To What Extent Can the Recognition of Unfamiliar Faces be Accounted for by the Direct Output of Simple Cells?

Peter Kalocsai, Irving Biederman, and Eric E. Cooper

University of Southern California
Hedco Neuroscience Building
Los Angeles, CA 90089-2520

e-mail: kalocsai@rana.usc.edu

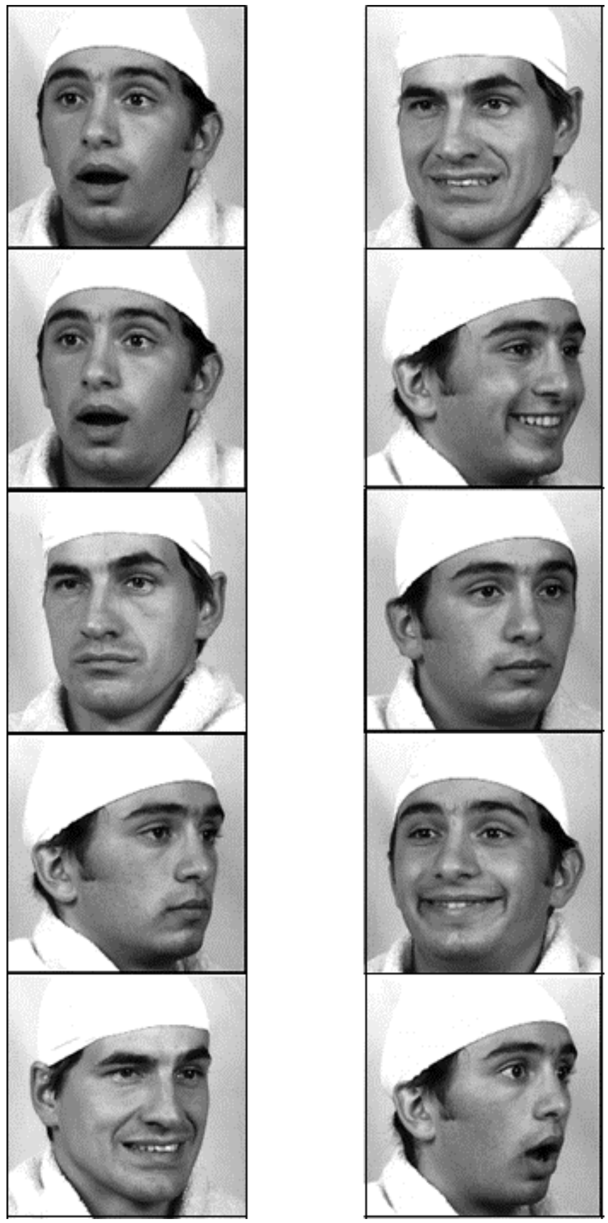
Poster presented May, 1994 at the annual meeting of the Association for Research in Vision and Ophthalmology, Sarasota, Florida Sponsored by AFOSR 88-0231 and McDonnell-Pew Cognitive Neuroscience T89-01245-029.

The authors wish to express their appreciation to Professor Christoph von der Malsburg for his help in making the Face Recognition System available and to the significant contributions of Jean-Marc Fellous and József Fiser in this research.

Abstract

Purpose. What computations are performed by the specialized areas tuned to the identification of faces? One possibility is that these areas transform the outputs of a lattice of hypercolumns of Gabor-like simple cells that characterize the early stages in the ventral pathway so that an intermediate representation which alters the simple cell similarity space is created. This intermediate representation might then be employed for the activation of identity. Another possibility is that these areas map the outputs of these simple cells without much change onto recognition units, similar to a two-layer network. The degree to which a two-layer network could account for the effects of rotational and emotional expression changes on the speed and accuracy of the recognition of unfamiliar faces was assessed. Method. Subjects judged whether a pair of brief (100 msec), masked, sequential presentations of face images were of the same or different individuals. The images could differ in orientation in depth (0˚-60˚) and expression (neutral, smiling, surprised). The similarity of each pair of faces was assessed by the Buhmann, Lades, & von der Malsburg's (1990) two-stage face recognition system. The system develops links between adjacent columns in the lattice so that relations among the filter activation values are coded. Results. Both orientation and expression differences produced highly reliable (and approximately additive) effects on RTs and error rates. The effects of rotation were highly correlated with the similarity values calculated by the model, r = -.90. The correlation for the different expressions was lower, r = -.59, but still highly reliable. Conclusion. A metric based on the activation of a lattice of simple cells correlates highly with real-time human face recognition performance.

Which of the following pairs of faces (left and right) are the same or different individuals?



To what extent is performance on a face recognition task similar to this one predictable from a simple cell similarity space?

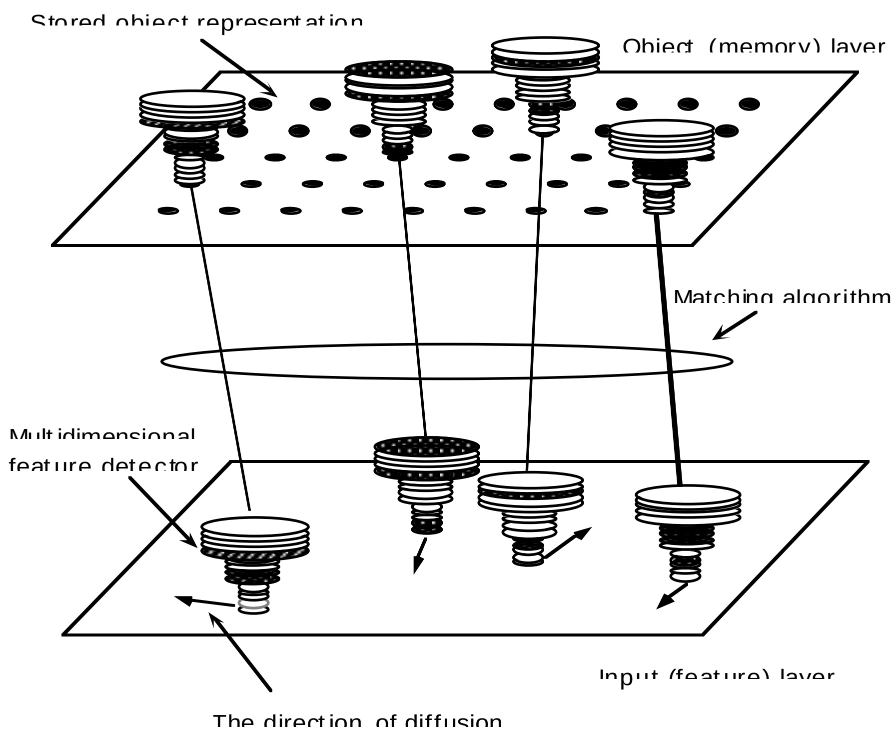
Simple Cell Similarity Space of Images of faces

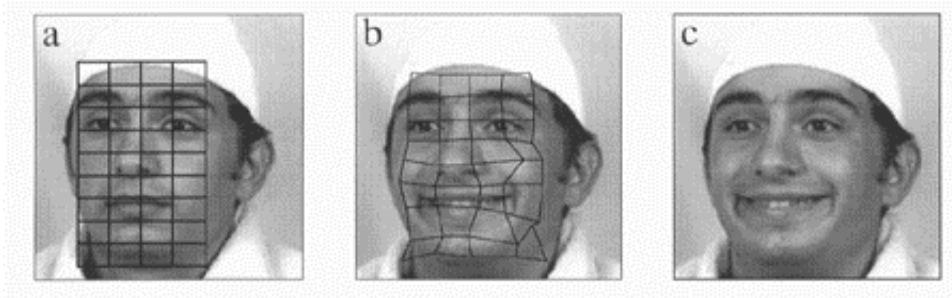
The simple cell similarity of the images was determined with the Buhmann, Lange, & von der Malsburg's (1990) simple cell (SC) model for face recognition.

General Structure of the SC Model

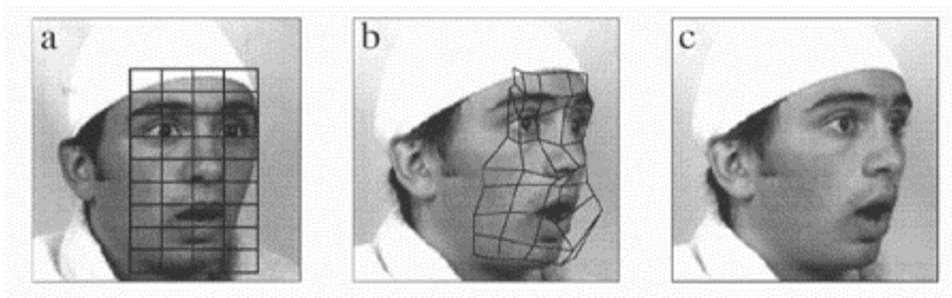
Architecture

The SC model first convolves each input image with a set of Gabor kernels at five scales and eight orientations arranged in a 5 x 9 lattice (figure below). The positioning of the lattice over kernels at each node in the lattice is termed a “Gabor jet.” The activation values of the kernels in each jet along with their positions are stored for each of the images to form a “gallery”. an image is shown in the left hand column, labeled (a) of the figure with faces below.

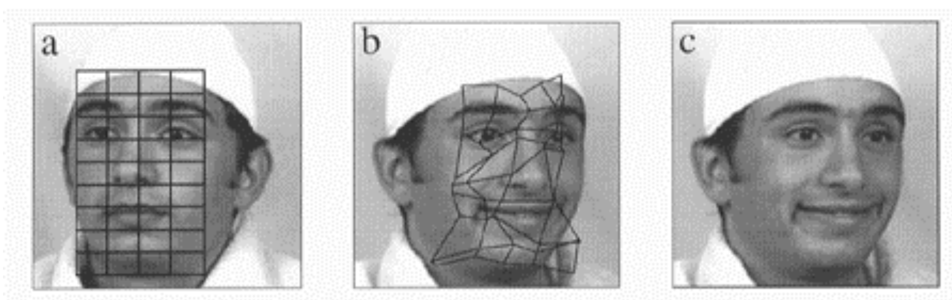




Same orientation-different expression



Different orientation-same expression



Different orientation-different expression

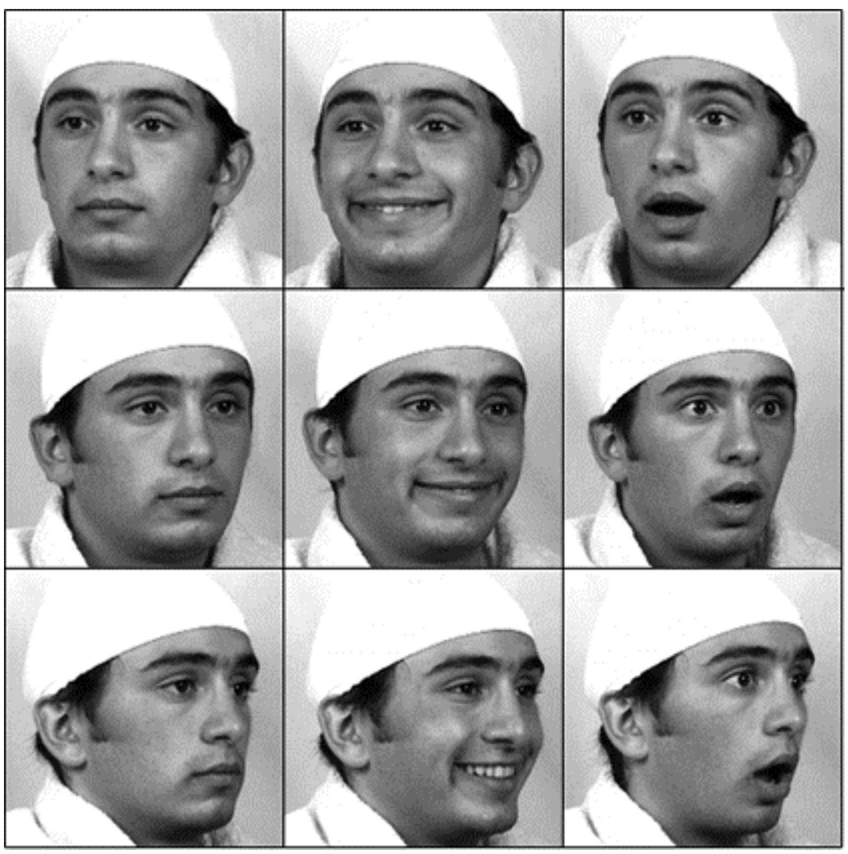
Matching (Determination of Similarity)

The similarity between a test image and the various stored images (gallery) is calculated by stochastic optimization that allows each of the jets to diffuse (gradually change its position) to optimize the similarity in kernel values and distances relative to adjacent jets. The result of the diffusion over a pair of faces is shown in the middle column (b) in the above figure (the test faces without the distorted grids are presented in the right hand column (c)). To the extent that the jets move independently, the resultant positions will no longer produce a rectangular lattice, as illustrated in the figure. In general the more distorted the lattice, the less the similarity of the image to the original. The most similar match of the test image is interpreted to be the recognition response of the model. The model achieves 83% accuracy in correctly recognizing a second image of an individual (out of a gallery of 160 individuals), even with considerable variation in facial expression but only slight differences in orientation. When the correct face does not receive the highest rank, it is almost always among the next two faces.

Experiment 1: Same-Different Judgment of Identity of Faces

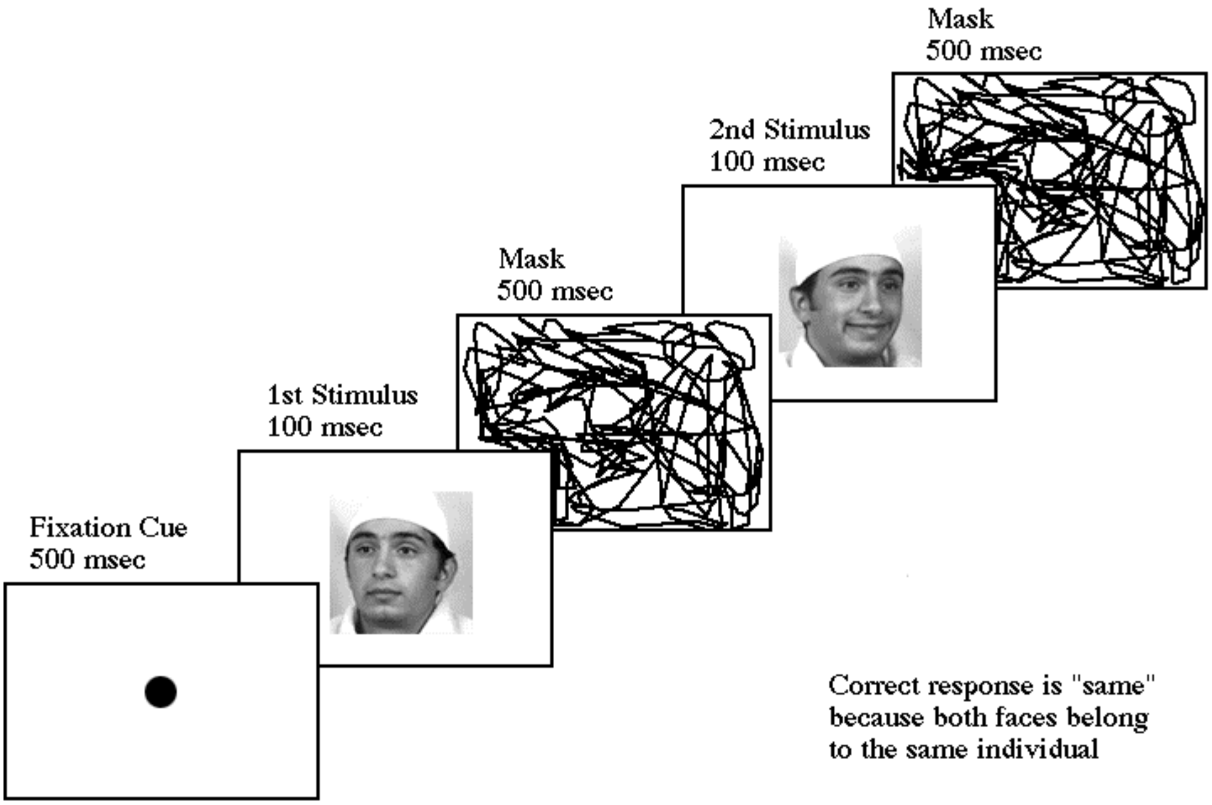
Method

Subjects judged whether two highly similar faces such as those shown above (same sex and age, no hair or clothing or easy features), were the same or different. The faces were viewed briefly (100 msec) and sequentially (with masks after each image). The two faces could differ in orientation by 0 to 60 deg (from 20 deg left to 40 deg right) and emotional expression (neutral, happy, surprise).

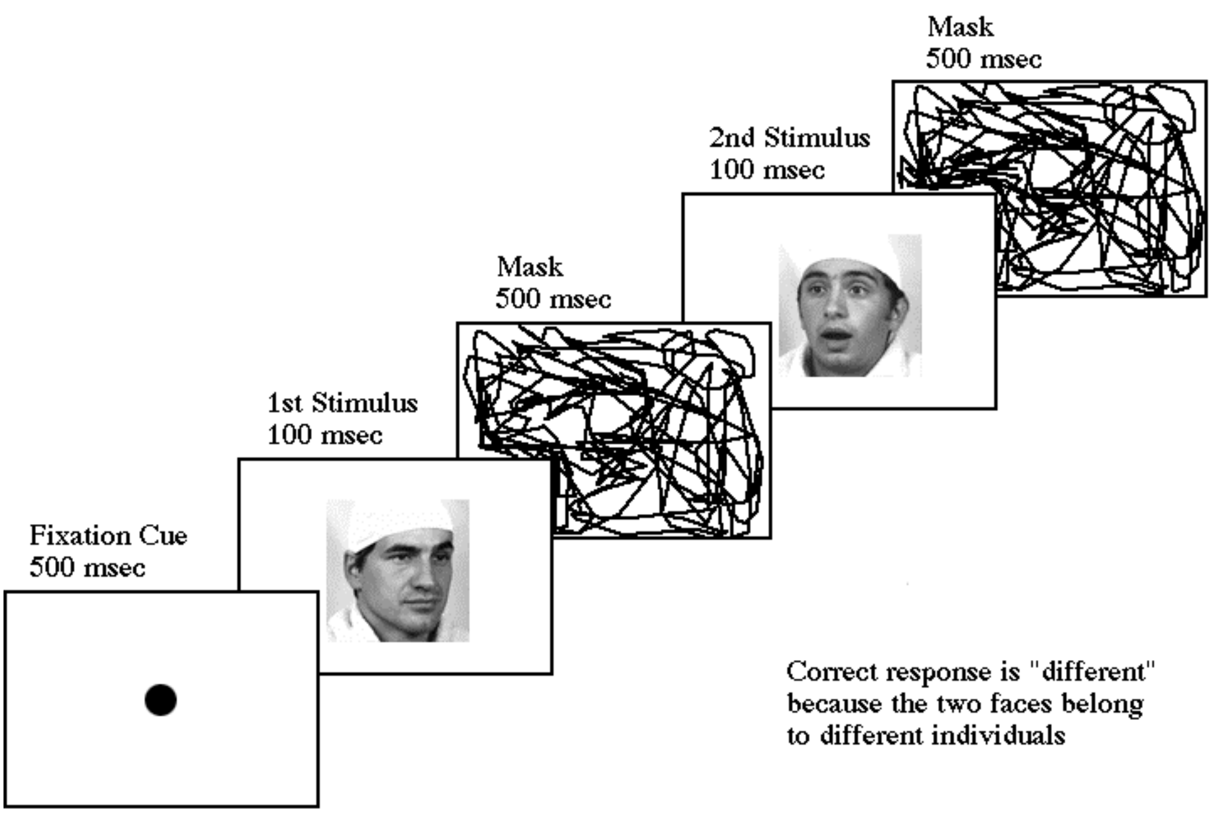


The stimuli set: nine pictures were taken of each individual.
Three orientations: 20˚ left, 20˚ right, and 40˚ right (vertically).
Three emotional expressions: neutral, happy, and surprised (horizontally).

Exp. 1 Trial Sequence

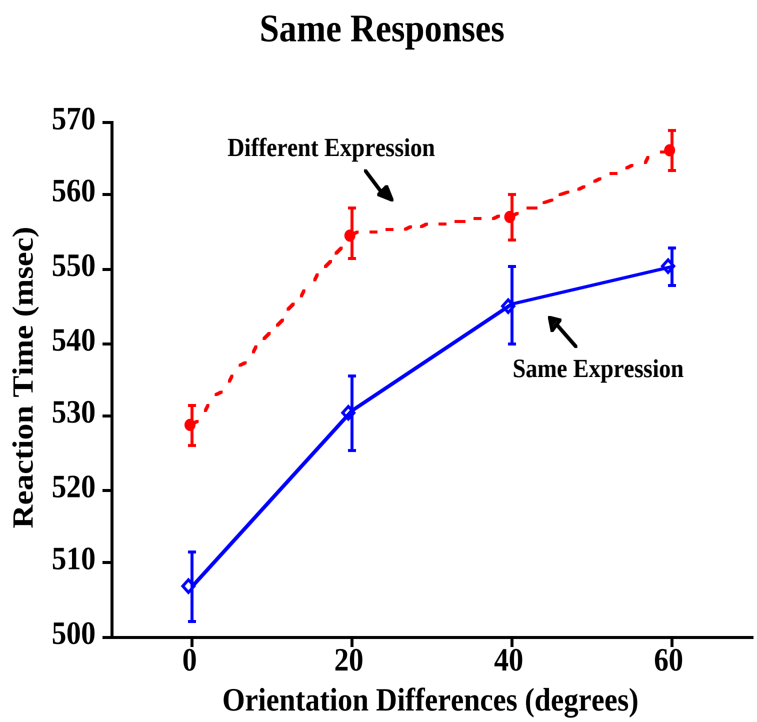


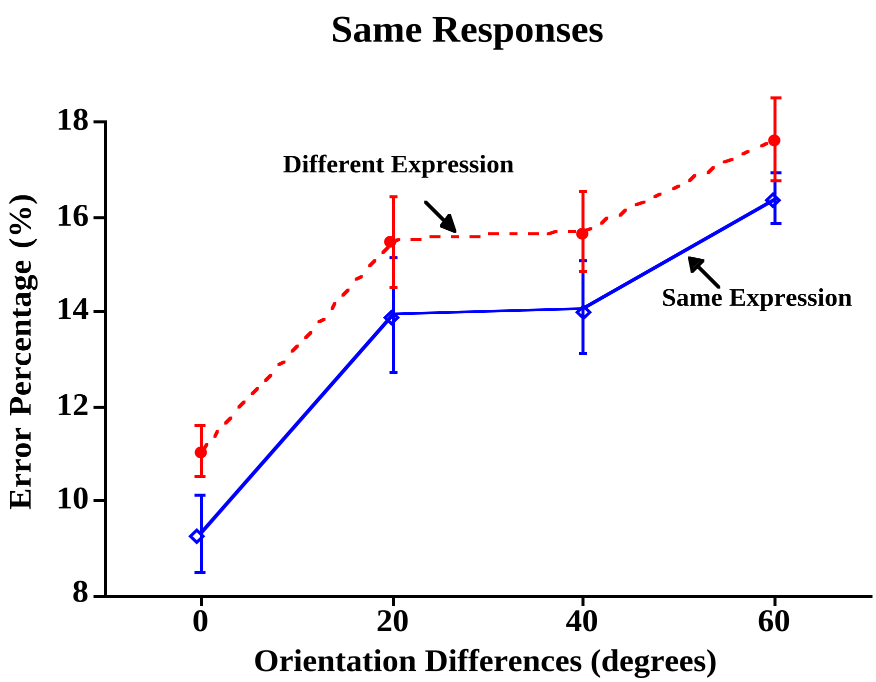
Exp. 1 Trial Sequence



Results

Differences in orientation and expression led to increases in reaction times and error rates in judging that two pictures were of the same individual.

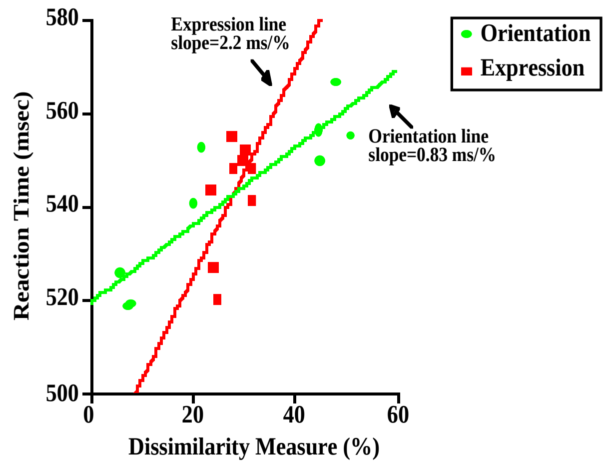




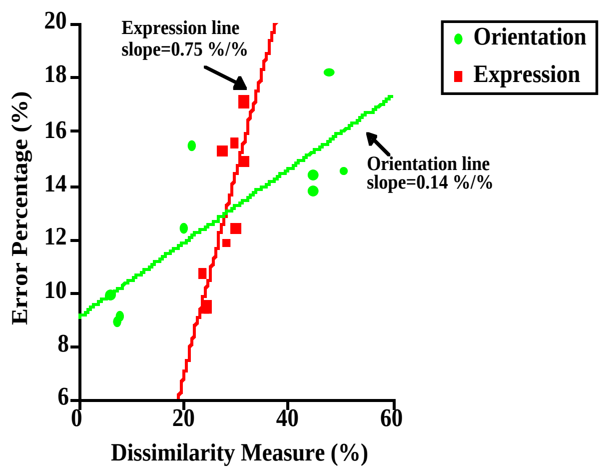
Were these effects predictable from the same Gabor jet similarity space? A measure based on the Gabor jet similarity space correlated .90 with the effect of rotation in depth. Although the correlation of differences in emotional expression was lower, .59, differences of emotional expression occupied a much smaller range of similarity values than orientation differences which would be expected to reduce the magnitude of the correlation. Within this limited range, it is apparent that the slopes of the regression functions, shown below, relating similarity values of emotional expression differences to RTs and error rates are steeper than those for orientation differences. That is, differences in Gabor similarity produced by different emotional expressions have a considerably larger effect on recognition speed than what would be predicted from the effect of orientation.

The similarity of each pair of face images was determined and expressed as a percent of the best possible match. 0% dissimilarity would be the value for matching identical images. 100% dissimilarity would be the value for matching random images. The datapoints on the graphs for orientation and expression below show the mean values (RTs, error percentages, and dissimilarity measures) for all possible combinations of the first and second face being any of three orientations and any of three expressions.

Same Responses
Regression of Reaction Time Against Dissimilarity



Same Responses
Regression of Error Rate Against Dissimilarity

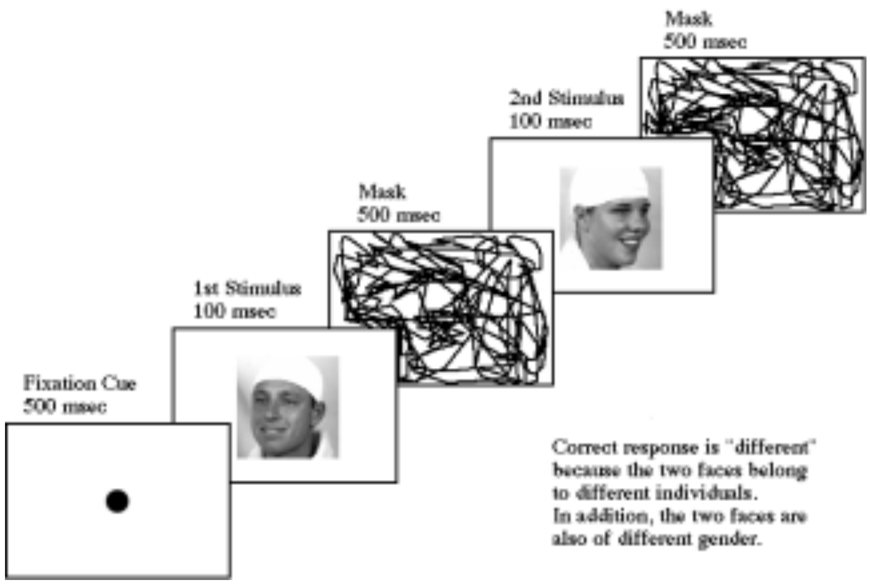


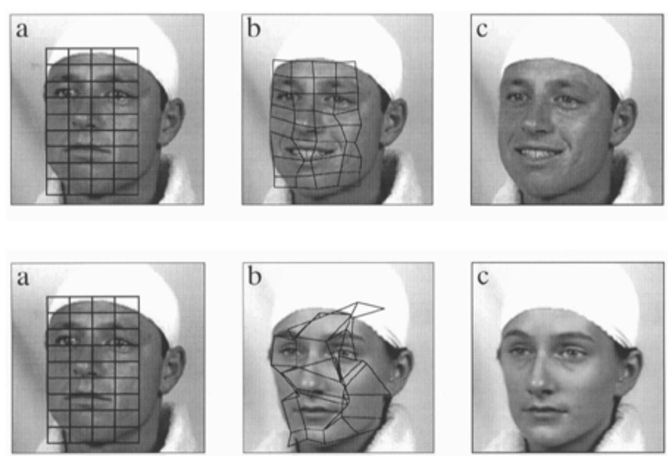
Experiment 2: Gender Effects on Face Judgments

Method

In Exp. 1 when the two sequentially presented faces were of different individuals they were always of the same sex. In Exp. 2, if the two faces were of different individuals, then they were also of different genders (figure below). In this way it was possible to test the effect of gender information on face recognition.

Exp. 2 Trial Sequence





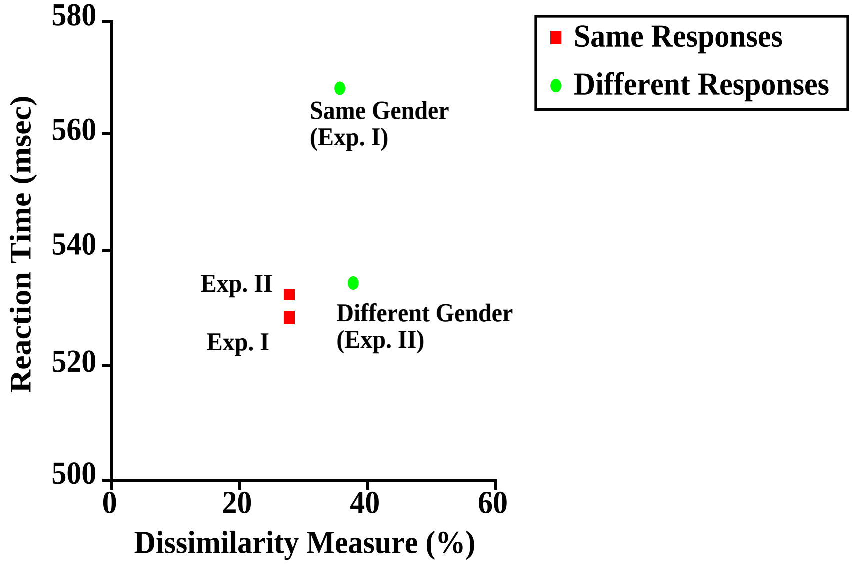
Examples of same and different "response" of the SC model with different gender on the different response.

a, Face in the gallery with the original grid
b, Test face with distorted grid. (Less distortion for same individual-different expression-same orientation and more distortion for different-individual-same expression-same orientation.)
c, Test face without grid.

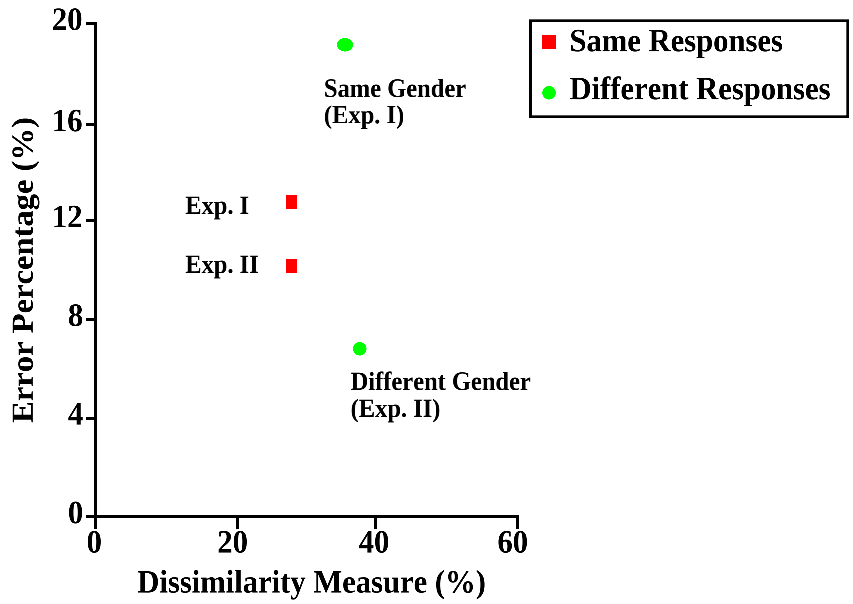
Results

There was virtually no difference in similarity of negative trial stimuli between same gender (Exp. I) and different gender (Exp. II) faces according to the SC model. Yet performance on negative trials in Exp. II was markedly facilitated for humans (by a 34 msec reduction in RTs and a 12.4% drop in error rates) by the introduction of gender difference on those trials.

Effect of Gender and Response Type on Dissimilarity and Reaction Time



Effect of Gender and Response Type on Dissimilarity and Error Rate



Conclusions

The effects of differences in orientation on the recognition of faces can largely be accounted for in terms of activation of representations based on simple cell similarity space. This representation can be well approximated by columns of Gabor-type kernels of multiple scales and orientations at specified locations.

The similarity space that can predict most of the variance due to differences in orientation, does not suffice for discrimination of emotional differences and gender. Small differences in the simple cell similarity space produced by variations in emotional expression or gender have disproportionaltely large effects on performance.

References:

Buhmann, J., Lades, M., & von der Malsburg, C. (1990). Size and Distortion Invariant Object Recognition by Hierarchical Graph Matching. In *International Conference on Neural Networks*, (pp. 411-416). San Diego.

Buhmann, J., Lange, J., von der Malsburg, C., Vorbrüggen, J. C., & Würtz, R. P. (1991). Object Recognition in the Dynamic Link Architecture: Parallel Implementation of a Transputer Network. In B. Kosko (Eds.), *Neural Networks for Signal Processing*, (pp. 121-159). Englewood Cliffs, NJ: Prentice Hall.

Guilford, J. P. (1956). Fundamental Statistics in Psychology and Education. In *McGraw-Hill Series in Psychology*, McGraw-Hill Book Company, Inc.

Lades, M., Vortbrüggen, J. C., Buhmann, J., Lange, J., von der Malsburg, C., Würtz, R. P., & Konen, W. (1993). Distortion Invariant Object Recognition in the Dynamic Link Architecture. *IEEE Transactions on Computers, 42*, 300-311.