Perception, 2006, volume 35, pages 761–773

Learning the moves: The effect of familiarity and facial motion on person recognition across large changes in viewing format

Dana A. Roark*, Alice J. O'Toole*, Hervé Abdi*, & Susan E. Barrett †

* School of Brain and Behavioral Sciences, The University of Texas at Dallas, Richardson, TX 75083-0688, USA.
† Lehigh University.

Abstract.
Familiarity with a face or person can support recognition in tasks that require generalization to novel viewing contexts. Using naturalistic viewing conditions requiring recognition of people from face or whole body gait stimuli, we investigated the effects of familiarity, facial motion, and direction of learn/test transfer on person recognition. Participants were familiarized with previously unknown people from gait videos and were tested on faces (Experiment 1a) or were familiarized with faces and were tested with gait videos (Experiment 1b). Recognition was more accurate when learning from the face and testing with the gait videos, than when learning from the gait videos and testing with the face. The repetition of a single stimulus, either the face or gait, produced strong recognition gains across transfer conditions. Also, the presentation of moving faces resulted in better performance than static faces. In Experiment 2, we investigated the role of facial motion further by testing recognition with static profile images. Motion provided no benefit for recognition, indicating that structure-from-motion is an unlikely source of the motion advantage found in the first set of experiments.

1. Introduction

Individuals become familiar as we meet and re-meet them across a wide range of contexts, under different viewing conditions, and at times when different portions of their face and body are in view. Moreover, in all cases, except when the person is introduced to us through a photograph, we learn the identity of a person in motion. We eventually come to recognize individuals under a varied set of viewing conditions that include motions of all sorts. How is it possible to extract the individuating information about a person from diverse encounters that seemingly share few similarities and are punctuated with motion information that carries its own social message?

Familiarity with a face or person seems to be a key prerequisite for the impressive recognition skills we display under variable viewing conditions. In fact, the robust recognition invariance that characterizes *familiar* face recognition does not hold for *unfamiliar* faces. Changes in the facial image between learning and test, including changes in viewpoint (e.g., Troje & Bülthoff, 1996, O'Toole, Edelman, & Bülthoff, 1998), illumination (e.g., Braje, Kersten, Tarr, & Troje, 1999; Hill, Schyns, & Akamatsu, 1997), and distance of the viewer (Wagenaar & van der Schrier, 1996), although easily-handled for familiar faces, are especially problematic when participants must recognize newly learned faces (Bruce & Burton, 2002; see Hancock, Bruce, & Burton, 2000; Zhao, Chellapa, Phillips, & Rosenfeld, 2003 for reviews).

There are analogous differences when recognizing familiar versus unfamiliar people in motion. For example, Burton, Wilson, Cowan, and Bruce (1999) compared recognition accuracy between participants who were either familiar or unfamiliar with people presented in video clips. Participants initially viewed the targets from low-resolution video clips. When asked subsequently to pick out faces from high quality color photographs, participants who were unfamiliar with the targets performed poorly, whereas participants who were familiar with the targets performed the task easily. In trying to understand the source of the information used by participants familiar with the targets, Burton et al. edited the videos to obscure either the face or body/gait. They reported more accurate recognition from the "gait-obscured" videos than from the "face-obscured" videos. This suggests that participants relied primarily on the face for recognition, despite the poor-quality video in which the target faces were small and difficult to see. However, because the familiar participants in Burton et al.'s study entered the experiment already knowing the faces well, it is not possible to determine the kinds or amount of experiences that led to success in matching people across these large changes in viewing conditions. A similar caveat applies to studies that use famous faces as "familiar" stimuli. Both famous faces and faces of friends and acquaintances are likely to have been experienced by the participants across a wide range of viewing conditions, expressions, and motions. Other transformations with a much longer timeline, such as those involved in ageing, may also help the viewer extract information that aids face recognition across successive encounters.

There are also qualitative differences between the recognition of familiar and unfamiliar faces in motion. Several studies have indicated that although motion is beneficial for recognizing familiar faces, (e.g., Knight & Johnston, 1997; Lander, Christie, & Bruce, 1999; Lander & Bruce, 2000; Lander, Bruce, & Hill, 2001; Lander & Chuang, 2005), it is of questionable value for learning and recognizing unfamiliar faces (cf. Pike, Kemp, Towell, & Phillips, 1997; Lander & Bruce, 2003; and Thornton & Kourtzi, 2002 who reported a motion advantage and Bruce, Henderson, Greenwood, Hancock, Burton, & Miller, 1999; Bruce, Henderson, Newman, & Burton, 2001; and Christie & Bruce, 1998, who reported no motion advantage). Combined, the studies suggest that familiarity may mediate the usefulness of motion as an identity cue, such that motion may take on a progressively more important role in recognition as one becomes familiar with a person. Thus, a qualified conclusion regarding the role of motion in recognition is that participants require sufficient experience with a face to benefit from motion information (O'Toole et al., 2002; Roark, et al., 2003).

Despite the empirically established importance of familiarity with faces and people for predicting the extent to which recognition can operate invariantly over a wide range of viewing conditions, familiarity is rarely manipulated in the learning phase of laboratory studies. This is somewhat surprising given the presumed importance of familiarity for recognizing people in motion and over large changes in viewing conditions. Although two previous experiments using *static* faces have included familiarity as a variable (Liu & Chaudhuri, 2002; see also Clutterbuck & Johnson, 2002), only one previous study has varied familiarity using moving faces. In that study, Bruce, Henderson, Newman, and Burton (2001) varied the number of times viewers saw a 30 sec. video clip of each target during the learning phase. In these clips, the targets rotated their heads, nodded, and smiled. Participants were asked subsequently to make match/no-match decisions from still images extracted from the videos that were paired with high quality photographs of targets or distractors. Viewing each video twice was not sufficient to improve participants' performance over a single exposure to the video. Thus, either pure repetition is an ineffective familiarizing experience, or more exposures, or experience with different kinds of images/videos may be needed to bridge these photometric gaps.

Interestingly, in an additional experiment, Bruce et al. (2001) showed that socially engaging with the face during a familiarization phase improved matching performance. In that experiment, some participants viewed the videos in pairs and were instructed to “chat about the faces” while other participants viewed the faces in isolation. The “deep” versus “shallow” processing variable affected matching performance over changes in photometric viewing conditions in much the same way as it affects recognition of high resolution still images (Bower & Karlin, 1974). A brief social engagement may enhance familiarity over what might be predicted from simple exposure time (see Roark, Barrett, Spence, Abdi & O’Toole, 2003, for a discussion of the role of social engagement in face learning.)

In the present study, we explored how familiarity and facial motion affect recognition of newly-learned faces. We also looked at the direction of transfer between face and gait stimuli. Thus, three variables were manipulated: level of familiarity with the target images (number of viewings during learning), presentation format of the face stimuli (moving versus static images), and direction of transfer between learning and test (face to gait versus gait to face).

For the familiarity variable, we focused on quantitative variations in the form of pure repetition with a single familiarizing stimulus. These include qualitative variations in views, illumination conditions, and so forth, as well as quantitative variations in exposure frequency. Given the limited extent of previous data on this question, we chose to focus the present effort on quantitative variations in the form of pure repetition with a single familiarizing stimulus. This stimulus must serve as the basis for recognizing the person under substantively different viewing conditions. This provides a strong control on the kind of access individuals have to faces during learning.

For the motion manipulation, we varied the presentation of faces as moving versus static, following the study of Bruce et al. (2001). The face motions we used, however, were more natural than those used by Bruce et al. Their videos showed a person filmed in a rotating chair, which afforded continually changing views of the face across 360 degrees. In the final segment of the video, the model looked up, looked down, faced the camera, and smiled. It is possible that the more posed aspects of Bruce et al.’s videos may have limited the extent to which their participants may have socially engaged with the videos. In the present study we employed videos that depicted the targets engaged in an unscripted conversation with an “out of view” experimenter. As a result, the videos included natural facial expressions, lip and eye movements, along with some head tilting. By using this type of facial motion, our participants viewed faces in a context that approximates a real-life encounter with another individual (see samples in Figure 1).

For the direction of transfer variable, the learning and test conditions varied between high resolution close-up views of faces and surveillance-like videos taken at a distance. The distance videos included information about the face, body, and gait of an individual. We used “gait videos” because they closely resemble the information captured on surveillance videos. The close-up views of faces were used because they provide the kind of information one might expect from a mug-shot image available for a watch-list suspect**.** From an information perspective, the only common information about the person present in the learning and test stimuli is the face itself. Burton et al. (1999) found that people previously familiar with the targets relied primarily on the face in recognizing the targets from low-resolution gait videos. This was the case despite the fact that participants’ previous experience with the targets included information about both the face and the body/ gait. For the purposes of the present study, this finding suggests that learning the face from a close-up image might be a more effective strategy than learning the face from a distant, lower resolution gait video.

2. **Experiments 1a and 1b**

Experiments 1a and 1b differed only in the direction of transfer from learning to test stimuli. In Experiment 1a, participants learned targets from the gait stimuli and were tested for recognition

FIGURE 1. Examples of the moving and static stimuli used in the study. Top left: Static frontal view of a face. Top right: Static profile view of a face. Middle: Images from a facial speech video. Bottom: Images from a gait video.

with either face-only videos or face-only static images. In Experiment 1b, participants learned from either the static or moving face-only stimuli and were tested for recognition with the gait stimuli. For brevity and convenience, we describe and analyze these experiments as a single unit.

2.1. *Method.*

2.1.1. *Participants.* Eighty-four undergraduate students enrolled at The University of Texas at Dallas (UTD) participated in each experiment. Students were compensated with a research credit in one of the psychology courses they were taking at UTD.

2.1.2. *Design.* We implemented a standard old/new recognition task in both experiments. Familiarity was manipulated as a within-participants variable, defined as the number of exposures (1, 2, or 4) to the learning stimuli (gait videos for Experiment 1a; face-only images for Experiment 1b). Presentation type of the face-only stimuli (moving vs. static faces) varied as a between-participants factor for the test trials in Experiment 1a and for the learning trials in Experiment 1b. Recognition accuracy was measured as d'.

2.1.3. *Stimuli.* The stimuli used for this experiment were taken from a database of moving and static faces collected at the Vision Lab at UTD (O'Toole, Harms, Snow, Hurst, Pappas, Ayyad, & Abdi, 2005)[1]. The participants recruited for the database were undergraduate students enrolled in psychology courses at UTD.

Face stimuli. The face stimuli consisted of digitized color photographs and digital video clips of 60 Caucasians (30 men; 30 women) in their twenties. The photographs used for the "static" condition and the digital video used for the "moving" condition were taken with a Canon XL1 camera using 18% gray, high quality, wide toned seamless paper as a background. The camera was placed six feet in front of the participant with the height determined by the participant's eye-level. Multiple light sources were used to approximate ambient illumination. Participants were filmed wearing a gray drape over their clothing and were asked to remove any noticeable jewelry. Hairstyles of the participants were not altered. The participant filled approximately 50% of the

[1]For examples of the video stimuli refer to
`http://www.utdallas.edu/dept/bbs/FACULTY_PAGES/otoole/database.htm`.

frame, with the head entirely within the frame. The background was visible between the top of the head and the top of the frame, and some shoulder was also visible.

The static images show the person looking directly toward the camera with a neutral expression (see Figure 1 for an example). These images were captured and saved in an image file to a resolution of 720 × 480 pixels. The moving images—"facial speech videos"— were 5-second clips showing the person engaged in a conversation with an experimenter, who is standing out of view, positioned directly behind the camera. The experimenter asked the participants mundane questions to elicit the facial speech contained in these videos (e.g., "What classes are you taking this semester?"). This facial speech movement consisted primarily of non-rigid movements (e.g., lip and mouth movements, expressive gestures), although occasionally rigid movements (e.g., head tilting and nodding) were visible (see Figure 1). Video sequences were saved in digital video format with 720 × 480 pixels, using a video rate of 29.97 frames per second in NTSC (National Television System Committee) format.

Gait stimuli. The gait videos were filmed in a building foyer with high ceilings, enclosed entirely on one side with glass windows. This environment approximates outdoor lighting and makes for variable lighting conditions across the set of videos because the position and intensity of the light (mostly the sun) varies on a stimulus-by-stimulus basis. The 9-second videos depict participants walking parallel to the line of sight of the camera starting at a distance of 10 meters away. The person is shown walking toward the camera, but then veers off to the left in the final few paces (see Figure 1). Due to the lighting variability and relatively short temporal exposure to faces, recognition from these videos is challenging. See Figure 1 for sample stimuli.

2.1.4. *Procedure.*

Experiment 1a. Participants (N = 84: 72 females, 12 males) were given an overview of the experiment and were informed that they would be asked to recognize the people they viewed in the learning stage of the experiment. Participants viewed 30 (15 males; 15 females) gait videos during the learning phase. Each video lasted 9 seconds with an inter-stimulus interval of 200 ms. Participants viewed a third of the faces once ("1-exposure" condition), a third of the faces twice ("2-exposure" condition), and a third of the faces four times ("4-exposure" condition). Thus, each participant saw a total of 70 videos, comprised of 10 different faces in each of the 1-exposure, 2-exposure, and 4-exposure presentation formats (i.e., 10 videos viewed once; 10 videos viewed two times; 10 videos viewed four times). Counterbalancing was implemented to assure that, across all participants, each video appeared equally often in the 1-, 2-, or 4-exposure conditions.

The test trials began after a 10-minute break that followed the learning trials. Each participant was assigned randomly to one of two test conditions. For the "static" condition (n = 42), participants viewed 60 (30 old; 30 new) still frontal images of faces with a neutral expression presented for 5 seconds. For the "moving" condition (n = 42), participants viewed 60 (30 old; 30 new) 5-second video clips of speaking faces. After each face was presented, participants were instructed to indicate whether or not the person was one they had seen during the learning trials by pressing a key on the keyboard (1 for "old", 3 for "new"). The entire experiment took approximately 25 minutes. All experimental events were controlled by a Macintosh G4 computer programmed with Psyscope (Cohen, McWhinney, Flatt, & Provost, 1993).

Experiment 1b. This experiment was similar to Experiment 1a, except that participants (N = 84, 63 females, 21 males) were assigned randomly to learn either the moving or static faces and were tested with the gait videos (27 old, 27 new).

2.2. *Results.* Hit and false alarm rates were compiled for each participant in each condition and at each level of familiarity. We computed d's from these scores and submitted them to an analysis of variance (ANOVA) with direction of transfer (gait to face, face to gait), presentation type of the face stimuli (moving or static), and familiarity (1, 2, or 4 exposures during learning) as variables. An alpha level of.05 was adopted for all analyses.

Main effects were found for all three factors (see Figures 2a and 2b). Participants performed more accurately when they learned faces and were tested on the gait videos than when they learned from the gait videos and were tested on the faces, $F(1,164) = 9.21$, $MSE = 8.1$, $p < .01$. Participants performed more accurately when the face-only stimuli were in motion, $F(1,164) = 4.21$, $MSE = 3.71$, $p < .05$. Familiarity with the person improved performance, $F(2,328) = 83.0$, $MSE = 19.98$, $textitp$<.01.

None of the two-factor interactions was significant. At the simplest level of interpretation, the lack of an interaction between presentation type and familiarity fails to support the hypothesis that motion takes on an increasingly important role with increased familiarity. However, the triple interaction, between direction of transfer, presentation type, and familiarity proved significant, $F(2,328) = 3.1$, $MSE = .74$, $p < .05$, qualifying this conclusion. In fact, the pattern of means is consistent with an increased role for motion with familiarity only for the face to gait condition (see Figures 2a and 2b). We dissected the triple interaction more carefully by using pairwise comparisons with the Tukey test. Specifically, in the face to gait condition, where the pattern of means is consistent with the hypothesis that motion becomes more important with familiarity, we found that the difference between the moving and static conditions was significant for the four-exposure familiarity condition ($p < .01$), but not for the one- or two-exposure conditions. This suggests an increasing role for motion with familiarity for the face to gait recognition condition.

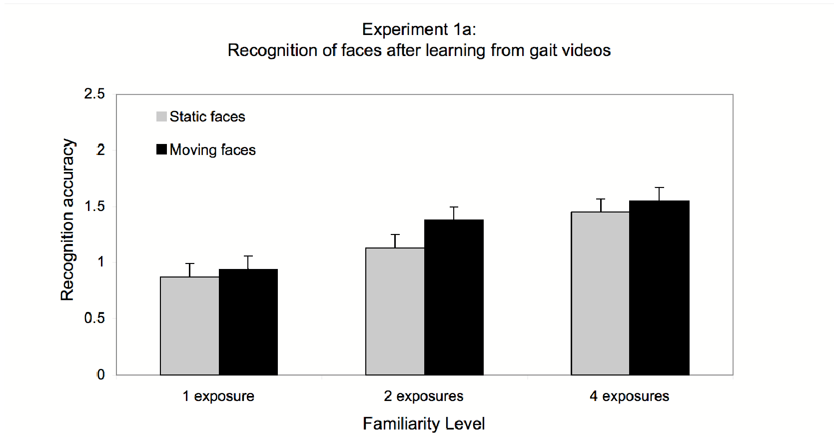
2.3 Discussion

The results of this first set of experiments indicate that additional exposures to the targets during learning improved recognition performance from a novel viewing format. Given that only a single set of viewing conditions was available from the learning stimulus, this familiarity advantage is noteworthy. This result also suggests that it may be necessary to include a 4-exposure condition in order to establish a familiarity benefit. Bruce et al. (2003) tested the effect of familiarity by using 1- versus 2-exposures to the faces during learning and reported no difference in subsequent recognition performance.

There was also a clear advantage for learning a person from the face and being tested with the gait video over learning the person from the gait video and being tested with the face. This result was obtained even given the shorter viewing times for the face (5-seconds) versus the gait video (9-seconds). Because the face is the only common component of the learning and test stimulus pairs, the face-to-gait advantage suggests that viewers benefit from the availability of high-resolution information about a face during learning. The results suggest further that the face can provide the viewer with a robust form of identity information capable of supporting recognition generalization to novel viewing conditions. The high-resolution face images we used contain fine-scale spatial information and facial details that are not available from the coarser-scale gait videos. This difference in image content likely contributed to the encoding advantage we found for the face-to-gait trials (see also, Liu, Collin, Rainville, & Chaudhuri, 2000).

The third finding was that presentation of the facial speech videos resulted in better recognition performance than presentation of the static facial images. The large differences in image format between the face and the gait stimuli make this result interesting from an information point of view.

What advantage do moving faces provide in a recognition task where the learn-test images differ so dramatically? From an information perspective, motion might benefit recognition in two non-exclusive and potentially complementary ways (O'Toole et al., 2002, Roark et al., 2003). First, it is possible that face/body motions contain "dynamic identity signatures" that can be matched when the learning and testing stimuli both contain these signatures. There is good evidence that dynamic identity signatures exist and can mediate face recognition (Hill and Johnston, 2001; Knappmeyer, Thornton, & Bülthoff, 2003). However, the facial details needed to generate an identity signature (e.g., idiosyncratic facial gestures contained in head, eye or lip movements)



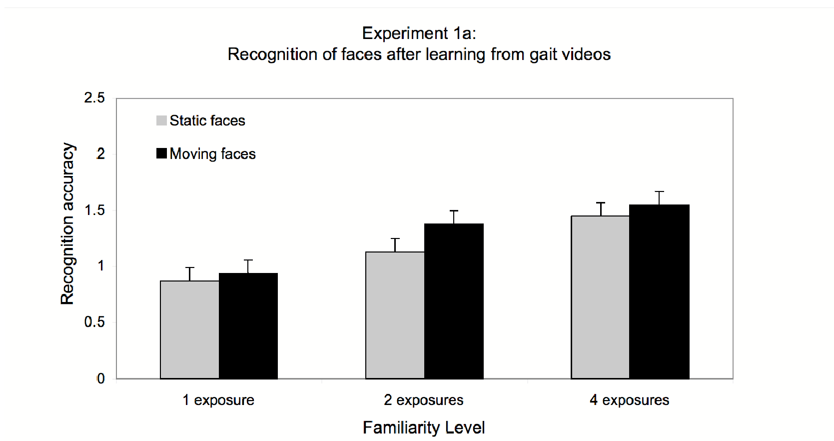


FIGURE 2. (A). Results from Experiment 1a. Recognition performance for the moving face and static face test conditions as a function of familiarity. Participants learned the targets from the gait videos. (B). Results from Experiment 1b. Recognition performance for the moving face and static face learning conditions as a function of familiarity. Participants were tested for recognition using the gait videos.

are available in the facial speech videos, but not in the lower-resolution gait videos. In fact, few facial movements are visible in the gait videos. Thus, the large image differences between these two types of stimuli make it unlikely that dynamic identity signatures can explain the advantage we found.

A second more likely source of the motion advantage is perceptually based structure-from-motion processing. Specifically, seeing the face in motion may allow for a more accurate perception of the three-dimensional structure of the head, resulting in better recognition performance. One caveat for this explanation is that these processes would have to apply and be useful both at learning *and* at test. The extra perceptual information in a moving image, which gives some indication of the three-dimensional structure of the person's face, might be particularly beneficial for recognizing the person over substantial changes in viewing conditions. If learning a face from a moving image provides qualitatively better information about the 3D structure of a face than learning from a static image, better recognition performance should be expected when the test image offers a novel perspective view of the face. We test this hypothesis in Experiment 2.

Before proceeding, we note that an additional alternative explanation of the motion benefit is that motion provides a social signal that helps participants attend to the face in ways that benefit recognition (cf., the "motion as social signal hypothesis" from Roark et al., 2003). We will consider this issue in the General Discussion.

3. **Experiment 2**

In Experiment 2, we investigated whether the motion advantage we found for generalizing recognition between face and gait videos in Experiment 1 would also apply to another type of recognition generalization: from full face to static profile. A common finding for unfamiliar face recognition is the decline in recognition accuracy with increasing differences in the pose or view between the learning and test stimuli (e.g., Liu & Chaudhuri, 2002; O'Toole et al., 1998; Troje et al., 1996). In this experiment, participants learned faces from either static frontal-view images or from facial speech videos. For the recognition test, all participants viewed static profile facial images of targets and distractors. We hypothesized that if the facial motion from speech provides structure-from-motion information that is beneficial for recognition, a motion advantage should occur in the current experiment, as was the case previously. Similar to Experiment 1, we also varied participants' familiarity with the faces. In sum, this experiment was identical to Experiment 1b, but with static high-resolution face profile images serving as the test stimuli.

3.1. *Method.*

3.1.1. *Participants.* Fifty-four (25 females, 29 males) undergraduate students enrolled at Lehigh University participated in the experiment as part of a class requirement.

3.1.2. *Stimuli.* . The learning stimuli were identical to those used in Experiment 1b. Participants learned faces from either the facial speech videos or the static frontal images. The test stimuli consisted of 54 digitized static images (27 "old," 27 "new"), which were close-up, left profile views (i.e., 90 degrees from frontal) of the subjects. The profile images were captured under the same controlled illumination conditions as the learning stimuli. The stimuli for all the experiments were created under identical conditions at UTD. Refer to Figure 1 for examples of the faces used in the learning and test trials.

3.2. *Procedure.*

The procedure was identical to Experiment 1b with the exception of the test stimuli.

3.3. *Results.*

A two-factor ANOVA was conducted, with one within-participants factor of familiarity and one between-participants factor of learning format (moving vs. static). Mean d' values obtained across conditions appear in Figure 3.

Learning from the moving faces versus the static faces offered no advantage for recognizing the profile images, $F(1, 51) = 1.21$, $MSE = 1.35$, *ns.* In fact, as can be seen in Figure 3, all three means favored the static faces over the moving faces. Similar to the familiarity advantage reported in Experiment 1, recognition performance from the profile images increased with face familiarity, $F(2, 102) = 34.7$, $MSE = 10.1$, $p < .01$, indicating that increased familiarity with frontal-view images of the faces transferred to recognition of the faces from profile views. The interaction between presentation type and familiarity was not significant, $F(2, 102) = .42$, $MSE = .12$, *ns.*

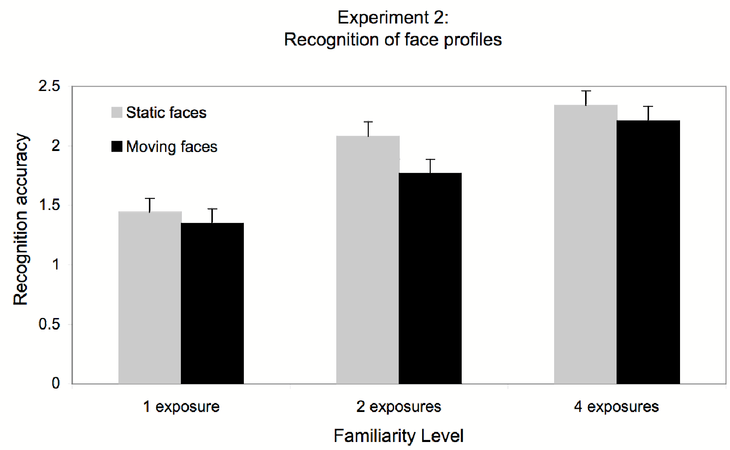


FIGURE 3. Results from Experiment 2. Recognition performance for the moving and static learning conditions as a function of familiarity. Participants were tested for recognition with static profile images.

3.4. *Discussion.*

We found no advantage for learning moving versus static faces in recognizing static profile images, making it unlikely that the benefit of facial motion observed in Experiments 1a and 1b stems from structure-from-motion processing. In fact, the animated facial speech videos tended toward less accurate recognition performance than the static facial images. Thus, either motion is not helpful in this particular task, or other factors (e.g., the social content of facial motions, task difficulty) mediate its potential benefits.

A recent study by Watson, Johnston, Hill, and Troje (in press) may shed light on our results. In that study, Watson et al., using animated synthetic 3D head models as stimuli, showed that participants could recognize, from a rotated view, facial motion that had previously been presented at a full face view. In addition, the authors reported that non-rigid motions resulted in slightly better performance than animations that included rigid motion. Unfortunately, a direct comparison between the results of Watson et al. and our results is difficult. We directly compared recognition performance across learning trials that included either moving faces or static faces, whereas Watson et al. compared recognition across two moving formats. Taken together, however, the two sets of results suggest that it may be easier to obtain a motion advantage in a recognition task when a moving image is available both at learning and at test. Recall that in Experiments 1a and 1b of the present study, recognition performance was better in trials where the transfer condition included two moving images (i.e., gait to facial speech or facial speech to gait) rather than when the transfer involved a moving and a static image (i.e., gait to static face or static face to gait). This pattern of results may indicate that bridging photometric gaps is easier between two moving images than between a static and a moving image. We will return to this point in the General Discussion.

The results of Experiment 2 also replicate the familiarity benefit found in the first set of experiments. Pure repetition of a single familiarizing stimulus, either a moving or static face, was sufficient to improve recognition from a novel test format, in this case, a static profile image. As was the case in Experiment 1a and 1b, the difficulty typically associated with recognizing a newly learned face from a novel viewing context is relatively easy to overcome with a simple repetition paradigm. Additional viewings seem to increase the strength of the memory trace of a face, which in turn can support large gains in recognition performance.

4. General Discussion

The novel findings of this study can be summarized as follows. First, repetition-based familiarity with a face/person can improve recognition generalization to novel viewing conditions. Second, recognition performance is better when we learn a person from a face image and are tested with a surveillance-style gait video than when we learn from a gait video and are tested with a face. Third, in recognition tasks that require participants to bridge photometric gaps between surveillance videos and face images, performance is better when the faces are viewed in motion. We discuss each of these results, in turn.

4.1. *Familiarity advantage.* Our results reveal a strong role for pure, un-elaborated repetition in improving recognition performance across large changes in viewing format. Recognition accuracy improved with increased exposure to the learning stimulus in three transfer conditions: gait-to-face (Experiment 1a), face-to-gait (Experiment 1b), and frontal face-to-profile (Experiment 2). This set of findings speaks to the relative ease with which a basic familiarization procedure can yield gains in recognition performance over photometric changes. Repeated exposure to a target stimulus during learning can significantly strengthen the memory trace for a face or person, enabling successful recognition, even from a novel test image.

The familiarity advantage we report here is based on repeated exposures to the same stimulus. Thus, in this set of experiments we defined familiarity *quantitatively*. Future work should examine face learning using a more *qualitative* definition of familiarity. For example, it would be useful to know whether the recognition gains we found with pure repetition are replicable in an experiment where participants learn faces from a variety of different viewing formats (e.g., a combination of different viewpoints, expressions, facial speech, etc). Moreover, designs in which participants learn the targets from multiple formats would emulate more precisely the way people encounter and learn to recognize others in natural settings.

4.2. *Face-to-gait versus gait-to-face.* We found that learning a person from a high quality face stimulus enabled better recognition from the lower resolution gait videos than vice versa. This result implies that faces contain a robust type of identity information that people may use in more general person recognition contexts. In fact, the unequal presentation times of the gait learning stimuli versus the face-only learning stimuli further testify to the superiority of the face over whole-body as a learning stimulus. The presentation time for the gait videos was nearly twice as long as the presentation time for the face-only stimuli, yet recognition performance was still better with the face-only stimuli.

This result is consistent with Burton et al.'s (1999) finding that people relied on the face, more so than cues from the body/gait, for recognizing individuals they knew well. The present finding extends Burton et al.'s work by providing an important control on the conditions through which familiarity was achieved. In Burton's study, participants were already familiar with the people in the videos and so may have had different experience with faces versus distant views of the individuals.

The face to gait advantage is also consistent with recent work by Liu, Seetzen, Burton and Chaudhuri (2003) in which the availability of a high-resolution image improved face-matching performance, even when the image-based properties of this stimulus were incongruent with the match image. Specifically, Liu et al. showed that the recognition of high- and low-resolution faces was generally equivalent when participants learned people from low-resolution gait videos. In fact, when differences were found in an analogous matching experiment, participants performed more accurately when the target and test images were mismatched in resolution than when they matched. This result occurred provided that the stimuli included a high-resolution face image.

Combined with the present results, it seems clear that a high-resolution face template in memory is useful, even if the stimulus to be recognized is degraded considerably. This indicates that the face itself, more so than a whole-body image, provides the viewer with an identity template that maps well onto other more complex scenes.

4.3. *The motion advantage.* The motion effects we found are less straightforward than the effects we found for familiarity and direction of transfer. The presentation of a moving face resulted in better recognition performance both when it appeared as a learning stimulus (Experiment 1b) and as a test stimulus (Experiment 1a). However, no advantage was found for learning moving faces when the test stimulus was a static profile image of the face (Experiment 2). Consistent with previous literature, obtaining a motion advantage in a recognition experiment seems to hinge on the types of stimuli and tasks used. Some studies with unfamiliar faces have reported a motion benefit (e.g., Lander & Bruce, 2003; Pike et al., 1997) and others have reported no benefit (e.g., Bruce et al., 2001; Christie & Bruce, 1998).

The context-specific nature of motion's benefit to recognition is perhaps the most important open issue relating to the topic of person recognition in naturalistic conditions. Why did facial motion help in Experiments 1a and 1b, but not in Experiment 2? Although the present data do not provide a direct answer to this question, it is possible to eliminate some possibilities and offer some speculative suggestions that help to place the present data into a more theoretical context.

From an information point of view, motion can improve recognition by providing an observer either with dynamic identity signatures about the person or by perceptual structure-from-motion processes (O'Toole et al., 2002; Roark et al., 2003). With respect to dynamic identity signatures, given the image-based differences between the gait and face stimuli in our experiments, it is unlikely that signature-like facial motions would be common across the kinds videos we used. A structure-from-motion explanation for the motion advantage is also unlikely given the results of Experiment 2.

Alternately, from a social signal point of view (Roark et al., 2003), it remains possible that part of the motion advantage we found here is related to unknown attentional benefits to recognition that viewing a face/person in motion provide. These may be similar to the recognition benefits that Bruce et al. (2001) found when participants, tested in pairs, were instructed to talk about the faces during the learning trials.

Finally, we consider an interesting, but speculative, neural account of the data that draws on the distributed theory of face processing (Haxby et al., 2000). This theory posits a dissociation between the processing of socially relevant motion information and static form information in faces. In their model, Haxby et al. propose separate processing of the invariant and changeable aspects of faces in different areas of the brain. The invariant facial features, useful for identifying a person, are processed in the fusiform gyrus ("fusiform face area," FFA; Kanwisher, McDermott, & Chun, 1997). The changeable, motion-based aspects of faces such as gaze direction, facial expression, and facial speech, however, are processed in the posterior superior temporal sulcus (pSTS).

In reconsidering the present experiments, it is possible to characterize the motion effects we found as follows. Participants performed more accurately when the recognition task involved a motion-to-motion transfer (i.e., moving face to gait or gait to moving face) than when the task involved either a static-to-motion transfer (i.e., static face to gait) or a motion-to-static transfer (i.e., gait to static face or moving face to static profile). Thus, we suggest that the inconsistent motion advantage we found might depend on the availability of a "motion match" between the learning and test trials. The logic of this explanation is that moving faces and people may be processed in the dorsal stream in the superior temporal sulcus and possibly also the ventral stream in the fusiform gyrus, whereas the processing of static faces may be limited to the ventral stream and traditional face areas in the fusiform gyrus. If this is the case, then recognition tasks that involve moving images both at learning and test may need only to access information from one channel (i.e., the motion stream). However, recognition tasks that require participants to bridge between a moving image and a static image require cross-access between the two channels (i.e., the motion and static information streams).

This motion match explanation is a speculative, but intriguing, account of these data that requires converging support to be accepted. It is worth noting that a recent proposal advanced by Thornton and Kourtzi (2002) and Knappmeyer et al. (2003), posits that motion information is represented in the brain in a qualitatively different way than static information. In particular, they suggest that moving faces generate "dynamic representations," which, unlike static face representations, contain a temporal dimension that makes them unique. Knappmeyer and her colleagues suggest that determining the identity of a moving face might therefore rely on functional connections between the neural system for form and the neural system for motion.

Although these accounts are speculative, they warrant consideration as the more traditional and likely accounts of the motion benefits we found fail to provide a coherent explanation for the data.

In summary, pure repetition with a single familiarizing stimulus can benefit recognition transfers even over dramatic changes in viewing conditions. The next question to address is how more diverse experience with a person might benefit recognition. Recognition is better when learning from the face than from a gait video, supporting the view that person recognition relies on the face more than on other identifying cues. Although the present data show some recognition benefits for moving faces, pinning down the exact conditions under which motion helps (or hurts) recognition performance is necessary before a comprehensive account of its role in memory can be advanced.

Acknowledgements. This work was supported by funding from ONR/DARPA and TSWG awarded to A.O'T. and H.A. The authors wish to thank Sarah Snow and Josh Harms for their careful assistance with stimulus preparation and in testing participants for Experiment 1 and to Marion Lapchak and Hillery Gross for their assistance with Experiment 2. Thanks are due to reviewers for helpful comments on an earlier draft of this paper. Address correspondence to:
Dana Roark
University of Texas at Dallas
School of Behavioral and Brain Sciences
GR 4.1

References

[1] Allison T, Puce A, McCarthy G, 2000 Social perception from visual cues: role of the STS region. *Trends in Cognitive Sciences* **4** 267-278

[2] Bower G H, Karlin M B, 1974 Depth of processing pictures of faces and recognition memory. *Journal of Experimental Psychology* **103** 751-757

[3] Braje W L, Kersten D, Tarr M J, Troje N F, 1999 Illumination effects in face recognition. *Psychobiology* **26** 371-380

[4] Bruce V, & Burton M A, 2002 Learning new faces, in *Perceptual Learning* Eds M Fahle, T Poggio (Cambridge, MA: MIT Press) pp 317-334

[5] Bruce V, Henderson Z, Greenwood K, Hancock P J B, Burton A M, Miller P, 1999 Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied* **5** 339-360

[6] Bruce V, Henderson Z, Newman C, Burton A M, 2001 Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental Psychology: Applied* **7** 207-218

[7] Burton A M , Wilson, S, Cowan M, Bruce V, 1999 Face recognition in poor-quality video. *Psychological Science* **10** 243-248

[8] Christie F, Bruce V, 1998 The role of dynamic information in the recognition of unfamiliar faces. *Memory & Cognition* **26** 780-790

[9] Clutterbuck R, Johnston R A, 2002 Exploring levels of face familiarity by using an indirect face-matching measure. *Perception* **31** 985-994

[10] Cohen J D, McWhinney B, Flatt M, Provost J, 1993 Psyscope: A new graphic interactive environment for designing psychology experiments. *Behavior Research Methods, Instruments and Computers* **25** 257-271

[11] Hancock P J B, Bruce V, Burton A M, 2000 Recognition of unfamiliar faces. *Trends in Cognitive Sciences* **4** 330-337

[12] Haxby J V, Hoffman E A, Gobbini M I, 2000 The distributed human neural system for face perception. *Trends in Cognitive Sciences* **4** 223-233

[13] Henderson Z, Bruce V, Burton A M, 2001 Matching the faces of robbers captured on video. *Applied Cognitive Psychology* **15** 445-464

[14] Hill H, Johnston A, 2001 Categorizing sex and identity from the biological motion of faces. *Current Biology* **11** 880-885

[15] Hill H, Schyns P G, Akamatsu S, 1997 Information and viewpoint dependence in face recognition. *Cognition* **62** 201-222

[16] Kanwisher N, McDermott J, Chun M M, 1997 The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience* **17** 4302-4311

[17] Knappmeyer B, Thornton I M, Bülthoff H H, 2003 The use of facial motion and facial form during the processing of identity. *Vision Research* **43** 1921-1936

[18] Knight B, Johnston A, 1997 The role of movement in face recognition. *Visual Cognition* **4** 265-273

[19] Lander K, Bruce V, 2000 Recognizing famous faces: exploring the benefits of facial motion. *Ecological Psychology* **12** 259-272

[20] Lander K, Bruce V, 2003 The role of motion in learning new faces. *Visual Cognition* **10** 897-912

[21] Lander K, Bruce V, Hill H, 2001 Evaluating the effectiveness of pixelation and blurring on masking the identity of familiar faces. *Applied Cognitive Psychology* **15** 101-116

[22] Lander K, Christie F, Bruce V, 1999 The role of movement in the recognition of famous faces. *Memory & Cognition* **27** 974-985

[23] Lander K, Chuang L, 2005 Why are moving faces easier to recognize? Visual Cognition **12** 429-442

[24] Liu C H, Chaudhuri A, 2002 Reassessing the 3/4 view effect in face recognition. *Cognition* **83** 31-48

[25] Liu C H, Chaudhuri A, 2003 Face recognition with perspective transformation. *Vision Research* **43** 2393-2402

[26] Liu C H, Collin C A, Rainville S J M, Chaudhuri A, 2000 The effects of spatial frequency overlap on face recognition. *Journal of Experimental Psychology: Human Perception and Performance* **26** 956-979

[27] Liu C H, Seetzen H, Burton A M, Chaudhuri A, 2003 Face recognition is robust with incongruent image resolution: Relationship to security video images. *Journal of Experimental Psychology: Applied* **9** 33-44

[28] O'Toole A J, 2005 Psychological and neural perspectives in face recognition, *The Handbook of Face Recognition* Eds S. Li, A. Jain (Springer-Verlag)

[29] 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

[30] O'Toole A J, Harms J, Snow S L, Hurst D R, Pappas M R, Ayyad J, Abdi H, 2005 A video database of moving faces and people. *IEEE Transactions in Pattern Analysis and Machine Intelligence* **27** 812-816

[31] O'Toole A J, Roark D A, Abdi H, 2002 Recognizing moving faces: A psychological and neural synthesis. *Trends in Cognitive Sciences* **6** 261-166

[32] Pike G E, Kemp R I, Towell N A, Phillips K C, 1997 Recognizing moving faces: The relative contribution of motion and perspective view information. *Visual Cognition* **4** 409-437

[33] Roark D A, Barrett S E, Spence M A, Abdi H A, O'Toole A J, 2003 Psychological and neural perspectives on the role of motion in face recognition. *Behavioral and Cognitive Neuroscience Reviews* **2** 15-46.

[34] Thornton, I M, Kourtzi, Z, 2002 A matching advantage for dynamic human faces. *Perception* **31** 113-132.

[35] Troje N F, Bülthoff H H, 1996 Face recognition under varying pose: The role of texture and shape. *Vision Research* **36** 1761-1771

[36] Wagenaar W A, van der Schrier J H, 1996 Face recognition as a function of distance and illumination: A practical tool for use in the courtroom. *Psychology, Crime, & Law* **2** 321-332

[37] Watson T L, Johnston A, Hill H C H, Troje N F, (in press) Motion as a cue for viewpoint invariance. *Visual Cognition.*

[38] Zhao W, Chellappa R, Phillips P J, Rosefeld A. 2003 Face Recognition: A Literature Survey. *ACM Computing Surveys* **35** 399-459