Exp Brain Res (2005)
DOI 10.1007/s00221-005-0283-8

RESEARCH ARTICLE

Karin S. Pilz · Ian M. Thornton · Heinrich H. Bülthoff

A search advantage for faces learned in motion

Received: 5 January 2005 / Accepted: 20 October 2005

Abstract Recently there has been growing interest in the role that motion might play in the perception and representation of facial identity. Most studies have considered old/new recognition as a task. However, especially for non-rigid motion, these studies have often produced contradictory results. Here, we used a delayed visual search paradigm to explore how learning is affected by non-rigid facial motion. In the current studies we trained observers on two frontal view faces, one moving non-rigidly, the other a static picture. After a delay, observers were asked to identify the targets in static search arrays containing 2, 4 or 6 faces. On a given trial target and distractor faces could be shown in one of five viewpoints, frontal, 22° or 45° to the left or right. We found that familiarizing observers with dynamic faces led to a constant reaction time advantage across all setsizes and viewpoints compared to static familiarization. This suggests that non-rigid motion affects identity decisions even across extended periods of time and changes in viewpoint. Furthermore, it seems as if such effects may be difficult to observe using more traditional old/new recognition tasks.

Electronic Supplementary Material Supplementary material is available for this article at http://dx.doi.org/10.1007/s00221-005-0283-8 and is accessible for authorized users.

K. S. Pilz (✉) · I. M. Thornton · H. H. Bülthoff
Max Planck Institute for Biological Cybernetics, Spemannstr. 38, 72076 Tübingen, Germany
E-mail: karin.pilz@tuebingen.mpg.de

K. S. Pilz
Graduate School of Neural and Behavioural Sciences, Tübingen, Germany

I. M. Thornton
Department of Psychology, University of Wales, Swansea, Wales, UK

Introduction

Non-rigid movements of the human face, such as when we smile, talk or cry, have long been of interest to researchers studying the communicative and emotional aspects of social interaction (Cunningham et al. 2003; Rosenblum et al. 2002; Campbell et al. 1996; Bassili 1978; Kamachi et al. 2001). More recently, several groups have also begun to explore the role that characteristic facial motion might play in the assignment of individual identity (Hill and Johnston 2001; Knappmeyer et al. 2003; Lander et al. 1999; Thornton and Kourtzi 2002; see O'Toole et al. 2002 for a review). In the current work we explore the impact of seeing a face move on the ability to match or search for the same individual in a subsequent static test array. In particular, we were interested in whether previously reported advantages for dynamic versus static presentations of faces (Thornton and Kourtzi 2002) would also extend over longer periods of time and changes in viewpoint.

Previous studies using famous or highly familiar faces have clearly shown that non-rigid motion can facilitate recognition under a variety of non-optimal viewing conditions (Knight and Johnson 1997; Lander et al. 1999). In such studies, effects of motion are typically only found when image sequences are degraded in some way, such as by blurring or pixilating each frame, to reduce the influence of static form cues. For unfamiliar faces, in contrast, it is still unclear whether non-rigid motion can reliably affect performance. While some studies have found an advantage for dynamic over static stimuli (Lander and Bruce 2003), other studies did not find this effect (Christie and Bruce 1998; Lander and Bruce 2000).

Such contradictory findings could have arisen for a number of reasons, including differences in stimuli, degradation technique or methodological detail. Another possibility is that the nature of the experimental task being used, makes it difficult to extract stable measures of dynamic performance. Specifically, most previous studies have adopted old/new recognition tasks to investigate the issue of non-rigid motion. Such explicit recognition paradigms may be, by design, inappropriate for studying dynamic aspects of object representation. That is, such tasks may strongly bias observers to adopt memorization strategies that directly favor static, pictorial content.

Recently, Thornton and Kourtzi (2002) tried a different approach. They used an immediate matching paradigm (Sekular and Palmer 1992) to examine the impact of priming with short movie clips versus traditional still images of faces. Their task was novel in the context of facial motion for two reasons. First of all, their images were non-degraded, as compared to previous work, in which methods of form degradation had typically been employed. Second, the timescale involved was much shorter than that associated with the old/new recognition paradigms from the face literature (e.g., Christie and Bruce 1998; Lander and Bruce 2003).

In their task, observers were asked to watch an initial "prime" face—showing either a dynamic facial expression of emotion, or a single static image from the expressive sequence—and to explicitly match it for identity with a subsequently presented static target face. Using this immediate matching task, Thornton and Kourtzi (2002) found a reliable motion advantage, such that responses following the dynamic prime were consistently faster than following a static prime. Importantly, this effect appeared to be specific to identity processing, as a task using the same set of stimuli that required observers to match expression rather than identity yielded no cost or benefit for the moving images, ruling out a general alerting or arousal effect of motion. These results suggest that at least over very brief intervals, information about facial change can be used to help establish and maintain the identity of an individual.

The purpose of the current work was to investigate whether such dynamically presented information survives over longer periods of time. Specifically, we used non-degraded, unfamiliar dynamic and static images to explore the long-term, rather than the short-term impact of facial motion on identity decisions. To avoid the possible pitfalls associated with old/new recognition tasks, here, we changed both the method of learning and the method of testing long-term retention, compared to previous studies.

To familiarize observers with the target faces, we used an incidental learning phase in which two faces were presented sequentially for an extended period of time (Knappmeyer et al. 2003). One face was always moving while the other was either a single static image (Experiment 2) or a concurrently presented sequence of static images (Experiment 3). Learning two faces ensured that observers quickly became familiar with the target identities. More importantly in the current context, we tested memory for faces using a perceptual rather than a standard recognition task. Specifically, once the faces had become familiar, we used a standard visual search paradigm (see Wolfe 1998 for a review) in which observers were required to indicate the presence or absence of the target faces among a variable setsize of novel distractor faces. In our opinion, this paradigm, with its real world parallel of "finding a friend in a crowd," provides a behaviorally more relevant task than recognition from a sequentially presented list. All faces in this test phase, both targets and distractors, were presented as static images to equate the perceptual and response characteristics of the task. Our main research question was whether dynamic exposure to a previously unknown target face during the familiarization phase would affect performance in such a delayed visual search task.

As an initial step, in Experiment 1 we wanted to replicate the basic immediate matching effect of Thornton and Kourtzi (2002) using a new set of stimuli, and including changes in viewpoint (i.e., head rotation) between the prime and target faces. Changes in viewpoint between study and test are a useful manipulation as they lessen the possibility of picture matching (Bruce 1982). In addition, showing that the matching advantage is maintained across changes in viewpoint, would strengthen the hypothesis that dynamic stimuli facilitate the encoding of identity. Having validated our new set of stimuli, Experiment 2 was used to test our main research question concerning the long-term effects of viewing moving faces. Observers were familiarized with two faces, one moving and one static, and the efficiency with which they could detect the two targets was assessed in a subsequent visual search task. To assess whether change-over-time or simply additional information influenced search performance (Wallis and Bülthoff 2001), Experiment 3 contrasted arrays containing multiple static views (still shots from the video sequence) with a single repeated image and Experiment 4 used a moving sequence that was alternated with individually presented, different static views.

General methods

For use in the current experiments we constructed a new database of full-color moving faces. To do this, faces were filmed in a purpose built digital video laboratory (Kleiner et al. 2003). Five synchronized digital video cameras [25 frames/second (fps)] were positioned so that faces could be filmed from frontal view and from 22° as well as 45° on either side of the face. Cameras were raised to a height of 130 cm from the floor and positioned in a circular pattern around the head of the actors, at a distance of 130 cm. This arrangement and example stimuli are shown in Fig. 1. Previous research has shown that external features of a face like hairstyle seem to be important features used to identify unfamiliar faces (Bruce et al. 1999; Ellis et al. 1979; Young et al. 1985). To avoid any effect of external features on recognition performance and to increase the saliency of dynamic internal features, hairstyle and clothing were obscured by a black cap and a black scarf so that only face, ears and neck could be seen in the final shots, as

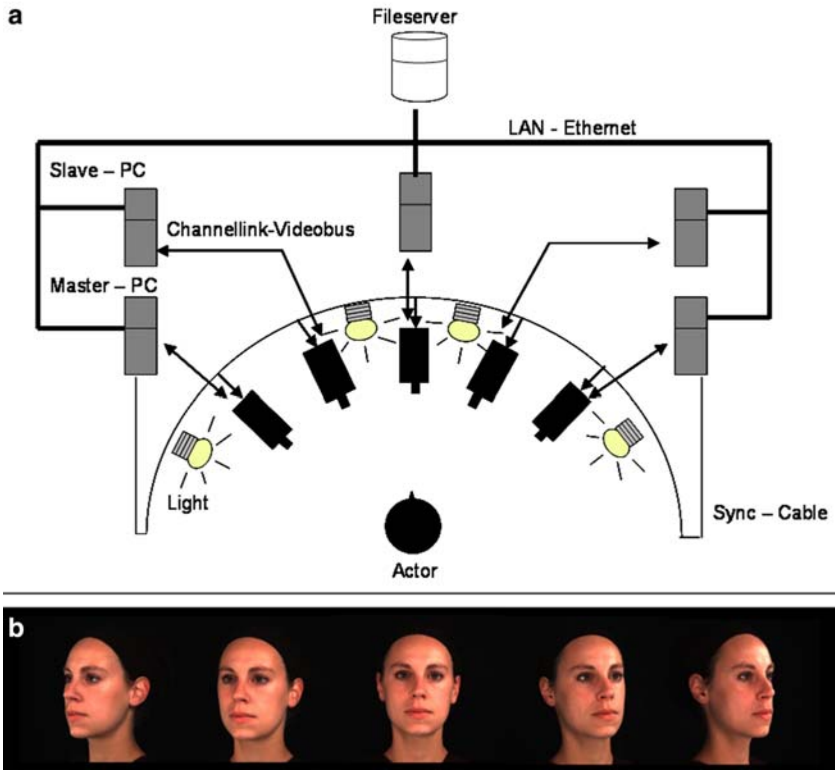

Fig. 1 **a** Schematic overview of the Max Planck video lab. **b** Example stimuli from one actor showing the five different perspectives used in the current experiments

can be seen in Fig. 1b (example stimuli can be downloaded from http://vdb.kyb.mpg.de).

Eight amateur actors were filmed making a range of isolated expressive gestures. In the current experiments, we made use of two types of clip, those relating to anger and surprise. To provide situational context for acting the different expressions, actors were asked to read short stories following the protocol described by Amaya et al. (1996).

While imagining the situation described in the stories, actors were asked to speak out loud a word fitting the emotional context. This was “what!” for the anger context and “wow” for surprise. To minimize the influence of rigid head movement, actors were trained to keep their head as still as possible throughout the recordings and only move relevant facial muscles. The final moving clips were edited to last 26 frames, i.e., 1,040 ms at a frame rate of 25 fps. The moving clips started with a neutral expression, ending with the peak of the expression in the last frame. They were cropped to be of equal size so that the head of the actor was always centered. The static pictures referred to in the experiments below are always taken from the last frame of the video sequences, except in Experiments 3 and 4, where observers were trained using multiple frames.

Participants

For all experiments students from the MPI subject pool served as observers in return of 8 euros/h. All observers had normal or corrected-to-normal vision and were naïve regarding the purpose of the experiment. Observers did not participate in more than one experiment. All observers gave their informed consent.

Apparatus

All experiments were conducted on a Macintosh G4 and a 21 in. monitor (1,152×875 pixels) with a frame rate of 75 Hz. Stimuli were presented using the Psychtoolbox extension for MATLAB (Brainard 1997). Observers were seated 60 cm from the screen.

Experiment 1

Thornton and Kourtzi (2002) found that when observers were shown a brief dynamic prime, they responded more quickly to a subsequent test image than if they had been presented with a static prime. The main purpose of Experiment 1 was to replicate this immediate matching effect with our new set of stimuli and to explore whether such effects survive changes in prime-target viewpoint. Generally, introducing changes between the nature of a prime and target image is useful to ensure that we are assessing face processing rather than identical picture matching. In the original study, all faces were shown from frontal viewpoints and the only change involved the expression that was depicted. Here, in addition to expression changes, we also varied the angle from which the prime face was shown.

Methods

Participants

Fifteen right-handed observers aged 16–39 (mean age 24.55 years) participated in this study (4 females and 11 males).

Stimuli

For this experiment we used 60 short video clips from the database, showing three male and three female actors expressing surprise and anger. Stimuli subtended a visual angle of 6.6°.

Task and design

On each trial a prime face appeared in the middle of the screen for 1,040 ms. The prime face was either a dynamic video sequence or a static picture, showing one of the two expressions, anger or surprise. The prime face was varied across one of the five different viewpoints. After a blank of 300 ms the target face appeared in the middle of the screen. The target face was always a still image of a frontal view face and showed the other of the two expressions relative to the prime.

The task of the observers was to determine if the prime and target faces belonged to the same individual. They were asked to respond as quickly and as accurately as possible using one of two marked keys. For half of the observers this was the "s" key for "same" and the "l" key for "different," with the assignment reversed for the remaining observers. Figure 2 shows the presentation sequence for a typical trial. A warning tone was used as an auditory feedback signal and appeared whenever the observers took longer than 800 ms to respond. The target face stayed on the screen until the observer responded.

The main experimental block consisted of 240 trials, half of which contained dynamic primes and half of which contained static primes. An equal number of same and different trials were distributed across these factors. For the primes, all of the six different faces and five different perspectives were used equally often. Different trials were constructed by randomly selecting pairs of sequences from different faces. The order of trials was randomized separately for each observer.

Procedure

Before running the experiment, observers were briefly familiarized with all of the faces using a picture-sorting task in which they had to sort four frontal view pictures per face according to identity. Pictures showed two neutral faces as well as one angry and one surprised face of each of the six actors. This was done to reduce error rates as our main measure of interest was response time.

The sequential matching experiment consisted of three distinct blocks, two training blocks of 20 trials and one experimental block of 240 trials. The first training block was designed to make observers comfortable with viewing the faces on the computer screen and only emphasized accuracy. A loud tone indicated whether the observer's response was correct or not. The second training block emphasized both speed and accuracy and the procedure was exactly the same as in the main experimental block. A loud tone indicated whether observers were either too slow (slower than 800 ms) or gave the wrong response. In this and all other experiments, data were only analyzed for correct responses that did not exceed more than 2.5 times the standard deviation of the mean reaction time. The whole experiment took about 45 min to complete. The order of the trials within each block was completely randomized. Each trial began automatically after the previous response.

Results

Reaction times

Figure 3 shows mean reaction times (RTs) for correct, same trials, plotted as a function of viewing angle for both the non-rigidly moving and the static primes. Here, as in all experiments, error bars denote between subject variance. Overlapping error bars are non-critical here, as all analysis relies on within-subject repeated measures design. As can be seen there is an overall reaction time

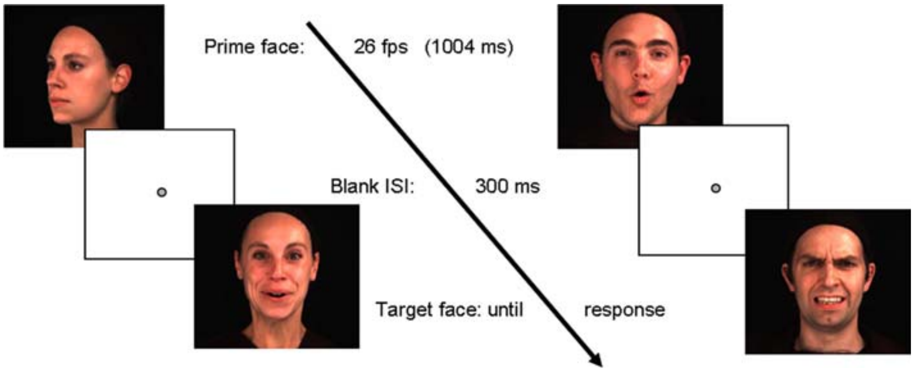

Fig. 2 Presentation sequence for Experiment 1. The prime face was either a still image shown for 1,004 ms or a 26-frame video sequence. It could be shown from one of five viewpoints. After a short retention interval, a static face was shown in frontal view. Prime and target faces always expressed different gestures

advantage of 20 ms for dynamically presented prime faces ($M = 554$ ms, SE = 18.2 ms), compared to static prime faces ($M = 574.3$ ms, SE = 19.6 ms). This advantage is consistent throughout different viewpoints. A repeated-measures ANOVA (target type × viewpoint) was carried out on RTs for correct responses as well as error rates for trials in which one of the targets was present. The ANOVA revealed a clear RT advantage for dynamic primes compared to static ones ($F_{(1,14)} = 9.3$, $P < 0.01$). There was no main effect of viewpoint and no interaction with movement condition.

Accuracy

Error rates were very low in both conditions (*dynamic*: $M = 7.4\%$, SE = 1.9; *static*: $M = 8.3\%$, SE = 1.9) and did not reveal any significant differences.

Discussion

In the condition in which observers viewed a short video sequence as a prime, their decision of whether the target face was the same or a different person was 20–30 ms faster. This finding supports the view that motion can facilitate the immediate matching of faces, as suggested by Thornton and Kourtzi (2002) for non-rigidly moving faces. Additionally, our results provide evidence that this advantage also holds across prime-target viewpoint changes (Watson et al. 2005).

It is interesting that the size of this advantage does not vary with the differences in prime-target viewpoint. However, overall reaction time scaled very little with this manipulation, as indicated by the lack of a viewpoint effect. Possibly the small angular deviations from frontal view were not sufficient to affect performance in such a simple task (see also Liu and Chaudhuri 2002; Troje and Bülthoff 1996; Hill et al. 1997). For instance, using a similar matching task with static faces, Troje and Bülthoff (1996) only found an effect of viewpoint change with angles exceeding 45°. Nevertheless, the presence of a dynamic advantage across such image changes underlines the robustness of this effect. Having replicated Thornton and Kourtzi's results with our new set of stimuli, in Experiment 2 we addressed the question of whether such dynamic advantages persist over longer periods of time.

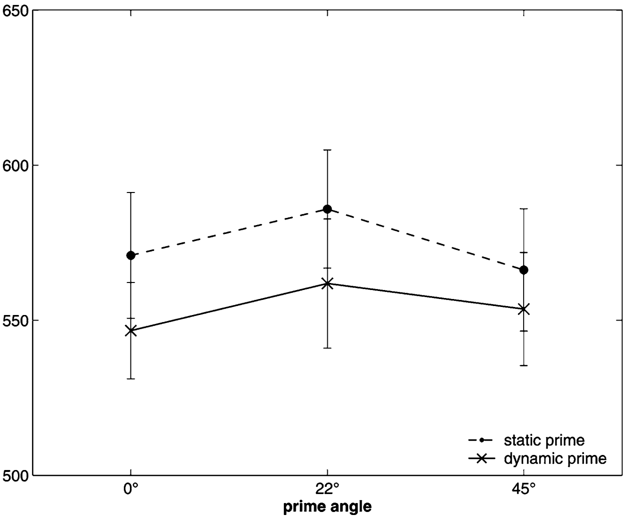

Fig. 3 Reaction time for same trials from Experiment 1

Experiment 2

Experiment 1 showed that a non-rigidly moving prime facilitates sequential matching. To explore whether such a dynamic advantage extends across time, Experiment 2 used a delayed visual search task in which observers were familiarized with two faces, one moving and one static.

In a typical visual search task, target items are surrounded by a variable numbers of distractor items. Examining the pattern of error or response times as a function of setsize can provide useful information about target properties. Visual search has been used to explore many aspects of perceptual processing, including the basic building blocks of vision (Treisman and Gelade 1980) and the relationship between vision and attention (e.g., Duncan and Humphreys 1989; Enns and Rensink 1990; Wolfe 1994, 1998). Several previous studies have also used visual search to directly explore face processing (e.g., Tong and Nakayama 1999; Nothdurft 1993; Eastwood et al. 2001, 2003). Our question was whether target faces that had been learned in motion would give rise to faster or more efficient search than the face learned as a static snapshot.

Methods

Participants

Eighteen observers (12 males and 6 females) between 20 and 33 years of age (mean age 24.7) participated in this experiment.

Stimuli

For each observer one male and one female face were randomly selected from the moving face database and served as stimuli in the familiarization phase of this experiment. These faces were shown from a frontal viewpoint, expressed surprise and extended a visual angle of 3.3°. For each observer, one face moved and the other was shown as a static snapshot.

The search arrays consisted of static views of both the target faces and one, three or five randomly chosen distractor faces. These static views could be shown in one of five viewpoints and were equally spaced in random position around an imaginary circle with a radius of 8° visual angle centered in the middle of the screen (see Fig. 4). Both video clips and pictures were presented on a black background, so that the image borders could not be differentiated from the screen background.

Test and design

The experiment consisted of two phases, learning and test. In the learning phase observers were familiarized with one male and one female frontal view face showing an expressive gesture of surprise. One face was shown as a short video clip, the other one as a static picture. Faces were alternated 100 times on the screen, each time presented for 1,040 ms with an inter stimulus interval of 2 s. While watching the faces, observers filled out a questionnaire. They were asked to rate factors such as the apparent attractiveness, age, kindness, aggressiveness and intelligence of the two faces, as well as to describe their prominent facial features. After the faces stopped alternating, observers were allowed to take a short break of approximately 3 min.

On each trial of the test phase, two, four or six static faces expressing the opposite emotion to that shown during learning were shown in a circular search array. Observers were asked to respond as quickly and accurately as possible to whether one of the learned faces was present in the search pattern or not. Targets were present on 66% of trials, with each of the familiarized faces appearing equally often. Observers responded "target present" by pressing the "s" key and "target absent" by pressing the "l" key. Auditory feedback was given for incorrect responses. Each trial started automatically after a response was given. The experiment consisted of 450 trials, in which each target was present on 150 of the trials. In the remaining 150 trials no target was presented. All setsize by target type and viewpoint trials occurred with equal frequency and were randomized for each observer individually.

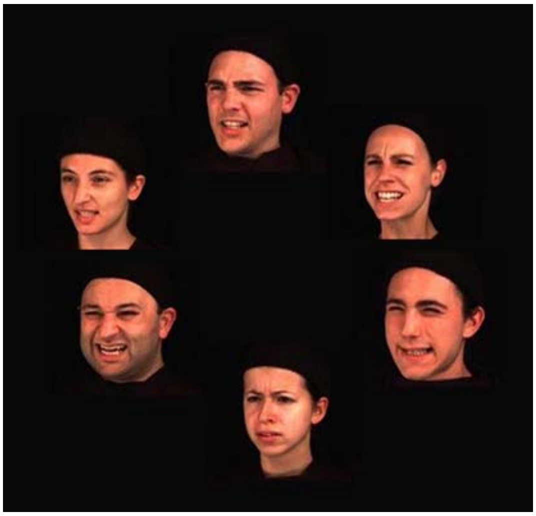

Fig. 4 Example of a visual search array used in the test periods of Experiments 2 and 3. Two, four or six static faces were shown equally spaced around an imaginary circle. On a given trial all faces had the same orientation, taken from one of the five recorded viewpoints

Results

Figure 5 shows mean RTs for target-present trials, plotted as a function of setsize for both the non-rigidly moving and the static target faces. As can be seen there is a reliable reaction time advantage of 280 ms for dynamically learned target faces ($M = 1{,}317.8$ ms, SE = 84.16 ms), as compared to static snapshots ($M = 1{,}595.5$ ms, SE = 190.19 ms). This advantage is consistent across all three setsizes. The search for absent trials shows a typical pattern with a steeper slope than target present trials and longer RTs ($M = 2{,}123.8$ ms, SE = 126.65 ms). A repeated-measures ANOVA (target type × viewpoint × setsize) was carried out on RTs for correct responses for trials in which one of the targets was present. In addition, error rates were examined for potential speed-accuracy trade-offs.

Reaction times

The ANOVA revealed significantly faster RTs for dynamically learned stimuli than for statically learned ones ($F_{(1,17)} = 4.7$, $P < .05$). There was a main effect of setsize ($F_{(2,34)} = 54{,}4$; $P < .0001$) and viewpoint ($F_{(2,34)} = 5{,}7$; $P < .01$). Across viewpoints, RTs were fastest for frontal view faces (*frontal view*: $M = 1{,}635$ ms, SE = 132.9 ms; *7/8 view*: 1,695.8 ms, SE = 136.2 ms; *3/4 view*: $M = 1{,}706$ ms, SE = 131.9 ms). There were no significant interactions, in particular, no setsize × motion interaction, i.e., no difference in the search slopes.

Accuracy

Observers were slightly more accurate for dynamically learned stimuli ($M = 94.3\%$), than statically learned stimuli ($M = 91\%$). While this difference did not reach significance, it does argue against a speed-accuracy trade-off. There was a main effect of viewpoint ($F_{(2,34)} = 5.3$; $P < .01$). Subjects performed slightly worse for 45° faces ($M = 90.3\%$, SE = 3.09%) as compared to 22° ($M = 93.6\%$, SE = 2.2%) and frontal view faces ($M = 93.9\%$, SE = 2.6%). An interaction between viewpoint and setsize ($F_{(4,68)} = 2.6$; $P < .05$) also appears to be driven by the 45° viewpoint. Unexpectedly, performance was slightly worse in setsize 2 of the 45° viewpoint. While there is no obvious interpretation for this result, the associated increase in variability suggests that this may simply be noise (see Table 1 for further detail).

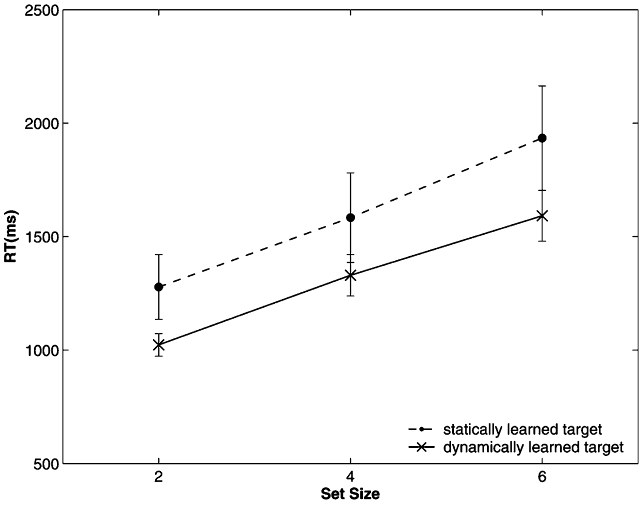

Fig. 5 Reaction time for same trials from Experiment 2

Discussion

Learning a face as a moving sequence facilitates the speed of search through a static array by approximately 300 ms compared to when learning involved a single static picture. This dramatic processing advantage does not reflect differences in the efficiency of the search—a pattern that would have resulted in condition × setsize slope differences—but appears to be a matching or recognition advantage once the target has been located. While changes in viewpoint did affect performance, this pattern was the same for both static and dynamic targets. The lack of a viewpoint × condition interaction for reaction time suggests that the underlying search behavior was the same for both dynamically and statically learned faces. This constant reaction time advantage may thus reflect the same matching mechanism that facilitated performance in Experiment 1. We return to this issue in the General Discussion.

The main effect of viewpoint and the interaction of viewpoint and setsize for the accuracy data indicate the general difficulty in generalizing to novel viewpoints. This pattern was the same for both dynamically and statically learned faces. Observers were more accurate in identifying frontal view faces, an advantage that increased with setsize. This advantage might reflect the fact that observers were trained on that view or that we frequently encounter frontal view faces during everyday communication.

Experiment 3

Experiment 2 showed a clear advantage for non-rigidly moving faces in a visual search task across viewpoints. In this and the following experiment we investigate whether this advantage is due to the motion of the face itself or simply due to the additional static information contained in the individual frames of the moving sequence. In Experiment 3, observers were trained on static faces presented in a 4×4 matrix. For one face this matrix contained the same static frontal view face, repeated sixteen times. For the other face sixteen different frames were taken from the moving sequence used in Experiment 2. Our question was whether this additional static information would lead to a speed or accuracy advantage.

Methods

Participants

Twelve observers (three males and nine females) between 20 and 37 years of age (mean age 24.1) participated in this experiment.

Stimuli, test, design and procedure

These were the same as in Experiment 2 except that observers were familiarized with a matrix of sixteen static faces in both conditions. In one condition the sixteen pictures showed the same static frontal view faces. In the other condition, the pictures showed sixteen different frames from the moving sequences used in Experiment 2. These sixteen frames were randomly arranged in a different order each time they were presented. This manipulation ensured that observers saw more than one static picture even if they tended to look at a preferred location in the array. Figure 6 gives an example of the stimuli used in the familiarization phase.

Results

Reaction times

Figure 7 shows mean RTs for target-present trials, plotted as a function of setsize. The search slopes for different target types overlap, indicating no main effect between conditions. A repeated-measures ANOVA (target type × viewpoint × setsize) revealed a main effect of setsize ($F_{(2,22)} = 76.3$; $P < .000$) but no other effects or interactions.

Accuracy

Observers performed very well in both conditions (*multiple*: $M = 94\%$; *single*: $M = 91\%$) and there were again no significant main effects or interactions.

Discussion

The results of Experiment 3 suggest that the provision of additional static information cannot account for the advantage observed in Experiment 2. However, the use

Table 1 Mean accuracy data for all setsizes and viewpoints collapsed across learning conditions (static/dynamic) from Experiment 2

	Setsize 2 (%)	Setsize 4 (%)	Setsize 6 (%)
0° face	$M = 94.2$ (SE = 2.6)	$M = 93.6$ (SE = 2.8)	$M = 93.9$ (SE = 2.6)
22° face	$M = 93.9$ (SE = 2.2)	$M = 94.0$ (SE = 2.3)	$M = 92.8$ (SE = 2.1)
45° face	$M = 87.8$ (SE = 3.9)	$M = 90.7$ (SE = 3.0)	$M = 92.4$ (SE = 2.5)

of a face matrix may not have been the optimal way to convey the variation contained in the moving sequence. Different observers may have employed very different viewing strategies. Also, it may be better to compare static variation to a moving baseline. We addressed both of these issues in Experiment 4.

Experiment 4

Experiment 4 used exactly the same design as Experiment 2 with alternating dynamic and static faces during learning. However, here, the static frame shown was not identical between repetitions but was randomly chosen from the moving sequence. Our question was whether the dynamic advantage would still be observed relative to variations in static information.

Methods

Participants

Fourteen observers (six males and eight females) between 20 and 32 years of age (mean age 24.7) participated in this experiment.

Stimuli, test, design and procedure

These were the same as in Experiment 2 except that observers were familiarized with different frames out of a moving sequence of one face that was alternated with a dynamic sequence of another face.

Results

Reaction times

Figure 8 shows mean RTs for target-present trials, plotted as a function of set size. The search slopes for different target types are clearly separated. A repeated-measures ANOVA (target type × viewpoint × setsize) was carried out on RTs for correct responses for trials in which one of the targets was present. There was a main effect of target type ($F_{(1,13)} = 3.8$, $P < 0.05$), with dynamically learned faces giving rise to faster responses (*dynamic*: $M = 1{,}303$ ms, SE = 77 ms; *static*: $M = 1{,}452$ ms, SE = 107 ms). There was also a main effect of setsize ($F_{(2,26)} = 51.218$; $P < .000$) but no other effects or interactions.

Accuracy

Observers performed very well in both conditions (*dynamic*: $M = 95\%$; *static*: $M = 95\%$) and there were no significant main effects or interactions.

Discussion

As in Experiment 2, we found a significant difference between the dynamic and the static condition. This direct comparison between variable static frames and the dynamic sequence, again suggests that the current learning advantage cannot be explained by additional static information. This finding is consistent with previous studies using both rigid (Wallis 2002; Wallis and Bülthoff 2001) and non-rigid (Lander and Bruce 2000) motion, which suggest that additional information alone cannot account for observed dynamic advantages. Together, the results of Experiments 3 and 4 would seem to suggest that motion itself is a crucial component for conveying this form of search advantage.

General discussion

In a series of experiments we asked whether non-rigid facial motion can provide an identity processing advantage. In Experiment 1 we used a new set of stimuli to replicated previous findings by Thornton and Kourtzi (2002) that non-rigid motion can facilitate immediate matching of non-degraded and unfamiliar faces. Additionally, this experiment also demonstrated that such an advantage survives changes in viewpoint, showing no modulation with prime/target differences of up to 45°. Crucially, Experiment 2 showed that the same non-rigid motion also enhances performance in a delayed facial search task. While searching identical static test arrays, faces that were learned in motion were responded to some 300 ms faster than statically learned faces across all setsizes. Experiments 3 and 4 showed that this effect was not simply due to the additional information provided by the individual frames of the moving sequence, suggesting instead that motion had influenced learning.

To our knowledge the current findings provide the first evidence that non-rigid facial motion can affect long-term memory performance for previously unfamiliar faces when static form cues to identity have been left completely unaltered. In previous studies, either the quality of the test images was degraded (e.g., Knight and Johnson 1997; Lander et al. 1999) such that color

Fig. 6 Example stimulus array from the learning phase of Experiment 3. Observers were sequentially shown a 4×4 matrix of 16 repetitions of one facial image (*left*), followed by a similar matrix of 16 different images taken from the video sequence of the other face (*right*)

information was deprived due to grayscale conversion (Lander and Bruce 2003; Bonner et al. 2003) or form cues were weakened or rendered uninformative in some way, for instance via the use of morphing (Knappmeyer et al. 2003) or averaging (Hill and Johnston 2001) techniques.

These results have two important implications. First, they strongly suggest that models of face recognition need to more accurately reflect a role for motion during the computation of identity (Haxby et al. 2000; O'Toole et al. 2002). Indeed, a strict separation between form and motion, as suggested by some models (e.g., Bruce and Young 1986), is becoming less and less tenable given the growing body of evidence that we retain and use patterns of characteristic motion to help individuate both people (Hill and Johnston 2001; Knappmeyer et al. 2003; Knight and Johnson 1997; Lander and Bruce 2000; Thornton and Kourtzi 2002) and other objects (Stone 1999; Vuong and Tarr 2004; Newell et al. 2004).

The second implication is that standard old/new recognition paradigms, which have often struggled to reveal any influence of motion even with degraded form cues (e.g., Christie and Bruce 1998; Lander et al. 1999), may be ill suited for studying dynamic aspects of face processing. As mentioned in the Introduction, the requirement to learn a list of sequentially presented items may bias observers to adopt strategies specifically evolved for static word or picture processing rather than for dealing with complex, dynamic patterns, such as faces. While we can currently only speculate about this issue, a useful direction for future research might be to more formally compare different tasks to establish why some expose dynamic advantages and some do not.

More generally, we believe that future research into the role of facial motion would be well served by adopting study and test regimes that more closely approximate our real world experience with faces. The current use of an incidental learning phase in which observers engage in a wide range of judgments on limited number of faces, and a test phase that approximates finding a friend in a crowd, go someway in this direction. We should note that learning faces to such a high degree also tends to shift the emphasis away from accuracy, the primary measure in old/new recognition tasks, onto response time. It is possible that dynamic manipulations influence the accessibility rather than the content of the underlying facial representations. Indeed, this shift from accuracy (e.g., Christie and Bruce 1998) onto response time (Thornton and Kourtzi 2002) may go some way to explain some of the contradictory findings with dynamic stimuli.

Clearly, the current task is still conducted under strict laboratory conditions and as such cannot claim to be completely "natural". One of the great potentials of virtual reality technology will be the development of environments in which observers can interact with dynamic virtual characters in much richer and more naturalistic settings.

The question that remains to be answered is exactly *how* the presentation of movement can come to affect performance. Experiments 3 and 4 suggest that it is not simply the additional information present in the individual snapshots. These experiments also go some way to ruling out an explanation based on differential viewing time or attentional allocation between the critical conditions. That is, the presentation of multiple frames is also likely to have increased looking time and to have required additional processing resources over and above those needed for looking at the same image. Of course, moving stimuli are particularly salient and are well

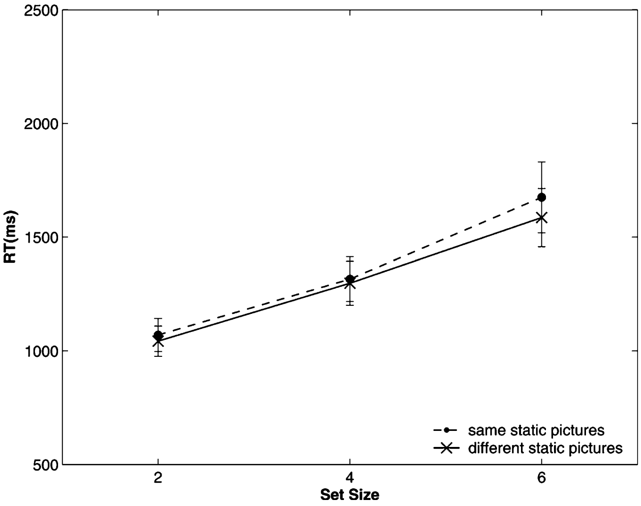

Fig. 7 Reaction time for same trials from Experiment 3

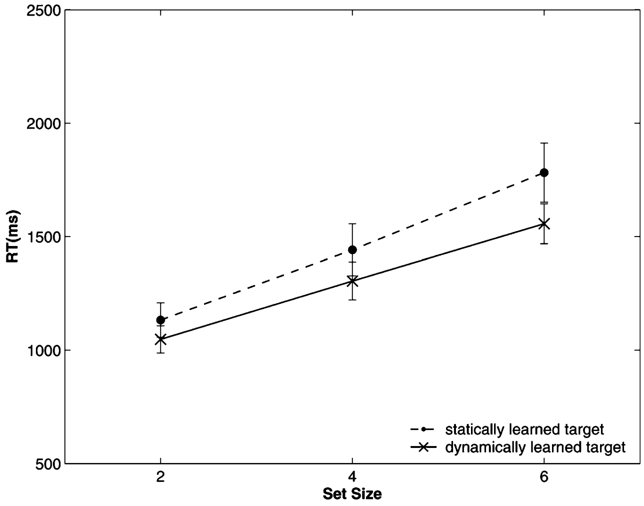

Fig. 8 Reaction time for same trials from Experiment 4

known to attract attention; thus, we cannot rule out the possibility that the current learning advantage arises as a side effect of "effective" exposure. To further explore this issue it would be interesting to explicitly vary exposure duration for two statically presented faces, showing one face many more times or for a longer duration than the other. The nature of any search advantage for the over-exposed face would then help to resolve this issue.

It has been well established that becoming familiar with a face involves a change of emphasis from attending to external features, such as hairstyle and head shape, to the processing of the internal configuration of facial features (Bruce et al. 1999; Ellis et al. 1979; Young et al. 1985). As non-rigid motion mainly reveals information about such internal features and their spatial relations, it is possible that in the current dynamic condition movement helps to speed familiarization relative to the static condition. In a recent study, Bonner et al. (2003) directly tested this hypothesis. They familiarized observers with either video-based images or static snapshots over successive days. While matching the external features remained constant throughout the whole experiment, matching the internal features improved with familiarity. However, in this study there was no evidence that studying moving images speeded or altered this effect of familiarization.

Previous research has suggested that motion can affect performance by directly providing an additional cue to identity. For example, a particularly distinctive smile or expression of surprise might become represented as a characteristic pattern of movement (e.g., Knappmeyer et al. 2003; Knight and Johnson 1997; Lander and Bruce 2000; Lander and Chuang 2005; see O'Toole et al. 2002 for a review). In the current work, however, it is less clear how such characteristic motion patterns might affect performance, as all test images were static. More generally, however, non-rigid facial motion during learning might provide dynamic cues that help to predict possible expressive configurations that a specific face can adopt. As Christie and Bruce (1998) suggest, facial motion might provide a better range of variation from which we can extract invariant characteristics of a face. In the current experiment, this may facilitate subsequent identification across the image variations between learning and test, e.g., different expressions or viewpoints.

Thornton and Kourtzi (2002) suggested that the motion advantage they found in their immediate matching paradigm reflected the creation and maintenance of working memory representations that directly reflected the dynamics of the stimuli. Such "dynamic mental representations" (Freyd 1987, 1993) have been postulated in a number of perceptual domains, including object localization (e.g., representational momentum, Freyd and Finke 1984), object recognition (Kourtzi and Nakayama 2001) and biological motion processing (Cavanagh et al. 2001). It is thought that retaining information about how an object moves can afford processing advantages in coping with subsequent viewpoint, configuration or position changes. If such dynamic representations underlie the current effects, then they can clearly survive for longer periods of time than has typically been supposed (Freyd 1993; Freyd and Johnson 1987; Thornton and Kourtzi 2002). The delay of several minutes between study and test in Experiment 2, a period that was not filled with explicit rehearsal or explicit maintenance, is more consistent with operation of long-term rather than working memory systems. Clearly, a useful avenue for future research will be to more systematically vary the retention interval between study and test to map out the temporal limits of the current dynamic face advantage.

Interestingly, the pattern of visual search data in Experiments 2 and 4 suggests a direct link between this long-term advantage and the immediate matching effect of Thornton and Kourtzi (2002). That is, the static and dynamic conditions differ not in the efficiency of the search process itself—which would be reflected in set-size functions with different slopes—but in the speed with which the target item is recognized. This gives rise to search offsets of between 150 and 280 ms. This suggests that observers adopt a single strategy for searching through the test array, but that when the target is encountered their ability to recognize the dynamically familiarized face is significantly enhanced. Thus, we appear to be observing a long-term matching advantage. As an aside, we should note that the lack of a search slope difference is not really that surprising, given that during the test phase observers are searching through identical static arrays. The only difference for the two target faces is the way in which they were familiarized. Any difference in search efficiency would thus be wholly driven by top-down effects. While such effects have been previously reported in the literature (e.g., Eastwood et al. 2001, 2003), they are by no means common.

Having suggested a link between the immediate and long-term matching of dynamic faces, we should also note that quantitatively they are quite different. Specifically, the matching advantage in Experiment 1, and in the previous work by Thornton and Kourtzi (2002), ranged from 16 to 30 ms, while the search advantage ranges from 150 to 280 ms. That the advantage is almost an order of magnitude bigger for the search task could arise for a number of reasons. To begin with, the matching task is much less demanding than the search task, with overall RTs of around 600 ms compared to the range of between 1,000 (setsize 2) and 1,600 ms (setsize 6) for the most efficient search. This speeded response may impose a limit on the observable difference between dynamic and static conditions in the matching task. Other aspects of the search task itself or the delay between study and test present in Experiment 2 may also have contributed to amplify a single underlying "matching" process. Clearly, however, we cannot rule out the possibility that the large difference between the search and matching conditions might reflect fundamental differences in the way dynamic information is being treated in the two tasks.

In conclusion, the current findings provide further evidence that non-rigid motion of a previously unfamiliar human face can affect identity decisions over extended periods of time. Here, we have shown that this is true, even when form cues to identity were undegraded and information content was equated. We suggest that the presence of non-rigid facial motion necessarily gives rise to the creation of dynamic representations, and that these representations can facilitate recognition, particularly across image transformations such as changes in expression or viewpoint. Additionally, we believe that such facilitation may be difficult to observe using more traditional old/new recognition tasks, as they may encourage coding strategies that are simply not appropriate for use with dynamic stimuli.

Acknowledgements The authors would like to thank Martin Breidt and Mario Kleiner for helping to produce the stimuli used in these experiments. Thanks also to all of the actors who allowed us to videocapture their facial motions. We are also grateful to Quoc Vuong, Vicki Bruce and one anonymous reviewer for their valuable and thoughtful comments on an earlier draft of this paper.

References

Amaya K, Bruderlin A, Calvert T (1996) Emotion from motion. GI '96 online papers (http://www.graphicsinterface.org/proceedings/1996/ABC/)

Bassili JN (1978) Facial motion in the perception of faces and of emotional expression. J Exp Psychol Hum Percept Perform 4:373–379

Bonner L, Burton AM, Bruce V (2003) Getting to know you: how we learn new faces. Visual Cognit 10:527–536

Brainard D (1997) The psychophysics toolbox. Spat Vis 10:433–436

Bruce V (1982) Changing faces: visual and non-visual coding processes in face recognition. Br J Psychol 73:105–116

Bruce V (1999) Complexities of face perception and categorisation. Behav Brain Sci 22:369

Bruce V, Young A (1986) Understanding face recognition. Br J Psychol 77:305–327

Bruce V, Henderson Z, Greenwood K, Hancock PJB, Burton AM, Miller P (1999) Verification of face identities from images captured on video. J Exp Psychol A 5:339–360

Campbell R, Brooks B, de Haan E, Roberts T (1996) Dissociating face processing skills: decision about lip-read speech, expression, and identity. Q J Exp Psychol A 49:295–314

Cavanagh P, Labianca A,Thornton IM (2001) Attention-based visual routines: sprites. Cognition 80:47–60

Christie F, Bruce V (1998) The role of dynamic information in the recognition of unfamiliar faces. Mem Cognit 26:780–790

Cunningham DW, Breidt M, Kleiner M, Wallraven C, Bülthoff HH (2003) How believable are real faces: towards a perceptual basis for conversational animation. Computer Animation and Social Agents 2003, 23–39

Duncan J, Humphreys GW (1989) Visual search and stimulus similarity. Psychol Rev 96:433–458

Eastwood JD, Smilek D, Merikle PM (2001) Differential attentional guidance by unattended faces expressing positive and negative emotion. Percept Psychophys 63:1004–1013

Eastwood JD, Smilek D, Merikle PM (2003) Negative facial expression captures attention and disrupts performance. Percept Psychophys 65:352–358

Ellis HD, Shepherd JW, Davies GM (1979) Identification of familiar and unfamiliar faces from internal and external features: some implications for theories of face recognition. Perception 8:431–439

Enns JT, Rensink RA (1990)Influence of scene-based properties on visual search. Science 9:721–723

Freyd JJ (1987) Dynamic mental representations. Psychol Rev 94:427–438

Freyd J (1993) Five hunches about perceptual processes and dynamic representations. In: Meyer D, Kornblum S (eds) Attention and performance XIV: synergies in experimental psychology, artificial intelligence, and cognitive neuroscience. MIT Press, Cambridge, pp 99–119

Freyd JJ, Finke RA (1984) Facilitation of length discrimination using real and imaged context frames. Am J Psychol 97:323–341

Freyd JJ, Johnson JQ (1987) Probing the time course of representational momentum. J Exp Psychol Learn Mem Cogn 13:259–268

Haxby JV, Hoffman EA, Gobbini MI (2000) The distributed human neural system for face perception. Trends Cogn Sci 4:223–233

Hill H, Johnston A (2001) Categorizing sex and identity from the biological motion of faces. Curr Biol 11:880–887

Hill H, Schyns PG, Akamatsu S (1997) Information and viewpoint dependence in face recognition. Cognition 62:201–222

Kamachi M, Bruce V, Mukaida S, Gyoba J, Yoshikawa S, Akamatsu S (2001) Dynamic properties influence the perception of facial expressions. Perception 30:875–887

Kleiner M, Wallraven C, Bülthoff HH (2003)"The MPI VideoLab." Technical report 104. Max-Planck-Institute for Biological Cybernetics, Tübingen, Germany

Knappmeyer B, Thornton IM, Bülthoff HH (2003) Facial motion biases the perception of facial form. Vision Res 43:1921–1936

Knight B, Johnson A (1997) The role of movement in face recognition. Vis Cogn 4:265–273

Kourtzi Z, Nakayama K (2001) Dissociable signatures of processing for moving and static objects. Vis Cogn 9:248–264

Lander K, Bruce V (2000) Recognizing famous faces: Exploring the benefits of facial motion. Ecol Psychol 12:259–272

Lander K, Bruce V (2003) The role of motion in learning new faces. Vis Cogn 10:897–921

Lander K, Chuang L (2005) Why are moving faces easier to recognize? Vis Cog 12:429–442

Lander K, Christie F, Bruce V (1999) The role of movement in the recognition of famous faces. Mem Cognit 27:974–985

Liu CH, Chaudhuri A (2002) Reassessing the 3/4 view effect in face recognition. Cognition 83:31–48

Newell FN, Wallraven C, Huber S (2004) The role of characteristic motion in object categorization. J Vis 4:118–129

Nothdurft HC (1993) Faces and facial expressions do not pop out. Perception 22:1287–1298

O'Toole AJ, Roark D, Abdi H (2002) Recognizing moving faces: a psychological and neural synthesis. Trends Cogn Sci 6:261–266

Rosenblum LD, Yakel DA, Baseer N, Panchal A, Nodarse BC, Niehus RP (2002) Visual speech information for face recognition. Percept Psychophys 64:220–229

Sekular AB, Palmer SE (1992) Perception of partly occluded objects: a microgenetic analysis. J Exp Psy gen 121:95–111

Stone JV (1999) Object recognition: view-specificity and motion-specificity. Vision Res 39:4032–4044

Thornton IM, Kourtzi Z (2002) A matching advantage for dynamic faces. Perception 31:113–132

Tong F, Nakayama K (1999) Robust representations for faces: evidence from visual search. J Exp Psychol Hum Percept Perform 25:1016–1035

Treisman AM, Gelade G (1980) A feature-integration theory of attention. Cognit Psychol 12:97–136

Troje NF, Bülthoff HH (1996) Face recognition under varying poses: the role of texture and shape. Vision Res 12:1761–1771

Vuong QC, Tarr MJ (2004) Rotation direction affects object recognition. Vision Res 44:1717–1730

Wallis G (2002) The role of object motion in forging long-term representations of objects. Vis Cogn 9:233–247

Wallis G, Bülthoff HH (2001) Effects of temporal association on recognition memory. PNAS 98:4800–4804

Watson TL, Hill HC, Johnston A, Troje N (2005) Motion as a cue for viewpoint invariance. Vis cogn (in press)

Wolfe JM (1994) Visual search in continuous, naturalistic stimuli. Vision Res 34:1187–1195

Wolfe JM (1998) Visual Search. In: Pashler H (eds) Attention. Psychology Press/Erlbaum (UK) Taylor & Francis, England, pp 13–73

Young AW, Hay DC, Ellis AW (1985) The faces that launched a thousand slips: everyday difficulties and errors in recognizing people. Br J Psychol 76:495–523