Multiple Representations in the Recognition of Faces at Novel Orientations

THOMAS A. BUSEY* & SAFA R. ZAKI
Indiana University, Bloomington, Indiana

* Please send correspondence to Thomas A. Busey, Department of Psychology, Indiana University, Bloomington, IN 47405, email: busey@indiana.edu
Abstract

Bilaterally-symmetric objects such as faces are easily recognized from an orientation that is symmetric with respect to the original view. Several authors have proposed that a simple image-based transformation acting on relatively raw image information can account for this effect. In a series of perceptual and memory experiments, we explore the nature of the abstractions that are used in addition to an image-based representation, and contrast the image-based approach against one based on structural descriptions. We demonstrate that by itself the image-based model is insufficient to account for the recognition of inverted and tilted faces, faces with texture asymmetries, and data from memory experiments where faces are studied in a smooth motion sequence. When used, these additional representations appear more robust against texture asymmetries, stress relative (rather than absolute) location of important features, provide robust generalization to the symmetric orientation view in memory paradigms, and may include local depth information provided by smooth motion. Extensions to other domains beyond symmetric orientation recognition are explored.
This article addresses the nature of the representations that allow us to recognize faces at novel orientations. This task is of everyday importance since we can recognize faces from a variety of viewpoints, even under conditions in which we have been exposed to only a single viewpoint of a novel face (e.g., Rock & DiVita, 1987; Tarr & Pinker, 1989; 1990; Bülthoff & Edelman, 1992). When generalizing to novel viewpoints, one particular orientation stands out as having special properties. Several authors, (see Vetter, Poggio & Bülthoff, 1994; Liu, et al., 1995, Troje & Bülthoff, 1996) have pointed out that the symmetric orientation could be used by the observer in conjunction with properties of the human head to enable good generalization and therefore accurate recognition. This explanation takes advantage of the approximate bilateral symmetry of the human head. For example, a head viewed at a 45° angle will appear as an approximate mirror image when viewed at a -45° angle, despite the fact that the other side of the head is now visible (Troje & Bülthoff, 1996). Human observers perform quite well at the symmetric orientation, and in fact perform better than at the frontal view. This strong recognition of this particular novel orientation is known as the symmetric orientation effect.

Vetter, Poggio & Bülthoff (1994) and Vetter & Poggio (1994) have pointed out that this process can be accomplished with image-based (also known as view-based) transformations that work in a two-dimensional coordinate system. The salient landmark features are reflected across the vertical meridian and then matched to the face shown at the symmetric orientation. No knowledge of the three-dimensional structure of the head is required, and the technique is specific to bilaterally-symmetric objects. Models that rely on this transformation accurately capture many of the aspects of human recognition, including the fact that the symmetric orientation is recognized better than the frontal view, and that changes in lighting prove devastating to performance, suggesting that humans do not extract enough three-dimensional structure from the head to compensate for lighting changes (Troje & Bülthoff, 1996).
This image-based transformation is an extension of a class of theories known as view-based models, which seek to explain as much of object recognition as possible without resorting to representations other than transformations applied to two-dimensional views. These two-dimensional views are seen as holistic representations that enable matching whole images to images stored in memory. The images in memory are not exact copies of the raw image, since preprocessing allows normalization for location and size on the retina. In addition, a hierarchy of views are possible, such that parts of objects are stored as separate views. However, these representations do not represent relational information except those represented implicitly by the stored image views. Even in the hierarchical view case, no knowledge of the relation between parts in the hierarchy is allowed under the view-based framework, because the addition of such relational knowledge would change the model into a structural description model (as described below).

A vigorous debate has existed between those who support view-based models and those who propose an additional set of processes that extract relational information about the relevant features in an object. This second class of models has been termed structural description models, because the models process relevant features into parts (e.g. geons or other feature primitives) and then represent the relations between these features independently of the features themselves (e.g. Biederman, 1987; Hummel & Stankiewicz, 1996, 1998; Stankiewicz & Hummel, 1996). This represents a critical difference between structural description models and view-based models, because in view-based models the only representation of relational information is implicit in the location of the features in the image. The view based models cannot represent relational information separately from the parts or features of an object.

The structural description models are also consistent with the symmetric orientation effect. For example, one instantiation by Hummel & Biederman (1992) proposes that the horizontal relations between relevant features are not coded as 'left-of' or 'right-of', but simply as 'beside'. Thus to the structural description model the original view at 45° would be
viewed as nearly identical (barring asymmetries in the head shape) to the -45° view, providing strong symmetric orientation recognition. The structural description model would parse a frontal view with a different set of relations, and thus would have more difficulty matching the frontal view to the original view, thus accurately accounting for the human difficulties with the frontal view. The structural description models might have difficulty accounting for the performance degradation seen with changes in illumination, since the structural description models are designed to be robust against these shading changes. At the very least there is no strong a priori prediction for cases in which illumination changes, while the view-based models clearly predict a decline in performance.

In the literature, the extant data has been portrayed as in support of the view-based account, and for several good reasons. First, the view-based models account for the symmetric orientation effect, as well as the disruption in performance brought on by changes in illumination. According to the view-based models, this disruption occurs because observers do not abstract three-dimensional information that would overcome changes in lighting. Second, the view-based approach is parsimonious, in that it can account for the symmetric orientation effects without additional processing of relational information independent of features. Finally, the template-match nature of view-based models is consistent with the general view that faces are processed holistically (Tanaka and Sengco, 1997). Authors of structural description models have remained relatively agnostic about whether structural description models would apply to faces, or remain restricted to object recognition. All of these suggest that view-based models provide a fairly complete account of symmetric orientation effects.

To provide a quantitative test of image-based transformation in the view-based approach, Troje (1998) developed an image-based transformation he termed the Symmetry Corrected Image Distance, which computed the distance in pixel difference space between two images for purposes of predicting recognition data. This transformation horizontally flips the original view and performs a pixel-based template match with the test image.
through a cross-correlation calculation. This image-based transformation predicts the recognition data quite well, accounting for the translation from one orientation to another. The model that includes the correction for symmetry in its calculation of image distance fails only with novel views of studied faces (other than the symmetric orientation) where it tends to reject studied faces at rates much greater rates than humans. These misses suggest that humans generalize to arbitrary novel orientations better than the template match model, and thus the model cannot model all aspects of face recognition. However, the model does capture the essential aspects of the generalization to the symmetric orientation.

Troje (1998) notes that additional image-based transformations could be applied, each with an associated cost. For example, scaling, translation, rotation, and compensation for illumination could all be image transformations that could be used to facilitate recognition at the symmetric viewpoint. However, prior work either did not test for these transformations or found that observers did not compensate for changes (as in the case of changes in illumination, see Troje and Bülthoff, 1998), and thus these may not be needed.

Despite these successes, conceding victory to the view-based account is perhaps premature. Many of the previous studies on symmetric orientation recognition of faces used testing conditions that may have favored a view-based representation. First, most experiments have used novel faces tested for only brief retention intervals on the order of a few seconds. In the experimental task the observer views two faces sequentially and asked to make a speeded decision about whether the two faces were the same or not. This same-different task may have not required anything other than a briefly stored mental image of the original face. If very little additional processing is required, the massively parallel architecture of the visual system is well-designed for such a task, and will behave as a template matcher. However, longer retention intervals may not be served well by these systems, and observers may be forced to perform abstractions that act as a form of compression in order to remember the faces. In addition, the faces in these tasks did not contain texture information, and were instead shaded busts. This lack of texture may have
made salient features difficult to extract by a structural description model, and forced the observer to rely on lower-level features such as the outline of the head from a particular view. Thus, evidence for the need for abstractions such as those found in a structural description model may be lacking simply because the tasks themselves did not require them.

These issues aside, none of the experiments have been specifically designed to test view-based models against structural description models. As Hummel & Biederman (1992) point out, it is not sufficient to show that performance degrades as the test image deviates in orientation from the study image, because both image-based and structural description models predict this degradation. Likewise, both approaches account for all major effects of symmetric orientation, except perhaps effects of illumination in the case of structural description models. Thus, while previous work has shown the sufficiency of image-based models for short retention intervals and non-textured faces, it has not ruled out the potential contributions of additional abstractions.

The present work is designed to address the role of representations that go beyond a view-based approach. For our purposes, the essential question is how much abstraction is performed on the image before it is matched to memory. We view image-based models as involving relatively little abstraction, which requires relatively little mental effort, and provides a parsimonious account of the existing face recognition data. A structural description model adds a layer of abstraction as parts are processed and relational information between salient features are calculated. This processing of spatial relations appears to require visual attention (Logan, 1994). Other types of abstractions could include matching individual features irrespective of their locations, and abstracting three-dimensional structural information from structure-from-motion displays. In our first experiment we directly test the view-based and structural description models, and then in three additional experiment we address the nature of the abstractions that appear to subserve performance in same/different recognition tasks and longer-term memory experiments. Our
final experiment addresses the role of structure-from-motion displays in the abstraction of limited depth information that assists recognition at the symmetric orientation view.

**Experiment 1**

The template match nature of view-based models makes a strong prediction: the more a test image deviates from the stored template in memory, the greater the image distance and therefore the less likely the model will conclude that the test image corresponds to the study image. Additional transformations such as rotation, translation, scaling or horizontal flipping could easily be brought to bear on the test image to bring it into correspondence with the study image, but once these have been applied the model behaves as described above. Note that the authors of view-based models have also alluded to more complex image warps such as deformable templates (Ullman, 1989, 1996), which we will return to later.

In contrast to view-based models, structural description models stress relational information over exact feature matches. This leads to a situation where deviations applied to the test image might not disrupt the structural description representation as long as the salient relations remain the same. Thus the absolute location of a feature is less important than its location relative to other salient features.

To compare the template-match nature of view-based models, which stresses exact pixel matches, and the structural description approach model, which emphasizes spatial relations, we adopt a manipulation introduced by Cooper and Wojan (2000). In their work, they took famous faces and shifted either one or both of the eyes vertically up the forehead. Examples of their manipulation applied to our stimuli are shown in Figure 1. In Experiment 1, the observer studied a face with neither eye raised (left panel of Figure 1) and is then asked to recognize the same face with either one or both eyes raised. The second face could be shown either at the original orientation or the symmetric orientation and the observer is asked to say 'yes' if it is the same face regardless of any eye raise manipulation. The template-match nature of view-based models provides the prediction that observers will be
more likely to correctly match a test image that has only a single eye raised (compared with both eyes raised). This results from the fact that view-based models make comparisons based on linear coordinates. Under this assumption (or any metric that is based on linear transformations of linear coordinates), the distance between the studied image and the one-eye raised test image is smaller than the distance between the studied image and the two-eyes raised test image (Hummel, 2000). However, a model that includes the categorical relations between salient features predicts better performance for the both eye raised condition, because the salient relation (‘eyes at the same height’) is preserved.

Cooper and Wojan (2000) found evidence for an image-based type representation for face recognition, which is consistent with the view-based approach. As a further test of the view-based models, we repeated the experiment for a new set of observers with inverted
faces. When faces are inverted, observers typically rely on individual features rather than treating the entire face as a single feature or template. Perhaps the best example of this is by Tanaka and Sengco (1997), who showed that the recognition of individual features from inverted houses and faces was not sensitive to the configuration of other features, while features from upright faces were sensitive to the configuration the other features. For example, for upright faces the recognition of the mouth feature was influenced by the inter-eye spacing, demonstrating the importance of metric spatial distances in face processing and consistent with a global template match nature of the view-based approach. The configuration of irrelevant features had no effect on the processing of individual features in the inverted faces, which suggests that for inverted faces, observers rely more on individual features and perhaps a few important spatial relations between salient features. This would tend to support the structural description approach.

To summarize the predictions for the various models, consider the comparison between the one-eye raised and both-eyes raised conditions. If we find better performance in the one-eye raised condition, this is consistent with the global template match nature of the view-based models. If performance in the two-eyes raised condition is superior, this is consistent with a structural description model that stresses relative spatial relations between salient features, rather than treating the entire face as a global template. Both of these predictions apply to upright and inverted faces, but from the literature we have evidence that inverted faces may be treated more as objects, which suggests that a structural description model may play more of a role for these stimuli.

**Method**

**Observers**

The observers were Indiana University undergraduate students. There were 51 observers in the upright face portion of the experiment, and 49 observers in the inverted face portion of the experiment. Observers saw either upright faces or inverted faces, but not both. The observers received course credit for their participation.
**Stimuli**

The stimuli were 16 laser scan models of human heads obtained from a Cyberware(TM) laser scanner. During scanning, the heads of the scanned faces were covered with a beige cap that hid most of the hair. The laser scanner produces a depth map of the head, mapping the deformations of the face from a cylinder at 512 (horizontal) by 256 (vertical) positions. Red, green, and blue color values are also measured at each point.

While designing the stimuli, we observed slight asymmetries in both the structure and texture of the heads. The structural asymmetries came about from the natural variations in the human head, but the slight texture asymmetries came from asymmetric lighting during the scanning process (over which we had no control). This produced a texture asymmetry that will rotate with the face (as opposed to the asymmetric lighting of Troje and Bülthoff (1998) which did not rotate as the face rotated). We did not want these asymmetries interacting with the asymmetries introduced by the one-eye-raised condition, and thus we removed both the texture and structural asymmetries as described below. Note that we evaluate the effects of symmetrization on recognition at the symmetric angle in Experiment 2. In addition, we use unsymmetrized faces in Experiments 3 and 4 because we felt that good performance at the symmetric orientation even in the face of this texture asymmetry would illustrate the strength of symmetric orientation effects in memory.

The heads were symmetrized and the eye movement manipulation was performed as follows. First, the depth map (which provides the three-dimensional structure of the head) and the corresponding luminance map were treated as 512 pixel by 256 pixel images. In the depth map, the value at each location provides the deformation from a cylinder that gives the face its structure. The corresponding pixel in the luminance map provides the color at that location once the luminance map has been molded to fit the depth map. We then used graphics programs to flip both texture and depth maps horizontally which creates what might be thought of as the mirror face. We then placed control points on the important features of both of the texture maps. Then, using morphing procedures (Beier & Neely,
1992) we combined the original face with the mirror face by warping the locations of the features onto an average set of locations (computed individually for each face) and then performing a cross-fade blend that combined the two images. This results in a completely symmetric face at both the level of the structure and the texture, which looks quite realistic. We noted a slight blurring effect in the texture as a result of the blending process, so we introduced a slight sharpening effect in the texture maps prior to 3D rendering of the head models.

Once we obtained the symmetrized face, we took advantage of the fact that our program put out the control points for the averaged features. To raise an eye, we merely moved the target positions for the control points of one eye (and eyebrow) upward and warped the texture and depth maps to produce a new image. This procedure was repeated three times for each of the 16 heads to produce models in which the left eye, right eye, or both eyes were raised.

The models were then converted into files that could be photo-realistically rendered in the POV-Ray rendering package. These were rendered using a light source positioned above and behind the location of the virtual camera. The heads were rotated around the vertical axis, which extends through the middle of each head.

The stimuli were presented in full color on two 21” monitors that mirrored the monitor of a Macintosh 7200 PowerMac computer. Timing and stimulus control was provided by the VideoToolbox library of c routines (Pelli & Zhang, 1991). Data collection was provided by a two-button keypad with millisecond resolution.

All within-subject manipulations involved random assignments across trials, and were not blocked. All experiments used within-subject designs, with the exception of Experiment 1, in which half the observers saw all upright heads and half saw all inverted heads.

**Design and Procedure**

The stimuli were presented in a sequential same-different task. A study face appeared for 1440 ms and was replaced by a pattern mask composed of parts of various faces. This
mask remained on the screen for 1000 ms, and then was replaced by a test face. The study face was positioned slightly to the left of the center of the monitor, and the test face was positioned slightly to the right of center. The test face remained on the screen until all subjects had responded.

The test face was the study face at some orientation (and perhaps with one or more eyes raised) on half of the trials, and a distractor face, chosen from the remaining 15, was presented on the remaining trials. All variables were counterbalanced, such that the study face appeared either at an orientation of 25° or -25°. At test it appeared at 25°, 0° or -25°. Observers were given instructions that they should respond ‘same’ if the same person appeared at test even if the person appeared at a new orientation or had one or more eyes raised. They were given practice trials and feedback to insure that they understood the instructions. Observers were also asked to respond as quickly as possible, maintaining a high degree of accuracy. We gave them feedback via a flashing LED if their response exceeded 800 ms on any particular trial.

Subjects completed 240 trials, which lasted approximately 45 minutes. This provided 10 trials per subject per condition.

**Results**

We did not observe differences in the results as a function of whether the study face was oriented to the left or the right, and so we collapsed over the two study angles. We also did not observe differences between left or right eyes raised, and so we combined these into a single one-eye-raised condition.

Individual d' values were computed for each subject for each condition and then submitted to a between-subjects analysis of variance. To avoid training effects on the 12 faces, we limited the number of observations per subject per condition. In Experiment 1, each observer gave 10 responses per condition, which can lead to difficulties when computing d' if the hit or false alarm rates are zero or one. To avoid this problem, we corrected the d' values according to the correction described in Macmillian and Creelman,
1991 (p.10). In brief, false alarm rates of 0 are set to $\frac{1}{2n}$ and hit rates of 1.0 are set to $1 - \frac{1}{2n}$ where $n$ is the total number of trials per condition per subject.

Figure 2 summarizes the sensitivity ($d'$) results for the relevant conditions. The comparisons of most interest for our current question are represented by the white bars. For upright faces, observers are better in the one-eye raised condition, while for inverted faces
observers are better in the two-eye raised condition. Recall that the image-based representation predicts better performance for the one-eye raised condition at the symmetric orientation, while a structural description representation predicts better performance for the two-eye raised condition when the salient relation (eyes at same level) is preserved. These effects hold only at the symmetric orientation, which leads to a significant interaction between the number of eyes raised (one vs. both), angle (same vs. symmetric) and inversion (upright vs. inverted) ($F(1,98)= 5.0, p<0.05$). This illustrates that the eye movement manipulation affects performance in different ways for upright and inverted faces, especially at the symmetric angle: one-eye raised is better for upright faces, and both-eyes raised is better for inverted faces. None of the two-way interactions were significant, which implies that performance at the symmetric angle differs from that at the same angle. One possible explanation for this is that when viewing the inverted heads at the same orientation, observers could still rely on a template-match representation which results in better performance for the one-eye raised condition and leads to the significant three-way interaction.

**Discussion**

The representations used to recognize faces at the symmetric orientation for upright and inverted faces is best illustrated by a comparison between performance in the one-eye and both-eye raised conditions. For upright faces, performance was better in the one-eye raised condition than the both-eye raised condition at the symmetric orientation. This is consistent with the template-match nature of view-based models that does not incorporate the kind of relational information that the one-eye raised condition breaks. This trend reverses with the inverted faces, which supports a model in which individual features and perhaps a few salient spatial relations are employed. Together, the results confirm the view-based approach for upright faces, but for inverted faces suggests a structural description representation that relies more on individual features and their relative relational information.
Extensions to View-Based Models

An extension to view-based models has been described that might save the view-based approach for the inverted faces. Ullman (1989, 1996) described a complex image warp called "deformable templates" that could be used to recognize non-rigid objects. This extension could be applied to our stimuli as well, through deforming the study-image template by raising one eye or both eyes. The problem with this approach, as described by Hummel (2000), is that a model that uses deformable templates must know what parts of the template are free to vary and what parts should remain fixed. The model must have knowledge of the parts of the object independent of their spatial locations. That is, in order to know that it should deform the template by raising only one of the eyes, the model must identify that only one eye was raised in the test image. This is knowledge of the identity of a feature independent of its location, and adding this information into the model turns it into a structural description model. Taking a difference between the study and test images to compute deformations will not work, because when distractors are used the study image will simply deform to the test image and give a false alarm. This approach only works if you give the model knowledge of which parts of the image are free to vary (the eyes in our case) and the direction they are allowed to move (up in our case). This knowledge turns the view-based model into a structural description model.

Even if deformable templates didn't violate the structure of view-based models, the deformable template approach would still fail to account for our data, because the models do not know about inversion. To a view-based model, the study image is just a template to be matched against the test image, and the 180° rotation in the inverted faces will have no effect whatsoever on its predictions. Thus the view-based account, even with deformable templates, cannot simultaneously account for the fact that for upright faces, one-eye raised performance is better, while for inverted faces the reverse is true. However, the structural description model, applied to the inverted faces where independent evidence suggests that
these stimuli are treated more as objects, provides a cogent explanation of the entire dataset when conjoined with a view-based approach for upright faces.

**Summary**

The eye-move manipulation in Experiment 1 illustrates that observers rely on a view-based template match approach for upright faces, which is consistent the view-based approach proposed by previous findings by Vetter, Poggio and Bülthoff (1994) and Vetter and Poggio (1994) and the symmetry corrected image distance calculation proposed by Troje (1998). For inverted faces, a structural description approach provides a better account, which suggests that observers treat inverted faces as objects and rely more on spatial relations between salient features. Thus observers have the ability to abstract information from a relatively raw copy of the image and use it during the recognition process. In Experiments 2-4 we examine the possibility that additional abstractions are at work in the recognition process of upright faces as well.

**Experiment 2**

To further test the view-based approach for upright faces, in Experiment 2 we devised two manipulations that we thought would disrupt an image-based template match system and therefore encourage observers to rely on an alternative representation such as higher-level abstract features, three-dimensional structure, or features and their relative spatial relations.

We again used the heads shown at the original, frontal or symmetric angles at test. Our first manipulation involves introducing a $30^\circ$ tilt applied in the image plane after rotation. This rotation was applied using Photoshop, and thus lighting and shadows move with the image as it rotates. A template match system will require a rectification procedure to match the study and test images, and although this would extract some cost in terms of performance, it should not otherwise affect the predictions of the model. That is, we would
not expect tilt to interact with other variables such as orientation as long as the observer can determine how to rectify the image equally well at all orientations.

We reasoned that if the template match were disrupted (rather than being engaged after rectification) an alternative representation, perhaps based on a comparison of individual features, would be more likely to play a larger role in the recognition process and thereby reveal some of its properties. We should stress that we consider a 30° tilt to be enough to disrupt an easy match of an image-based representation, but not so tilted that the face changes its properties in other ways. In the face processing literature, inverted faces are often described as having properties more like objects and less like faces (Tanaka & Sengco, 1997), but we consider a face tilted at 30° to still be processed as a face.

Our second manipulation was motivated in part by the asymmetric illumination condition of Troje and Bülthoff (1998). They found that asymmetric illumination drastically reduced recognition of the symmetric orientation view. Observers simply did not overcome the shading asymmetries brought on by an off-midline light source. As noted previously, our original unsymmetrized images contained asymmetries in the texture maps as well as in the structure of the head. The texture differences were introduced by subtle asymmetric lighting of the individual during the scanning procedure, while the structural asymmetries were specific to each individual's head. Both the texture and structural asymmetries were removed by the symmetrization procedure described in Experiment 1, but motivated by the effects of asymmetric lighting reported by Troje and Bülthoff (1998), we include both symmetrized and unsymmetrized faces in Experiment 2. Our central interest is whether and under what conditions observers could overcome these texture asymmetries.

Based on the results of Troje and Bülthoff (1998), we expect better symmetric orientation recognition with the symmetrized faces. However, we are most interested in interactions between symmetrization, test angle and tilt. Figure 3 shows examples of our stimuli and provides the logic by which we expect an interaction between the different independent variables. Assume for the moment that observers have access to two different
types of representations: an image-based representation and one in which more abstract feature elements or perhaps spatial relations provide the basis for comparison. When the test face shown at the symmetric orientation is not tilted, observers tend to rely on an image-based comparison because this requires little abstraction and therefore presumably less mental effort. The large storage requirements necessary to maintain such a representation are not an issue because the task requires relatively little long-term storage and therefore might be accomplished by massively parallel lower-level visual mechanisms, storing a relatively raw copy of the distal image.

In the case of unsymmetrized faces as shown on the left side of Figure 3 (conditions A-D), a reliance on the image-based representation leads to poor recognition performance at the symmetric orientation because of the texture asymmetries. Thus we predict a large
performance drop between conditions A and B. However, when the test image is tilted, we suggest that observers tend not to use an image-based representation and instead rely on information such as abstractions derived from individual features or representations in which relative spatial relations play a role. With unsymmetrized faces, abandoning the template match representation turns out to be a good strategy, because by relying less on image-based information and more on individual features or abstractions, the observer tends to overcome the asymmetries in the texture. While an image-based comparison would return the incorrect answer, a representation that relies on more abstractions will be robust against texture or structural asymmetries. Thus, as shown in the left side of Figure 3, we predict that the performance degradation going from the original to the symmetric angle should be smaller for tilted faces (conditions C and D).

Note that this logic only applies to unsymmetrized faces. As shown on the right panel of Figure 3 (conditions E-H), using an image-based representation with a symmetrized face should work just fine for the untilted, symmetric orientation view because there are no texture differences to disrupt the image-based representation. Thus, performance will decline only modestly for condition F relative to E and H relative to G.

Because the use of an image-based representation produces a large error rate only for untilted, unsymmetrized faces at the symmetric orientation (condition B in Figure 3), we expect a three-way interaction between tilt, test angle (original or symmetric angle) and symmetrization of the face. Another way of thinking about this prediction is that the interaction between test angle and tilt depends on the type of symmetrization (absent or present).

To summarize, condition B in Figure 3 represents the critical case where we expect an image-based system to be used and fail badly due to the texture asymmetries. Tilting the image in condition D may cause an alternative system to contribute, which may overcome the texture asymmetries that cause such difficulty for the upright face in condition B. These
effects depend critically on the presence of texture asymmetries, and so we expect no such interaction for the symmetrized faces (conditions E-H).

Method

Observers
The observers were Indiana University undergraduate students. There were 20 observers in Experiment 2. The observers received course credit for their participation. Note that although we used fewer observers in Experiment 2, each ran twice as many trials as those in Experiment 1.

Stimuli
The stimuli for Experiment 2 included the 16 head models from Experiment 1, to which we added the 16 unsymmetrized (original) head models. Thus, each face had an unsymmetrized and a symmetrized version. As noted above, the symmetrization procedure tended to slightly blur aspects of the texture, and of course the symmetrized faces looked more symmetric. In order to prevent subjects from using these artifacts as cues to the correct answer in the same/different task, we chose distractors from the same class of faces. Thus if an unsymmetrized face was used as the study face, the distractor test face would also be an unsymmetrized face.

For the unsymmetrized faces we took special care to insure that the center of rotation was placed at the center of the head as determined by the distance between the ears. We also used the tip of the nose to align the rotation of the head such that at 0° the head was pointing perfectly straight ahead. This checks are important so that additional asymmetries are not introduced by imperfect rotations or alignments.

Design and Procedure
The procedure was identical to that in Experiment 1. The study image was always an upright face presented at either 40° or -40°. The test face was either at the same angle, the opposite angle, or the frontal view. In addition, the test face could have been tilted 30° to the left or right. All conditions were counterbalanced.
Subjects completed 480 trials which lasted approximately 90 minutes.

**Results**

Figure 4 shows the sensitivity (d') data for Experiment 2. We found no differences between faces studied at 40° and -40°, and thus we average over these conditions. The left side of Figure 4 presents data from the unsymmetrized faces, which shows a very different pattern from the symmetrized faces (right side of Figure 4).

First consider the unsymmetrized faces (left side of Figure 4). For the upright unsymmetrized faces, the drop in performance between the original angle and the symmetric angle is much larger than for the tilted faces. This is consistent with the logic of Figure 3, and suggests that when a tilt is added to the face, an alternative representation contributes that makes the recognition process relatively robust against texture asymmetries. However, without the tilt, the recognition process relies on a template match process that is disrupted by the texture asymmetries, leading to a much larger drop in performance at the symmetric orientation. These effects lead to a significant interaction between tilt (upright vs. inverted) and test angle (save vs. symmetric orientation) for the unsymmetrized faces (F(1,19)=4.634, p < 0.05).

For the symmetrized faces, a complete different pattern of results occurs, as shown in the right side of Figure 4. Performance at the symmetric orientation is high in both untitled and tilted conditions, with tilt extracting a small performance decrement. This represents a completely different ordering of the symmetric orientation views than with the unsymmetrized faces, with additive effects for test angle and tilt. The interaction between tilt and test angle is not significant for the symmetrized faces (F(1,19) < 1),

Despite these differences between unsymmetrized and symmetrized faces, the overall three-way interaction between test angle (original or symmetric angle), tilt, and symmetrization did not quite reach significance (F(1,19)=2.76, p=0.113).
Discussion

The results of the sensitivity analyses confirm the predictions of Figure 3. When tilted, unsymmetrized faces do indeed exhibit a smaller decrease in performance at the symmetric orientation. Symmetrized faces produce a different pattern of results, with additive effects of both tilt (upright and tilted) and test angle (original vs. symmetric) and therefore no interaction between tilt and test angle.

Our interpretation of this data is that when viewing a tilted test face, observers reject an image-based comparison in favor of one that is more flexible, involves more abstractions, and is less reliant on surface texture or texture asymmetries. This representation might include matches of local features and their spatial relations, such as proposed by structural
description models. However, when viewing an upright test face observers instead choose the computationally easier image-based template match representation. This leads to a large decrease in performance for unsymmetrized faces at the symmetric test orientation, but a small drop in performance for the symmetrized faces at the symmetric orientation.

The results confirm, using different variables and methods, the results of Experiment 1 that demonstrate that while an image-based representation is used for recognizing faces at the symmetric orientation, additional representations do contribute and are capable of providing recognition of the symmetric orientation view. We conclude that the range over which a template-match system operates is somewhat limited: observers tend to rely on it for its computational ease when conditions appear to allow it (i.e. short storage period, upright faces) but are prepared to abandon it for a representation that includes more abstractions and is more robust against changes in surface properties such as texture asymmetries or tilt.

**Symmetric Orientation Recognition In Memory**

The existence of additional representations that work in addition to a template match image-based representation suggests a level of abstraction by the visual system. The purpose of this abstraction would be to create a representation that works with a wider class of visual stimuli for which a template-match image-based approach is not well suited, such as objects or inverted faces. In addition, such a representation would be more robust against texture or structural asymmetries. Given the large storage requirements mandated by an image-based representation, it seems likely that for longer-term memory tasks more abstraction (a form of compression) would take place. To address the nature of these abstractions that enable recognition at the symmetric orientation for longer-term recognition, we conducted two memory experiments. Our first memory study (Experiment 3) establishes the symmetric orientation effects in memory and addresses whether observers differentiate between the original and symmetric views. This experiment also allows a comparison of the magnitudes of the symmetric orientation effects in same-different and memory tasks, which
turn out to be quite different. The second memory study (Experiment 4) explores whether observers can use motion to improve recognition at the symmetric orientation, perhaps by extracting some three-dimensional information or using motion to extract a representation that is not bound up with texture or structural asymmetries.

The memory experiments use manipulations (motion and cognitive control) that differ somewhat from those used in the perceptual matching tasks (e.g., eye shifts, symmetrization and tilt). We felt that the nature of the representations that support recognition memory might be less influenced by variables such as tilt, and thus we chose manipulations that we deemed more likely to have longer-term effects and thus reveal the properties of the representation used for memory. The use of the identical faces allows for qualitative comparisons across same/different and memory paradigms.

**Experiment 3**

We have two goals in Experiment 3. First, to our knowledge the symmetric orientation effects have not been established in recognition memory, despite the general agreement that the results of perceptual matching tasks should be extended to memory (see Troje & Bülthoff, 1998).

Second, if we do find symmetric orientation effects, can the observer discriminate between the original view and the symmetric orientation view? Intuitively, one may think this trivial. However, in longer-term memory the directionality of the face may be poorly encoded as a result of the fact that orientation is usually irrelevant for identification. For example, in a classic memory demonstration, Nickerson and Adams (1979) produced drawings of pennies that included versions with Lincoln's face reversed. Fewer than half of the observers could pick the correct penny out of the 15 examples. In the neuropsychological literature, Turnbull and McCarthy (1996) describe a patient who could not discriminate between an object and its mirror image while both were clearly visible. The results suggest that orientation information is fragile or possibly stored in a location that can
be damaged, while leaving object recognition intact. Finally, Price and Gilden (2000) demonstrate that when observers are asked to remember a rotating, translating object, the observers decouple the two events and retained only the translation direction. The translation direction tends to be more informative because it indicates an object that may require interaction or uniquely identifies the object because two objects typically cannot occupy the same location simultaneously. Rotation direction may not require an adaptive behavior because any particular state of the object can be achieved by either clockwise or counterclockwise rotation. Thus rotation direction is relatively uninformative and observers apparently do not encode it. Similarly in the present experiments, the orientation (left- or right-facing) of the face is usually irrelevant for identification purposes, since faces can be observed from either orientation. Thus we may observe a similar dissociation between identity and orientation in face recognition. Note that some structural description models (e.g. Hummel & Biederman, 1992) replace 'left-of' and 'right-of' relations with a 'beside' relation, and thus a structural description model may not be able to discriminate between an object and its mirror image unless additional model components explicitly encode these relations.

To assess whether observers could distinguish between representations of the original and symmetric orientation view, we conducted Experiment 3 as a between-subjects design. Half of the observers were asked to say 'old' if they recognized the face, even if it was at a new orientation. The other observers were asked to say 'old' only if the face appeared at its original orientation. We term these instructions Inclusion and Exclusion, respectively.

We use an old/new recognition paradigm in which observers studied 12 faces. Each of these is studied in one of 5 possible study orientations, which were 70°, 40°, 10°, -20° and -50°. They were then tested with all 5 orientations, as well as with 12 distractor faces also shown at all 5 orientations.

We are interested in whether symmetric orientations are better recognized than other novel orientations, but we chose study and test angles that were asymmetric around the
frontal view. Our rational for this choice was as follows. Previous evidence for good recognition at symmetric orientations suggests that this involves a fairly low-level comparison of the 2D image information. However, Experiments 1 and 2 demonstrated that other representations involving more abstractions were also at work. We are looking for evidence that higher-level information is used, but we did not want our observers to be able to exploit the fact that exact mirror-symmetric angles were included in the test session. One particularly salient cue is the contour of the two-dimensional projection of the face, and this profile tends to change fairly dramatically as the face is rotated by only a few degrees. Thus by using images that were rotated away from the symmetric view by at least 10 degrees we could assess the robustness of symmetric orientation recognition effects without the data becoming contaminated by observers using a strategy that exploited the exact match between a studied view and its symmetric test view. Finding evidence for symmetric orientation effects even under these conditions would illustrate the strength of this phenomenon. In addition, to provide further evidence of the strength of the phenomenon, we used unsymmetrized faces in Experiments 3 and 4. Finding strong symmetric orientation effects even under these conditions would demonstrate that additional information was stored in memory that included enough information to overcome texture asymmetries and changes in orientation that affect the profile view of the face.

Method

Observers

There were 167 Indiana University students in the Inclusion task and 140 students in the Exclusion task. They received course credit for their participation.

Stimuli

The stimuli were 24 head models, which included the 16 unsymmetrized heads from Experiment 2 plus an additional 8 head models that shared similar characteristics.

As with the previous studies, the models were converted into files that could be photo-realistically rendered in the POV-Ray rendering package. These were rendered using a light
source positioned above and behind the location of the virtual camera. The heads were rotated around the vertical axis, which extends through the middle of each head. The study orientations were 70°, 40°, 10°, -20° and -50°. The same orientations were used at test.

The stimuli were presented in full color on two 21” monitors that mirrored the monitor of a Macintosh 7200 PowerMac computer. Timing and stimulus control was provided by the Video Toolbox library of c routines (Pelli & Zhang, 1991). Data collection was provided by a Macintosh Centris 610 computer with 6 external keypads. Up to 6 observers participated at one time.

**Design and Procedure**

Twelve faces were chosen as target faces for the entire experiment, and 12 were reserved as distractors. Each of the 12 faces was randomly assigned to one of 5 study orientations, which was rotated across groups so that each face appeared at each study orientation an approximately equal number of times. The faces were shown for 2 seconds, and were separated by 2 seconds. Following the presentation of the study faces, the observers were shown 120 test trials, which consisted of 60 target faces (12 targets at all 5 orientations) and 60 distractor faces. Those observers in the Inclusion condition were told to respond 'old' if a studied face appeared at any orientation. Those observers in the Exclusion condition were told to respond 'old' if a studied face appears only at the identical orientation. In order to prevent observers in the Exclusion condition from explicitly encoding the orientation at study, the response instructions were provided only after the conclusion of the study portion of the experiment for the observers in the Exclusion condition.

Prior to the study session, observers were given a brief practice session in which they viewed 3 practice faces at various orientations. These faces were not included in the later study or test session. Prior to the test session, the observers were given 6 practice test trials. In particular, observers were shown two test trials early in the practice test session, one that had an old face at its original orientation, and one in which an old face appeared at a new
Observers in the Inclusion condition were told to respond 'old' to both images of this person, since a view of this person was shown in the practice study session. Observers in the Exclusion condition were told to respond 'old' only to the face at the identical orientation and the practice test session was given at the end of the study session. Feedback was given during the practice test session to ensure that observers understood that they should respond 'old' to a studied face only to orientations that conform to their instructions.

**Results and Discussion**

We discuss the data from the Inclusion and Exclusion instruction observers separately.

**Inclusion Instructions**

Sensitivity data (d') from the Inclusion instructions portion of Experiment 3 are shown in the left panel of Figure 5 for those study angles that we expect to show symmetric orientation effects. For those faces studied at a -50° angle, performance at the 40° test angle...
shows improvement over the 10° test angle (which is near the frontal view). This improvement is significant (F(1,166)=5.9, p<0.05). Likewise for those faces studied at a 70° angle, performance at the -50° test view is vastly improved over performance at the -20° test view (also near the frontal view). This improvement is also significant (F(1,166)=19.4, p<0.01).

These comparisons illustrate that symmetric orientation effects do occur in recognition memory, despite the fact that the test angles did not correspond exactly with the symmetric orientations (for example, observers studied a face at 70°, but -50° was the closest test angle to the symmetric orientation of -70°).

**Exclusion Instructions**

If observers can accurately distinguish between the original study view and the symmetric orientation view, then we would not expect the symmetric orientation effects described for the Inclusion data above. Note that the Exclusion instructions were only given after the study session to avoid problems with the subject explicitly encoding the orientation of the face. The right panel of Figure 5 shows the d' results for the Exclusion task. As with the p('old') analysis, we find a complete absence of symmetric orientation effects for the Exclusion task. Thus it appears that observers can distinguish between symmetric orientation. That is, whatever representation is used to recognize faces at the symmetric orientation in memory tasks, it appears to be distinguishable (and rejected when necessary) by the subjects from the representation used to recognize the original orientation.

In summary, we find evidence for large symmetric orientation effects in recognition memory. That we find such effects even when using unsymmetrized faces and test angles that are 10-20° away from the symmetric study angle with unsymmetrized faces suggests that these effects are fairly robust and that observers have abstracted enough information about the faces to overcome differences in texture asymmetries and orientation changes.

The data from the Exclusion instruction task suggests that observers seem to be able to distinguish between representations that support recognition of the original view and the
symmetric orientation view. Structural description models would require explicit storage of left-of and right-of relations in order to discriminate between the original and mirror-image views.

**Experiment 4**

The results of Experiment 3 demonstrate that in recognition memory observers rely on a representation that is robust against texture asymmetries, perhaps by abstracting higher-level features and/or spatial relations from the raw image-based code. When designing Experiment 4 we introduced an additional manipulation at study that we hoped would further encourage the use of representations that involve similar kinds of abstractions that create a more robust representation. During the study portion we rotated the heads around the vertical axis ±15° away from the study orientation in a smooth motion sequence. This gave the appearance of a person shaking their head back and forth. We anticipated that such a manipulation might provide enough depth-from-motion to allow observers to encode structural depth information. This abstracted information could then complement an image-based system and improve recognition at the symmetric view.

Two lines of evidence suggest that motion can affect the nature of the information stored in memory. McNamara, Diwadkar and Blevins (1998) addressed the viewpoint-dependent nature of scene perception using displays that provided (or inhibited) apparent motion. They showed observers a perspective display of 5 colored dots on a computer monitor. The dots could be rotated to simulate a new viewpoint, and when alternated with the appropriate ISI, strong apparent motion was reported by observers. In a control condition, the dots jumped from one side of the screen to the other as well as rotated, which prevents a percept of apparent motion. They then tested the observer's memory for the configuration of dots at novel orientations using an old/new recognition paradigm. Distractor stimuli were composed of the same colored dots in new spatial configurations. They found that recognition for views that fell in between the two views (i.e. the interpolated views) was as
good as recognition for the actual study views. These did not hold for the interpolated views in the condition in which apparent motion was prevented. McNamara et al. (1998) concluded that the motion created ‘virtual views’ in memory in between the endpoint views that facilitated recognition for these interpolated views. Their data suggest that the apparent motion allowed the observers to extract enough structural information about the locations of the dots to infer their relative placement for the interpolated views.

Our face stimuli are arguably more complex than the 5 dots used by McNamara et al. (1998). This rich structure adds additional information that could be used by observers, especially information defined by *constraints* in which information from one source guides the processing of information from another source. A large literature suggests that depth relations (e.g., Hildreth, Grzywacz, Adelson & Inada, 1990) and possibly the three-dimensional structure of objects can at least partially be recovered from an object by placing...
it in motion (e.g., Wallach & O'Connell, 1953; Todd & Bressan, 1990). However, the recovery of depth from motion depends in part on constraints provided by the stimulus. One source of information that may provide these constraints is a line of coplanar points. The structure of faces is for the most part bilaterally symmetric, and as a result faces have a series of features run down the center of the forehead and down the nose and chin that make up a particularly salient line of coplanar points. Pizlo and Stevenson (in press) have demonstrated that coplanar lines, like the one running down the middle of faces, can provide rich structural information when placed into motion. This requires that the observer identify the points as coplanar, which should not be a issue given the fact that humans have vast experience with faces. Pizlo et al (in press) find that shape constancy is best achieved by a stimulus containing planar contours, and that other relations such as symmetry and topological stability also contribute. Human faces contain both bilateral symmetry and a salient line of coplanar points, suggesting that motion may provide strong structural information through the use of these invariants. Thus motion information may provide at least limited structural information about the face that may assist the recognition of faces at symmetric orientations. Note that merely placing an object into motion may not automatically provide additional structural information that would allow recognition of novel viewpoints: Bülthoff and Edelman (1992) compared motion versus static study conditions for wire-frame and amoeba objects, and found no benefit for the motion conditions. These objects were without constraints such as symmetry and readily-identifiable coplanar points, and as a result the motion information did not prove particularly useful when generalizing to novel viewpoints. This reinforces the conclusion by Pizlo et al (in press) that it is the interaction between regularities in the stimuli and motion that provides structural information.

Although relatively little structure-from-motion work has been done on faces, Hill, Schyns and Akamatsu (1997) used motion sequences with untextured 3D laserscan models of faces, and their results suggest that motion enhances face recognition. They rotated faces
from one profile to another and back through a sequence of 5 frames, presenting each frame for 100 ms. In a control condition the ordering of the frames was randomized. Their motion experiment did not allow an investigation of symmetric orientation effects (or performance at any novel orientations), since all 5 test views spanning both sides of the frontal view were shown in the study sequence. The authors report only an overall increase in performance in the motion condition relative to the random motion control, which could imply only that motion stimuli are easier to look at or more informative than random motion displays. This leaves open the question of how motion enhances the recognition of faces: does it provide more overall information by virtue of the smooth nature of the motion, or does motion enable the acquisition of specific types of information (such as limited 3-dimensional structure) that can assist recognition at particular novel viewpoints such as the symmetric orientation? If it does, this would suggest that motion allows observers to extract additional information than that provided by a single image-based representation.

We constructed smooth motion sequences by rotating the face around the vertical axis and compared this condition to two control conditions designed to reduce or eliminate the motion percept. For each of the two study views, five frames were generated for each face, which includes two views that are ±7.5° from the study view, and two views that were ±15° away from the study view. Rather than use all 5 study views from Experiment 3, we limited our study views to 70° and 35°, and generated frames around these views to create motion sequences. The top images in Figure 6 show the 5 views centered around the 35° study angle, and the bottom row shows the 5 test orientations.

We used three motion sequences. The first provides a smooth motion percept by sequentially showing the 5 frames for 180 ms each for a total of 5760 ms. This provides a rich sensation of the head smoothly rotating back and forth for 4 complete cycles. A random motion control condition was generated by taking the frames from the smooth motion condition and randomizing the sequence such that no two frames repeated. While some accidental sequences will produce a slight perception of motion, overall the sequence
shows a face randomly jerking from one view to the next, preventing acquisition of a smooth motion percept. The second control condition reorders the frames so that a slow motion sequence is shown; each view is visible for 720 ms, allowing the face to move through only one complete cycle in 5760 ms. It should be stressed that only the ordering of the frames changed across the motion conditions; the timing and duration parameters were identical, as is the static information available to the observer if the order of the frames is ignored.

We anticipated that the differences between the control and experimental conditions might be somewhat small, and to ensure that we measured any effects of motion on the perception of the symmetric view, we tested the exact symmetric orientation. Thus the five test orientations were 70°, 35°, 0°, -35° and -70°. A face was shown either at 70° or 35° in one of the 3 motion conditions, and tested using static images at all 5 orientations.

Method

Observers

Observers were 149 Indiana University students who received course credit for their participation.

Stimuli

The head models were identical to those used in Experiment 4. Movie sequences were created by rendering 4 views around the 70° and 35° study views that were ±7.5° and ±15° away from the study views. These were presented in full color on the PowerMac 7200, which is able to write the individual images in less than a screen refresh, making the effective ISI between frames 0 ms. Each frame of the motion sequence was on display for 180 ms, for a total of 5760 ms per sequence.

Design and Procedure

Twelve faces were used in the study session, and shown in one of the 6 conditions (2 study orientations x 3 motion conditions) according to a randomized counterbalancing schedule. The instructions were similar to those in Experiment 3, except that the practice
study period contained faces that were placed into motion. Observers were told that they would have to recognize the faces at novel orientations, and the practice test session verified that they understood that a previously-seen face at test at a novel orientation was still an old face.

Following the practice study and practice test session, the observers viewed the 12 study faces and then made old/new judgments on all 120 test faces. These faces come from the 12 target faces shown at 5 test orientations (70°, 35°, 0°, -35° and 70°), as well as the 12 distractor faces also shown at the 5 test orientations.

**Results and Discussion**

Observers studied only left-facing faces in Experiment 4, and this may introduce small criterion shifts for right-facing faces. Memory performance is expected to be lower for right-facing test angles, and observers might shift their decision criterion to compensate. As anticipated, there were small but significant differences between the false alarm rates, which were .42, .45, .43, .40 and .39 with a common SEM = 0.010. As a result, we report only sensitivity (d') data for Experiment 4.

The data for Experiment 4 are shown in Figure 7 for the two study orientations, and show similar effects. The abscissa shows the 5 test orientations, and the three motion conditions produce the three curves. The error bars represent one standard error of the mean.

*Data from the 70° Study View*

The left panel of Figure 7 shows the data from faces that rotated around the 70° study view. Performance is highest for angles tested at the study view of 70°, and falls for the frontal view, and then shows a recovery for the smooth motion condition. A repeated-measures ANOVA revealed an interaction between motion condition (Smooth, Random or Slow) and test angle ($F(8, 1312)=2.4, p<0.05$). The critical comparison between smooth and random motion sequences reveals a significant interaction ($F(4, 656)=3.8; p<0.01$); in addition, the smooth and slow motion condition also interact with test angle ($F(4,
656)=2.66; p<0.05). Inspection of Figure 7 reveals that these interactions do not result from better performance for the smooth motion condition at the study orientation of 70°, where there were no differences between the three conditions (F(2, 328) = 2.08, p = .13). In addition, there was no main effect of motion condition (F(2,328)<1), indicating that the smooth motion condition did not provide better performance overall.

Instead, the interactions between motion condition and test angle derive from differences between the three conditions at the -35 and -70 degree test angle, which are at and near the symmetric orientation. Restricting the analyses to the -70° test angle reveals a significant difference between the Smooth and Random conditions (F(1, 164)=4.54, p<0.05), although not between the Smooth and Slow conditions (F(1,164)=1.55, p>0.05). When the nearby test angle of -35° is included in the analysis, the smooth motion condition
dominates both the Random condition (F(1,164)=4.81, p<0.05) and the Slow condition (F(1,164)=6.85, p<0.01).

All of these findings are consistent with process that uses aspects of the motion to derive a representation that assists recognition of the symmetric orientation view. Interestingly enough, the beneficial effects of smooth motion appear to be restricted to angles at or near the symmetric orientation, and do not facilitate the original view, which might instead still rely on relatively raw image-based information such as a template match.

Data from the 35° Study View
The right panel of Figure 7 shows the data from the 35° study view, and show effects similar to those from the 70° study view in the left panel of Figure 7. As with the 70° data, the interaction between motion and test angle is significant (F(8, 1312)= 3.16, p<0.01). The critical comparison between smooth and random motion sequences again reveals a significant interaction (F(4, 656)=4.3; p<0.01); in addition, the smooth and slow motion condition also interact with test angle (F(4, 656)=4.45; p<0.01).

As with the 70° study view data, there is no main effect of motion condition, indicating that the smooth motion condition does not provide more information overall (F(2,328)=1.29, p<0.05). Nor are there significant differences between the three motion conditions at the 35° study view (F(2,328)<1.0). Instead, the interactions result in part from superior performance of the smooth motion condition at the symmetric test angle of -35°, although this effect was only marginally significant (F(1,164)=2.93, p = .089).

Motion Assists the Recognition of the Symmetric Orientation View
The recovery of performance at the symmetric orientation view only for the smooth motion and the interaction between motion and test angle are both consistent with a process that uses motion to derive a representation that assists in the recognition of the symmetric orientation view. How might this be accomplished? We offer one possible mechanism that is based on the recovery of local three-dimensional information in the form of the contour of the face. There are likely other possible explanations based on image-based
transformations such as flow-fields or improved image interpolation. Further work is needed to distinguish between alternative representations.

**Extraction of Local 3D Structure**

One process that might provide good recognition at the symmetric orientation is the possibility that motion may enhance the perception of lines of *coplanar points*. Faces are (mostly) bilaterally symmetric, and as a result have a series of features run down the center of the forehead and down the nose and chin that make up a particularly salient line of coplanar points. Pizlo and Stevenson (in press) have demonstrated that coplanar lines like the one running down the middle of faces can provide rich structural information when placed into motion. This requires that the observer identify the points as coplanar, which should not be a issue given the fact that humans have vast experience with faces. Pizlo et al (in press) find that shape constancy is best achieved by a stimulus containing planar contours, and that other relations such as symmetry and topological stability also contribute. Our faces contain both bilateral symmetry and a salient line of coplanar points, suggesting that motion may provide strong structural information through the use of these invariants. Thus motion information may provide at least limited structural information about the face that may assist the recognition of faces at symmetric orientations. Note that merely placing an object into motion may not automatically provide additional structural information that would allow recognition of novel viewpoints: Bülthoff and Edelman (1992) compared motion versus static study conditions for wire-frame and amoeba objects, and found no benefit for the motion conditions. These objects were without constraints such as symmetry and readily-identifiable coplanar points, and as a result the motion information did not prove particularly useful when generalizing to novel viewpoints. This reinforces Pizlo et al (in press) conclusion that it is the interaction between regularities in the stimuli and motion that provides structural information.

Our test stimuli were static, thus requiring observers to extract information about the contour of the face from a single image. The frontal view provides essentially no
information about the contour of the face, but the symmetric orientation view does allow some contour information to be inferred, thus supporting good recognition at the symmetric orientation view, but only for the smooth motion condition.

Performance at the study view
For both the 70° and 35° study views, performance at the study view did not differ for the three motion conditions. Why is there no benefit overall for the smooth motion condition? One explanation is that all three motion conditions provide enough image-based information that allows an image-based representation such as a template match to recognize the original view. However, this representation does not work well for the symmetric orientation due to texture or structural asymmetries, and thus the observer relies on additional information, possibly including motion-assisted contour shape information. The data from Experiment 4 are consistent with a model in which observers rely on information derived from smooth motion to enhance the recognition at the symmetric orientation view, which augments the image-based representation.

In summary, the results of Experiment 4 suggest that observers not only extract information from a raw image-based representation and form abstractions that overcome the texture asymmetries (as was seen in Experiment 3) but also use information derived from smooth motion to enhance performance at the symmetric test angle. One possible explanation for this selective improvement is that observers use regularities in coplanar points to estimate depth parameters. This depth information could be used at test, but only for test angles that provide similar estimates of depth. The frontal view, for example, is very difficult to extract depth estimates from, and therefore we see no improvement at this test angle. The original study angle can still rely on the raw image-based representation for all three conditions, which leads to no improvement at the original angle. In combination these two effects provide no overall benefit of smooth motion, but improved performance at the symmetric test angle.
**Accounting for Additional Effects**

In addition to the differences between motion conditions described above, Figure 7, right panel, shows that performance at the 70° test view for items studied at 35° also shows large differences between the three conditions (F(2,328)=8.25, p<0.01). Surprisingly, the smooth motion condition is much worse than either the random or slow motion conditions. We did not anticipate this result, and we offer only this admittedly post-hoc explanation.

For our stimuli, information in the 70° view about the ridge line of the eyes or the protrusion of the mouth (as estimated by the corner of the mouth) suggests more depth in the face than is actually present. If observers try to obtain contour information from the static 70° view they will not match the contour provided by smooth motion rotations around 35°. That is, the smooth motion condition places a representation into memory that includes shape information that is difficult to match to the static 70° test angle, and this produces decrements in performance at the 70° test angles for the smooth motion condition.

Although it may be difficult to go from a 35° study stimulus to a 70° test stimulus, the reverse may be possible. The rotating 70° study stimulus may provide enough contour information to enable good recognition at the -35° test stimulus, which would account for the slightly better recognition performance at -35° than -70° in Figure 7.

**General Discussion**

The results of these four experiments demonstrate that observers exhibit a great deal of flexibility in the nature of the representations that are at work when recognizing faces at symmetric orientations. When conditions appear to make a template match mechanism easy and accurate (short memory delays and the same orientation), observers make use of it, sometimes causing large performance drops for untilted, unsymmetrized faces at the symmetric orientation (Experiment 2). However, when pressed with inverted or tilted stimuli, observers generate representations that rely on relative spatial relations and overcome texture asymmetries to some degree. Longer-term memory demands also may cause observers to
generate representations that are more robust against texture asymmetries and may include motion-assisted depth information derived from contours.

Whatever the nature of these representations, the processes that enable recognition of the symmetric orientation are under cognitive control, since observers can and do differentiate between real and symmetric orientations if necessary, as in Experiment 3. This occurs even when orientation is not explicitly encoded (or at least observers were not instructed to do so).

An open question is how recognition of the symmetric orientation view maps on to related fields such as the identification of symmetry within an object (such as a face seen in a frontal view). Despite a fairly well-developed field (See Tyler, 1994, 1995, Wagemans, 1995, for reviews), no consensus has been reached about the underlying neurophysiological underpinnings of symmetry detection. However, such a model, when developed, may also be able to account for the image-based representation proposed by Troje and Bülthoff (1998). Whether it can be extended to include elements of motion and categorical relations between individual features, or whether an additional representation is required, must then be addressed. This remains a rather active area of research in our laboratory.

Author Notes

The authors gratefully acknowledge Alice O'Toole, who provided the 3D lasercan models. Portions of this research were supported by an NIMH grant to Thomas Busey. Safa Zaki was supported in part by the NIMH Mathematical Modeling Training Grant. The authors also thank John Hummel for his assistance with the predictions of the structural description models.

References


