

Prototype-referenced shape encoding revealed by high-level aftereffects

David A. Leopold¹, Alice J. O'Toole², Thomas Vetter^{1,3} and Volker Blanz^{1,3}

¹ Max Planck Institut für biologische Kybernetik, Spemannstraße 38, 72076 Tübingen, Germany

² School of Human Development, GR4.1, The University of Texas at Dallas, Richardson, Texas 75083-0688, USA

³ Present Address: Institut für Informatik, Am Flughafen 17, 79110 Freiburg, Germany

Correspondence should be addressed to D.A.L. (david.leopold@tuebingen.mpg.de)

We used high-level configural aftereffects induced by adaptation to realistic faces to investigate visual representations underlying complex pattern perception. We found that exposure to an individual face for a few seconds generated a significant and precise bias in the subsequent perception of face identity. In the context of a computationally derived 'face space,' adaptation specifically shifted perception along a trajectory passing through the adapting and average faces, selectively facilitating recognition of a test face lying on this trajectory and impairing recognition of other faces. The results suggest that the encoding of faces and other complex patterns draws upon contrastive neural mechanisms that reference the central tendency of the stimulus category.

Our ability to distinguish among similar visual patterns or objects is thought to rest upon finely tuned neural mechanisms in the brain. This is especially true for object classes whose members share the same basic features and differ only in subtle aspects of shape, texture and color. Recognition at this subordinate level is thought to engage specialized neural mechanisms, a notion supported by physiological findings^{1,2}. For primates, faces represent arguably the most relevant example of subordinate recognition. Neural structures apparently specialized for face analysis have been identified with electrophysiological recordings, as well as with neuroimaging, in both monkeys^{3–5} and humans^{6,7}. However, despite the abundance of brain regions that are visually responsive to faces and other complex patterns, the nature of their contribution to perception and recognition remains poorly understood.

Psychological studies of face representations have offered a multidimensional 'face space' framework for investigating the neural encoding of faces. Faces are represented as points (or vectors) in a space with the average of all faces at the center⁸. This heuristic has proved useful for unifying face recognition findings that involve distinctiveness, prototype-generation and the other-race effect^{9–12}. The study of caricatures, which are faces with exaggerated features, provides an example. In face space terminology, caricatures represent an extrapolation of the vector joining the average face to a particular individual, along what we term the face's 'identity trajectory.' By manipulating the position of a face along its identity trajectory, one can systematically vary a face's 'individuality,' which falls to zero with the average prototype in the center of face space (Fig. 1b). Studies have shown that caricatures retain the identity of the original face, but are perceived as more distinctive and recognized more accurately than veridical faces^{13–15}. Not surprisingly, faces lying between an individual face and the mean (termed 'anti-caricatures') are recognized less accurately¹⁶. And those lying on the

same axis, but on the 'other side of the mean' (so-called 'anti-faces'), have the appearance of an entirely different individual (Fig. 1a). Thus, the transition across the prototype between a face and an anti-face represents a point of discontinuity in perceived identity¹⁷.

A recent study introduced figural aftereffects as a tool to investigate face perception¹⁸. Aftereffects are perceptual distortions that arise following extended periods of visual exposure to a stimulus, and have often been used to infer mechanisms underlying basic visual processing^{19–21}. In that study, subjects were adapted for several minutes to configurally distorted faces (for example, grotesquely compressed or expanded faces), and face perception was evaluated thereafter. Following adaptation to a distorted face, normal faces appeared distorted in an opposite manner. This effect depended on the nature of the gross distortion, and was nonspecific with regard to the identities of the faces.

We considered whether adaptation could be used to tap into more refined shape representations in the brain relevant for the extraction of identity. Because discrimination and recognition ultimately rely upon networks of neurons to differentiate between highly similar patterns, might it be possible to adapt highly specific neural populations, and thereby generate identity-based aftereffects? Here we examined how adaptation to one realistic face, generated in the context of a computationally derived face space, affected the perception of another. We show that this adaptation led to misperceptions in face identity that were very precise, biasing perception along an identity trajectory away from the adapting stimulus.

RESULTS

Images were generated using a three-dimensional face-morphing algorithm developed previously²². Face/anti-face pairs (Fig. 1a) were generated in the context of a schematized face space (Fig. 1b). Within this space, the 'identity trajectory' of each

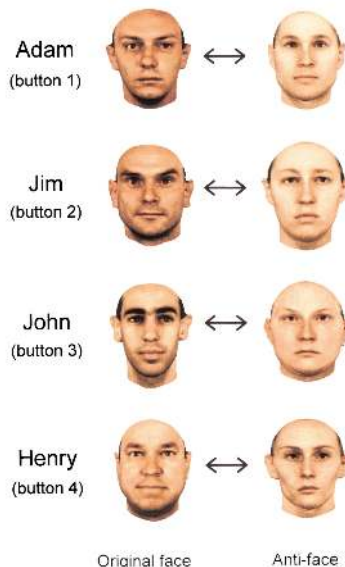
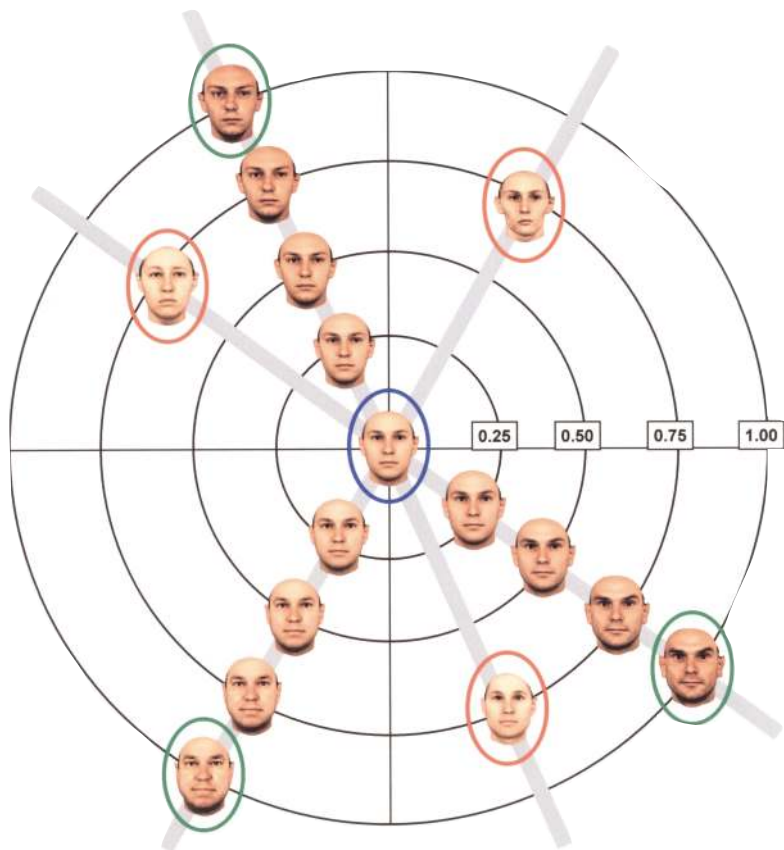
**a** "The usual suspects"**b**

Fig. 1. Examples of stimuli and the multidimensional face space. **(a)** Face/anti-face pairs. Whereas the features of faces and anti-faces are in some sense 'opposite,' the anti-faces appeared simply as entirely different individuals. **(b)** Computationally derived face space in which the stimuli were generated. The original faces (green ellipses) are connected to the average face (blue ellipse) by an 'identity trajectory.' Numbers refer to the 'identity strength' possessed by the given face. (For example, the veridical face equals 1.0, and the mean face equals 0.0.) Caricatures (not depicted) would lie further from the mean, beyond the original faces on each identity trajectory. Anti-faces on the opposite side of the mean from the original face have negative identity strengths.

face (gray lines) passed through the original face (green ellipses) as well as the average or prototype face (blue ellipse). Lying on the same axis, but on the opposite side of the average face was the anti-face (red ellipses). Morphed faces, varying in 'identity strength,' lay between the average and original faces. Anti-faces, lying on the opposite side of the mean from the original face, possessed negative values of identity strength.

We reasoned that the behavioral and biological relevance of a contrastive coding of faces defined relative to a prototype-centered face space might be probed by adaptation, with adaptation to an anti-face biasing subsequent perception toward the original face. Thus, we measured subjects' baseline identification thresholds for several faces, and then examined the impact of anti-face adaptation on their identification performance.

Baseline face identification thresholds

In measuring thresholds, we asked how much 'identity' a morphed face must contain to be correctly identified. Subjects viewed briefly (200 ms) flashed faces that varied in their identity strength, and identified each as one of four previously learned individuals (**Fig. 2a**, see Methods). Identity strength was varied from -0.10 (a slight anti-face) to 0.35 (35% of the distance from the mean to the original face), in steps of 0.05 .

Subjects identified the individuals perfectly when approximately one-third or greater of the identity information was present

(**Fig. 3**, ○). This pattern was nearly identical for all but one subject, whose performance never exceeded 90%. An ogival function fit the psychometric data exceptionally well with a coefficient of determination equal to 0.9983 (**Fig. 3b**, ○), suggesting that even increments of physical morphing correspond to equal psychological increments in face identity. The average identification threshold was 0.11 , or 11% of the distance between the average and original faces. At the 0.0 identity strength (average face), performance dropped to chance levels (0.25). For negative values of identity strength (that is, in the direction of the anti-face) performance dipped slightly below chance, reflecting the pronounced difference in appearance for faces on opposite sides of the mean¹⁷.

Adaptation to anti-faces

Identification thresholds were measured a second time, but in this case, following a five-second exposure to an anti-face (identity strength, -0.8). Two conditions were analyzed separately. First, we analyzed the matching condition, where the adapting and test faces lay on the same identity trajectory. (The identity trajectory passed through the average face.) Second, we analyzed the non-matching condition, where the adapting and test face lay on different identity trajectories. Data are pooled for all anti-face/face pairs.

Previous exposure to a matching anti-face shifted the entire psychometric function to the left (**Fig. 3**, ●), increasing sensitivity to the original face. The mean shift in threshold was 12.5% of the identity strength. This shift placed the threshold on the other side of the average face. For the average face, performance increased from chance (25%, no bias toward a particular face) to greater than 60%. (For zero strength, one-fourth of the stimuli were randomly designated beforehand as lying on each axis, allowing for a 'performance' to be calculated.) These results demonstrate that not



only does adaptation alter sensitivity, but it actually spurs the visual system to 'create' a new identity for the average face.

In the non-matching case (■), performance declined for test stimuli approaching the mean face, likely reflecting competition between the structure of the weakened test face and the distortions introduced by adaptation. This is evident when the data for non-match trials are expressed as the fraction of trials for which subjects responded according to the adapting stimulus (Fig. 4). This fraction is a measure of the strength of the adaptation to 'override' a dissimilar test stimulus. The peak of this function occurs at zero identity strength (average face). At higher fractions of test-face identity, the adaptation effects were overwhelmed by the physical structure of the test face, and the adaptation-based responses fell to zero.

The adaptation effects were not contingent on the faces being presented upright. Psychometric functions were obtained for two subjects with upright and inverted faces, with and without adaptation (Fig. 5). The orientation of the adapting stimulus always matched that of the test stimulus. Although recognition performance was affected by inversion, the influence of adaptation was not. This result, although somewhat puzzling given the profound effects of inversion on face recognition, agrees with the result of a previous study¹⁸, where configural effects also did not depend on face orientation.

To control for any general effects that the adaptation protocol might have on the discriminations, we also used the average face rather than the anti-faces as the adapting stimulus. Although there was a slight tendency for this 'nonspecific' adaptation to lower thresholds (mean shift, 0.024), the effect was minimal compared to the effect with anti-faces (mean shift, 0.11), illustrating that the adaptation protocol by itself did not significantly alter face sensitivity.

Finally, we tested the robustness of the aftereffect to a delay between adaptation and testing to verify that the effect was due to adaptation, rather than a particular spatiotemporal sequence. We introduced blanking times of 150, 300, 600, 1200 and 2400 ms and found that although the effects were somewhat diminished for intervals greater than 300 ms, adaptation continued to influence perception even after delays of 2400 ms. The survival of the aftereffect with blanking suggests that adaptation itself is responsible for the effect. This result, combined with the fleeting nature of the percept itself (see Methods), may be related to the phenomenon of 'storage', described for simple aftereffects, in which the potency of adaptation survives a period of non-stimulation, but is quickly spent when a test stimulus appears²¹.

Translation invariance

It is possible that the above results could be explained by a conjunction of localized aftereffects for orientation, spatial frequency and color. Such low-level distortions, acting together, could potentially create systematic changes in the perceived identity of a subsequently presented test face. This was unlikely, given that the subjects freely scanned the adaptation images, continually varying

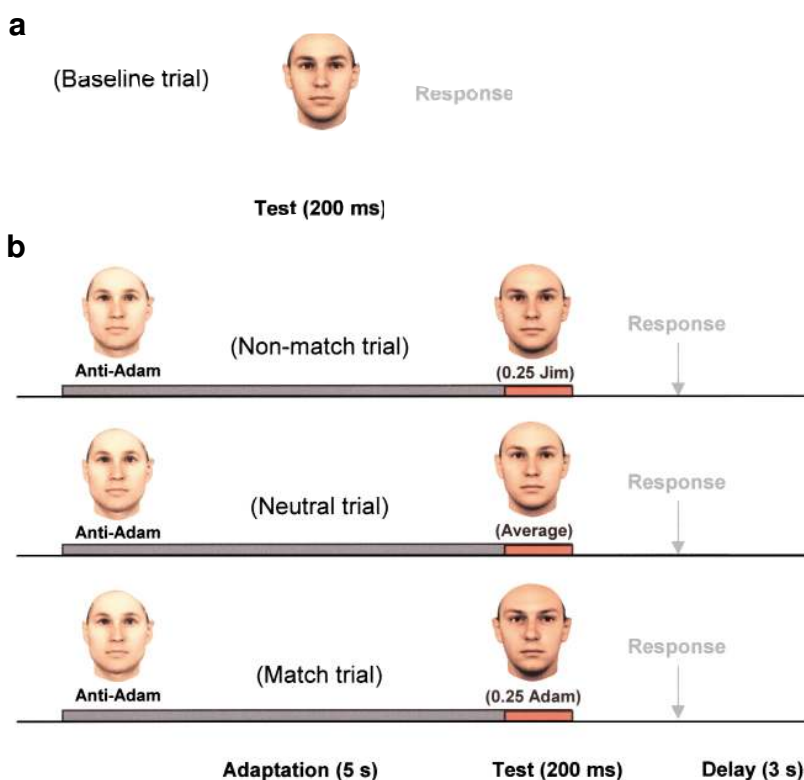


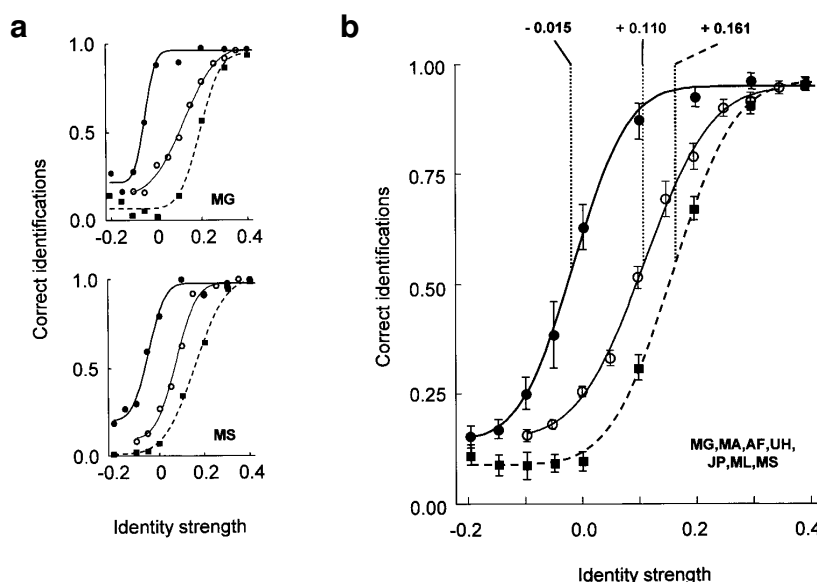
Fig. 2. Testing protocol. (a) Measurement of baseline identification thresholds. Subjects viewed a test face for 200 ms and indicated which of four memorized faces it most closely resembled. (b) Effects of adaptation. Thresholds were reassessed following a 5.0-s period of adaptation to an anti-face. In the match condition, the test face was on the same identity trajectory as the anti-face, whereas in the non-match condition it was on a different trajectory. The average face lay on all four identity axes, and was therefore neutral. Subjects pressed one of four buttons in response to the test stimulus, to signal which of the memorized faces it most closely resembled.

the retinal locations of the various facial features. Nonetheless, we decided to investigate this possibility systematically. We reasoned that if the neural mechanisms underlying the adaptation effects were involved in encoding face identity, they would most likely be invariant to the types of transformation encountered in normal object vision, such as translation and scaling. We first investigated this possibility by simply changing either the position or size of the test stimulus with respect to the adapting stimulus. This yielded some invariance, as the adaptation effects remained significant with shift or size changes of up to two degrees; however, the results were contaminated by the perception of apparent motion caused by the physical stimulus changes. Such motion might either disrupt or mask the aftereffect.

To combat this problem, we designed an experiment in which the stimulus remained stationary, but the retinal projection was translated by the subject's own saccadic eye movement. This achieved a stimulus shift on the retina, without introducing perceptual transients that might mask the stimulus. Both the adapting and test stimuli were large (11.25°). Throughout adaptation, subjects fixated a small red dot at one of several locations along the midline of the face. A tone 500 ms before the change to the test stimulus alerted the subjects to redirect their gaze with a saccade to a green dot positioned halfway between the eyes. The largest saccade, between the chin and the eyes, was 6.0° . Eye position was verified by means of a camera-based eye tracking sys-



Fig. 3. Sensitivity to face identity with and without previous adaptation to an anti-face. Three conditions are shown: no adaptation (\circ), adaptation to the matching anti-face (\bullet) and adaptation to the non-matching anti-face (\blacksquare). Subjects are referred to by their initials (MG, MA, AF, UH, JP, ML, MS). Data are shown for two individual subjects (a) and for all subjects (b). The ordinate refers to the fraction of trials in which subject responses matched the identity trajectory of the test face. The data are modeled with a cumulative normal (ogival) function, with mean, standard deviation, base and maximum value. The base and maximal values were free parameters. The recognition threshold was taken to be the inflection point (the mean) of the cumulative normal, and served as a comparison across conditions. The minimum number of trials used in computing each point in (a) was as follows. No adaptation, MG, 100; MS, 120. Match condition, MG, 30; MS, 35. Non-match condition, MG, 90; MS, 120.



tem (Eyelink, SensoMotoric Instruments, Teltow, Germany).

The results (Fig. 6) are shown for several saccade sizes (\bullet), and compared with the data obtained from the same subjects without an adaptation stimulus (\circ). Even for large retinal shifts (6.0°), the effect of adaptation was nearly equivalent to the no-shift condition (solid line). This confirms that our aftereffects do not arise as a combination of localized adaptation effects, and suggests that the underlying neural mechanisms are involved in the analysis of complex object form, in this case, faces.

Single-session testing

Pilot studies suggested that the aftereffect was present from the first exposure. However, the primary experiments used an extensive training regime to gauge a parametric effect of the manipulation. To be certain that our effects were not the result of this training, we tested eight subjects with only minimal exposure (total < 1 h) to the stimuli. Data were collected with the same faces and timing parameters, but with different subjects. Immediately after a quick familiarization to the four faces, subjects viewed the anti-face adaptation trials, with the testing face appearing at one of five levels. Again, adaptation biased perception away from the anti-face, improving performance for matching conditions and degrading it for the non-match case (Table 1). Although the fraction of adaptation-based responses for the average face trials (-0.45) declined somewhat from that seen with more extensive training (> 0.60), it was still significantly above the 0.25 chance level ($t_8 = 2.95$, $p = 0.02$). Thus, although training improved performance, no doubt through increasing recognition sensitivity for faces near the average, the prototype-referenced aftereffects were present from the beginning.

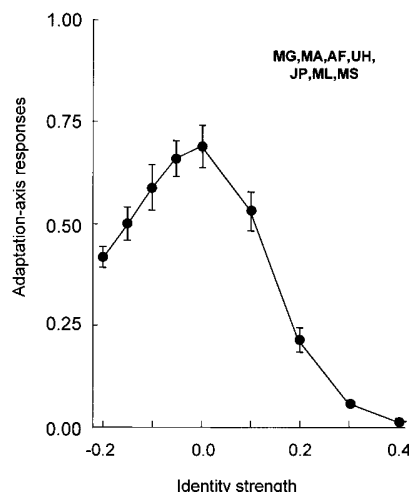
DISCUSSION

Adaptation has often been termed the 'psychophysicist's electrode' because of its power to isolate and temporarily diminish the contribution of specific neural populations, and thereby provide inferences about mechanisms related to visual function. Here we demonstrated that adaptation can be used to tap into mechanisms underlying subordinate-level encoding of shapes as complex as faces. Adaptation to a face that was normal in appearance had a significant and predictable impact on the perceived identity of subsequently presented faces. The effect was present even when the

adapting and test faces were structurally very different, but depended critically upon the relative positions of the two faces in face space. The direction of the perceptual shifts, away from the adapting face and along an axis passing through the center of face space, suggests that the perception of face structure rests upon contrastive mechanisms. The shifts presumably resulted because adaptation temporarily altered the balance of responsiveness among neurons specialized in the analysis of facial structure. Most striking was the observation that adaptation to a realistic face could 'endow' an otherwise neutral test stimulus (the average face) with a new identity, in accordance with the particular trajectory in face space.

Our results suggest that the average face (the center of face space) is particularly important in the interpretation of face structure. This is consistent with the well-studied 'prototype effect,' in which the

Fig. 4. Face identification performance with non-matching adapting and test faces. The ordinate refers to the fraction of trials in which responses matched the identity axis of the anti-face, rather than the identity axis of the test face. For large identity strength values, the number falls to zero, as the judgment is overwhelmed by the identity of the test face.



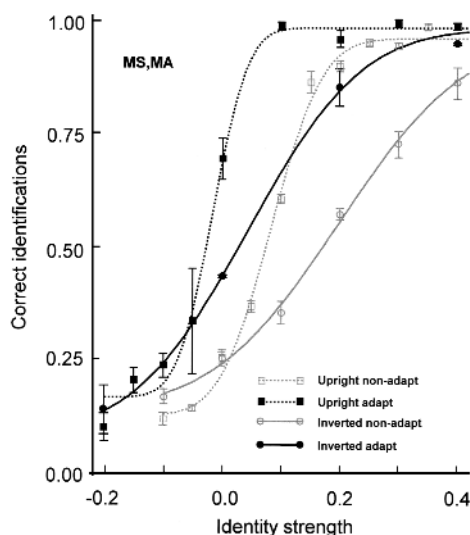


Fig. 5. Effect of face inversion. Sensitivity was measured for two subjects with and without adaptation for both upright (dotted lines) and inverted (solid lines) faces. The adapting anti-face always matched the test stimulus in both identity axis and orientation. Performance was strongly impaired by inversion, flattening the psychometric function and shifting it toward higher identification thresholds. Following adaptation (black, closed symbols), performance improved comparably for upright and inverted faces compared to the unadapted condition (gray, open symbols). Thresholds shifted from 0.84 to -0.016 for upright faces and from 0.199 to 0.043 for inverted ones.

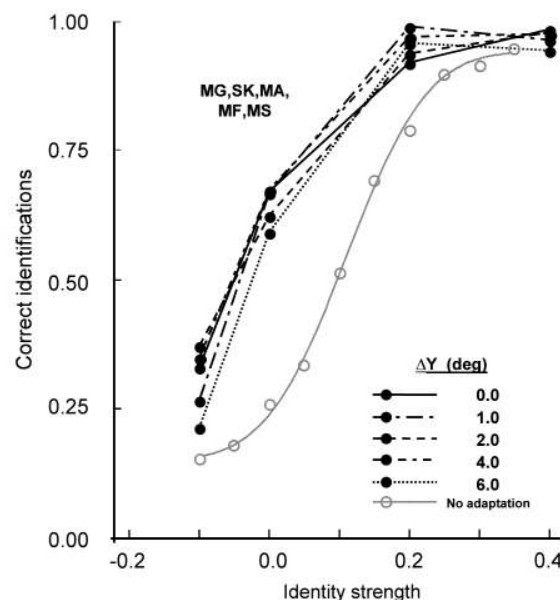


Fig. 6. Translation invariance of face adaptation effects. The ordinate refers to the fraction of identified test faces during the matching condition with (●) and without (○) adaptation. In the adaptation trials, subjects redirected their center of gaze by a distance ΔY just before the appearance of the test stimulus. The solid line refers to condition in which there was no saccade, whereas the various dashed and dotted black lines represent performance following saccades of different sizes.

central tendency of a collection of stimuli, though never actually seen, is classified or recognized more easily than particular exemplars^{10,23,24}. The underlying cause of this effect, however, is less clear at both the neural and conceptual levels. Theoretical work has demonstrated that prototype-like effects can emerge from a diversity of encoding schemes. (For a sampling of the relevant issues in the categorization and face literatures, see refs. 25–28.) The anti-face adaptation effect differs in at least two ways from many studies exploring prototype and exemplar theory. First, the anti-face adaptation suggests a function for the face category prototype in identifying the uniqueness of exemplar faces, whereas most other studies consider the potential function of prototypes in categorizing rather than individuating exemplars. Second, the current experiments involved a change in the visual impression of the test stimulus following adaptation. Temporary visuo-perceptual distortions are a trademark of adaptation aftereffects, but are generally absent or not evaluated in shape recognition and classification tasks.

Table 1. Effects of face adaptation for subjects with no previous experience with the adapting or test stimuli.

Identity fraction	Adaptation-based responses	
	Match condition	Non-match condition
0.00	0.45*	0.45*
0.05	0.50	0.36
0.15	0.56	0.34
0.25	0.68	0.29
0.35	0.72	0.22

Chance performance, 0.25. *Calculated from the same trials.

Which brain areas might be responsible for the observed after-effects? The translation invariance of this effect allows one to eliminate the topographic visual areas, at least as the primary site. Aftereffects were not rooted in the local adaptation of visual primitives, but rather appeared to tap directly into complex shape representations. The inferotemporal cortex (IT) is a more likely candidate for such comparisons, because neurons there are selective for shapes—faces in particular—and have large receptive fields, often exhibiting a high degree of shift invariance in their responses^{29,30}. A previous study describing the ‘shape-contrast’ effect, in which the perception of a quickly flashed simple shape (such as a circle) is distorted after previously viewing another simple shape (such as a line), similarly pinpointed IT as the area most likely underlying the perceptual distortions³¹. In that study, the test stimulus was also briefly flashed, and the effects also tolerated several degrees of displacement between the first and second stimulus.

The current results may predict that responses of neurons or populations of neurons in IT reflect relative rather than absolute coding of complex stimulus attributes. Physiological studies suggest that responses of neurons in this region are unlikely to reflect either generic shape analysis nor highly specific object detection^{32–34}. Most ‘face cells,’ for example, will respond to several faces, albeit with different magnitudes for different individuals or facial expressions³⁵. The hypothesis that neural encoding of faces adheres to particular, well-defined trajectories in face space can be explored with parametric stimulus sets as were used here. Direct measurements of neural responses, in combination with precise psychophysical tools such as the aftereffects introduced here, offer promise to unveil strategies by which the brain discriminates, recognizes, and remembers highly similar complex patterns.



METHODS

Apparatus and stimuli. Morphed face images were presented in color on a computer monitor with spatial resolution of 1024×768 and an eye-screen distance of 110 cm. The faces were generated by a face morphing algorithm, in which point by point correspondence between face pairs is established using both three-dimensional structure and reflectance maps²². The average face was based on 200 scans (100 male and 100 female). The morphing algorithm operated based on correspondences between all data points (rather than a small number of reference points) on each face, yielding a high-resolution representation of face structure. The dimensions of the face space in the creation of the stimuli were therefore a natural consequence of the statistics of a large number of sample faces. The individual faces used in the study were selected randomly from a large database. Unless otherwise mentioned, faces were 7.5 degrees in height, with an aspect ratio of 2:3, and presented to the center of the screen with no fixation spot. Subjects held their head comfortably in a chin rest and answered using a button box.

Training and testing protocol. In the 'training' sessions, subjects learned to associate both a name and a response button to each of the four veridical faces (Fig 1). Subjects then learned to identify the faces with diminished features (that is, closer to the mean). During all training (but not testing), performance feedback was given. Sensitivity was measured then with different identity strengths using the method of constant stimuli with between 100–200 trials per level, depending on the subject. For some subjects, adaptation trials began after the baseline sensitivity testing was complete, whereas for others, baseline and adaptation blocks were interleaved. Pilot studies were run to optimize the duration of the adapting and test stimuli, and were determined to be 5000 and 200 ms, respectively. Shorter adaptation times failed to produce the observed effects, whereas longer times did not significantly change the outcome. Longer test times led to a confused percept, because the adaptation effect was fleeting, disappearing after a few hundred milliseconds. Pilot experiments also suggested that strict fixation did not alter the adaptation effects. Thus, subjects were instructed to 'inspect' the adapting stimulus, perhaps to recognize it later. There was no fixation point, and eye-movement recordings verified that subjects scanned different features of the face. Each level of the test stimulus appeared at least 30 times with each of the four adapting stimuli, resulting in well over 1000 adaptation trials to compute the sensitivity changes for each subject.

Subjects. The subjects were 4 men and 5 women between 22 and 30 years old. Although five subjects had previous experience as psychophysical observers, all were naive to the objectives of the experiment. They were paid for their participation, which averaged approximately twenty one-hour sessions across several weeks. Absolutely critical in this investigation was the naiveté of each subject to the goals and/or design of the experiment. For example, if a subject learned that adaptation with a given anti-face always gave rise to the perception of a particular face, the adapting stimulus could influence the subject's decision at a cognitive rather than perceptual level. For that reason, we went to extreme measures to ensure that subjects could never make associations between faces and anti-faces. First, the subjects were always presented with extra, unanalyzed trials in which the adapting stimulus was an anti-face that did not correspond to any of the memorized stimuli. Second, the test stimuli in a given session always consisted of many levels of identity strength. An interview and small 'examination' at the end of the final experimental session verified the naiveté of seven of the nine subjects on this point (two were no longer available).

ACKNOWLEDGEMENTS

The authors would like to thank N. Logothetis, S. Edelman, M. Wenger and H. Abdi for comments on the manuscript and S. True for discussion regarding the project. We also thank U. Helmich and M. Sauer for help conducting the experiments, and N. Bartolozzi and C. Moya for technical assistance. Finally, we acknowledge the support of the Max Planck Institute for Biological Cybernetics in the development of stimuli. This work was supported by the Max Planck Society. In addition, A.J.O. was supported by the National Institute of Standards and Technology.

RECEIVED 17 AUGUST; ACCEPTED 14 NOVEMBER

- Logothetis, N. K. & Pauls, J. Psychophysical and physiological evidence for viewer-centered object representations in the primate. *Cereb. Cortex* **5**, 270–288 (1995).
- Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P. & Gore, J. C. Activation of the middle fusiform "face area" increases with expertise in recognizing novel objects. *Nat. Neurosci.* **2**, 568–573 (1999).
- Perrett, D. I., Rolls, E. T. & Caan, W. Visual neurones responsive to faces in the monkey temporal cortex. *Exp. Brain Res.* **47**, 329–342 (1982).
- Desimone, R., Albright, T. D., Gross, C. G. & Bruce, C. Stimulus-selective properties of inferior temporal neurons in the macaque. *J. Neurosci.* **4**, 2051–2062 (1984).
- Logothetis, N. K., Guggenberger, H., Peled, S. & Pauls, J. Functional imaging of the monkey brain. *Nat. Neurosci.* **2**, 555–562 (1999).
- Ojemann, J. G., Ojemann, G. A. & Lettich, E. Neuronal activity related to faces and matching in human right nondominant temporal cortex. *Brain* **115**, 1–13 (1992).
- Kanwisher, N., McDermott, J. & Chun, M. M. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* **17**, 4302–4311 (1997).
- Valentine, T. A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Q. J. Exp. Psychol. A* **43A**, 161–204 (1991).
- Rhodes, G., Carey, S., Byatt, G. & Proffitt, F. Coding spatial variations in faces and simple shapes: a test of two models. *Vision Res.* **38**, 2307–2321 (1998).
- Cabeza, R., Bruce, V., Kato, T. & Oda, M. The prototype effect in face recognition: Extension and limits. *Mem. Cognit.* **27**, 139–151 (1999).
- O'Toole, A. J., Vetter, T., Nikolaus, T. F. & Bulthoff, H. H. Sex classification is better with three-dimensional head structure than with image intensity information. *Perception* **26**, 75–84 (1997).
- O'Toole, A. J., Deffenbacher, K. A., Valentin, D. & Abdi, H. Structural aspects of face recognition and the other-race effect. *Mem. Cognit.* **22**, 208–224 (1994).
- Benson, P. J. & Perrett, D. I. Visual processing of facial distinctiveness. *Perception* **23**, 75–93 (1994).
- Deffenbacher, K. A., Vetter, T., Johanson, J. & O'Toole, A. J. Facial aging, attractiveness, and distinctiveness. *Perception* **27**, 1233–1244 (1998).
- Rhodes, G., Brennan, S. & Carey, S. Identification and ratings of caricatures: implications for mental representations of faces. *Cognit. Psychol.* **19**, 473–497 (1987).
- Stevenage, S. V. Can caricatures really produce distinctiveness effects? *Br. J. Psychol.* **86**, 127–146 (1995).
- Blanz, V., O'Toole, A. J., Vetter, T., & Wild, H. A. On the other side of the mean: the perception of dissimilarity in human faces. *Perception* **29**, 885–892 (2000).
- Webster, M. A. & MacLin, O. H. Figural aftereffects in the perception of faces. *Psychon. Bull. Rev.* **6**, 647–653 (1999).
- Koehler, W. & Wallach, H. Figural after-effects: an investigation of visual processes. *Proc. Am. Phil. Soc.* **88**, 269–357 (1944).
- Blakemore, C. & Campbell, F. W. On the existence of neurons in the human visual system selectively sensitive to the orientation and size of retinal images. *J. Physiol. (Lond.)* **203**, 237–260 (1969).
- Mather, G., Verstraten, F. & Anstis, S. *The Motion Aftereffect: A Modern Perspective*. (MIT Press, Cambridge, Massachusetts, 1998).
- Blanz, V. & Vetter, T. in *1999 Symposium on Interactive 3D Graphics—Proceedings of SIGGRAPH '99*. 187–194 (ACM Press, New York, 1999).
- Posner, M. I. & Keele, S. W. Retention of abstract ideas. *J. Exp. Psychol.* **83**, 394–308 (1970).
- Posner, M. I. & Keele, S. W. On the genesis of abstract ideas. *J. Exp. Psychol.* **77**, 353–363 (1968).
- Ashby, F. G. & Maddix, W. T. Relations between prototype, exemplar, and decision bound models of categorization. *J. Math. Psych.* **37**, 372–400 (1993).
- Nosofsky, R. M. Tests of an exemplar model for relating perceptual classification and recognition memory. *J. Exp. Psychol. Hum. Percept. Perform.* **17**, 3–27 (1991).
- Estes, W. K. *Classification and Cognition*. (Oxford University Press, Oxford, 1994).
- Lewis, M. B. & Johnston, R. A. A unified account of the effects of caricaturing faces. *Vis. Cognit.* **6**, 1–41 (1999).
- Logothetis, N. K., Pauls, J. & Poggio, T. Shape representation in the inferior temporal cortex of monkeys. *Curr. Biol.* **5**, 552–563 (1995).
- Ito, M., Tamura, H., Fujita, I. & Tanaka, K. Size and position invariance of neuronal responses in monkey inferotemporal cortex. *J. Neurophysiol.* **73**, 218–226 (1995).
- Suzuki, S. & Cavanagh, P. A shape-contrast effect for briefly presented stimuli. *J. Exp. Psychol. Hum. Percept. Perform.* **24**, 1315–1341 (1998).
- Yamane, S., Kaji, S. & Kawano, K. What facial features activate face neurons in the inferotemporal cortex of the monkey? *Exp. Brain Res.* **73**, 209–214 (1988).
- Sugase, Y., Yamane, S., Ueno, S. & Kawano, K. Global and fine information coded by single neurons in the temporal visual cortex. *Nature* **400**, 869–873 (1999).
- Hasselmo, M. E., Rolls, E. T. & Baylis, G. C. The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behav. Brain Res.* **32**, 203–218 (1989).
- Young, M. P. & Yamane, S. Sparse population coding of faces in the inferotemporal cortex. *Science* **256**, 1327–1331 (1992).