Visual Cognition
Publisher: Psychology Press, part of the Taylor & Francis Group
Issue: Volume 12, Number 7 / October 2005
Pages: 1291 - 1308
URL: http://taylorandfrancis.metapress.com/link.asp?id=24grbt7pk7dvq0jf
DOI: 10.1080/13506280444000526

Motion as a cue for viewpoint invariance

Tamara L. Watson and Alan Johnston
Department of Psychology, University College London, UK

Harold C. H. Hill
ATR Human Information Science Laboratories, Kyoto, Japan

Nikolaus F. Troje
Fakultät für Psychologie, Ruhr-Universität-Bochum, Germany and Department of Psychology, Queen's University, Kingston, Ontario, Canada

Natural face and head movements were mapped onto a computer rendered three-dimensional average of 100 laser-scanned heads in order to isolate movement information from spatial cues and nonrigid movements from rigid head movements (Hill & Johnston, 2001). Experiment 1 investigated whether subjects could recognize, from a rotated view, facial motion that had previously been presented at a full-face view using a delayed match to sample experimental paradigm. Experiment 2 compared recognition for views that were either between or outside initially presented views. Experiment 3 compared discrimination at full face, three-quarters, and profile after learning at each of these views. A significant face inversion effect in Experiments 1 and 2 indicated subjects were using face-based information rather than more general motion or temporal cues for optimal performance. In each experiment recognition performance only ever declined with a change in viewpoint between sample and test views when rigid motion was present. Nonrigid, face-based motion appears to be encoded in a viewpoint invariant, object-centred manner, whereas rigid head movement is encoded in a more view specific manner.

The recognition of static faces has been shown to exhibit viewpoint dependence. Studies using unfamiliar faces demonstrate that for both recognition memory (Newell, Chiroro, & Valentine, 1999; O'Toole, Edelman, & Bülthoff, 1998) and matching tasks (Hill, Schyns, & Akamatsu, 1997; Newell et al., 1999; Troje & Bülthoff, 1996, 1998) judgements for faces seen from a novel viewpoint, whether measured by speed or accuracy, are typically impaired in proportion to the difference in angle of view.

The symmetry of the face gives symmetrical views a special status since generalization between symmetric views is better than would otherwise be expected from the angular difference between the views (Troje & Bülthoff, 1998). This is an image-based symmetry effect rather than a geometrical symmetry effect since illumination changes between the views dramatically reduced performance. Hill et al. (1997) found that the addition of cues that do not vary over view, such as facial colouring, greatly enhanced the accuracy of the results for cross-view matches. However when presentation times were reduced viewpoint dependencies reappeared, except in the case of three-quarters learnt views. These viewpoint effects suggest that generalized prior knowledge of the three-dimensional (3-D) structure of faces does not allow a view invariant representation of the face to be accessed from a single static view.

View invariance for faces has been interpreted within the more general context of object constancy (Tarr & Bülthoff, 1998; Tarr & Cheng, 2003). However, although an object-based description would provide an efficient internal representation, the balance of evidence points towards view-based encoding for faces and objects in human vision. In Tarr and Cheng's review they judge that the cases in which viewpoint invariance has been reported result from subjects focusing on unique, local features that are diagnostic for identity. However, the great majority of experiments on viewpoint dependence have not taken into account the dynamic aspect of faces and objects. Faces in particular are normally in constant flux. Although the low-level motion fields generated by facial movement are quite different for different views, high-level encodings of object-based motion might generalize well to different views. Properties of timing, for example tempo and rhythm, may also be recoverable independently of viewpoint and facilitate generalization.

Patterns of facial movement provide useful information that can support face identification and classification. Knight and Johnston (1997) showed that the recognition of familiar faces presented as photographic negatives can be significantly enhanced by the addition of facial and head motion. Note that negation degrades facial recognition but leaves the low-level motion information in moving images unchanged. Lander, Christie, and Bruce (1999) have also demonstrated a motion advantage when recognizing famous faces degraded by negation or binary thresholding the face. Christie and Bruce (1998), however, did not find as pronounced a motion advantage for recognition of unfamiliar faces that were initially studied in motion and argued that motion mainly helps when accessing an established face representation. Recently Thornton and Kourtzi (2002) found motion provides unfamiliar faces with a matching advantage without image or motion degradation. They used a sequential matching task rather than a recognition task minimizing the memory load required. They demonstrated that the presentation of a short video sequence improved matching, as compared to a static prime, for a face that differed in expression or viewpoint between prime sequences and test images. Rigid motion of a head alone can also provide useful information for the viewer. Pike, Kemp, Towell, and Phillips (1997) have shown that this additional motion information presented at learning can enhance recognition.

Bassili (1979) utilized markers on the face to investigate facial motion under conditions for which spatial cues have been seriously degraded and found that subjects could make reliable judgements about emotional expression. Hill, Jinno, and Johnston (2003) demonstrated that a few well-placed markers highlighting facial features could be as effective as solid body animations in a sex judgement task; however, greater in number but randomly placed dots did not support accurate sex judgements. Point light displays have traditionally been used to investigate whole body motion (Johanssen, 1975). Primed detection of point light walkers is dependent on walking direction occurring when the walkers have the same spatial orientation (left or right facing) for walker identification (Verfaillie, 1993) and the same orientation and direction of articulation (forwards vs. backwards walking) for an articulation discrimination (Verfaillie, 2000). That both walker detection and direction identification judgements are facilitated when the primed stimulus is the same as the preceding walker suggests that the stored representation used for both the identification of and discrimination between walkers is viewpoint dependent. The motion of a point light walker is nonrigid suggesting that recognition of nonrigid motion in general may be viewpoint dependent. However, nonrigid motion of biological entities can be classified into two of three types of object-based motion (Aggarwal, Cai, Liao, & Sabata, 1998). A point light walker is an example of ''articulated motion'' since the motion of each constituent part is rigid but the motion of the whole is nonrigid. Faces on the other hand can be thought of as an example of ''elastic motion''. Here the object deforms within certain continuity constraints. Therefore, while priming of articulated motion may have previously been shown to be viewpoint dependent in the context of point light walkers it is still possible that processing of elastic motion, as displayed by a face, will exhibit viewpoint invariance. It has recently been suggested that the discrimination of direction of walkers occurs due to the sequential matching of form templates such that recognition is achieved through a motion form analysis (Beintema & Lappe, 2002). If this is the case the view-dependent effects may reflect the differences in the forms projected into the different viewplanes.

Studies of facial motion have typically presented spatial cues alongside motion cues and have therefore not studied the role of facial motion in isolation, leaving open the question of whether facial motion alone can be used for recognition. Hill and Johnston (2001) mapped facial motion captured from a number of ''actors'' onto a 3-D computer generated model of an average head. They showed that both rigid head movements and nonrigid facial movements in the absence of spatial cues provide sufficient information to allow observers to categorize faces on the basis of both identity and gender. Nonrigid motion was more useful for classification of the sex of the actor, whereas rigid motion was more effective for categorization on the basis of identity.

Nonrigid facial motion, generated by changes in expression or speech, arises from dynamic changes in object shape, whereas rigid motion arises from changes in head orientation with respect to the viewpoint. Geometrically, changes in object shape can be dissociated from changes in viewpoint, whereas rigid changes in pose are essentially defined relative to the viewpoint. Pose and viewpoint are confounded since the same effect can result from either transformation. Viewpoint experiments that exclusively study static faces have not provided an opportunity for subjects to establish an object-based encoding of shape parameters and shape change from experience of the pattern of variation in object shape over time (Johnston, 1992).

Since an object-based encoding could in principle support generalization to different viewpoints, it is possible that recognition may be less sensitive to changes in viewpoint for moving faces as compared to static faces, and the degree of sensitivity may also depend on the type of facial motion. The experiments described below are designed to test these possibilities. In the three experiments we adopt similar methods to those described in detail in Hill and Johnston (2001). Markers are place on the faces of performers who are asked to tell simple jokes to a confederate. We then use 3-D motion tracking to record the position of the markers. The sequence of three 3-D positions is used to drive the animation of a 3-D average head. Four dots placed on a headband, which were relatively unaffected by changes in facial expression, were used to track rigid head movements. The 3-D head is rotated and rendered in 3D Studio Max (3DS Max) to generate the final animation. All the experiments use variants of a self-paced delayed match to sample paradigm in which a sample of facial motion is presented and subjects have to decide which of two alternative stimuli, usually rendered from a different viewpoint, is the facial movement they have seen previously.

STIMULI

Stimuli used in each experiment consisted of a total of 64 animations based on motion capture recordings of eight males and eight females, each telling four question-and-answer jokes, a technique adopted to illicit natural expressive facial movement. Recordings were made with an eight camera Oxford Metrics' Vicon motion capture system with the cameras placed in a semicircle at different heights in front of the head. Forty markers were used to capture facial movement and a headband with four markers was used to capture rigid movements.

The resulting motion information was used to animate an average 3-D facial model created from 100 male and 100 female laser-scanned faces (Vetter & Troje, 1997). Animation of the 3-D model was achieved in Famous Animator where ''areas of influence'' around each marker placed on the face inherit the movement of the marker (see also Hill & Johnston, 2001). As no eye movements were captured the eyes were made to ''look at'' a point straight ahead of the face. The 3-D head model was texture mapped with a corresponding average texture and the resulting animated sequences rendered using 3DS Max. Rigid motion of the head was identified from four dots placed on a headband. Rigid or nonrigid motion could be subtracted from total movement to isolate either type of motion. Three versions of each sequence were rendered; one with just rigid head movements, one with just nonrigid facial movements, and the last with both types of movements combined. On average, the animation lasted 6.75 s. All stimuli were presented centrally on a 17-inch monitor using the multimedia presentation application Macromedia Director. Each face was approximately 10 $\times$ 8 cm viewed at a distance of approximately 57 cm.

EXPERIMENT 1

This experiment was designed to assess view dependence when matching nonrigid facial movement as opposed to rigid and nonrigid movement together. Our prime interest was in the viewpoint generalization of facial movement and so isolated head movement was not included as a separate condition in Experiments 1 or 2.

Method

Participants. A total of 40 people participated for monetary reward; 20 each in the ''rigid + nonrigid'' (combined motion or CM) and ''nonrigid alone'' (NM) movement conditions.

Design. The experiment consisted of a 2 $\times$ 7 $\times$ 2-way design. Type of motion was presented between subjects and consisted of two levels, CM (combined motion) and NM (nonrigid motion).

Test view was presented within subjects and consisted of seven levels: viewpoint in depth of 0° (full-face view), 15°, 30°, 45°, 60°, 75°, or 90° (profile). Orientation was also varied within subjects and consisted of two levels, upright and inverted. One group of subjects was presented with stimuli containing only nonrigid motion and another group was presented with stimuli containing combined rigid and nonrigid motion.

Trials were blocked for the inversion condition, i.e., during the upright condition all faces were presented upright and during the inverted condition all faces were presented upside down. The initial state of inversion was randomized across subjects. Each condition contained a total of 128 trials, 64 upright and 64 inverted. All test views were presented nine times (except 0°, which was presented 10 times) in each orientation condition.

Procedure. Each trial consisted of a learning and test phase presented sequentially. During each trial participants were first shown the learning phase consisting of the learning animation sequence oriented at 0° (lasting approximately 6.75 s). This was immediately followed by a target and distractor animation (Figure 1A) presented sequentially both at the same rotation in depth. The viewpoint of the test stimuli was 0°, 15°, 30°, 45°, 60°, 75°, or 90°. The order of presentation of the target and distractor animations was randomized on a trial by trial basis. The target animation was the same as the learning animation (rotated) while the distractor was randomly chosen with the constraint that it contained an actor telling the same joke as the test stimulus. Both target and distractor animations were shortened such that the video sequence started at a random point within the first half of the animation and then ran for half the length of the full animation. This procedure ensured subjects could not use just the first or last frames as a basis for their judgements and had the effect of lowering performance, which a pilot study showed would otherwise have been at ceiling. Participants were asked to indicate which animation was the same as had

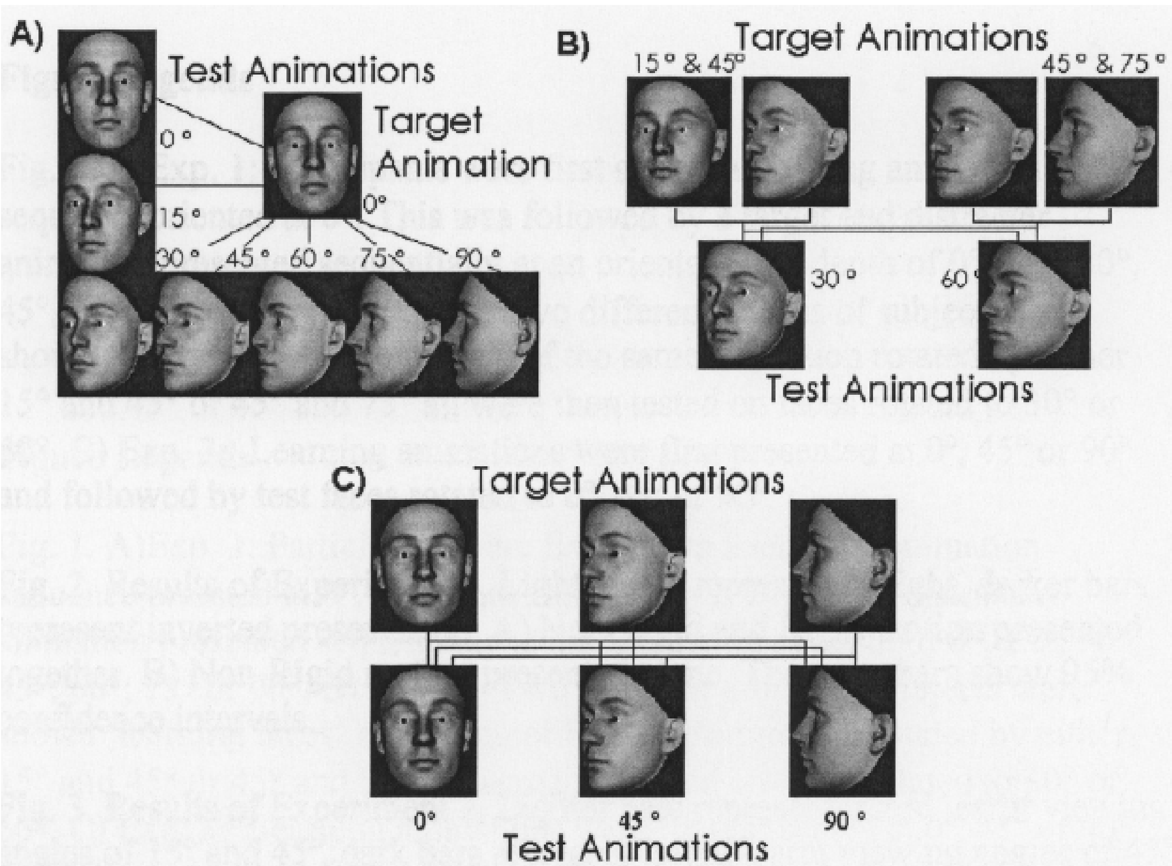

Figure 1. (A) Experiment 1: Participants were first shown a learning animation sequence oriented at 0°. This was followed by a target and distractor animation presented sequentially at an orientation in depth of 0°, 15°, 30°, 45°, 60°, 75°, or 90°. (B) Experiment 2: Two different groups of subjects were shown ''learning faces'' consisting of the same animation rotated by either 15° and 45° or 45° and 75°; all were then tested on faces rotated to 30° or 60°. (C) Experiment 3: Learning animations were first presented at 0°, 45°, or 90° and followed by test faces rotated to 0°, 45°, or 90°.

been shown in the learning phase by clicking on buttons on the monitor that appeared after each trial. Subjects controlled presentation of each trial via buttons on the screen. However, each animation could only be viewed once and all animations had to be viewed before a response could be made.

Results

The results for Experiment 1 are shown in Table 1 and Figure 2. In each condition recognition performance is better than chance (50%). A 2 (type of motion—between) × 7 (test view—within) × 2 (inversion—within) mixed-design ANOVA was carried out on the accuracy data. Significant main effects of test view, $F(6, 228) = 2.7$, $p < .02$, and inversion, $F(1, 38) = 11.6$, $p < .002$, were found. No significant interactions were found, Inversion × Type of motion, $F(1, 38) = 0.07$, $p = .8$; Test view × Inversion, $F(6, 228) = 1.4$, $p = .2$; Test view × Inversion × Type of motion, $F(6, 228) = 0.6$, $p = .7$. Although not statistically significant, a marginal Test view × Type of motion interaction, $F(6, 228) = 1.9$, $p < .1$, was evident. As such a within-subjects ANOVA was carried out for each type of motion separately as planned. Combined motion displayed a significant effect of test rotation (Figure 2A), $F(6, 114) = 2.3$, $p < .04$, which can be described as an approximately linear reduction in performance with viewpoint difference, $F(1, 19) = 12.4$, $p < .002$, as shown by tests of trends. The effect of test viewpoint for nonrigid motion (Figure 2B), although not statistically significant, showed a marginal reduction in performance with viewpoint, $F(6, 114) = 2.2$, $p = .052$. The effect of inversion for combined motion, $F(1, 19) = 5.8$, $p < .03$, and nonrigid motion, $F(1, 19) = 5.8$, $p < .03$, was found to be statistically significant as expected.

An additional ANOVA was carried out on a sample of the data testing for a bias produced by controlling for joke told, but not the gender of the speaker. No significant difference was found in the number of correct responses whether the gender of the target and distractor animation was the same or different, $F(1, 11) = 0.04$, $p > .8$.

TABLE 1
Mean percentage correct with standard error, Experiment 1

Target rotation	*0°*	*Std err.*	*15°*	*Std. err.*	*30°*	*Std. err.*	*45°*	*Std. err.*	*60°*	*Std. err.*	*75°*	*Std. err.*	*90°*	*Std. err.*
Nonrigid motion														
Upright test	82.5	2.6	77.2	3.2	77.8	2.8	75.0	3.2	78.3	2.8	74.5	3.8	73.9	4.0
Inverted test	76.5	3.4	65.0	4.8	78.3	3.8	67.8	4.6	68.1	5.2	71.1	3.1	77.2	2.6
Rigid + Nonrigid motion														
Upright test	81.5	3.2	81.7	2.7	81.7	2.5	77.8	3.1	75.6	3.5	75.0	3.0	69.0	4.7
Inverted test	75.4	2.9	74.4	3.7	75.1	2.4	74.5	4.5	72.0	2.6	70.4	3.2	70.0	3.0

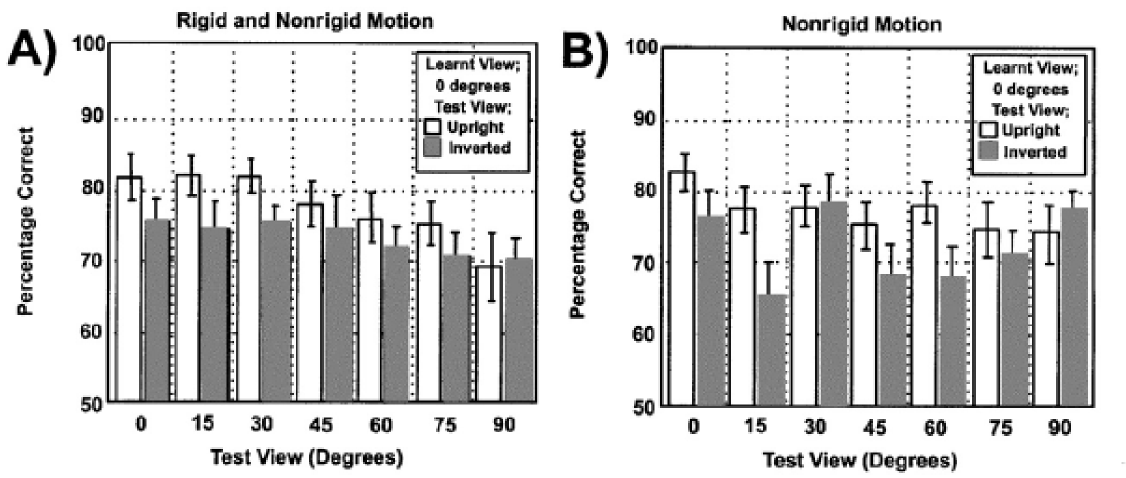

Figure 2. Results of Experiment 1. Lighter bars represent upright presentation; darker bars represent inverted presentation. (A) Nonrigid and rigid motion presented together. (B) Nonrigid motion presented alone. The error bars show 95% confidence intervals.

Discussion

These results show that nonrigid facial motion is marginally less viewpoint dependent than combined motion when generalizing from a full-face view. Both the statistics and inspection of the graph suggest that for nonrigid animation there is some decline in performance as test viewpoint rotates away from the target view. However, this is not as pronounced as when rigid motion information is also added to the animation.

The fact that both data sets show an inversion effect indicates that optimal performance is based on processing in a face-based analysis system rather than the extraction of low-level motion cues or temporal pattern cues that would not be affected by face inversion. Note performance is better than chance (50%) in the inverted condition demonstrating that low-level motion or temporal cues can support matching of dynamic information across changes in viewpoint.

Experiment 1 investigated viewpoint generalization with respect to a full-face learning view. It is possible that the full-face view may be considered a special case when analysing facial motion as it is the view most seen when interacting directly with another person. Therefore it is important to extend the study to other views.

EXPERIMENT 2

This experiment extends the investigation to rotated learning views and assesses generalization for test views lying between (interpolation) or outside (extrapolation) pairs of learning views (cf. Bülthoff & Edelman, 1992). In this case subjects are given more information at encoding (two views) and one might expect improved viewpoint generalization.

Method

Participants. We used 20 paid participants in each condition who had not been part of Experiment 1.

Design. Type of motion was presented between subjects with two levels, CM and NM. Learning view was presented between subjects with participants viewing animations rotated by either 15° and 45° or 45° and 75°. Test view was presented within subjects at either 30° or 60° (Figure 1B). Inversion was included as a blocked within-subjects factor as in Experiment 1. During each condition animations were all presented either upright or inverted, with initially presented orientation randomized. A total of 32 trials were presented in each orientation condition (16 test animations were presented at 30° and 16 at 60°).

One half of the participants were included in the NM condition while the rest viewed the CM condition.

Procedure. Each trial consisted of a learning and a test phase. One group of subjects were shown "learning faces" consisting of the same animation rotated by 15° and 45°; another group was shown the same animations rotated by 45° and 75°. Both groups were then tested on faces rotated to 30° or 60°. The two learning and two test animations (target and distractor) were presented sequentially during one trial and were randomized as to which was viewed first within the learning and test phase. The target face was identical to the learning face (rotated) while the distractor was chosen randomly with the constraint that it would be at the same rotation as the test and of the same gender. Participants were asked to choose which of the test animations was the same as the learning animation by clicking on buttons on the monitor screen that appeared after each trial.

Both test and distractor animations were shortened as in Experiment 1. Presentation of each trial was controlled by the participant; however, each animation could only be viewed once and all four animations of each trial had to be inspected before responding.

Results

The results for Experiment 2 are shown in Table 2 and Figure 3. A mixed-design fourway ANOVA was carried out on the recognition performance data, 2 (inversion—within) × 2 (type of motion—between) × 2 (learning view—between) × 2 (test view—within). This test shows a significant main effect of viewpoint, $F(1, 76) = 45.0$, $p < .001$, a marginal four-way interaction, $F(1, 76) = 3.7$, $p = .06$, and an interaction between learning view and test view, $F(1, 76) = 6.8$, $p < .02$. All other interactions did not reach significance levels below .15

TABLE 2
Mean percentage correct with standard error, Experiment 2

Target learnt	*15° + 45°*	*Std. err.*	*45° + 75°*	*Std. err.*
Nonrigid motion				
Upright test				
30°	81.3	3.1	81.6	2.6
60°	79.4	3.0	81.2	2.2
Inverted test				
30°	74.3	3.6	75.0	3.2
60°	70.0	4.2	77.8	3.3
Rigid + nonrigid motion				
Upright test				
30°	85.0	4.1	88.1	2.3
60°	76.3	4.7	90.6	2.2
Inverted test				
30°	74.4	3.2	78.1	2.4
60°	71.9	3.9	78.4	3.1

Due to the marginal four-way interaction found in the full design a within-subjects ANOVA was carried out for each type of motion, CM and NM, as planned. For NM (Figure 3B) we did not find an effect of learning viewpoint, $F(1, 38) = 0.6$, $p > .4$, test view, $F(1, 38) = 0.3$, $p > .5$, or an interaction, $F(1, 38) = 1.6$, $p > .2$. We did find an effect of inversion, $F(1, 38) = 14.3$, $p < .001$. Inclusion of rigid motion information (CM) provided a significant interaction

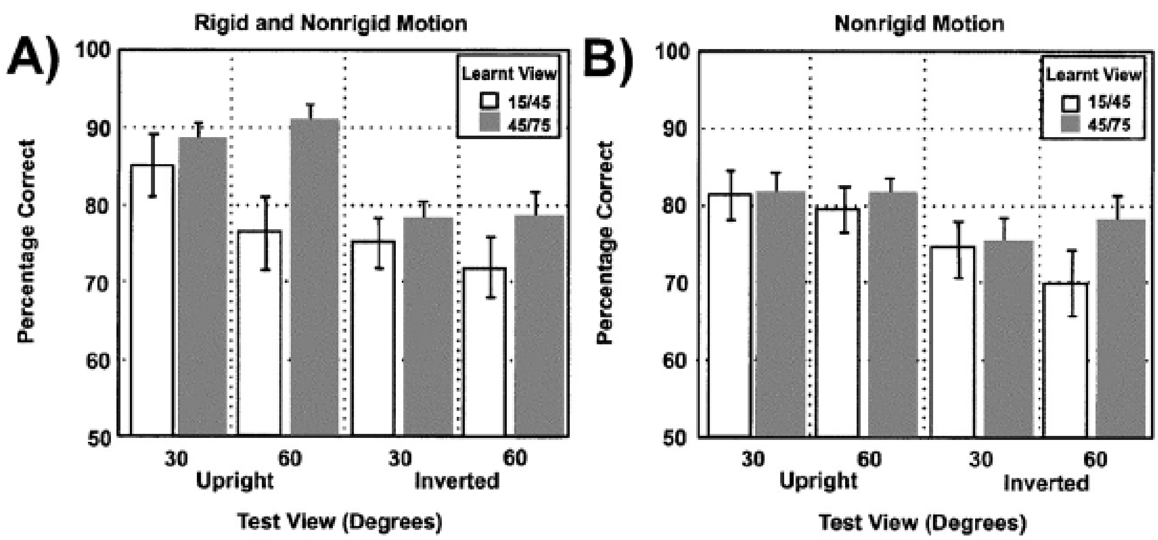

Figure 3. Results of Experiment 2. Lighter bars represent initial learnt viewing angles of 15° and 45°; darker bars represent initial learnt viewing angles of 45° and 75°. Results are grouped by U and I for upright and inverted test conditions respectively. (A) Nonrigid and rigid motion presented together. (B) Nonrigid motion presented alone. The error bars show 95% confidence intervals.

(Figure 3A) between the learning and test rotation, $F(1, 38) = 7.1$, $p < .02$. An inversion effect was also found in this CM condition, $F(1, 38) = 33.5$, $p < .001$. The statistically significant inversion effects indicate processing of face-based cues rather than low-level motion cues.

An additional ANOVA was carried out on a sample of the data to asses the impact on accuracy when the joke told was the same or different at learning and test. This difference did not reach significance, $F(1, 15) = 4.51$, $p = .06$. However, as only gender or the joke could be controlled with these stimuli, and as a difference in joke told had a greater overall effect on the stimuli than a change in gender, it was concluded that controlling for the joke would be the best option for further experiments.

Discussion

As in Experiment 1, view dependence is stronger when there is rigid motion of the head. We expected view dependence to be reduced in this study due to the presentation of two views at the learning stage. Nonrigid motion delivers view generalization; however, an interaction between learnt and test view remains in the complete motion condition, which is indicative of view dependence. Inspection of the data in Figure 3 shows the inclusion of rigid motion does not undermine viewpoint generalization for the 30° target whether it was inside or outside the learning views. The asymmetry probably results from the difficulty of generalizing movement seen close to profile to the 30° test. It could be that the full-face view and those close to it do hold more ecological value and as such we may build up a better object-centred representation for patterns of rigid motion views close to the full-face view. This could be due to the prevalence of this viewpoint when actively interacting with another person. We explore this asymmetry further in the next experiment.

EXPERIMENT 3

Here we examine whether there is a general benefit for similar views at learning and test by permuting three target and test views. We also include a condition of rigid motion alone but drop the inversion condition since we have established in the first two experiments that subjects are using face-based cues to perform the motion-matching task.

Method

Participants. Twenty-five participants each took part in all conditions for a small monetary reward.

Design. Type of motion was included as a within-subjects three-level factor containing NM, CM, and, additionally, rigid motion (RM).

Learning view was presented within subjects with three levels, 0°, 45°, or 90° of rotation in depth. Test view was also presented within subjects with three levels, 0°, 45°, or 90° rotation in depth (Figure 1C). All animations were presented upright. There were 63 trials in each motion condition, 7 of each combination of learnt and test view.

Procedure. Each trial consisted of a learning and test phase. Learning animations were first presented at 0°, 45°, or 90° with presentation order randomized. These were directly followed by target and distractor animations presented sequentially, both rotated to 0°, 45°, or 90°. Presentation order was also randomized such that the number of trials with each combination of views was balanced. The order of target and distractor animations was randomized on a trial by trial basis. The target animation was the same as the learnt animation (rotated) while the distractor was randomly chosen with the constraint that the actors were telling the same joke in both cases. Test sequences were shortened as in Experiments 1 and 2. Subjects were asked to indicate which of the test animations were the same as the learnt animation by clicking on buttons on the monitor that appeared after each trial. The subjects initiated each trial and had unlimited time to respond. However, each animation could only be viewed once and all animations in each trial had to be viewed before responding as in Experiments 1 and 2.

Results

The results are shown in Table 3 and Figure 4. A within-subjects ANOVA was carried out on the performance data for all motion conditions. Significant interactions were seen between type of motion and learnt view and between learnt view and test view (Figure 4D), $F(4, 96) = 2.7$, $p < .05$, $F(4, 96) = 2.8$ $p < .05$. A main effect of type of motion, $F(2, 48) = 5.5$, $p < .01$, and of learnt view, $F(2, 48) = 10.4$, $p < .001$, was also found. An ANOVA examining the interaction between learnt view and test view showed no significant simple main effect of test view for all of the levels of learned view.

To further explore the learnt view by test view interaction an ANOVA was carried out with effects of learnt view found for RM, $F(2, 48) = 8.5$, $p < .005$, and CM, $F(2, 48) = 8.3$, $p < .05$. *T*-tests were then carried out to further investigate the effect of learnt view for RM and CM. At a corrected alpha level of .003, RM LV0° vs. LV90°, $t(24) = 4$, $p < .003$, CM LV0° vs. LV90°, $t(24) = 3.8$, $p < .003$, and CM LV45° vs. LV90°, $t(24) = 3.6$, $p < .003$, were found to be significant. The *t*-tests suggest profile disadvantages against full face for RM and against full face and three-quarters face for CM with no differences for NM.

TABLE 3
Mean percentage correct with standard error, Experiment 3

Target learnt	*0°*	*Std err.*	*45°*	*Std. err.*	*90°*	*Std. err.*
Nonrigid motion						
0°	76.0	4.0	70.3	4.3	75.2	3.6
5°	70.3	3.8	72.0	3.9	63.4	4.3
90°	72.0	4.4	66.3	3.5	78.0	4.0
Rigid motion						
0°	84.0	2.9	78.9	3.0	76.0	3.5
45°	74.9	3.6	74.9	3.9	71.4	4.7
90°	70.8	3.9	67.4	4.2	69.7	4.7
Rigid + nonrigid motion						
0°	82.3	3.4	80.0	4.0	78.9	3.2
45°	74.9	3.5	80.6	3.3	78.3	3.5
90°	68.6	4.4	72.0	3.7	75.4	4.0

Discussion

The main effect of learnt viewpoint suggests generalization appears to be most accurately achieved from the frontal view and least from the profile view. Figure 4D shows that when the data are pooled over type of motion, subjects were generally most accurate when matching to the learnt viewpoint but accuracy does not systematically and consistently drop as a function of the difference between the test and learnt view. The main effect of type of motion shows a difference in accuracy while inspection of the graphs suggests that overall accuracy was worst for nonrigid motion and best when both types of motion were presented together. This is consistent with previous findings showing that rigid movements are more useful than nonrigid movements when discriminating identity (Hill & Johnston, 2001). There is a suggestion of a combined motion advantage for upright faces in Experiment 2, and it seems to be most clear in the data for the 45° learnt view. It may be that the 45° view provides for optimal encoding of both forwards, lateral, and rotational head movements and that subjects utilize head movement information more in this situation.

The motion by initial view interaction supports a difference between the pattern of results for nonrigid and rigid motion. When viewing rigid motion or both nonrigid and rigid motion together (but not nonrigid motion alone) there is an overall advantage of initially viewing a full-face animation, followed by the three-quarters view, then profile. Again, there is no dependency on relative angle of view for nonrigid motion.

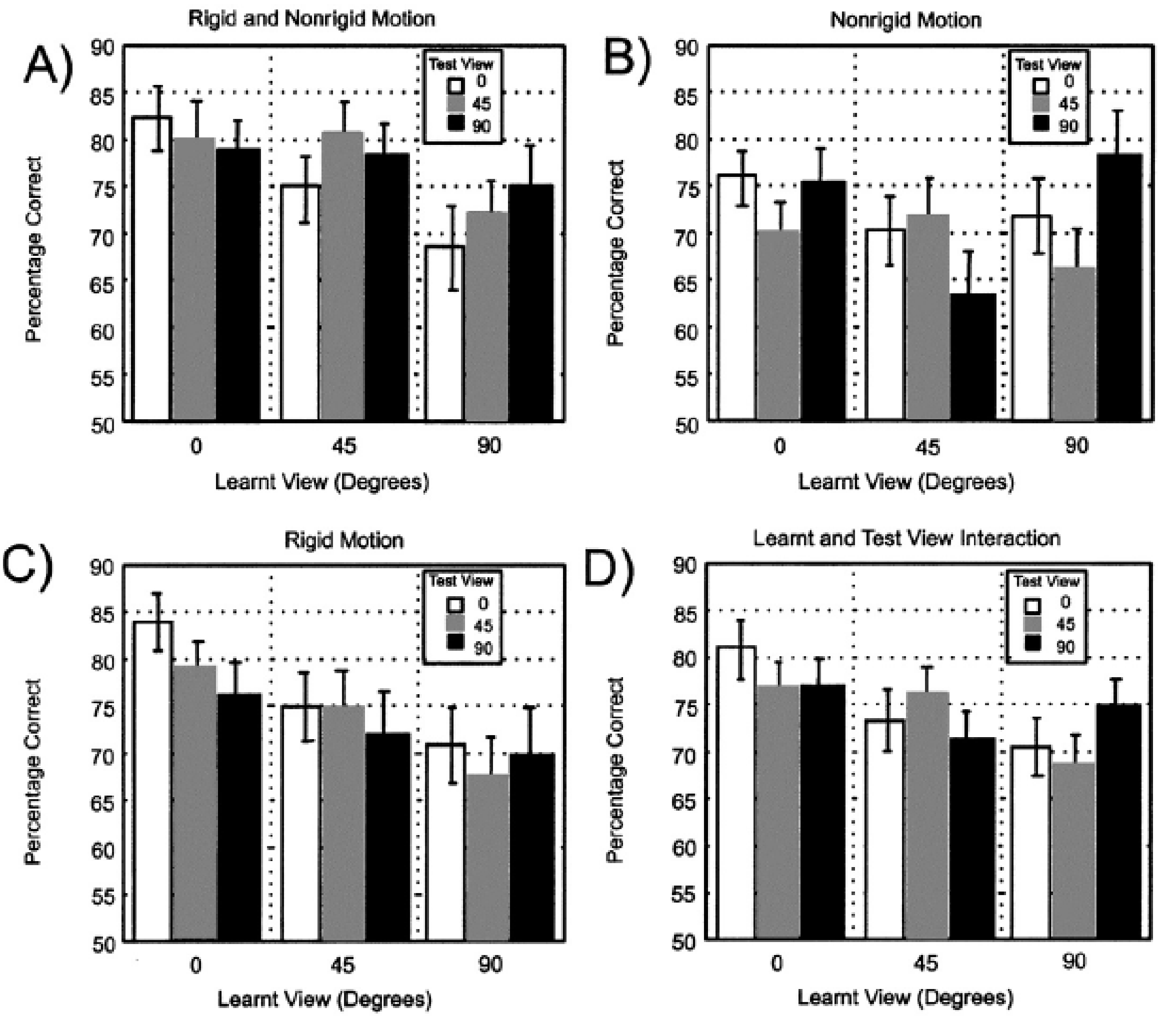

Figure 4. Results of Experiment 3. Lighter bars represent a test view at 0°; grey bars represent 45°; and black bars represent 90°. (A) Nonrigid and rigid motion presented together. (B) Nonrigid motion presented alone. (C) Rigid motion presented alone. (D) All motion conditions collapsed together. The error bars show 95% confidence intervals.

GENERAL DISCUSSION AND CONCLUSION

Each experiment has shown that, in general, matching of facial motion is robust to changes in viewpoint. Although viewpoint dependence was found for rigid motion, variation in matching performance across viewpoint does not seem to be as pronounced as that typically found for static views of the face (Troje & Bülthoff, 1996, 1998). Additionally, in all experiments nonrigid transformations of a face are encoded in a less viewpoint-dependent manner than rigid transformations.

The inversion effects found in Experiments 1 and 2 indicate that the results for upright faces reflect the encoding of dynamic changes in facial configuration rather than low-level cues that are unaffected by inverting the stimulus. It is noted that the results with the inverted faces also show a high degree of viewpoint invariance. It may be that for inverted faces, since configural cues are unavailable, subjects are forced to rely on temporal pattern or motion cues that will be invariant with respect to viewpoint such as the overall magnitude of the movement, the frequency of head nodding or the temporal pattern of slow and fast movements.

Interestingly, Jordan and Thomas (2001) have found that rotating a head in depth does not affect visual speech influences on auditory speech perception. However, rotating a head in the picture plane away from the upright position reduces interference from incongruent visual speech on auditory speech perception, with the notable exception of a fully inverted head in which case mouth movements are similar to the fully upright case (Jordan & Bevan, 1997). These results parallel the present study in showing that facial motion remains an effective signal under rotations in depth but is diminished by rotation in the picture plane.

Overall, matching is less accurate for nonrigid motion when compared against conditions in which both rigid and nonrigid motion is presented. Optimal performance, particularly when viewing faces closer to a full-face view, requires the utilization of all available motion cues, demonstrating that both nonrigid and rigid motion contribute to face recognition. For rigid motion subjects found generalization from profile views more difficult than generalization from full-face views. There was no effect of learnt view in the case of nonrigid motion although overall accuracy appears to be less than for the other types of motion tested here. It is worth considering why the profile view is less effective at the encoding stage in the case of rigid motion. A profile view would seem to give as much information about rigid movements of the head as full face and so the reasons for the asymmetry are not immediately clear. In everyday human interaction the full-face and nearby views are of primary interest, as people talking to us are usually facing us. Therefore expertise may allow us to build up a view invariant representation more quickly. Another possibility is that lateral and rotation head movements (e.g., shaking), which would be less discernable from profile than from a full-face view, vary more between individuals and are therefore more informative and better encoded than head movements in the sagittal plane (nodding movements).

Detection and discrimination of direction of point light walkers has been found to produce viewpoint-dependent results using a priming task (Verfaillie, 2000). Evidence that the perception of point walkers is not disrupted by adding randomized stereo depth information but leaving the viewed 2-D structure intact has also been used to support the proposal that a 3-D structural description is not useful (Bülthoff, Bülthoff, & Sinha, 1998) for recognition of this class of biological stimuli. These results suggest perception of the articulated nonrigid motion of biological walkers is viewpoint dependent. One possible explanation for the difference in view dependence for faces and point walkers is that facial expressions involve elastic deformations, whereas point walkers articulate but remain locally rigid. However, the point light studies to date have not required subjects to perform a task in which subtle patterns of movement have to be encoded and distinguished from similar complex patterns of movement. We can certainly encode subtle dynamic information from point light displays (Pollick, Paterson, Bruderlin, & Sanford, 2001) and it would be interesting to see whether delayed match to sample tasks for subtle gestures reveal the view independent encoding in point light displays that we found for the 3-D face animations in this study.

It has also been suggested recently on the basis of experiments with limited lifetime point walkers that motion from form is used to recover the direction of walkers while image motion information can be used to carry out detection of point walkers (Beintema & Lappe, 2002). Beintema and Lappe propose that we encode the structural changes in the posture of the body—variation in the structure or form. We take a similar view in emphasizing the encoding of change within a parameterizable model of face shape in which the time dependent sequence of parameters encodes the facial expression.

It has been shown recently that temporal correlation of views of the face undergoing rigid rotation can affect judgements of facial identity. When a face morphs as it rotates, such that identity changes with pose, observers often miss the identity transformation. Views of faces associated in this way are more difficult to distinguish as separate faces than if the same views are presented in a random order (Wallis & Bülthoff, 2001). It has been suggested that temporal correlation may play a large role in associating separate views of the face and building up spatiotemporal signatures in general (Stone, 1999). We have found that generalization of rigid facial motion is dependent on the similarity of viewpoint, which is consistent with view-based generalization through spatio-temporal association. It is possible the continuity of action over viewpoint, for example from a speaking face that rotates, might play a similar role in linking views of the rotating face. However, in the experiments reported here subjects are asked to match patterns of movement in two views without the opportunity to see a linking rigid rotation. The problem in matching the motion fields generated by both rigid and nonrigid motion seen from two views is conceptually no simpler a problem than that of matching spatial pattern from two views. Nevertheless, nonrigid motion is distinguished from rigid motion in supporting viewpoint invariance in most instances. This would suggest that there is less need to link experienced views of nonrigid motion through spatiotemporal association. As nonrigid motion is an object-centred deformation our great expertise with faces may allow us to encode nonrigid deformations as referenced to the face, giving rise to an essentially 3-D representation of the patterns of motion available in individual faces.

Object-based motion can be described as object change defined with respect to the parameter space that determined the allowable transformations of that object. In this perspective one learns the natural patterns of variation of objects in an abstract view-independent manner and what allows generalization between views is recognition of the pattern of change that has a common cause in the structural change of the object (Johnston, 1992). This approach can be contrasted with the idea that viewpoint invariance results from the association of the different appearances of the object when seen from different views. The finding of viewpoint invariance that is specific to object-based motion supports the existence of these abstract codes for facial motion.

To conclude, it seems that the currently prevailing theories of object recognition will in the future need to account for not only patterns of static object recognition but also those of object motion and the interaction between motion and spatial information. In many instances object motion is as important a vehicle of information as object shape and indeed representations of object shape, even those applied to static views, might rely on knowledge about object transformations for their establishment.

REFERENCES

Aggarwal, J. K., Cai, Q., Liao, W., & Sabata, B. (1998). Nonrigid motion analysis: Articulated and elastic motion. *Computer Vision and Image Understanding*, *70*, 142–156.

Bassili, J. N. (1979). Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face. *Journal of Personality and Social Psychology*, *37*, 2049–2058.

Beintema, J. A., & Lappe, M. (2002). Perception of biological motion without image motion. *Proceedings of the National Academy of Science, USA*, *99*, 5661–5663.

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

Bülthoff, I., Bülthoff, H. H., & Sinha, P. (1998). Top-down influences on stereoscopic depth perception. *Nature Neuroscience*, *1*, 245–247.

Christie, F., & Bruce, V. (1998). The role of dynamic information in the recognition of unfamiliar faces. *Memory and Cognition*, *26*, 780–790.

Hill, H., Jinno, Y., & Johnston, A. (2003). Comparing solid body with point-light animations. *Perception*, *32*, 561–566.

Hill, H., & Johnston, A. (2001). Categorizing sex and identity from the biological motion of faces. *Current Biology*, *11*, 880–885.

Hill, H., Schyns, P. G., & Akamatsu, S. (1997). Information and viewpoint dependence in face recognition. *Cognition*, *62*, 201–222.

Johansson, G. (1975). Visual motion perception. *Scientific American*, *232*, 76–88.

Johnston, A. (1992). Object constancy in face processing: Intermediate representations and object forms. *Irish Journal of Psychology*, *13*, 426–439.

Jordan, T. R., & Bevan, K. (1997). Seeing and hearing rotated faces: Influences of facial orientation on visual and audiovisual speech recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 388–403.

Jordan, T. R., & Thomas, S. M. (2001). Effects of horizontal viewing angle on visual and audiovisual speech recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 1386–1403.

Knight, B., & Johnston, A. (1997). The role of movement in face recognition. *Visual Cognition*, *4*, 265–273.

Lander, K., Christie, F., & Bruce, V. (1999). The role of movement in the recognition of famous faces. *Memory and Cognition*, *27*, 974–985.

Newell, F. N., Chiroro, P., & Valentine, T. (1999). Recognizing unfamiliar faces: The effects of distinctiveness and view. *Quarterly Journal of Experimental Psychology*, *52A*, 509–534.

O'Toole, A. J., Edelman, S., & Bülthoff, H. H. (1998). Stimulus specific effects in face recognition over changes in viewpoint. *Vision Research*, *38*, 2351–2363.

Pike, G., Kemp, R., Towell, N., & Phillips, K. (1997). Recognizing moving faces: The relative contribution of motion and perspective view information. *Visual Cognition*, *4*, 409–437.

Pollick, F. E., Paterson, H., Bruderlin, A., & Sanford, A. J. (2001). Perceiving affect from arm movement. *Cognition*, *82*, B51–B61.

Stone, J. V. (1999). Object recognition: View-specificity and motion-specificity. *Vision Research*, *39*, 4032–4044.

Tarr, M. J., & Bülthoff, H. H. (1998). Image-based object recognition in man, monkey and machine. *Cognition*, *67*, 1–20.

Tarr, M. J., & Cheng, Y. D. (2003). Learning to see faces and objects. *Trends in Cognitive Sciences*, *7*, 23–30.

Thornton, I., & Kourtzi, Z. (2002). A matching advantage for dynamic human faces. *Perception*, *31*, 113–132.

Troje, N. F., & Bülthoff, H. H. (1996). Face recognition under varying poses: The role of texture and shape. *Vision Research*, *36*, 1761–1771.

Troje, N. F., & Bülthoff, H. H. (1998). How is bilateral symmetry of human faces used for recognition of novel views. *Vision Research*, *38*, 79–89.

Verfaillie, K. (1993). Orientation-dependent priming effects in the perception of biological motion. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 992–1013.

Verfaillie, K. (2000). Perceiving human locomotion: Priming effects in direction discrimination. *Brain and Cognition*, *44*, 192–213.

Vetter, T., & Troje, N. (1997). Separation of texture and shape in images of faces for image coding and synthesis. *Journal of the Optical Society of America*, *14*, 2152–2161.

Wallis, G., & Bülthoff, H. H. (2001). Effects of temporal association on recognition memory. *Proceedings of the National Academy of Science, USA*, *98*, 4800–4804.

Manuscript received May 2003
Manuscript accepted September 2004
PrEview proof published online 12 August 2005