THE ROYAL SOCIETY biology letters

Face processing in humans is compatible with a simple shape-based model of vision

Maximilian Riesenhuber[1,2,3*], Izzat Jarudi[2], Sharon Gilad[2] and Pawan Sinha[2]

[1]*McGovern Institute for Brain Research and Center for Biological and Computational Learning, and* [2]*Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02142, USA*
[3]*Department of Neuroscience, Georgetown University Medical Center, Washington, DC 20007, USA*
[*] *Author for correspondence* (*mr287@georgetown.edu*).

Recd 04.03.04; *Accptd* 04.05.04; *Published online* 06.07.04

Understanding how the human visual system recognizes objects is one of the key challenges in neuroscience. Inspired by a large body of physiological evidence, a general class of recognition models has emerged, which is based on a hierarchical organization of visual processing, with succeeding stages being sensitive to image features of increasing complexity. However, these models appear to be incompatible with some well-known psychophysical results. Prominent among these are experiments investigating recognition impairments caused by vertical inversion of images, especially those of faces. It has been reported that faces that differ 'featurally' are much easier to distinguish when inverted than those that differ 'configurally'; a finding that is difficult to reconcile with the physiological models. Here, we show that after controlling for subjects' expectations, there is no difference between 'featurally' and 'configurally' transformed faces in terms of inversion effect. This result reinforces the plausibility of simple hierarchical models of object representation and recognition in the cortex.

Keywords: computational neuroscience; neuroscience; object recognition; faces; psychophysics; inversion effect

1. INTRODUCTION

Since its discovery by Yin (1969), the face-inversion effect (i.e. the observation that faces are surprisingly more difficult to recognize when turned upside down versus right-side up) has acquired the status of a cornerstone finding in the domain of visual recognition. The dominant explanation of this effect is that the human visual system's strategy for facial representation is primarily 'configural', i.e. it involves encoding the second-order spatial relationships between face parts such as the eyes, nose and mouth (Carey & Diamond 1986; Tanaka & Farah 1993; Farah *et al.* 1995; Freire *et al.* 2000; Le Grand *et al.* 2001; Mondloch *et al.* 2002). Configural analysis is believed to be compromised with vertically inverted faces, and, under these circumstances, the visual system is forced to resort to a 'featural' mode of processing. Directly testing this hypothesis, several studies have reported that changes to the 'features' of a face (commonly defined as consisting of the eyes, mouth and nose) can be detected equally well in upright faces as in inverted faces, while changes to the 'configuration' of a face (defined as the 'distinctive relations among the elements that define the shared configuration' (Carey & Diamond 1986) of face features) cause an inversion effect, with much better detectability in upright than in inverted faces (Freire *et al.* 2000; Le Grand *et al.* 2001).

However, the models suggested by physiology (Hubel & Wiesel 1962; Livingstone & Hubel 1988; Felleman & Van Essen 1991; Zeki 1993; Tso *et al.* 2001) are agnostic about the source of differences between two faces; they make no explicit distinction between featural and configural changes (Selfridge 1959; Reisenhuber & Poggio 1999). According to the models, therefore, if two modifications to the shape of a face—be they a result of changes in the 'configuration' or in the 'features'—influence discrimination performance to an equal degree for upright faces, they should also have an equal effect on the discrimination of inverted faces; i.e. there is no special role for 'configuration' or 'features'. There is, thus, an important inconsistency between reported psychophysical data and predictions from the hierarchical models of recognition.

A potentially significant shortcoming of the psychophysical studies mentioned above is that they have used blocked designs (trials were either grouped by change type (Mondloch *et al.* 2002) or used a different subject group for each change type (Freire *et al.* 2000)) where 'featural' and 'configural' changes were separated into different groups, making it possible for subjects to use change type-specific recognition strategies different from generic face-processing strategies. For instance, in a blocked design, it is conceivable for subjects in featural trials to use a strategy that does not rely on the visual system's face representation but rather focuses just on detecting local changes in the image, for example, in the eye region. This would then lead to a performance less affected by inversion. However, such a local strategy would not be optimal for configural trials since, in these trials, the eye itself does not change, only its position with respect to the rest of the face. Thus, configural trials can profit from a 'holistic' strategy (i.e. looking at the whole face, which for upright but not for inverted faces presumably engages the learned (upright) face representation), which would in turn predict a strong effect of inversion.

We therefore performed a same or different face-matching experiment using face pairs differing in features or configuration, in which subjects were not able to predict change type (see figure 1 and § 2; for additional details, see electronic Appendix A).

2. MATERIAL AND METHODS

Subjects performed a same or different task using pairs of faces differing either by a 'configural' (figure 1*a*) or a 'featural' (figure 1*b*) change. Photorealistic stimuli were created using a custom-built morphing system (see electronic Appendix A) that allowed us to freely move and exchange face parts of 200 face prototypes (Blanz & Vetter 1999). Subjects performed a total of 160 trials, based on 80 image pairs, each presented upright and inverted on different trials. Forty face pairs were associated with 'featural trials'; 20 face pairs with the faces in each pair differing by a feature change (replacement of eyes and mouth regions with those from other faces prototypes; Freire *et al.* 2000; Le Grand *et al.* 2001; Mondloch *et al.* 2002), and 20 'same' face pairs composed of the same outlines and face component positions as the corresponding 'different' faces, with both faces having the same eye and/or mouth regions. Another 40 face pairs were used in the 'configural change' trials, consisting of 20 face pairs

DOI 10.1098/rsbl.2004.0216

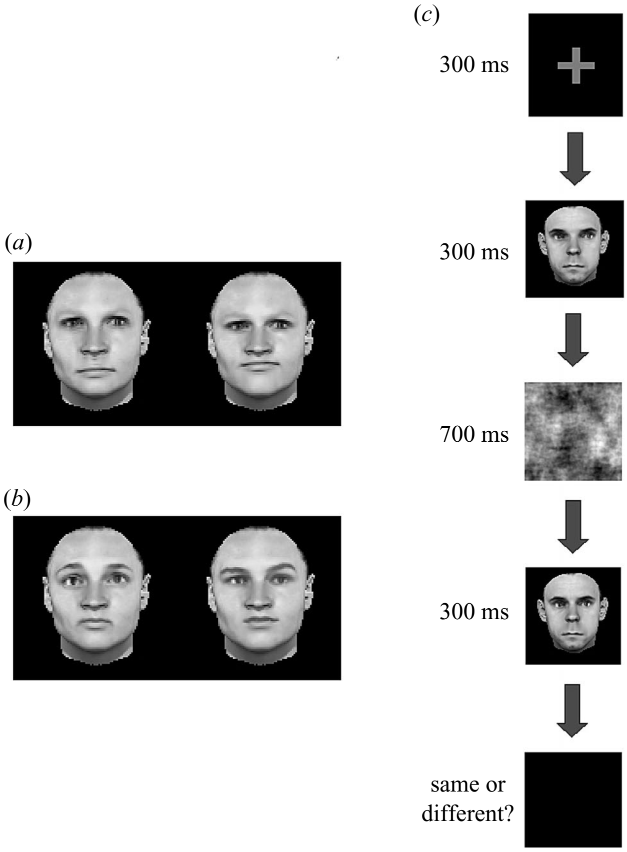

Figure 1. Example stimuli and task. (*a*) Example of a 'configural' change stimulus pair. The two images show two versions of the same face, with the right face's mouth moved up and the eyes moved together relative to the left face. (*b*) Example of a 'featural' change stimulus pair. The two images show two versions of the same face, with the right face's mouth and eyes replaced by the mouth and eyes from another randomly selected face. The face outline and the nose are the same as in (*a*). (*c*) Experimental paradigm. Subjects first fixated on a cross for 300 ms, then viewed one of two pictures in a face pair (here, a 'configural change' pair) for 300 ms, followed by a noise mask for 700 ms, and the second picture in the pair for 300 ms, and finally, a blank screen until subjects made their 'same' or 'different' judgement by pressing a specific keyboard button.

differing by a configural change (displacement of the eyes and/or mouth regions; Freire *et al.* 2000; Le Grand *et al.* 2001; Mondloch *et al.* 2002), plus 20 'same' face pairs composed of faces with the same face outlines and parts as the corresponding 'different' pairs, with both faces in the pair having the same configuration. Faces were selected so that performance in upright featural and configural trials was comparable (see experiment 1 in electronic Appendix A). In the 'unblocked' version of the experiment (experiment 2 in electronic Appendix A and § 3), configural and featural trials were presented in a random order (counterbalanced across subjects). In the 'blocked' version (experiment 3 in electronic Appendix A and § 3), trials were grouped by change type.

3. RESULTS

Results for 15 subjects are shown in figure 2. As expected, inversion adversely affects performance. More importantly, we find the effect of inversion to be comparable for 'featural' and 'configural' changes. This is also borne out by an ANOVA that shows a highly significant main effect of orientation ($p < 0.0001$) but no main effect for change type ($p > 0.29$) and no interaction ($p > 0.17$).

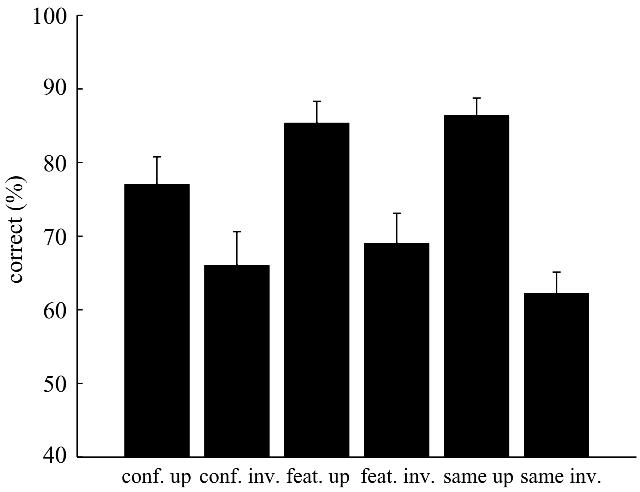

Figure 2. Subject performance in the unblocked discrimination experiment. The bars show subjects' performance ($n = 15$) for the different trial types (conf., trials where the two images differed by a configural change; feat., trials with featural changes; same, trials with two identical images; up, upright images; inv., inverted images). Error bars show standard error of the mean.

These results are compatible with a shape-based representation specialized for upright faces, but not with a representation that explicitly encodes facial 'configuration'.

These results are thus in notable contrast to previous experiments that have found an inversion effect for featural but not for configural changes (Freire *et al.* 2000; Le Grand *et al.* 2001; Mondloch *et al.* 2002). To investigate whether the earlier results might have been an artefact of blocking trials by change type, we tested additional subjects on a modified version of our experiment with identical trials, but this time blocked according to change type (see experiment 3 in electronic Appendix A). Thus, one group of subjects ($n = 12$, the 'configural first' group) was first exposed to all trials containing the 'configural' images (including 'same' and 'different' trials; see § 2), and then all 'featural' trials, whereas the blocks were reversed for another group of subjects ($n = 13$, the 'featural first' group). Subjects were not informed that images in the two blocks differed in any way. None of the subjects had participated in the unblocked experiment.

Interestingly, blocking the trials caused subjects' performance to vary substantially depending on which group they belonged to: in the 'configural first' group, subjects showed no difference in performance over all 'configural' versus all 'featural' trials (*t*-test: $p = 0.19$), compatible with the hypothesis that subjects used the same holistic, face-based strategy for all trials, as in the original unblocked experiment (where there was no difference between average subject performance on featural and configural trials: $p > 0.7$). The situation was very different in the 'featural first' group, where performance on featural and configural trials was highly significantly different ($p = 0.001$). This was owing to poor performance on the configural trials (63% versus 73% on featural trials), as would be expected if subjects used a strategy based on local, part-based image comparisons. Indeed, ANOVAs for the different subject groups showed a significant main effect of orientation in both groups ($p < 0.001$), but a

highly significant effect of change type only in the 'featural first' group ($p < 0.001$). The effect of change type missed significance for the 'configural first' group ($p > 0.05$), similar to the original, unblocked design. This suggests that blocking trials can cause subjects to adopt artefactual visual strategies.

4. DISCUSSION

Our psychophysical results therefore, help reconcile an important inconsistency between past experimental data and predictions from modelling and physiology. They strongly support a simple shape-based model of visual processing, in agreement with physiological data. Furthermore, they suggest that the representation of facial shape information, while holistic, is not explicitly configural. This hypothesis is supported by a recent paper by Sekuler *et al.* (2004), which showed that there might not be a qualitative shift in face processing strategy in going from upright to inverted faces. This hypothesis also makes interesting predictions regarding the response properties of face-selective neurons in the primate inferotemporal cortex, a brain area crucial for object recognition in primates (Logothetis & Sheinberg 1996). Many 'face neurons' have already been shown to exhibit 'holistic' tuning (as defined by Tanaka & Farah 1993; Farah *et al.* 1995), given that they 'require nearly all the essential features of a face' for activation (Tanaka 2003). Based on our experimental results here, we would predict that 'featural' and 'configural' changes of a face stimulus that cause an equal activation change for upright faces should also have an equal effect for inverted faces (but probably of lower magnitude given the preferred tuning of most face neurons to upright faces; Tanaka *et al.* 1991), in marked contrast to configural theories.

Acknowledgements

The authors thank Vikash Gilja, Daniel Ramage and Antonio Torralba for help with the pilot studies, Christoph Zrenner for developing the face morphing software, Thomas Vetter for the face prototypes and Tomaso Poggio for discussions and comments on an earlier version of the manuscript. M.R. was supported by a McDonnell-Pew award in Cognitive Neuroscience. I.J. was supported by the Meryl and Stewart Robertson UROP fund at MIT. P.S. is supported by DARPA's HumanID program and an Alfred P. Sloan fellowship.

Blanz, V. & Vetter, T. 1999 A morphable model for the synthesis of 3D faces. In *SIGGRAPH '99*, pp. 187–194. New York: ACM Computer Society Press.

Carey, S. & Diamond, R. 1986 Why faces are and are not special: an effect of expertise. *J. Exp. Psychol. Gen.* **115**, 107–117.

Farah, M. J., Tanaka, J. W. & Drain, H. M. 1995 What causes the face inversion effect? *J. Exp. Psychol. Hum. Percept. Perform.* **21**, 628–634.

Felleman, D. J. & Van Essen, D. C. 1991 Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex* **1**, 1–47.

Freire, A., Lee, K. & Symons, L. A. 2000 The face-inversion effect as a deficit in the encoding of configural information: direct evidence. *Perception* **29**, 159–170.

Hubel, D. H. & Wiesel, T. N. 1962 Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* **160**, 106–154.

Le Grand, R., Mondloch, C. J., Maurer, D. & Brent, H. P. 2001 Neuroperception. Early visual experience and face processing. *Nature* **410**, 890.

Livingstone, M. S. & Hubel, D. H. 1988 Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science* **240**, 740–749.

Logothetis, N. K. & Sheinberg, D. L. 1996 Visual object recognition. *A. Rev. Neurosci.* **19**, 577–621.

Mondloch, C. J., Le Grand, R. & Maurer, D. 2002 Configural face processing develops more slowly than featural face processing. *Perception* **31**, 553–566.

Riesenhuber, M. & Poggio, T. 1999 Hierarchical models of object recognition in cortex. *Nature Neurosci.* **2**, 1019–1025.

Sekuler, A. B., Gaspar, C. M., Gold, J. M. & Bennett, P. J. 2004 Inversion leads to quantitative, not qualitative, changes in face processing. *Curr. Biol.* **14**, 391–396.

Selfridge, O. G. 1959 Pandemonium: a paradigm for learning. In *Mechanisation of thought processes: proceedings of a symposium held at the National Physics Laboratory* (ed. D. Blake & A. Uttley), pp. 511–529. London: HMSO.

Tanaka, J. W. & Farah, M. J. 1993 Parts and wholes in face recognition. *Q. J. Exp. Psychol.* A **46**, 225–245.

Tanaka, K. 2003 Columns for complex visual object features in the inferotemporal cortex: clustering of cells with similar but slightly different stimulus selectivities. *Cerebr. Cortex* **13**, 90–99.

Tanaka, K., Saito, H., Fukada, Y. & Moriya, M. 1991 Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *J. Neurophysiol.* **66**, 170–189.

Tso, D. Y., Roe, A. W. & Gilbert, C. D. 2001 A hierarchy of the functional reorganization for color, form, and disparity in primate visual area V2. *Vision Res.* **41**, 1333–1349.

Yin, R. K. 1969 Looking at upside-down faces. *J. Exp. Psychol.* **81**, 141–145.

Zeki, S. 1993 *Vision of the brain*. Oxford UK: Blackwell Science.

Visit www.journals.royalsoc.ac.uk and navigate to this article through *Biology Letters* to see the accompanying electronic appendix.