1991) A unified account of the effects of distinctiveness, inversion and race on face recognition. *Q. J. Exp. Psychol. A.* 43, 161–204
- 7 Webster, M. and MacLin, O. (1999) Figural aftereffects in the perception of faces. *Psychon. Bull. Rev.* 6, 647–653
- 8 Leopold, D. *et al.* (2001) Prototype-referenced shape encoding revealed by high-level aftereffects. *Nat. Neurosci.* 4, 89–94
- 9 Loffler, G. *et al.* (2005) fMRI evidence for the neural representation of faces. *Nat. Neurosci.* 8, 1386–1390
- 10 Leopold, D.A. *et al.* (2003) Prototype-referenced encoding of faces in the monkey inferotemporal cortex. *Soc. Neurosci. Abstr.* Program No. 590.7. (<http://sfn.scholarone.com/tin2003/>)
- 11 Webster, M. *et al.* (2004) Adaptation to natural face categories. *Nature* 428, 557–560
- 12 Anderson, N. and Wilson, H. (2005) The nature of synthetic face adaptation. *Vision Res.* 45, 1815–1828
- 13 Calder, A. and Young, A. (2005) Understanding the recognition of facial identity and facial expression. *Nat. Rev. Neurosci.* 6, 641–651
- 14 Freiwald, W.A. *et al.* (2005) Single-unit recording in an fMRI-identified macaque face patch. II. Coding along multiple feature axes. *Soc. Neurosci. Abstr.* Program No. 362.6 (<http://sfn.scholarone.com/>)
- 15 Kayaert, G. *et al.* (2005) Tuning for shape dimensions in macaque inferior temporal cortex. *Eur. J. Neurosci.* 22, 212–224

1364-6613/\$ – see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tics.2006.07.009

Reproduction of material from Elsevier articles

Interested in reproducing part or all of an article published by Elsevier, or one of our article figures?
If so, please contact our *Global Rights Department* with details of how and where the requested material will be used. To submit a permission request online, please visit:

www.elsevier.com/locate/permissions