TR - H - 117

Viewpoint Dependence in Face Recognition

Philippe G. Schyns
Harold Hill

1995. 1. 5

ATR人間情報通信研究所
〒619-02　京都府相楽郡精華町光台2-2　☎ 0774-95-1011

ATR Human Information Processing Research Laboratories
2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02 Japan
Telephone: +81-774-95-1011
Facsimile: +81-774-95-1008

VIEWPOINT DEPENDENCE IN FACE RECOGNITION

Philippe G. Schyns and Harold Hill

ATR Human Information Processing Research Laboratories
2-2 Hikari-dai, Seika-cho, Soraku-gun
Kyoto 619-02, JAPAN

schyns@hip.atr.co.jp hill@hip.atr.co.jp

ABSTRACT

Face recognition has attracted the attention of vision researchers for the observation that although most faces are very much alike, people discriminate between them very well. Even though little is known about the mechanisms of face recognition, a recurrent phenomenon of general object recognition (viewpoint dependency) could illuminate the way faces are recognized. In this paper, we investigate in detail the conditions for viewpoint-dependent face recognition. The first experiment tested whether a particular view of a face was better and more reliably recognized than other views. Results indicate that when all views were shown (randomly or in an animated sequence), all were equally well recognized. The second experiment tested generalization from single views of a face. Different learning views were found to produce different patterns of generalization. For full-face, performance fell off with increasing angle of rotation, while for three-quarter there was a peak for the opposite three-quarter. Profile performances dropped off steeply and there was no recovery for the opposite profile. Results are discussed in the context of recent psychological, computational and physiological accounts of viewpoint-dependent object recognition.

Object recognition and categorization research normally assume that within-class discriminations are more difficult than between-class discriminations. For example, while people would experience little or no difficulty distinguishing a car from a tree, it would be comparatively more difficult to distinguish among brands of cars or species of trees. Researchers explain this discrepancy by the nature of the comparisons being made: Within-class judgments must distinguish objects generally more similar than between-class judgments. Face recognition is thought to be a within-class discrimination that can be done quickly and accurately, at least under favorable conditions. Although faces are mostly very similar--they share roughly the same overall shape, configuration, textures and features--we are able to recognize individual examples with amazing efficiency. Of course mistakes are made, as the fallibility of eyewitness testimony reveals (Loftus, 1979), but the human system remains of utmost interest to psychologists, computer scientists and neurophysiologists, as a successful system for making within-class discriminations.

In the real-world, objects are seen from a variety of viewing conditions. Changes of viewpoint pose a serious challenge to object recognition mechanisms because of the large changes they produce in the overall appearance of objects. However, the effect of viewpoint on recognition performance is probably not absolute, but instead relative to the type of object considered and to the nature of the task. To illustrate, consider the simple task of determining whether a particular object is a face. In these circumstances, it is quite likely that any view of a face could provide sufficient information to confidently reach a decision. Imagine another task in which you must decide whether the same face is X. In this case, it is more important that you get a good view of the face, to insure that you gather sufficient discriminating information to identify the input as X. Finally, imagine you must recognize X at the airport terminal, with only of one view of his face, for example from a picture. This latter task is particularly difficult because X's real face can appear in a pose quite different from the picture you have in your hand, but this picture is your only way to identify X. Under these constraints, which picture would be the best?

This paper studies which viewpoints are best for face recognition. Viewpoint preference has recently come under the close scrutiny of object recognition researchers. Recent studies in experimental psychology (Palmer, Rosch & Chase, 1981; Bülthoff & Edelman, 1992; Tarr & Pinker, 1989; among others) and neurophysiology (Logothetis, Paul, Bülthoff, Poggio, 1944; Perrett, Mistlin, & Chitty, 1989; among others) suggest that object recognition is viewpoint-dependent and that few views are required to represent an object in memory. Just exactly which views, or more precisely which view-specific information is used for the particular task of identifying faces is the focus of this research.

Phenomenological evidence for viewpoint-dependent recognition. In their seminal study, Palmer, Rosch and Chase (1981) demonstrated that the relationship between the observer and the object was a major determinant of recognition performance (see also Rock & Di Vita, 1984). The authors found that human subjects consistently labeled certain views of common objects as "better" than other views. Furthermore, a naming task revealed that subjects' reaction time was quicker when the input object was presented from a *canonical*, or "best" perspective. The degree of misorientation between the canonical perspective and other perspectives was responsible for monotonically increasing reaction times. Evidence of such "viewpoint-dependent recognition" were reported under a variety of experimental conditions, using familiar, unfamiliar, realistic or comparatively simpler stimuli (Bülthoff & Edelman, 1992; Edelman & Bülthoff, 1992; Farah, Rochlin & Klein, 1994; Rock & Di Vita, 1984; Tarr & Pinker, 1989, 1990).

Much of the face recognition literature has concentrated on the full-face view, emphasizing its two-dimensional configuration, and ignoring other views. For example, the well-known effect of inversion on face recognition shows that it is not a view-independent process (Yin 1969, Valentine 1988). Moreover, rotations in depth between study and test, for example from full-face to profile views, have also been shown to produce decrements in recognition performance (Bruce, 1982). There is also evidence for a decrement associated with the profile view (e.g., Bruce, Valentine & Baddeley, 1987). With respect to canonical

views, paintings and pictures of faces are often represented in the 3/4 view, and research has attempted to establish if this view is best for recognition. While advantages have been found for the 3/4 view (e.g., Krouse, 1981; Logie, Baddeley & Woodhead, 1987) these may be limited to the task of matching unfamiliar faces rather than reflecting more fundamental properties of their representations (Bruce et al., 1987).

In summary, many recognition studies of objects and faces suggest that viewpoint affects recognition performances (though, see Biederman, 1987). Although further research will probably reveal that conditions of viewpoint dependence are not absolute, but instead relative to the requirements of different object categorization tasks, there is little doubts that viewpoint dependence is an important characteristic of recognizing everyday objects (see Tarr & Bülthoff, in press, for a review).

Viewpoint-dependent processes and representations. Viewpoint-dependent recognition was interpreted in Tarr and Pinker (1989, 1990) as evidence that objects are represented in memory with a collection of views depicting the appearance of objects from specific viewpoints, and that recognition first normalizes the input object to the stored views by a process of mental rotation (see also Palmer, 1983). An example of this normalization to stored models is provided in Ullman's (1989) *alignment* research. Alignment is a three-stage process. In the first stage, a correspondence is established between a few key features of the input view and a stored representation. In the second stage, a 3D transformation (derived from the correspondence between key features) is applied to align the input and the stored model. In the third stage, a matching process finds which stored model best fits with the input. In Ullman and Basri (1991), the explicit alignment stage is not necessary: The input view is compared to the views obtained by interpolating from a small number of views stored in memory. Under the specific condition of orthographic projection, the 2D coordinates of an object's points can be rewritten as the linear combinations of a small basis of 2D views of the same object.

Poggio and his collaborators provided a general formalization of the multiple-view approach to object representation (Poggio, 1990; Poggio & Edelman, 1990). Poggio and Edelman (1990) proposed a theory of object recognition in which a 3D object is represented with two 2D views plus an interpolation mechanism. The authors formulate the problem of recognizing an object from all viewpoints in the general approach of approximating a smooth hypersurface from sparse data points (two views of the object). This approach was successfully implemented with a two-layer network in which the units of the first layer are each centered on a stored view. They act as "viewer-centered neuron" by responding maximally to the stored view (more precisely, their response is a Gaussian function of the Euclidean distance between the stored and the input views). The second stage linearly combines the responses of the viewer-centered neurons and fires if the linear combination is above a criterion threshold. To circumvent the difficulties caused by occlusions and opacities, Poggio and Edelman suggest to segment the object's viewpoint space into a few "stable" views revealing the important features of the object (see Koenderink & van Doorn, 1979).

The "multiple-view approach" finds structural support in neurophysiology. For example, Perrett and his collaborators discovered cells of the macaque superior temporal sulcus (STS) which were preferentially tuned to respond to specific views of a head (Perrett, Smith, Potter, Mistlin, Head, Milner & Jeeves, 1985). Most of the cells were viewer-centered, responding unimodally to one view (either the frontal, the two profiles, or the two back views); few cells were tuned to other views of the 360 degree range (Perrett et al., 1985, 1989; Perrett, Oram, Harries, Bevan, Benson & Thomas, 1991). In a related vein, Logothetis, Pauls, Bülthoff and Poggio (1992, 1994) trained monkeys on a recognition task of new wire objects. It was discovered that cells in IT cortex had a viewpoint-dependent response profile for some of the trained views. That is, they were tuned to the trained view with a gradual decrement of response with degree of misorientation.

A recent computational extension of the multiple view theory has investigated the possibility that object properties such as bilateral symmetry are exploited to minimize the number of effectively stored views, and facilitate generalization from single views (Vetter, Poggio & Bülthoff, 1994). For objects, it has been demonstrated that "virtual views" can be generated from a single view so long as it is not head on the axis of symmetry and that pairs of symmetric feature points can be identified on the object (such as, e.g., two eyes). These views are obtained by image plane transformations and do not require knowledge of the three-dimensional structure of the object. They are not mirror images but legal views of the object as would result from a rotation in depth. Experimentally, it has been suggested that the virtual views of wire objects are better recognized than other unseen views (Vetter et al., 1994).

In summary, the idea that objects are represented in memory with a collection of views seems to receive support from psychology, neurophysiology and computation. The major strength of this approach is sparse representations, because generalization to all views can in principle be achieved with very few views The difficulty is to specify among the set of all possible views the subset giving rise to the best generalization performances, for a given object category, and for a particular recognition task. As explained earlier, there may not be a unique, canonical and task-independent view-based representation of an object. Instead, the multiple-view approach leaves open the possibility that the selection of the views representing an object is sensitive to the functional role of the view with respect to the possible categorizations of the object (see, e.g., Schyns & Murphy, 1994).

Phenomenology and representations. Viewpoint-dependent recognition is characterized by faster recognition times and fewer errors in recognition/generalization. The multiple-view theory suggests specific representations and processes to account for the phenomenon--i.e., by collections of 2D views and interpolation. There are at least two possible interpretations of viewpoint-dependent recognition. A strong interpretation establishes a direct mapping between viewpoint-dependence and a representational scheme by

suggesting, for example, that a few two-dimensional views of an object are stored in memory for later recognition (Bülthoff & Edelman, 1992). Another, perhaps less adventurous interpretation proposes a mapping between viewpoint-dependence and the information transmitted by preferred views. Viewpoint-dependence would then be equivalent to "information-dependence" in the sense that some views transmit diagnostic features for specific classification tasks. Changing the conditions of the task would then change the preferred views of a given object, because it would not be the view itself, but the features it represents, that are responsible for viewpoint dependence. Note that this latter interpretation is compatible with different types of representations, collections of 2D views and others. It is not clear that experiments on viewpoint dependence are made sufficiently constrained to distinguish between information and detailed representations of the information.

In this paper, we investigate which, if any, particular view of a face is best used for face identity decisions--in particular, for recognizing a face learned from a single view. Specifically, we investigate whether any view or views are inherently easier to recognize. In all experiments full-face and both left and right profile and three-quarter views were used. In Experiment 1 all views were presented an equal amount and we tested for subsequent differences in recognition performance. In Experiment 2 we additionally investigated whether any view led to better encoding by testing generalization from single views.

EXPERIMENT 1

Experiment 1 was intended as a simple test of whether any view of a face is especially well recognized, or canonical (Palmer et al., 1981). For example, we might expect a 3/4 advantage as reported in Krouse (1981). To look for evidence of a viewpoint preference, recognition performance was tested as a function of view when the complete set of views had been previously learned (-90, -45, 0, 45, 90 degree rotation in depth--see Figure 1). The experiment also provided baseline information for the second experiment which tested generalization from single views.

Two groups of subjects were run, one of which saw all the views in sequence, producing an animation of the head apparently rotating in depth. The other group saw the views that composed the sequence, but order of presentation was randomized, disrupting the impression of rotation. We expected better performance in the animation condition and also thought that the coherent motion of the animation might affect the pattern of recognition performance: The animated sequence might give a good overall impression of overall shape, resulting in all views being well recognized, while a preference for a particular view might be brought out in the more difficult random condition.

Methods.

The experiments reported in this paper investigate the recognition of three-dimensional shape across changes in viewpoint. Faces, however, also contain features such as hair color, hairstyle, texture or color of the skin, type and size of eyebrows that could be relatively invariant under rotations in depth. When they are diagnostic (and this is essentially relative to the targets and distractors composing the stimulus set), these cues can be used to achieve viewpoint invariant face recognition (e.g., Bruce et al., 1987). To provide a better control of the available information, all faces were presented as gray-level images of 3D shape models, allowing shape perception to be investigated in isolation (see Figure 1). That is, obvious viewpoint invariant features were removed from the stimuli and we only tested shape-based face recognition.

INSERT FIGURE 1 AROUND HERE

Subjects. Subjects were 24 ATR employees with normal or corrected vision who volunteered their time to participate to the experiment. Equal numbers of subjects were assigned to each condition. All subjects were unfamiliar with the faces used, allowing a control of subjects' exposure to the stimuli.

Stimuli. Experiment 1 and Experiment 2 used the same set of stimuli. Stimuli were 256 gray-level views of 3D Japanese face models presented on the monitor of a Silicon

Graphics workstation. There were 30 different face models with equal numbers of males and females. Twenty faces were used as targets and ten as distractors. Face data were Cyberware laser-scanned three-dimensional coordinates of real faces. Each head was reconstructed by approximating the face data with a bicubic BSpline surface. Laser head scans have many measurement errors around the hair and so we trimmed them to produce 3D masks. There were five views of each face: -90, -45, 0, 45, 90 degree rotations in depth (0 degree is the frontal, or full-face view). Faces were illuminated by a directional light source located 30 degrees above the line of sight with a relative intensity of 1 and an ambient light source from directly above with a relative intensity of .3. A Phong Shading model with a matte mid-gray reflectance was used as shader. Stimulus size was of 300 x 300 pixels.

Design. Test View was the single within subjects factor which had levels Left Profile (LPR), Left Three-Quarter (LTQ), Full-face, Right Three-quarter (RTQ), and Right Profile (RPR). Sequence was a between subjects factor with two levels (animated or random). We measured the average of hits and correct rejections although response times were also recorded. Comparisons were planned to test for differences between levels of test view. The order of trials was randomized for each subject as was the pairing of distractors to targets.

Procedure. Subjects were told that they would be presented with faces to learn and that their task was to decide (as quickly and as accurately as possible) for each of the following two faces whether it was the same or different from the one that they had just seen. The experiment was composed of 20 trials consisting of a learning stage and a testing stage. In the learning stage, subjects were presented with a single target face to learn. In the animated group, the target face rotated through the five views, once clockwise and once counterclockwise, starting randomly from one of the five possible views. Each view was shown twice and there were thus ten possible animation sequences. Apparent rotation was produced by showing the face views in rapid succession with a frame rate of 100 ms per view making a total presentation time of 1 sec per face. In the random group, the same views

were presented the same number of times and for the same durations but in a random order. The learning stage for each target face was immediately followed by a testing stage. This consisted of presenting two faces, one at a time, in the same orientation. Unknown to subjects, one was the target face and the other the distractor face. Order of presentation of target and distractor testing faces was randomized. For each test face subjects had to decide whether or not it was a view of the target face by pressing the appropriate response-key on the Silicon Graphics' keyboard. Each of the five views was tested four times giving a total of 20 trials for each subject. Each learning face was only tested once, ensuring that no learning occurred during the testing stage.

Results and discussion. A 5(Test View) x 2(Sequence) ANOVA showed a main effect of sequence, $F(1, 22) = 7.5, p < .01$. Subjects who saw the animation sequence during the learning stage performed significantly better than subjects who saw the random sequence (86% and 78%, respectively). The results for the animated group are shown in Figure 2. Contrary to any particular view being preferred for recognition there was no effect of Test View $F(4, 88) = 1$, n.s. or any interaction $F(4, 88) = .7$, n.s.

The results of this experiment showed that presenting the learning views in sequence did facilitate recognition compared to when they were presented randomly. It also showed that all views were equally well recognized when they had been presented the same amount. Thus, it appears that under the precise conditions of this experiment no view is canonical, or inherently easier to recognize. This was true even when views were presented randomly which might not have given such a good impression of overall shape, and therefore could have induced the system to learn its "preferred" views.

EXPERIMENT 2

Experiment 2 sought to test generalization from single views. Testing was the same as in Experiment 1 but only one view was shown during the learning stage. It was assumed that any differences in the recognition of the test views in this experiment must be a function

of generalization from the learning view as Experiment 1 had shown that the test views did not produce inherently different performance.

It is possible that although all views may be equally well recognized they may not be equally good for learning. For example it has previously been reported that learning the three-quarter view leads to better recognition subsequently (Krouse, 1981). We also expected that there would be an interaction between learning view and test view on this experiment if only because the view which is learned would be expected to be best recognized (cf. Bruce, 1982). Such an interaction would be exaggerated if different views also showed different patterns of generalization. For example, left and right three-quarter and profile views are almost mirror images because of the vertical symmetry of the face and so might be expected to generalize well to each other. The full-face view is its own mirror image and so would not generalize well to other views on this basis. Also the full-face view may contain less information about shape important for generalizing to other views whose projections are a function of that shape. Thus we expected that performance would be best for the view that had been shown at learning, resulting in an interaction between test view and learned view. It was also possible that certain views would be better learned and show different patterns of generalization.

Methods.

Procedural details were the same as for Experiment 1 except for the learning phase. In this experiment subjects only saw one view at the learning stage, presented for the same total time of 1 s as the animation. Learning view was a between subjects factor with 5 groups of subjects, one for each view. Five subjects learned each of left profile, left three-quarter, right three-quarter and right profile and ten subjects learned full-face. This was to ensure equal group sizes after collapsing across left and right pairs of views, having tested that these were not significantly different. Comparisons were planned to test if symmetric views were better recognized than other unseen views.

Results and Discussion. The average hits and correct rejections in this experiment were 76%, worse than the animated group in the previous experiment but at the same level as the random group. A comparison of left against right views showed no difference, $t(11) = 0.43$, ns., and so results were collapsed across left and right views as planned (see Figure 2). Training views were distinguished as being the Same or Opposite side to the learned view.

INSERT FIGURE 2 AROUND HERE

As can be seen from Figure 2 the pattern of generalization was very different for the three views. In all cases, performance was best when the learning view was the view shown at test, but generalization to unseen views was dependent on learning view. For the full-face view there was an inverted U shape pattern of generalization with performance falling off with increasing angle of rotation. For the three-quarter view the most noticeable feature is a peak for the opposite three-quarter. For the profile view there was no such peak, generalization being equally bad to all unseen views. A 3(Learning View) x 5(Testing View) analysis of variance gave a Training View x Testing View interaction, $F(8, 132) = 3.2$, $p < .01$, confirming that the pattern of generalization to test views depended on training view. Analysis of simple main effects showed effects of learning view for learned Full-face $F(4, 132) = 4.1$, $p < .01$ and Three-quarter $F(4, 132) = 4.0$, $p < .01$ but not for Profile, $F(4, 132) = 1.8$, n.s. Planned comparisons showed that the three-quarter view learning view group recognized the opposite three-quarter significantly better than the other unseen views $t(11) = 2$, $p < .05$ but there was no such advantage for the opposite profile $t(11) = .2$, n.s.

These results indicate that learning a single view of a face produces viewpoint-dependent generalization to other views. In all conditions, performances dropped off sharply from the learned view. In the 3/4 condition, however, the opposite 3/4 was better recognized than the other unseen views. We discuss implications of this result in the General Discussion.

GENERAL DISCUSSION

This paper investigated viewpoint-dependent face recognition in the context of findings and theories about view dependent object recognition. Experiment 1 showed that, of the views used, none was inherently easier to recognize. That no view is canonical contrasts with reports in the object recognition literature of common objects having specific views that lead to better recognition performance (Palmer et al, 1981). It should be noted that we showed only a limited range of views of the face, all of which are fairly familiar, and advantages might have been found if more unusual and less informative views like upside-down faces or the back of the head had been included in a face identification task. Again, this stresses the importance of the task and of availability of information for the phenomenon of viewpoint dependence.

Animated presentation might have lead to good generalization because structure from motion cues gave a good overall impression of the shape of the face not available from the random sequence. If this was the case the comparatively more difficult random condition could have elicited preferences for specific views. However, the results showed that all views were still equally well recognized.

Experiment 2 demonstrated that learning a single view produced viewpoint-dependent generalization. First, face recognition is trivially viewpoint-dependent in that the view that has been learned will be best recognized. In general, our results indicate that recognition performances drop off quite sharply with angle of rotation from the learned view. These results on face recognition are compatible with the general phenomenology of viewpoint dependence observed for other objects (e.g., Tarr & Pinker, 1989; Bülthoff & Edelman, 1991). Note, however, that the latter results were obtained in conditions where more than one view of an object are experienced, while our results show no viewpoint-preference in this condition. There is one notable exception to the drop off in performances: When subjects learned a 3/4 view, the symmetric 3/4 view was recognized equivalently well. But

symmetry does not apply to the profile view. That is, subjects learning the profile did not recognize equivalently well the symmetric profile.

There is an interesting analogy between these results and Poggio and Vetter's (1992) symmetry argument. We saw earlier that an important condition of Poggio and Vetter's (1992) theory is that pairs of symmetric points must be identifiable in the single view of an object to generate a virtual view (e.g., two eyes, two corners of the mouth, and so forth). Occlusions may hide important symmetric feature points of the pairs. The 3/4 view is only partially occluded, and so subjects could find symmetric feature points to generate a virtual view (the reflection of the 3/4 view they learned) and use two views to generalize from. The profile, however, is fully occluded (with respect to the other profile) and so falls outside the conditions of application of the theory. Incidentally, although a mirror image could perfectly well be generated from a profile view, subjects did not seem to do so (as is indicated by the absence of a peak on the other profile of their generalization patterns). We do not claim that our results confirm Poggio and Vetter's (1992) theory, but they do at least provide suggestive evidence that symmetry is being used in specific circumstances, which in our experiment are equivalent to those specified in Poggio and Vetter (1992) (see also Vetter et al., 1994). Note that subjects might not actually generate virtual views, but instead use features present in the single view that are invariant under mirror symmetry.

Our research showed that the 3/4 view was special because it was the only view giving rise to less viewpoint-dependent recognition. An important aspect of future research should be to understand which information, or which facial features are present in the 3/4 view that are invariant for mirror symmetry. This may be more difficult than it initially appears, because it is not clear whether isolated features, or feature configurations are what actually matters, from a psychological viewpoint. In any case, our experiments suggest that if you want to be identified from a single picture, give a 3/4 view of your face. If you prefer anonymoucy, profile and full-face views are more appropriate choices.

REFERENCES

Bruce, V. (1982). Changing faces: Visual and nonvisual coding processes in face recognition. *British Journal of Psychology*, **73**, 105-116.

Bruce, V., Valentine, T., & Baddeley (1987). The basis of the 3/4 view advantage in face recognition. *Applied Cognitive Psychology*, **1**, 109-120.

Bülthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view theory of object recognition. *Proceedings of the National Academy of Science USA*, **89**, 60-64.

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

Ellis,H.D., Shepherd, J.W., & Davies,G.M. (1979). An investigation into the use of the Photofit technique for recalling faces. *British Journal of Psychology*, **66**, 29-37.

Farah, M. J., Rochlin, R., & Klein, K. L. (1994). Orientation invariance and geometric primitives. *Cognitive Science*, in press.

Koenderink, J. J., & van Doorn, A. J. (1979). The internal representation of solid shape with respect to vision. *Biological Cybernetics*, **32**, 211-216.

Krouse, F.L. (1981). Effects of pose, pose change, and delay on face recognition performance. *Journal of Applied Psychology*, **66**, 651-654.

Loftus, E. F. (1979). *Eyewitness testimony.* Cambridge, MA: Harvard University Press.

Logie, R.H., Baddeley, A.D., & Woodhead, M.M. (1987). Face recognition, pose, and ecological validity. *Applied Cognitive Psychology*, **1**, 53-69.

Logothetis, N. K., Pauls, J., Bülthoff, H. H., & Poggio, T. (1992). Evidence for recognition based on interpolation among 2D views of objects in monkeys. *Investigative Ophtalmological Vision Science Supplement*, **34**, 1132.

Logothetis, N. K., Pauls, J., Bülthoff, H. H., & Poggio, T. (1994). Viewpoint-dependent object recognition by monkeys. *Current Biology*, **4**, 401-414.

Palmer, S., Rosch, E., & Chase, P. (1981). Canonical perspective and the perception of objects. In J. Long & A. Baddeley (Eds.), *Attention and Performances IX*. Hillsdale, NJ: Lawrence Erlbaum.

Palmer, S. (1983). The psychology of perceptual organization: A transformational approach. In J. Beck, B Hope, & A. Rosenfeld (Eds.), *Human and machine vision*. Academic Press: NY.

Perrett, D. I., Mistlin, A. J, & Chitty, A. J. (1989). Visual neurons responsive to faces. *Trends in Neuroscience,* **10**, 358-364.

Perrett, D. I., Oram, M. W., Harries, M. H., Bevan, R., Benson, P. J., & Thomas, S. (1991). Viewer-centered and object-centered coding of heads in the macaque temporal cortex. *Experimental Brain Research*, **86**, 159-173.

Perrett, D. I., Smith, P. A. J., Potter D. D., Milstin, A. J., Head, A. S., Milner A. D. & Jeeves, M. A. (1985). Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proceedings of the Royal Society of London*, **B233**, 293-317.

Poggio, T. (1990). A theory of how the brain might work. In *Cold Spring Harbor Symposium on Quantitative Biology*, **55**, 899-910. Cold Spring Harbor Laboratory Press.

Poggio, T., & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature*, **343**, 263-266.

Poggio, T., & Vetter, T. (1992). *Recognition and structure from one 2D model-view: Observations on prototypes, object classes, and symmetries.* AI MEMO # 1347, Artificial Intelligence Laboratory, MIT: Cambridge, MA.

Rock, I., & Di Vita, J. (1987). A case of viewer-centered object representation. *Cognitive Psychology*, **19**, 280-293.

Schyns, P. G., & Murphy, G. L. (1994). The ontogeny of part representation in object concepts. In Medin (Ed.). *The Psychology of Learning and Motivation*, **31**, 301-354.

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

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, **21**, 233-282.

Tarr, M. J., & Pinker, S. (1990). When does human object recognition use a viewer-centered reference frame? *Psychological Science*, **1**, 253-256.

Ullman, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition*, **32**, 193-254.

Ullman, S. & Basri, R. (1991). Recognition by linear combinations of models. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, **13**, 992-1005.

Valentine, T. (1988). Upside-down faces: a review of the effects of inversion on face recognition. *British Journal of Psychology*, **61**, 471-491.

Vetter, T., Poggio, T., & Bülthoff, H. H. (1994). The importance of symmetry and virtual views in three-dimensional object recognition. *Current Biology*, **4**, 18-23.

Yin, R. (1969). Looking at upside-down faces. *Journal of Experimental Psychology,* **81**, 141-145.

FOOTNOTES

Title note: This research describes a full-fledged version of the pilot experiments presented in Schyns and Bülthoff (1994), MIT AI MEMO #1432 and MIT CBCL Paper #81, which was also presented at the XVI Meeting of the Cognitive Science Society, Atlanta. The experiments were done within ATR Human Information Processing Laboratories, when Philippe Schyns was an invited scientist.

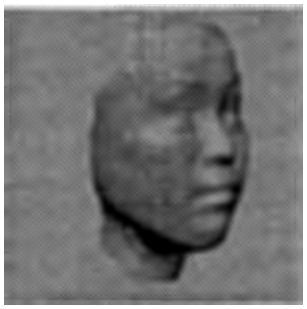
FIGURE CAPTIONS

Figure 1: Figure 1 shows examples of the stimuli used in Experiment 1 and 2. The top row shows the five views used in both experiments, left profile, left 3/4, full-face, right 3/4 and right profile. In Experiment 1 all views were presented in the learning stage either in sequence or randomly. In the second experiment each group of subjects were trained on only one of the five views. The testing stage, shown below, was the same for both experiments. A target and a distractor face shown in the same view were presented sequentially with order randomized and subjects were asked to respond "Yes" if they thought it was the learned face and "No" otherwise.

Figure 2: Figure 2 shows the results of Experiment 2 and the animated condition of Experiment 1, for comparison. The learning conditions for Experiment 2, FF, TQ and PR, are shown collapsed across left and right views. The conditions of test view are SPR (Same PRofile), STQ (Same Three-Quarter), FF, (Full-Face), OTQ (Other Three-Quarter), and OPR (Other PRofile). For the full-face and animated condition "Same" views correspond to left views. Error bars show standard errors.

LEARNING STAGE

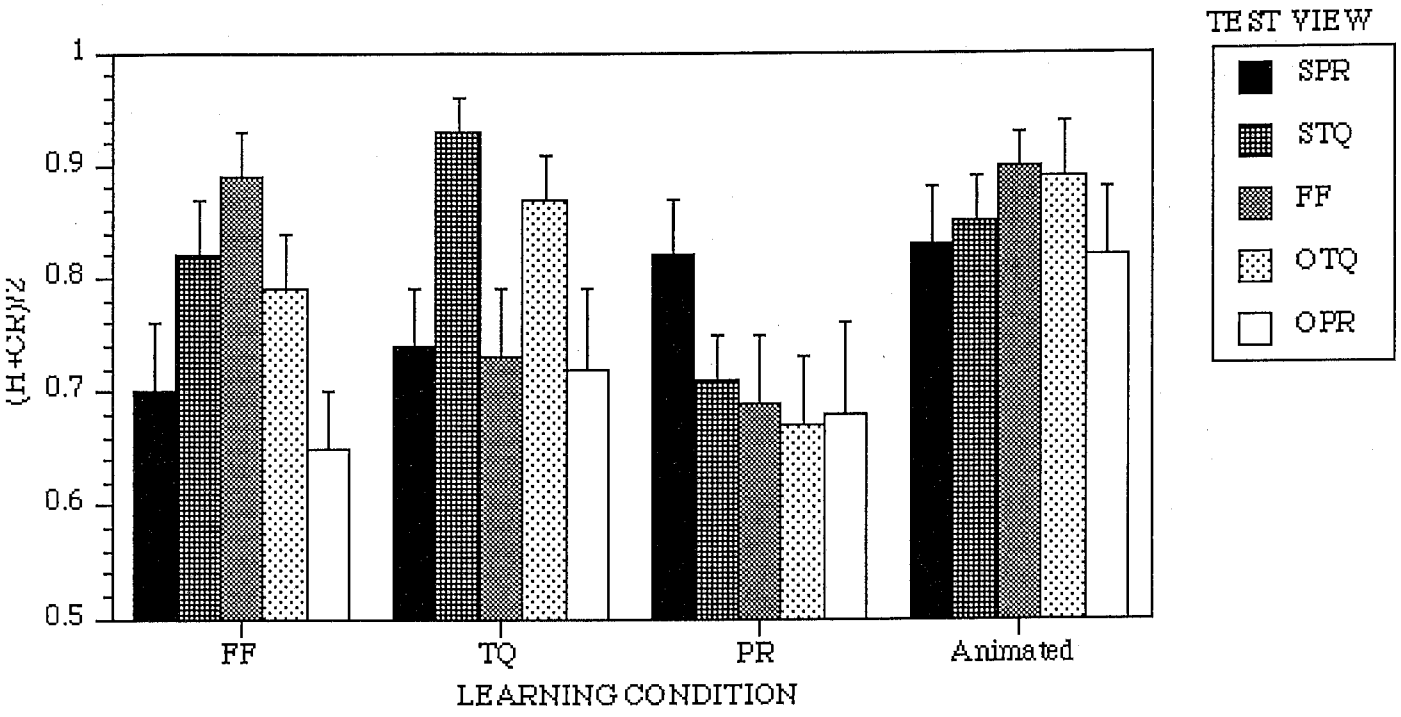
TESTING STAGE



+

SAME FACE?

YES/NO


TEST VIEW
SPR
STQ
FF
OTQ
OPR
1
0.9
0.8
0.7
0.6
0.5
(H+CR)/2
FF
TQ
PR
Animated
LEARNING CONDITION