Max-Planck-Institut
für biologische Kybernetik
Arbeitsgruppe Bülthoff
Spemannstraße 38 • 72076 Tübingen • Germany

Technical Report No. 21 September 26, 1995

Face Recognition Across Viewpoint

Alice J. O'Toole, Heinrich H. Bülthoff & Candice L. Walker

Abstract

In two experiments we examined the ability of human observers to recognize faces from novel viewpoints. Previous work has indicated that there are marked declines in recognition performance when observers learn a particular view of a face and are asked to recognize the face from a novel viewpoint. We replicate these findings and extend them in several ways. First, we replicate the well-known 3/4 view advantage for recognition and extend it to show that this advantage is stronger than would be expected simply due to the 3/4 view being the center of the learned views. In the second experiment, we found little evidence for advantageous transfer to a symmetric view of the other side of the face, in all cases, observers were much better at recognizing a face from the side learned. Third, we extended past results to explore the consistency of face recognizability for individual faces across different views and view transfer conditions. We found only a modest relationship between the recognizability of individual faces in the different view conditions. These data give insight into the organization of memory for faces and its stability across changes in viewpoint.

Many thanks are due to Niko Troje and Isabelle Bülthoff for the stimulus creation and processing. Alice O'Toole gratefully acknowledges support by the Alexander von Humboldt Stiftung and the hospitality of the Max-Planck Institut für biologische Kybernetik. Please direct all correspondence to A. J. O'Toole, School of Human Development, GR4.1, The University of Texas at Dallas, Richardson, TX 75083-0688, TEL: (214) 883-2486, Email: *otoole@utdallas.edu*

This document is available as `/pub/mpi-memos/TR-21.ps.Z` via anonymous ftp from `ftp.mpik-tueb.mpg.de` or from the World Wide Web, `http://www.mpik-tueb.mpg.de/projects/TechReport/list.html`.

1 Introduction

People are generally able to recognize familiar faces accurately from any of a variety of viewpoints. In the present study we address a number of issues related to this ability. Before proceeding, it is necessary to clarify the primary task used in this study and to distinguish it from other related, but different tasks. By face recognition, we mean the classification of a face as "known" or "unknown". We distinguish this from the task of face identification, which usually involves naming a face. This is a natural distinction in human memory. Often, we are absolutely certain that we recognize a face, but cannot identify it in any other way, e.g., retrieve a name or context of knowing.

Relating this to the object recognition literature, it is important to note that the term "object recognition" has been used to refer to tasks that draw on two quite different kinds of information about objects. "Object recognition" has frequently been used to refer to a basic level categorization task involving the classification of a particular object as an instance of a class of objects (e.g., "This object is a chair.")[1]. In this case, the task is to produce the correct category name (e.g., Bartram, 1974; Biederman & Gerhardstein, 1994, Exps. 1 & 2) when presented with some exemplar object. This is a useful psychological task, because in many cases the problem is to recognize an object for some general functional purpose (e.g., in searching for a pen, we are generally interested only in knowing that something is a pen, not in whether or not we "know" the pen or have used it before). For faces, this basic level categorization task is simply the determination that a particular object is a face, and is obviously only the first step.

Other studies have used the term object recognition to refer to tasks that involve explicitly distinguishing among, recognizing, or matching individual instances of objects within a particular category (cf., Tarr, 1989; block configurations; Bülthoff & Edelman, 1992; Edelman & Bülthoff, 1992; paper-clip and amoeboid-like objects, respectively; Biederman & Gerhardstein, 1995; Exps. 3, 4 &5; nonsense-objects[2], single volumes, and charm-bracelets). This is also a useful psychological task, because in many cases the problem is to recognize a particular object as "known" or "unknown", e.g., "that is my suitcase on the conveyer belt."

Both kinds of object recognition studies have addressed the question of "recognition" across viewpoint change, because a major part of both tasks is to recognize an object from whatever view of it you happen to have. An important practical difference, however, is that in the former case, you need to recognize the object (i.e., pen) among other dissimilar objects, but in the latter case, you need to recognize the object (i.e., your suitcase) among other quite similar objects (i.e., other suitcases on the airport conveyor belt). Additionally, successful completion of the former task requires an ability to extract the information that the particular object has in common with the object category (i.e., what this object has in common with all/most pens). In the latter case, one must extract information that makes the particular exemplar unique among other objects of that category (i.e., what makes your suitcase different from everybody else's suitcase).

Faces make a useful compare-and-contrast stimulus to the familiar and unfamiliar objects that have been used in previous work for two interconnected reasons. First, while both faces and objects need to be recognized in the above two ways (as the class of objects they represent and as individuals), the relative balance of the two tasks for faces and generic objects is different. For objects, when the latter type of individual exemplar recognition is required, it is usually limited to keeping track of a small number of individual exemplars. Thus, you usually need only to distinguish/remember a few individual suitcases from among all of the other suitcases in the world. For faces, we must be able to distinguish among and remember hundreds, if not thousands, of people by their faces. Second, given the number of faces we know individually, and the extensive experience we have in getting to know new faces, learning to recognize new faces is something most of us can do with very little difficulty/training. It is also interesting to note that for other-race faces we often lack expertise equivalent to that which we have with faces of our own race. In this case, learning to recognize new faces is apparently much more difficult, even though there is good reason to think that faces of all races are equally distinguishable (see Shapiro and Penrod, 1986 for a review).

In the present study, we examined the ability of observers to recognize faces across viewpoint changes. Additionally, we looked at the consistency of the recognizability of individual faces in

[1]Biederman & Gehardstein (1994) prefer the term "entry-level" categorization to basic level.

[2]It is not clear that this is reasonably thought of as a class of objects, though they are nameless and have the appearance of wooden toys.

Figure 1: Recognition accuracy data as a function of learn and test pose in Experiment 1.

Several points are worth noting from the data. First, the strong interaction between learn and test pose is an indication that there is a cost in recognizing faces from a novel view. Second, as found in previous work (cf., e.g., Bruce & Valentine, 1987, for a review), the three-quarter view was the best view to learn. Part of this might be due to it being the central view of the three tested. However, this would not seem to account for the entire effect, since the recognition of three-quarter face views in the no view-change condition, ($d' = 1.55$), was markedly better than the analogous no view-change conditions for profile ($d' = 1.10$) and full face ($d' = 1.16$) views. Finally, for the most part, performance on learn-test conditions and their inverses, (e.g., learning full-face $\rightarrow$ testing three-quarter and learning three-quarter face $\rightarrow$ testing full), were comparable with one exception. Learning three-quarter face $\rightarrow$ testing profile resulted much better performance than its inverse. One interpretation of this finding can be made as follows. It is possible that the profile

No-transfer Condition

View combinations	*Correlation Coefficients*
Learned Full & Tested Full correlated with Learned 3/4 & Tested 3/4	.35 ($p < .01$) explained variance $r^2 = .12$
Learned Full & Tested Full correlated with Learned Profile & Tested Profile	.41 ($p < .01$) explained variance $r^2 = .17$
Learned Profile & Tested Profile correlated with Learned 3/4 & Tested 3/4	.34 ($p < .01$) explained variance $r^2 = .115$

Table 1: Correlations between the recognizability of faces in the no-transfer conditions.

Transfer-Condition: full-face learned

View combinations	*Correlation Coefficients*
Learned Full & Tested Full correlated with Learned Full & Tested 3/4	.25 ($p < .05$) explained variance $r^2 = .06$
Learned Full & Tested Full correlated with Learned Full & Tested Profile	.18 (*ns*) explained variance $r^2 = .03$

Transfer-Condition: 3/4-face learned

View combinations	*Correlation Coefficients*
Learned 3/4 & Tested 3/4 correlated with Learned 3/4 & Tested Full	.36 ($p < .01$) explained variance $r^2 = .15$
Learned 3/4 & Tested 3/4 correlated with Learned 3/4 & Tested Profile	.26 ($p < .05$) explained variance $r^2 = .07$

Transfer-Condition: profile-face learned

View combinations	*Correlation Coefficients*
Learned Profile & Tested Profile correlated with Learned Profile & Tested Full	.18 (*ns*) explained variance $r^2 = .03$
Learned Profile & Tested Profile correlated with Learned Profile & Tested 3/4	.37 ($p < .01$) explained variance $r^2 = .14$

Table 2: Correlations between the recognizability of faces in transfer conditions when full-face was learned (top), 3/4-face was learned (center), profile-face was learned (bottom).

view, while not a bad view for recognition (as evidenced by the fact that its *d'* in the no-transfer case), is not a good view to transfer from.

Stimulus Measures. The recognizability of each face in each pose transfer condition was assessed by collapsing data across observers and computing a *d'* for each face in each transfer condition. Two points are interesting. First, while statistically significant, the inter-correlations of face recognizability among the no-transfer conditions were quite modest, explaining only 17 percent of the variance in the very best case (see Table 1). A similar situation is seen for the transfer conditions (see Table 2). Thus, the difficulty of recognizing individual faces in different views and view transfer conditions would seem to be only minimally related.

3 Experiment 2

The purpose of this experiment was to look at the role of view symmetry in recognition transfer. The design of this experiment was similar to Experiment 1 with only the following changes. Observers learned only three-quarter and profile face views from one side of the face and were tested for recognition using full, three-quarter, and profile face views of the other side of the face.

Observers. Thirty-six volunteers roughly half male and half female, between the ages of 18 and approximately 45 years old were tested.

3.1 Results & Discussion

Again the *d'* data were submitted to an ANOVA. No main effect of learn pose, $F(2, 34) = 3.01$, $p > .05$, or test pose was found, $F(2, 68) < 1$. There was a significant interaction between learn and test

Figure 2: Recognition accuracy data as a function of learn and test pose in Experiment 2. In all cases (except the full-face test cases), the learned pose was on one side of the face and the test pose was on the other side. Arrows indicate the means for the two duplicate control conditions in Exp. 1, indicating that the large accuracy cost with symmetric views cannot be attributed to less motivated, or accurate observers in this second study.

This result is interesting to compare to data by Troje and Bülthoff (1995), who found evidence for reasonably good transfer to a symmetric view using these stimuli without texture maps (i.e., pure shaded busts of the heads) in conjunction with a perceptual matching task for successively presented faces varying in view. They also found a trend, though not significant, for a similar symmetric transfer with identical stimuli to those used here, again in a perceptual matching task. The contrast may indicate an important difference in the strategies available to observers when they are forced to remember a large number of faces when views are varied, versus when they are required only to discriminate pairs of individual exemplar faces. A second very important factor might be the quality of the match in symmetric views between texture-based and image-covered face views, which these data would indicate favors better head symmetry for the three-dimensional structure information than for the texture maps.

4 Summary

These data replicate the findings of past studies and also extend that work to begin to look at the relationship between the recognizability of individual faces in different transfer conditions. We think this latter approach has the potential to shed light on how related representations of individual faces can be when they are created from only a single view. An early indication of the utility of this approach was the finding that relationships between the recognizability of faces in the different conditions, explained only small fraction of the variance on the task. This suggests that the quality and/or accessibility of the information used by observers to recognize the faces (i.e., what makes them unique among the set of faces) changes markedly with view. Thus, faces unusual from one view may be quite typical from another, indicating that the view, rather that surface it suggests may dominate in the face code.

An additional surprising result was the difficulty observers had in recognizing faces across symmetric view changes. While from an information point of view, it is clear that nearly any computational model would be successful at a recognition task under this transformation, human observers seem to be much less successful. The problem of symmetric transfer may be added to a list of other equally easy problems for computational models of faces that people seem to fail at. Examples include the well-known difficulties people have in recogning faces in the photographic negative and in recognizing upside-down faces. These simple failures may be informative about the nature of the human face code.

References

[1] Bartram, D. J. (1974). The role of visual and semantic codes in object naming. *Cognitive Psychology, 6*, 325-356.

[2] Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*, 115-147.

[3] Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experiment Psychology: Human Perception and Performance, 19*(6), 1162-1182.

[4] Bülthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proc Natl Acad Sci USA, 89*, 60-64.

[5] Edelman, S., & Bülthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research, 32*(12), 2385-2400.

[6] Edelman, S., Bülthoff, H. H. & Weinshall, D. (1989). Stimulus familiarity determines recognition strategy for novel 3D objects. it AI Lab. Tech Report 1138, Cambridge: MA.

[7] Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review, 99*(3), 480-517.

[8] Shapiro, P. N., & Penrod, S. D. (1986). Meta-analysis of face identification studies. *Psychological Bulletin, 100*, 139-156.

[9] Tarr, M. J. (1989). *Orientation dependence in three-dimensional object recognition.* Unpublished doctoral dissertation. Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA.

[10] Tarr, M. J., & Pinker, S. (1991). Orientation-dependent mechanisms in shape recognition: further issues. *Psychological Science, 2*(32), 207-209.

[11] Tarr, M. J. & Bülthoff, H. H. (1995) Is human object recognition better described by geon-structural-descriptions of by multiple views? *Journal of Experimental Psychology: Human Perception and Performance.*

[12] Troje, N. & Bülthoff, H. H. (in press). *Vision Research.*