Walk this way: Approaching bodies can influence the processing of faces ☆

Karin S. Pilz [a],*, Quoc C. Vuong [b], Heinrich H. Bülthoff [c,d], Ian M. Thornton [e]

[a] Department of Psychology, Neuroscience and Behaviour, McMaster University, Hamilton, Canada
[b] Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, UK
[c] Max Planck Institute for Biological Cybernetics, Tübingen, Germany
[d] Department of Brain and Cognitive Engineering, Korea University, Seoul 136-713, Republic of Korea
[e] Department of Psychology, Swansea University, Swansea, UK

ARTICLE INFO

ABSTRACT

A highly familiar type of movement occurs whenever a person walks towards you. In the present study, we investigated whether this type of motion has an effect on face processing. We took a range of different 3D head models and placed them on a single, identical 3D body model. The resulting figures were animated to approach the observer. In a first series of experiments, we used a sequential matching task to investigate how the motion of an approaching person affects immediate responses to faces. We compared observers' responses following approach sequences to their performance with figures walking backwards (receding motion) or remaining still. Observers were significantly faster in responding to a target face that followed an approach sequence, compared to both receding and static primes. In a second series of experiments, we investigated long-term effects of motion using a delayed visual search paradigm. After studying moving or static avatars, observers searched for target faces in static arrays of varying set sizes. Again, observers were faster at responding to faces that had been learned in the context of an approach sequence. Together these results suggest that the context of a moving body influences face processing, and support the hypothesis that our visual system has mechanisms that aid the encoding of behaviourally-relevant and familiar dynamic events.

1. Introduction

Movement can cause frequent and often quite dramatic changes to the visual information available to us. Such changes arise not only because of our own actions, but also because other humans, animals, objects and natural phenomena move when we stand still. A vast amount of research has concentrated on the low-level or early visual processing of motion (e.g., Bülthoff & Bülthoff, 1987; Carney, 1997; Derrington, Allen, & Delicato, 2004; Hassenstein & Reichardt, 1956; Krekelberg & Albright, 2005; Poggio & Reichardt, 1973). In terms of "visual cognition", however, the impact of motion is still relatively under-explored, and the use of static stimuli is still far more common than the use of dynamic stimuli.

In recent years, studies of object (e.g., Chuang, Vuong, Thornton, & Bülthoff, 2006; Stone, 1998, 1999; Vuong & Tarr, 2004, 2006), face (e.g., Christie & Bruce, 1998; Knappmeyer, Thornton, & Bülthoff, 2003; Lander & Bruce, 2000, 2003; O'Toole, Roark, & Abdi, 2002; Pike, Kemp, Towell, & Phillips, 1997; Pilz, Thornton, & Bülthoff, 2006; Wallis & Bülthoff, 2001), body (e.g., Johansson, 1973; Knoblich, Thornton, Grosjean, & Shiffrar, 2006) and scene perception (Matthews, Benjamin, & Osborne, 2007; Vuong, Hof, Bülthoff, &

Thornton, 2006) have begun to take into account the consequences of motion. Precisely how the cognitive system treats "dynamic" versus "static" information, however, remains an open question (Freyd, 1987; Matthews et al., 2007).

The focus of the current paper is on the perception and representation of facial identity. Until the 1990s, it was commonly assumed that facial form and facial motion served complementary functions during face processing. Facial features and their configuration provided static, invariant cues to identity, gender, race and age. The movement of the head and face were concerned with emotion and communication, via expressions, gestures and visible speech (e.g., Bruce & Young, 1986). More recently, however, such a simple dichotomy has been called into question. For example, it has been shown that both rigid and non-rigid facial motion can serve as reliable cues to individual identity (e.g., Hill & Johnston, 2001; Knappmeyer et al., 2003; Lander & Bruce, 2000, 2003; O'Toole, Roark, & Abdi, 2002; Pilz, Bülthoff, & Vuong, 2009; Pilz et al., 2006). Findings suggest that facial motion not only provides additional cues (e.g., structure from motion) but can also generally influence the efficiency with which we encode and process facial information (Pilz et al., 2006; Thornton & Kourtzi, 2002).

In the current work we take one particular dynamic context – a person walking towards you – and explore how this might affect face processing, in particular, the processing of facial identity. Although the vast majority of research treats faces as isolated objects, our real-world experience typically involves seeing a head in the context of a body that moves in space. Does this dynamic "body context" have any influence on the way we process faces? Could variations in viewpoint, changes in information availability (e.g., as people approach and recede) or simply the familiarity of certain events have a reliable impact on face processing? Here, we take a very simple and highly familiar event and examine whether seeing a face in this dynamic context, as compared to static snapshots, affects performance in an immediate face matching task (Experiments 1–3) or a long-term memory visual search task (Experiments 4–5).

Schiff, Banka, and Galdi (1986) previously compared recognition performance for individuals learned from video sequences of a liquor store hold-up to static shots taken from the same videos. They found better recognition performance for people that had been learned from dynamic video sequences compared to people that had been learned from static mug shots. Vicki Bruce and colleagues (e.g., Bruce, Henderson, Newman, & Burton, 2001; Bruce et al., 1999; Burton, Wilson, Cowan, & Bruce, 1999; Henderson, Bruce, & Burton, 2001) have used a range of paradigms to test observers' ability to identify people depicted in CCTV footage. They found that it is often difficult to match or recognize previously unfamiliar people from such videos. Familiar individuals, however, can be recognized quite successfully, suggesting that familiar gait and body information may play an important role in person recognition. Roark, O'Toole, and Abdi (2003) used whole body videos to test observers' ability to recognize individuals whom they had previously learned from either static photographs or facial speech videos. They found that observers were better at recognizing individuals

from the body videos when these had been learned from dynamic video clips rather than static snapshots. In a similar study O'Toole and colleagues showed observers real videos of moving faces, static faces and whole bodies with faces. Observers had to match the identity in pairs of images or videos. They demonstrated that human identification is best when the whole person was seen in motion (O'Toole et al., in press).

These studies represent an important advance in studying person recognition by taking into account the moving body instead of exclusively concentrating on the isolated face. Their findings clearly suggest that observers are influenced by seeing faces in the context of a moving body. However, although the use of video-based stimuli has the advantage of capturing naturalistic movement sequences, such stimuli also have some important limitations. That is, it becomes very difficult to separately assess or control for the variety of identity cues that may be available in such clips. Recognition may be based on facial change, on individual body motion or walking style, or other individually distinct features such as hairstyle or clothing.

Another general goal of the current paper is to demonstrate that computer animated, or virtual stimuli, can be used to overcome such limitations. Near-photorealistic virtual characters or avatars are quickly becoming commonplace on game consoles and movie screens. Previous research on facial motion has already shown how similar stimuli can be used to explore individual identity (e.g., Hill & Johnston, 2001; Knappmeyer et al., 2003). Here, we combined different 3D head models (Blanz & Vetter, 1999; Troje & Bülthoff, 1996) with a single 3D body model taken from a popular animation package Poser©. This single body model was used to create animation sequences in which the avatar approached the observer using an identical walk pattern. In this way we were able to vary the facial information but to precisely equate the body and how it moved. This is a manipulation that is simply not possible with real-world stimuli.

To explore the impact of viewing faces in the context of body motion, we chose two different paradigms that had already been successfully used to study facial motion. The first, an immediate matching task (Experiments 1–3), explored the brief, short-term representations we might form during a natural approach sequence (Pilz et al., 2006; Thornton & Kourtzi, 2002). Because the nature of an immediate matching paradigm as applied in Experiments 1–3, makes it possible for observers to adopt a speed-accuracy tradeoff as part of their response strategy, we also decided to use an additional task to investigate the effects of body motion on face recognition. In Experiments 4 and 5 we used a delayed visual search paradigm to test whether an advantage for encoding approaching over static presentations persisted over longer periods of time (Pilz et al., 2006, 2009).

2. General methods

2.1. Stimuli

Fifteen male heads from the MPI 3D head database (Blanz & Vetter, 1999; Troje & Bülthoff, 1996) were

mounted onto a 3D avatar using Poser©, a commercially available animation package. The resulting figures were animated to approach the observer on a straight walk path using the Poser© built-in walk designer. The virtual camera took perspective views of the animated figures. Since all stimuli had identical bodies and were animated with the same walk pattern and walk path, the only feature that distinguished one stimulus from another was the head itself. The final movie clips were converted to grayscale and edited to contain 30 frames. For each initial or 'prime' sequence 15 consecutive frames were taken from the clip (i.e., 600 ms at a frame rate of 25 fps) according to procedures described in the specific methods section for each experiment. The display area subtended a visual angle of $15.7° \times 11.7°$ (width × height). Fig. 1 gives an example of the stimuli. Dynamic prime bodies varied in size from $6.2° \times 6.2°$ (Fig. 1A) to $13.3° \times 4.3°$ (Fig. 1C) during the course of the animation, i.e., the whole upper part of the body in the initial frames up to only the head and shoulders in the final frames. The heads of those animated figures varied from $2.0° \times 2.9°$ to $5.0° \times 7.2°$. The static prime frame was a frame taken from the dynamic sequence. It was always chosen to match the end-point of the dynamic sequence. This was either the last frame of the animation sequence, when only the approach sequence was shown, or the middle frame, when both approaching and receding sequences were used. In addition to animating the avatars, we rendered the same fifteen 3D heads without the body. These images were shown in frontal view (Experiments 1–5), as well as 22° to the right and left (Experiments 2–5) with their original color pigmentation and served as target stimuli for all five experiments. The target display area subtended a visual angle of $10.5° \times 11°$ (width × height). The heads themselves subtended a visual angle of $6.2° \times 8.6°$. The slight variations in size and color between prime and test images were designed to increase the difficulty of the task and to minimize the possibility of picture matching.

2.2. Apparatus

All of the following experiments were conducted on a Macintosh G4 computer under the control of customized software using the PsychToolBox extension for MATLAB (Brainard, 1997; Pelli, 1997). Stimuli were presented on a 21 in. monitor with a resolution of 1152×864 pixels and a frame rate of 75 Hz. Observers were seated 60 cm from the screen.

2.3. Data analysis

In all experiments we examined both the speed and accuracy of responses. We used the median reaction time (RT) of single subjects for each condition, because the median is less affected by outliers than the mean and provides a better estimate of the true average for skewed distributions, which RT data often have (Ratcliff, 1979a, 1979b, 1993). As match and non-match decisions were likely to represent different processes, our experimental predictions relate to same trials. Complete RT data are presented in Tables 1 (Experiments 1–3) and 3 (Experiments 4–5). For Experiments 1–3, accuracy data were transformed into d' values, which measures sensitivity to our stimulus manipulations. It is computed from the hits (correctly detected same trials) and false alarms (incorrectly denoted different trials). For completeness, we report the accuracy data to show performance for same (hits) and different trials (correct rejections) separately for Experiments 1–3 in Table 2.

The main focus of analysis for this and all experiments of the current paper will reaction time (RT). We did not anticipate differences in accuracy, because in both tasks

Fig. 1. Example pictures of (A) the first (B) the 15th and (C) the 30th frame of the rendered movie sequence.

Table 1
Reaction time (ms) data for same and different trials for Experiments 1–3. SE refers to 95% confidence intervals calculated with the method described by Loftus and Masson (1994).

	Same trials						Different trials					
	Looming probe		Receding probe		Static probe		Looming probe		Receding probe		Static probe	
	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE
Experiment 1	577	9	–	–	614	13	594	11	–	–	609	13
Experiment 2	633	6	658	8	696	7	697	8	696	6	718	5
Experiment 3	610	4	611	8	631	8	646	7	640	9	658	9

Note. Receding probe stimuli were not tested in Experiment 1.

Table 2
Accuracy (% correct) data for same and different trials for Experiments 1–3. The data shown in the table are combined in the d'-analysis described in the text as hits (correctly detected same trials) and false alarms (incorrectly denoted different trials). SE refers to 95% confidence intervals calculated with the method described by Loftus and Masson (1994).

| | Same trials | | | | | | Different trials | | | | | |
| | Looming probe | | Receding probe | | Static probe | | Looming probe | | Receding probe | | Static probe | |
	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE
Experiment 1	79.8	0.01	–	–	81.0	0.01	79.9	0.01	–	–	81.0	0.01
Experiment 2	79.8	2.0	79.7	1.8	80.7	2.0	70.8	1.6	69.3	2.5	83.6	2.4
Experiment 3	81.9	0.8	86.0	0.5	87.9	0.9	79.2	0.5	79.2	1.2	79.8	1.3

Note. Receding probe stimuli were not tested in Experiment 1.

used here, target faces are clearly visible and often repeated. Furthermore, in the case of sequential matching, recognition occurs directly after exposure to the two faces with participants only having to indicate same or different. In the delayed visual search task, faces are learned to such a high degree that differences in accuracy between conditions are not usually detectable. In previous studies using similar paradigms we did not find accuracy differences between conditions (Pilz et al., 2006; Thornton & Kourtzi, 2002).

The current RTs were recorded using a standard USB keyboard. It has been shown previously that the delay between the actual key press and the finger touching the key is about 2 ms. In addition, the scanning time of a standard USB keyboard is between 18 and 32 ms (Shimizu, 2002) and may vary depending on key position. This scanning time might potentially introduce noise to the data. In addition, it has been shown that faster movements and RTs occur with the dominant hand (e.g., Shen & Franz, 2005).

In the current work, it is important to note that RTs for the critical conditions were all measured with the same key to minimize the noise introduced by the choice of answering device and handedness. The only differences between hand and key were for target conditions (same/different (Experiment 1–3), present/absent (Experiment 4–5) but not between the critical prime/learning conditions.

3. Experiment 1

In a first series of experiments, we used an immediate matching paradigm to investigate whether the motion of an approaching person affects the short-term encoding and recognition of unfamiliar faces. The purpose of Experiment 1 was to directly compare matching performance in trials with dynamic approach sequences to those in which the prime was a single static snapshot. Based on previous research using moving faces (Pilz et al., 2006; Thornton & Kourtzi, 2002), our prediction was that dynamic prime stimuli would lead to better matching performance than static prime stimuli.

3.1. Materials and methods

3.1.1. Participants

Twelve right-handed observers aged 21–39 (mean age: 27.0 years) participated in this study (five females and seven males). Observers were either volunteers from the MPI

community or were recruited from the MPI subject pool in return for 8 €/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 informed consent.

3.1.2. Stimuli

In Experiment 1, 15 video clips of an approaching person, as well as the corresponding 15 target faces in frontal view were used. In addition, 15 static target images of different identities served as distractors for 'different' trials. The static prime face was chosen to be the ending frame of the dynamic sequence as described above.

3.1.3. Procedure

Observers were seated in front of a computer screen at a viewing distance of 60 cm. They were told that each trial would involve the presentation of two faces, a prime face followed by a target face. They were instructed to pay close attention to the identity of the prime face so that they were able to decide if the target face showed the same or a different person. Observers were told that the target face would always be a static image, but the prime face would sometimes be a short video clip of an approaching person (dynamic prime) and sometimes a single static frame (static prime). It was emphasized that this video/static manipulation was not relevant to the identity decision they were required to make. The target face stayed on the screen until the observer responded and auditory feedback was given whenever observers made an incorrect response or took longer than 800 ms to respond. The whole experiment took about 10 min to complete.

3.1.4. Task and design

The experiment consisted of 144 trials divided into four blocks of trials. Each block consisted of 36 trials, half of which contained dynamic stimuli (18 trials) and half of which contained static stimuli (18 trials). After each block, observers were encouraged to take a small break. They initiated the next block by pressing 'n' on the keyboard.

Within this motion factor, there were equal numbers of same trials (nine trials) and different trials (nine trials). 'Same' trials were constructed by showing the same identity for prime and target face. 'Different' trials were constructed by randomly selecting different identities for prime and target. The order of trials within each block was randomized separately for each observer on a block-by-block basis. On
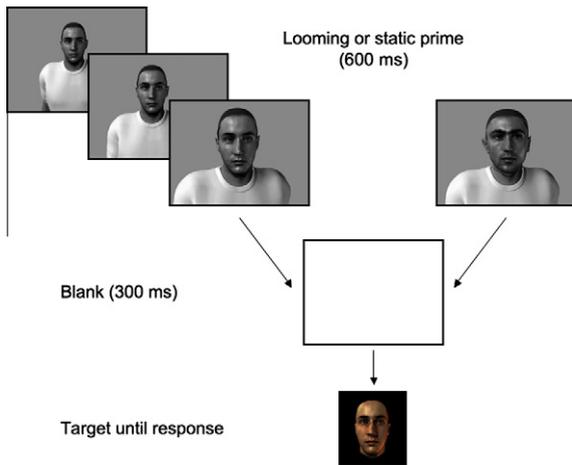
Fig. 2. Example of a trial sequence in Experiment 1. Observers saw either an approaching or a static probe person for 600 ms. After a blank of 300 ms, the target face (shown from frontal view) was presented until the observer responded to whether the probe and the target face showed the same or different identities.
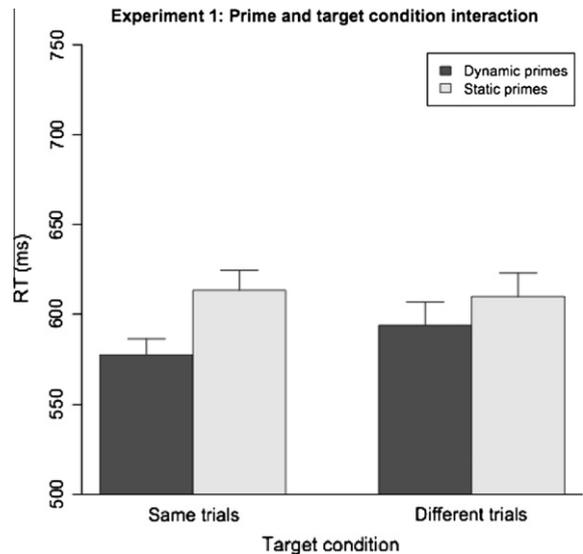

Fig. 3. Reaction times for dynamic and static same and different trials for Experiment 1. Error bars represent 95% confidence intervals calculated with the method described by Loftus and Masson (1994).

each trial a prime appeared in the middle of the screen for 600 ms. The prime was either a dynamic animation sequence of a person approaching the observer or a static picture. After a blank of 300 ms the target face appeared in the middle of the screen. The target face was always a static image showing the head from a frontal view, rendered without hair in front of a black background. On each trial the prime stimulus was randomly drawn from the stimulus set. The task of the observers was to determine if the prime and target faces was of same individual or of different individuals. They were asked to respond as quickly and as accurately as possible using one of two marked keys. This was the 's' key for 'same' and the 'l' key for 'different'. An example of the trial sequence can be seen in Fig. 2.

3.2. Results

3.2.1. Reaction times

Fig. 3 shows mean RTs for correct responses for both same trials and different trials. A 2 (prime condition (static, dynamic)) × 2 (target condition (same, different)) repeated measures ANOVA showed a main effect of prime condition, $F(1, 11) = 15.11$, $p < 0.01$, with a reliable overall RT advantage of 25.8 ms for dynamic compared to static primes, and no effect of target condition $F(1, 11) = 0.18$, $p = 0.6$. The interaction between prime and target condition was marginally significant, $F(1, 11) = 4.32$, $p = 0.06$. Post hoc tests showed that for same trials, there was a consistent RT advantage of 36 ms for dynamic compared to static primes, $t(11) = 22.0$, $p < 0.001$. The difference between static and dynamic prime stimuli in the different trials of 15 ms was not reliable, $t(11) = 3.2$, $p = 0.1$.

3.2.2. Sensitivity

Accuracy remained at about 80% in all conditions as can be seen in Table 2. There was no variation in sensitivity be-

tween conditions with an average d' of 1.8, $t(11) = 1.9$, $p > 0.1$.

3.3. Discussion

The results of Experiment 1 showed that dynamic stimuli can lead to faster matching responses than static stimuli. This advantage was more pronounced when the prime and the target showed the same person. This suggests that the advantage does not arise solely from a general increase in arousal or alertness for moving primes, but rather relates to the representation of a particular individual. Such an identity specific RT advantage has previously been shown with non-rigid facial motion (Pilz et al., 2006; Thornton & Kourtzi, 2002).

The next two experiments investigate the origin of this dynamic advantage in further detail.

4. Experiment 2

The results of Experiment 1 showed that an animated figure approaching in depth can speed identity decisions relative to a static snapshot. Does this dynamic advantage relate directly to the salience of an approach sequence, or would other body movements in depth have a similar effect? The purpose of Experiment 2 was to directly contrast the dynamic approach sequence used in the Experiment 1 with the same type of sequence played backwards. Such a control condition closely equates the amount of available facial information, while varying the familiarity and salience of the observed action.

Previous research has shown a particular sensitivity for approaching – but not receding – stimuli in a number of animal species (e.g., Ball & Tronick, 1971; Bower, Broughton, & Moore, 1970; Maier, Neuhoff, Logothetis, & Ghazanfar, 2004; Schiff, 1965; Schiff, Caviness, & Gibson, 1962;

Tronick, 1967). For humans, assessing the identity and/or the intentions of an approaching person is also highly significant. Compared to a sequence where a person is walking backwards and away from the observer – the control condition introduced here – such an approach sequence also varies in terms of the frequency, naturalness and salience of the action. We thus predicted that responses to dynamic approach sequences would be faster than responses to dynamic receding sequences. If, on the other hand, the two conditions were to give rise to similar patterns of performance, this would suggest the results of Experiment 1 relate more to general alerting or arousal effects of motion (e.g., Driver & Baylis, 1989; Franconeri & Simons, 2003; Hillstrom & Yantis, 1994), rather than being specific to the act of approach.

In addition to the new receding motion condition, we also repeated the static prime condition, to provide a general baseline, and introduced viewpoint changes between prime stimulus and test. Target faces could now appear in either frontal view or rotated by 22° to either the left or right. Changes in viewpoint are a useful manipulation, as they increase the difficulty of the task and lessen the possibility of picture matching. Although Experiment 1 had involved changes from grayscale to color and the format of the prime and test images were quite different, we felt that increasing the task demands even further may help provide more evidence for a dynamic advantage for face recognition. More generally, showing that a matching advantage is maintained across viewpoints would also strengthen the hypothesis that motion in depth facilitates the encoding of identity.

4.1. Methods

4.1.1. Participants

Sixteen right-handed observers aged 19–26 (mean age: 22 years) participated in this study (seven females and nine males).

4.1.2. Stimuli

Video clips of 15 approaching avatars, and their corresponding target faces in frontal view, as well as 22° on either side of the face were used in the current experiment. In addition, static target images from 15 individuals shown from the three different viewpoints served as distractors for 'different' trials.

Twenty-nine consecutive frames were taken from the video clips of the approaching figures. For each presentation of an approaching stimulus, the first 15 sequential frames were selected out of these 29 and played forwards. For the receding stimuli, the last 15 frames were taken out of the 29 and played backwards. Thus, the end frame of both the approaching and the receding sequences was the same as the static frame, i.e., the middle frame of the 29 rendered ones. The facial image in the stopping frame for both conditions subtended 3.3° × 4.3° (width × height) (Fig. 1B), the body subtended 8.6° × 5.7°. The static prime stimulus was chosen to be the stopping frame of both sequences as described in the methods section of Experiment 1.

4.1.3. Procedure

The procedure was the same as in Experiment 1 except that observers were told that the prime would sometimes be a short video clip of an approaching person (approaching prime), a person moving away from the camera (receding prime) or a single static frame (static prime).

4.1.4. Task and design

The experiment consisted of 144 trials, separated into three blocks of 48 trials each. After each block, observers were encouraged to take a small break. They initiated the next block by pressing 'n' on the keyboard. Across the three blocks, observers completed 72 same and 72 different trials of which a third (24 trials) showed approaching, receding, or static primes. Out of these 24 trials, eight showed frontal view faces as targets (frontal targets), eight showed left view faces (left targets) and eight showed right view faces (right targets). 'Same' trials were constructed by showing faces of the same identity for prime and target. 'Different' trials were constructed by randomly selecting from the 15 additional target faces. The order of trials within each block was randomized separately for each observer on a block-by-block basis. On each trial a prime appeared in the middle of the screen for 600 ms. The prime was either an approaching animation sequence, a receding animation sequence or a static picture. After a blank of 300 ms the target face appeared in the middle of the screen. The target face was always a still image showing the head from one of the three viewpoints. On each trial the prime stimulus was randomly drawn from the stimulus set. The task of the observers was to determine if the prime and target faces came from the same individual. They were asked to respond as quickly and as accurately as possible using one of two marked keys. This key was the 's' key for 'same' and the 'l' key for 'different'. Fig. 2 shows the presentation sequence for a typical trial.

4.1.5. Data analysis

Repeated measures analyses of variance (ANOVAs) were used to compare the factors of interest (prime condition (approaching, receding, static) × target condition (same, different) × target viewpoint (front, left, right)). Additional post hoc contrasts were used to test for specific differences across conditions.

4.2. Results

4.2.1. Reaction times

Fig. 5 shows RT across set size for both target types. A 3 (prime condition) × 2 (target condition) × 3 (target viewpoint) repeated measures ANOVA revealed main effects of prime condition, $F(2, 30) = 8.3$, $p < 0.01$, target condition, $F(1, 15) = 5.4$, $p < 0.05$, target viewpoint, $F(2, 30) = 9.5$, $p < 0.001$, and an interaction between prime condition and target condition, $F(2, 30) = 3.7$, $p < 0.05$. There were no further interactions. To investigate the prime × target condition interaction in further detail, we performed separate ANOVAs on same and different trials.

For same trials, a two-way repeated measures ANOVA (prime condition × target viewpoint) revealed main effects of both prime condition, $F(2, 30) = 13.9$, $p < 0.001$, and
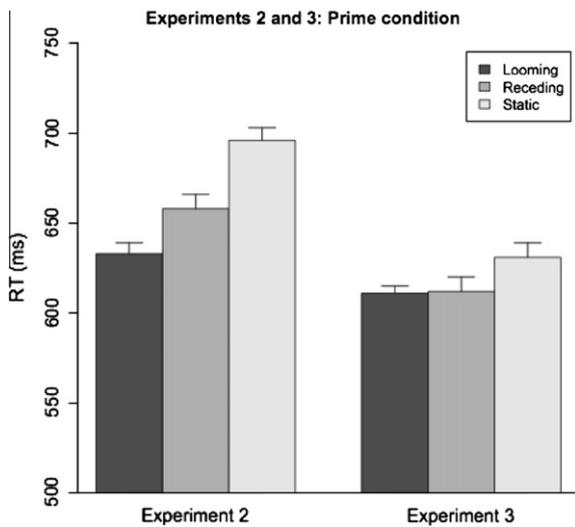
Fig. 4. Reaction time differences in ms for the three probe conditions 'looming' (white bars), 'receding' (dark gray bars) and 'static' (black bars) for Experiments 2 (left panel) and 3 (right panel). All RT results represent same trials. Error bars represent 95% confidence intervals calculated with the method described by Loftus and Masson (1994).

Fig. 5. Reaction times for correct trials in Experiment 4 for the target learned in motion (gray) and the target learned as static (black) across set size. Error bars represent 95% confidence intervals calculated with the method described by Loftus and Masson (1994).

target viewpoint, $F(2, 30) = 3.5$, $p < 0.05$, but no interaction between these two conditions. Importantly, post hoc analyses of the prime condition showed that the approaching prime stimuli gave rise to faster responses than the receding trials, $F(1, 15) = 4.3$, $p < 0.05$. Additionally, both approaching, $F(1, 15) = 27.6$, $p < 0.001$, and receding, $F(1, 15) = 10.2$, $p < 0.001$, trials were faster than static trials. These effects are illustrated in Fig. 4. The main effect of target viewpoint appeared to be driven by responses to frontal view faces ($M = 649$ ms, $SE = 49$ ms) being faster than responses to target faces oriented 22° to the right ($M = 684$ ms, $SE = 56$ ms) or 22° to the left ($M = 683$ ms, $SE = 43$ ms).

For different trials, there was no effect of prime condition, $F(2, 30) = 1.14$, $p = 0.2$, but a main effect of target viewpoint, $F(2, 30) = 4.4$, $p < 0.05$. The effect of viewpoint arose because responses to target faces oriented 22° to the right ($M = 733$ ms, $SE = 38$ ms), were slower than responses to target faces oriented 22° to the left ($M = 711$ ms, $SE = 40$ ms), $F(1, 15) = 9.9$, $p < 0.01$, and to frontal view faces ($M = 699$ ms, $SE = 34$ ms), $F(1, 15) = 7.1$, $p < 0.01$. There was no interaction between prime condition and viewpoint for the different trials.

4.3. Sensitivity

Accuracy was again about 80% on average (see Table 2). A 2 (prime condition) × 3 (viewpoint) repeated measures same ANOVA on d' showed a significant effect of prime condition, $F(2, 30) = 4.8$, $p < 0.05$. This was due to a slight drop in sensitivity for approaching ($M = 1.5$, $SE = 0.2$), $F(1, 15) = 5.0$, $p < 0.05$, and receding prime stimuli ($M = 1.5$, $SE = 0.2$), $F(1, 15) = 6.8$, $p < 0.05$, compared to static prime stimuli ($M = 1.9$, $SE = 0.2$). Importantly, there was
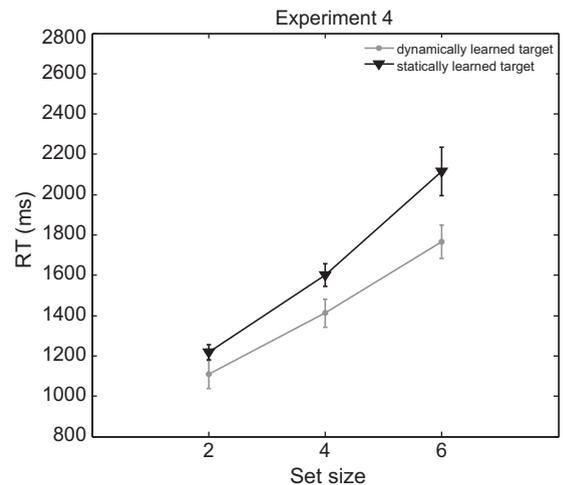
no difference in sensitivity between approaching and receding prime stimuli, $F(1, 15) = 0.02$, $p = 0.9$.

4.4. Discussion

The novel finding from Experiment 2 was that approaching trials gave rise to a reliable 25 ms advantage compared to receding trials. Although both types of dynamic primes were responded to more quickly than the static prime, suggesting that any form of motion in depth may be helpful to encode identity, the familiarity and/or salience of looming motion clearly has an additional effect. We note that as the final frame was equated between the approach and the receding sequence, the receding condition provided larger and higher resolution faces on average than the former during the initial phase of the prime stimulus. Thus, if performance depended on the quality of facial snapshots seen during the prime sequence, the receding condition would have been expected to lead to better performance.

Alternatively, because we used an identical stopping point, the images within the receding sequence were subjected to a larger overall change in size compared to the looming sequence. Such a change between prime and target could have disadvantaged receding trials (Kolers, Duchnicky, & Sundstroem, 1985). To explore this issue, we conducted a control experiment in which the looming and receding sequences were identical but were simply played forwards or backwards. That is, they neither started nor finished in the same place, but contained equal image variation. All six of the additional, naïve participants we ran in this condition continued to show an advantage of looming ($M = 562$ ms) over receding stimuli ($M = 618$ ms), ($F(1, 5) = 29$, $p < 0.01$).

Finally, some caution needs to be observed when interpreting the general advantage of the dynamic primes over the static prime conditions in this experiment. Although

Table 3
Reaction time (ms) data for Experiments 4 and 5. 95% Confidence intervals calculated with the method described by Loftus and Masson (1994) in brackets.

	Target learned as looming			Target learned as static			Target absent		
	2	4	6	2	4	6	2	4	6
Experiment 4	1108 (71)	1412 (75)	1768 (77)	1218 (40)	1601 (60)	2115 (84)	1458 (45)	2331 (67)	3217 (132)
	Target learned from multiple views			Target learned from a single image					
Experiment 5	1432 (129	1871 (115)	2120 (58)	1200 (119)	1688 (105)	2138 (181)	1654 (67)	2584 (112)	3593 (187)

Table 4
Accuracy (% correct) data for Experiments 4 and 5. 95% Confidence intervals calculated with the method described by Loftus and Masson (1994) in brackets.

	Target learned as looming			Target learned as static			Target absent		
	2	4	6	2	4	6	2	4	6
Experiment 4	89 (3)	92 (2)	91 (3)	84 (2)	85 (2)	86 (2)	85 (2)	76 (3)	71 (3)
	Target learned from multiple views			Target learned from a single image					
Experiment 5	87 (3)	85 (3)	90 (2)	84 (4)	87 (3)	88 (2)	84 (2)	78 (2)	69 (4)

the size of this advantage – 63 ms for approaching, 38 ms for receding primes – was considerably larger than that observed in Experiment 1, it was accompanied by a slight drop in sensitivity. This pattern may simply reflect an overall increase in the difficulty associated with the variety of prime type and target viewpoints, a notion supported by a general slowing of RT relative to Experiment 1. However, the nature of the immediate matching task does make it possible for observers to adopt a speed-accuracy tradeoff as part of their response strategy, a limitation that partly motivated our decision to seek converging evidence from a different task in Experiments 4 and 5.

5. Experiment 3

Experiments 1 and 2 suggest that body motion can affect the processing of faces. More specifically, the highly familiar event of seeing a person approach can improve identity matching performance relative to either static (Experiment 1) or receding (Experiment 2) baselines. An obvious question to ask next is whether this advantage relies on the presence of the body, or whether simply moving an isolated head in depth would produce a similar effect. To test this idea, in Experiment 3 we used exactly the same stimuli and design as in Experiment 2, but simply masked the bodies so that they were invisible. As mentioned in Section 1, the use of computer-generated figures provides the opportunity to easily design and create control conditions, such as this.

We should note that as the heads were identical to those used in Experiment 2, they not only scaled in size and resolution as they approached, but also translated with the slight variations in horizontal and vertical positions that are characteristic of human walking. Phenomenologically, even in the absence of body form information, these stimuli continue to convey a compelling sense of a person moving in depth. Our question was whether this sense of

human motion would still influence the pattern of matching results in the absence of an explicit body.

5.1. Methods

5.1.1. Participants

Fourteen right-handed observers aged 20–36 (mean age: 24.7 years) participated in this study (six females and eight males).

5.1.2. Stimuli

The same set of stimuli was used as in Experiment 2. However, rather than presenting the whole figures we occluded the bodies so that only the heads could be seen from the original sequences. To do this, we simply masked the bodies so that they were invisible.

5.1.3. Task, design and data analysis

The task, design and data analysis were identical to Experiment 2.

5.2. Results

5.2.1. Reaction times

The data were analyzed using the same ANOVA model as in Experiment 2. A 3 (prime condition) $\times$ 2 (target condition) $\times$ 3 (target viewpoint) repeated measures ANOVA revealed main effects of target condition, $F(1, 13) = 8.1$, $p < 0.05$, and target viewpoint, $F(2, 26) = 14.5$, $p < 0.001$, and a target condition $\times$ target viewpoint interaction, $F(2, 26) = 14$, $p < 0.001$, but no effect of prime condition, $F(2, 26) = 3$, $p = 0.07$. Analysing same and different trials separately to investigate the origin of the interaction term we found no effect of motion in the same, $F(2, 26) = 1.7$, $p = 0.5$, or different, $F(2, 26) = 1.8$, $p = 0.5$, trials. The ANOVAs did not show a viewpoint effect in the different trials, $F(2, 26) = 0.02$, $p = 0.9$, but in the same trials, $F(2, 26) = 35.0$, $p < 0.001$. This effect was due to an advantage of
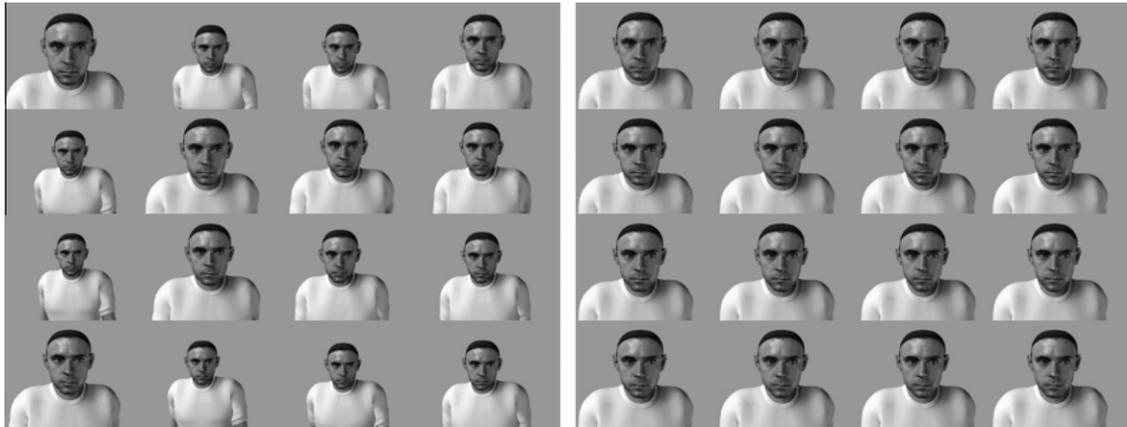
Fig. 6. Example stimulus as shown in the learning phase of Experiment 5. The matrix on the left depicts the one consisting of multiple static images, the one on the right the one of a single image shown repeatedly.

frontal targets (M = 579 ms, SE = 56 ms), over targets oriented 22° to the right (M = 653 ms, SE = 54 ms), $F(2, 26)$ = 69.7, $p < 0.001$, and targets oriented 22° to the left (M = 621 ms, SE = 57 ms), $F(2, 26)$ = 22.0, $p < 0.001$.

5.2.2. Sensitivity

The ANOVA on $d′$s revealed both prime condition, $F(2, 26)$ = 5.00, $p < 0.05$, and target viewpoint effects, $F(2, 26)$ = 6.9, $p < 0.01$, but no interactions between these factors. The viewpoint effect was due to decreasing sensitivity with increasing viewpoint differences between prime and target. The effect of prime condition was due to a disadvantage in performance for looming primes, M = 1.9, SE = 0.2, compared to both receding (M = 2.1, SE = 0.2), $F(1, 13)$ = 9.3, $p < 0.01$) and static primes (M = 2.2, SE = 0.2), $F(1, 13)$ = 9.5, $p < 0.01$. See Table 2 for details.

5.3. Discussion

The isolated heads used in this experiment failed to show the dynamic advantage observed in Experiments 1 and 2, despite the use of identical motion patterns. This suggests that the context of the body plays a vital role in the previous matching advantages. Fig. 4 indicates faster responses to static compared to moving prime stimuli. This trend, however, is not reliable. More generally, Fig. 4 indicates that observers are faster and more accurate when the body is absent (Experiment 3) than when it is present (Experiment 2). The additional RT cost in Experiment 2 may reflect some sort of mandatory processing of the body information (Thornton & Vuong, 2004). An alternative explanation might be that the absence of the body in Experiment 3 reduced the image difference between learning and test stimuli. Hence, the matching of the learning and test stimuli might be facilitated in Experiment 3, leading to faster overall responses. However, given that reaction times were fastest in Experiment 1, in which the body was also present, another explanation for the RT cost in Experiment 2 is simply that this set of observers was

generally responding more slowly compared to those in Experiments 1 and 3.

The drop in sensitivity in Experiment 2 compared to Experiment 3 could be due to a decreased signal-to-noise ratio as all stimuli had the same body, thus providing no useful matching information for the observer. Interestingly, in the current experiment, observers were less accurate at matching looming primes compared to both receding and static primes. The difference is small but reliable and might arise due to the fact that the average size of the looming heads was generally smaller than the size of both the static and the receding heads. This did not affect observers' performance in Experiments 1 and 2, in which the bodies were present but might have led to the decreased accuracy for looming primes in the current experiment, in which the heads were presented without the body.
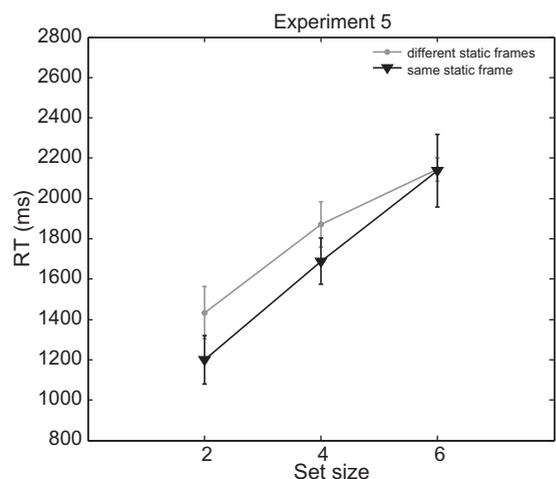

Fig. 7. Reaction times for correct trials in Experiment 5 for the target learned from different static pictures (gray) and the target learned from same static pictures (black) across set size. Error bars represent 95% confidence intervals calculated with the method described by Loftus and Masson (1994).

6. Experiment 4

In the first series of experiments (Experiments 1–3), we used an immediate matching task to investigate whether the approach of a person facilitates the short-term encoding of their identity. The purpose of the next series of experiments (Experiments 4–5), was to determine if this dynamic advantage also holds across long-term encoding of identity. Previously we used a delayed visual search paradigm to show that the matching advantages seen for nonrigid facial motion (Thornton & Kourtzi, 2002) extended across time (Pilz et al., 2006, 2009). Here, we used the same delayed visual search paradigm to further investigate the advantage seen in Experiments 1–2.

In the delayed visual search paradigm, observers are first familiarized with two target faces using an incidental learning technique (Knappmeyer et al., 2003). In the current experiments, this involved images of two avatars alternating on the screen while observes filled out a detailed questionnaire about them. One sequence always consisted of a single static image, the other was in motion or contained relevant control manipulations. During test, a visual search array of two, four or six *static* faces was presented. The task of the observer was simply to decide whether one of the previously learned faces was present.

Previously, we have suggested that such delayed visual search paradigms may be very useful for studying the dynamic aspects of face recognition over time (Pilz et al., 2006, 2009). That is, extended exposure to a small set of target identities may provide a better opportunity for dynamic information to have an effect, compared to brief exposure to multiple targets, as is usually the case with traditional old-new recognition tasks.

In Experiment 4, one target was presented in the context of the approach sequence used in Experiments 1–3. The other target was presented as static snapshots taken from the approach sequence. Our question was whether the dynamic context during learning would affect subsequent search behaviour.

6.1. Methods

6.1.1. Participants

Twenty-two right-handed observers aged 22–29 (mean age: 23.5 years) participated in this study (16 females and six males). None of the observers had participated in any of the other experiments presented in this paper.

6.1.2. Stimuli

The stimuli used were as described in Section 3.1. For learning, two avatars were randomly selected out of the set of 15 rendered animation sequences for each observer, one as a static frame, the other, as a video sequence of an approaching person. For testing, 15 heads, those of the two targets and thirteen additional ones, were used as distractor faces, each rendered from frontal view and 22° to the right and 22° to the left.

6.1.3. Task and design

The delayed visual search task used in this experiment consisted of a learning phase and a test phase. In the learning phase, observers were familiarized with two avatars as targets. One of the targets was presented dynamically in an approach sequence, whereas the other target was presented as static snapshot. These avatars alternated 100 times on the screen, each time presented for 600 ms with an inter stimulus interval of 2 s. While watching the avatars, observers filled out a questionnaire. They were asked to rate factors such as the apparent attractiveness, age, kindness, aggressiveness and intelligence of the two persons, as well as to describe their prominent facial features. After the avatars stopped alternating, observers were required to take a short break of approximately 3 min. This method of learning has been shown to be effective in several previous studies (Knappmeyer et al., 2003; Pilz et al., 2006, 2009).

On each trial of the test phase, two, four or six static faces were shown in a circular search array. Observers were asked to respond as quickly and accurately as possible to whether either one of the learned faces (targets) was present in the search pattern or not. The targets were present on 66% of trials, with each of the familiarized faces appearing equally often. Distractor faces were randomly selected from the set described above. 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 set size by target-type and viewpoint trials occurred with equal frequency and were randomized for each observer individually.

6.1.4. Data analysis

In this and subsequent experiments, we examined both the speed and accuracy of responses as a function of set size on target-present trials. RTs from correct trials are reported for trials in which one of the target faces was present in the visual search array. Repeated measures were used to compare the factors of interest (target type × viewpoint × set size). Table 3 shows correct RTs and Table 4 the accuracy data for both target-present and target-absent trials for Experiments 4 and 5.

6.2. Results

6.2.1. Reaction times

A 2 (target type) × 3 (set size) ∗ 3 (viewpoint) repeated measures ANOVA was used to explore correct target-present trials. There was a main effect of target type, $F(1, 21) = 5.3$, $p < 0.05$, with the target learned from approaching sequences leading to faster visual search responses than the target learned from static snapshots. There was also a main effect of set size, $F(2, 42) = 90.0$, $p < 0.001$, with longer responses for arrays containing more items. There was no main effect of viewpoint, $F(2, 42) = 1$, $p = 0.3$, but a significant target type × set size interaction,

$F(2, 42) = 3.0$, $p < 0.05$. This interaction was due to a steeper search slope for the target learned from static snapshots.

6.2.2. Accuracy

A 2 (target types) × 3 (set size) ∗ 3 (viewpoint) repeated measures ANOVA on accuracy did not reveal any main effects of target condition, $F(1, 21) = 3$, $p = 0.1$, viewpoint, $F(2, 42) = 2$, $p = 0.1$, or set size, $F(2, 42) = 1$, $p = 0.3$, but an interaction between viewpoint and set size, $F(2, 42) = 5.5$, $p < 0.01$, which was mainly due to lower accuracy scores for faces angles to the right at set size 2 and faces angles to the left for set size 6.

6.3. Discussion

Observers were significantly faster at finding a target face in the search array if that face had been learned in the context of a moving rather than a static avatar. This advantage held across viewpoints. In addition, we found a target type × set size interaction. Search slopes for statically learned targets were steeper than slopes for dynamically learned targets. This difference in slopes suggests that the search for the dynamically learned target was more efficient than the search for the statically learned one. Taken together these results suggest, that in addition to short-term advantages, as shown in Experiments 1–2, looming can also facilitate long-term encoding of identity and can facilitate later recognition.

As the stimulus parameters for both learned avatars were identical during the test phase, the origin of any difference between conditions must have occurred during learning. As in Experiments 1 and 2, the stimulus motion per se did not contain any information about the identity of the moving person. Therefore, the present results underline the hypothesis that our visual system has mechanisms that facilitate the encoding of stimuli that move in a relevant and familiar way without containing relevant dynamic information about the specific identify of the stimulus.

7. Experiment 5

Experiment 4 showed a clear advantage for recognizing targets learned from an approach sequence in a visual search task across viewpoints. In Experiment 5, we investigate whether this advantage is due to the motion of the face and body or simply due to the additional static information contained in the individual frames of the moving sequence. Observers were trained on static frames from the moving sequence presented in a 4 × 4 matrix. For one target identity, this matrix contained the same static view repeated 16 times. For the other target, 16 different frames were randomly selected from the moving sequence used in Experiment 4. Our question was whether this additional static information would lead to a speed and/or accuracy advantage.

7.1. Materials and methods

7.1.1. Participants

Thirteen right-handed observers aged 19–30 (mean age: 27 years) participated in this study (ten females and three males). One observer had to be excluded from further analysis, because he had already participated in one of the previous experiments.

7.1.2. Stimuli

The stimuli were the same as in Experiment 4 except that observers were familiarized with a matrix of 16 static pictures during the learning phase. In one condition, the 16 pictures showed the same static frame as used in Experiment 4 (single-picture condition). By comparison, in the other condition, the matrix showed 16 different frames from the moving sequences used in Experiment 4 (multiple-pictures condition). These 16 frames were randomly arranged in a different order on each presentation. This manipulation ensured that observers saw more than one static picture even if they tended to look at a preferred location in the array. Fig. 6 gives an example of the stimuli used during the learning phase.

7.1.3. Task, design and data analysis

The task, design, and data analyses were the same as in Experiment 4.

7.2. Results

7.2.1. Reaction times

A 2 (target type) × 3 (set size) ∗ 3 (viewpoint) repeated measures ANOVA revealed no main effects of target type, $F(1, 11) = 0.8$, $p = 0.4$, or viewpoint, $F(2, 22) = 2.3$, $p = 0.1$, but an effect of set size, $F(2, 22) = 18$, $p < 0.001$. Fig. 7 shows reaction times across set size for both target types.

7.2.2. Accuracy

2 (target type) × 3 (set size) ∗ 3 (viewpoint) repeated measures ANOVA revealed no main effects of target type, $F(1, 11) = 0.05$, $p = 0.8$, or viewpoint, $F(2, 22) = 0.9$, $p = 0.4$, but an effect of set size, $F(2, 22) = 3.5$, $p < 0.05$. Observers' performance was best for set size 6 compared to set sizes 2 and 4 (see Table 4 for further detail).

7.3. Discussion

The results of Experiment 5 suggest that the advantage for learning approaching stimuli over static ones found in Experiment 4 cannot be solely due to additional static information provided in the looming sequence as compared to a single static snapshot. Observers did not show any difference in performance for the individuals they learned from multiple static pictures versus the individual they learned from a single static snapshot.

8. General discussion

In the current paper, we investigated whether the context of an approaching person affected subsequent identity

decisions. We examined both short retention intervals, using a sequential matching paradigm (Experiments 1–3), and long retention intervals, using a delayed visual search paradigm (Experiments 4–5). In the first series of experiments, we obtained a response time advantage for trials in which a human figure was animated to approach the observer, compared to static (Experiment 1) or receding (Experiment 2) trials. We did not find an advantage of approaching versus receding primes in which only isolated heads were used (Experiment 3). The movement of the whole body in depth thus seems to be an important factor in obtaining the dynamic reaction time advantage.

The second series of experiments showed that the dynamic advantage for approaching figures obtained in Experiments 1–3 persists over long retention intervals. Viewing an approach sequence in the learning phase speeded up performance in a subsequent visual search task compared to a static learning phase (Experiment 4). Search performance was also not influenced when additional static information was provided in the learning phase, equating the views seen during static and dynamic approach sequence (Experiment 5).

While the majority of research continues to treat faces as isolated objects, the evidence presented here suggests that the context of a moving body can directly influences the processing of facial identity. If such context effects can be found with a wider range of stimuli – using real video and more natural actions, for instance – and if such effects also generalize to other tasks and dependent measures (e.g., old/new recognition, accuracy), then this could have important implications for the way faces are studied, both in the laboratory and in applied, forensic settings. We return to these issues below. First though, it is important to consider how the dynamic effects observed in the current experiments may have come about:

Previous studies have suggested that the addition of motion might affect face processing in at least two different ways (Lander & Bruce, 2000; O'Toole et al., 2002). The "supplemental information hypothesis" suggests that advantages for a moving face may arise when a particularly distinctive smile, expression of surprise, or a nod becomes represented as a characteristic pattern of movement (e.g., Knappmeyer et al., 2003; Knight & Johnston, 1997; Lander & Bruce, 2000). In the current study, however, the moving faces were not animated with expressive gestures and the body motion was identical for all stimuli. Thus, there were no characteristic dynamic cues to identity. The "representational enhancement hypothesis" suggests that the addition of motion indirectly assists recognition by facilitating the recovery of facial structure. Changes of head position during walking, and scaling of feature resolution during approach, could potentially provide additional information, compared to a single snapshot. However, the lack of an advantage for isolated approaching heads (Experiment 3) or multiple snapshots (Experiment 5) would seem to argue against this explanation.

It seems almost certain that attention plays an important role in the current findings. Increased deployment of attention in dynamic conditions could affect both the quality and quantity of information encoded about the target faces, thus leading to the observed pattern of results.

Motion alone is known to be a very effective cue for attention (e.g., Driver & Baylis, 1989; Franconeri & Simons, 2003; Hillstrom & Yantis, 1994), although this explanation might also predict an advantage for isolated moving heads (Experiment 3). The human body, however, is likely to be very effective at attracting and holding attention. It has been shown that bodies are processed very rapidly (e.g., Johansson, 1975; Thorpe, Fize, & Marlot, 1996), and even when the figure is not relevant to the current task (e.g., Bosbach, Prinz, & Kerzel, 2004; Thornton & Vuong, 2004), which is also the case in the current study. In addition, body orientation has been shown to be one of the primary cues for joint attention (e.g., Lawson, Clifford, & Calder, 2009; Nummenmaa & Calder, 2009) and the extraction of such cues may well demand attention (Cavanagh, Labianca, & Thornton, 2001; Chandrasekaran, Turner, Bülthoff, & Thornton, 2010; Thornton, Pinto, & Shiffrar, 1998; Thornton, Rensink, & Shiffrar, 2002). For the stimuli in the current study, this might suggest that if the moving figure is attracting and holding attention, the related head and face may also benefit relative to the static condition.

Aside from attracting and holding attention, the moving body may also provide a context that influences the encoding of information about the face. There are many examples in the literature showing that context can improve the recognition of objects (e.g., Biederman, Glass, & Stacy, 1973; Chun & Jiang, 1999; Davenport & Potter, 2004; Palmer, 1975; see Oliva & Torralba, 2007 for a recent review), particularly when image resolution is low or target objects are degraded (e.g., Biederman, 1981; Torralba, 2009). While "context" typically refers to the background of a scene, the functional relationship between co-occurring objects has also been shown to be relevant (e.g., Chun & Jiang, 1999; Green & Hummel, 2006). For example, a table can facilitate recognition of a chair, and a nail, the recognition of a hammer (Green & Hummel, 2006). Within the context of faces, de Gelder and colleagues have demonstrated that whole body signals help to facilitate the processing of expressions (de Gelder, 2006, 2009). In another study O'Toole and colleagues showed observers real videos of moving and static faces and whole persons and demonstrated that human identification is best when the whole person was seen in motion (O'Toole et al., in press). Even though the moving body in our experiments is not informative, its presence may help guide face perception in some way, particularly in early frames of the animation where image resolution would be relatively low.

There is an additional sense in which a moving body may provide a contextual advantage: Considerable evidence suggests that the motor system is directly involved in the visual perception of other people's bodies. Theoretically, studies of "embodied cognition" have long postulated that we use our bodies and actions to make sense of the world (e.g., Clark, 1997; Lakoff & Johnson, 1999; see Wilson, 2002, for a review) and the tight coupling between perception and action is extremely well documented (e.g., Humphreys & Riddoch, 2001; Prinz, 1997; Riddoch, Humphreys, Edwards, Baker, & Willson, 2003; Schutz-Bosbach & Prinz, 2007). The discovery of so-called "mirror neurons" in primates – cells that respond both to the observation and execution of actions – have provided

particularly important evidence in this regard (Kurata & Tanji, 1986; Rizzolatti et al., 1988; for review see Rizzolatti & Craighero, 2004). Functionally equivalent networks of areas have also been proposed in humans (e.g., Decety et al., 1997; Gazzola & Keysers, 2009; Kilner, Neal, Weiskopf, Friston, & Frith, 2009; Rizzolatti et al., 1996). If additional action-related processing occurred in the current work when a moving body was present, this could have strengthened or supplemented face-related information contained in our dynamic stimuli.

The results of Experiment 2 showed that a looming human figure leads to better performance than a receding figure. Such an advantage is consistent with the finding that many animals have evolved a bias for detecting and responding to looming events due to their relevance for survival (e.g., Maier & Ghazanfar, 2007; Maier et al., 2004; Schiff, 1965; Schiff et al., 1962). Both the familiarity and predictability of looming events could contribute to this processing bias. For example, in primates, it has been shown that more neurons are tuned to familiar views of objects and faces than unfamiliar views (Wachsmuth, Oram, & Perret, 1994). There is also evidence that the selectivity of cells in temporal cortex is biased towards stimuli experienced as an adult (Logothetis, Pauls, & Poggio, 1995, Perrett, Oram, & Ashbridge, 1998) and that spatio-temporal predictability of such representations may enhance firing rates (Perrett, Xiao, Barroclough, Keysers, & Oram, 2009). Similar experience-based neuronal plasticity might also contribute to the current dynamic advantage.

Elsewhere, we have suggested that advantages for moving over static stimuli might reflect the involvement of what Freyd (1987) termed, "dynamic mental representations". The central notion here is that by retaining both spatial and temporal dimensions relating to an event some behavioural advantage may be achieved (e.g., allowing you to anticipate the arrival of a dangerous projectile). Such dynamic mental representations have been introduced in a number of perceptual domains including face recognition (Thornton & Kourtzi, 2002), object recognition (Kourtzi & Nakayama, 2001; Stone, 1998) and biological motion processing (Cavanagh et al., 2001). Hubbard (2005) presents a very comprehensive review of many of the issues relating to this area.

Traditionally, such dynamic mental representations were thought to influence short-term rather than long-tem memory processes (Freyd, 1987; Freyd & Johnson, 1987; Kourtzi & Nakayama, 2001; Thornton & Kourtzi, 2002). Recently, however, Matthews et al. (2007) reported memory advantages for dynamic versus static scenes that persisted over 7 and 28-day retention intervals. The delay of several minutes that we introduced between study and test in the current delayed visual search paradigm, although not nearly as dramatic, also suggests the involvement of long-term memory systems. Matthews et al. (2007) suggest that a spatiotemporal version of "long-term" object file theory (Hollingworth & Henderson, 2002; Kahneman, Treisman, & Gibbs, 1992) may be a useful framework within which to model and explore the influence of dynamic mental representations. Relating this idea to the current work, it suggests that during encoding information about position and motion are explicitly stored along with other object properties. During retrieval, such information may act as additional cues to the identity of an object, leading to performance advantages.

In this paper we have presented some initial evidence that the motion of an approaching person can affect later facial identity decisions. While we have tried to suggest several potential explanations for this effect, clearly, more research is necessary. It is quite possible, for example, that other complex, multi-part objects, such as animals, bicycles, or cars, would show a similar dynamic context effect. We do believe that our results relate to the close functional relationship between the body and the face. Other relationships such as between a bicycle and its saddle, or a boat and its sails might produce a similar advantage. This is clearly a useful area of future research.

There are several other directions in which we feel this work could be usefully extended. For example, it would be interesting to establish whether other types of human action afford a similar advantage. Would observing faces in the context of bodies engaged in sports activities or everyday actions, such as making a cup of tea, also lead to dynamic advantages? While we have tried to focus on tasks that are appropriate for use with dynamic stimuli, it would also be interesting to explore other paradigms, such as traditional old/new recognition. Tasks that focus more on accuracy, rather than speed of response, would be particularly interesting with a view to developing forensic or other real-world applications.

On a related note, we mentioned in the introduction that another important goal of the current paper was to demonstrate the potential of using computer graphics and virtual reality techniques in an experimental context. We firmly believe that the added control and flexibility offered by these methods make them invaluable tools. Clearly though, our stimuli cannot be considered natural. While photo-realistic animated figures may soon be available, it may always make sense to confirm novel findings, such as ours, with live-action video (e.g., Burton et al., 1999; Roark, O'Toole, & Abdi 2003; Schiff et al., 1986), even though the range of control conditions cannot be replicated. Alone, neither approach may be sufficient to fully understand the interaction between face and body.

Finally, it will be interesting to probe into the neural underpinning of the current dynamic advantage. The superior temporal sulcus (STS) has already been implicated in studies of facial motion (e.g., Haxby, Hoffman, & Gobbini, 2000; Hoffman & Haxby, 2001; Puce, Allison, Bentin, Gore, & McCarthy, 1998; Schultz & Pilz, 2009), body motion (e.g., Beauchamp, Lee, Haxby, & Martin, 2003; Grossman & Blake, 2002; Pelphrey et al., 2003; Puce & Perrett, 2003; Saygin, Wilson, Hagler, Bates, & Sereno, 2004), action understanding, and social attention (Allison, Puce, & McCarthy, 2000). In both tasks as used in the current paper, we might predict increased activation in STS during the presentation of the moving body, relative to the static conditions. Of additional interest would be whether the static facial images used as targets in both sets of studies would also evoke responses in STS, even when there is no physical motion present.

Acknowledgements

References

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: Role of the STS region. *Trends in Cognitive Science, 4*(7), 267–268.

Ball, W., & Tronick, E. (1971). Infant responses to impending collision – Optical and real. *Science, 171*(3973), 818–820.

Beauchamp, M. S., Lee, K. E., Haxby, J. V., & Martin, A. (2003). FMRI responses to video and point-light displays of moving humans and manipulable objects. *Journal of Cognitive Neuroscience, 15*(7), 991–1001.

Biederman, I. (1981). On the semantics of a glance at a scene. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 213–263). Hillsdale, NJ: Erlbaum.

Biederman, I., Glass, A. L., & Stacy, W. (1973). Searching for objects in real-world scenes. *Journal of Experimental Psychology, 97*, 22–27.

Blanz, V., & Vetter, T. (1999). A morphable model for the synthesis of 3D faces. In *Proceedings, SIGGRAPH*, Vol. 99, pp. 187–194.

Bosbach, S., Prinz, W., & Kerzel, D. (2004). A Simon-effect with stationary moving stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 30*(1), 39–55.

Bower, T. G. R., Broughton, J. M., & Moore, M. K. (1970). Infant responses to approaching objects: An indicator for response to distal variables. *Perception and Psychophysics, 9*, 193–196.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436.

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*(4), 339–360.

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*(3), 207–218.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77*(Pt. 3), 305–327.

Bülthoff, H. H., & Bülthoff, I. (1987). Combining neuropharmacology and behavior to study motion detection in flies. *Biological Cybernetics, 55*, 313–320.

Burton, M. A., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video. *Psychological Science, 10*, 243–248.

Carney, T. (1997). Evidence for an early motion system which integrates information from the two eyes. *Vision Research, 37*, 2361–2368.

Cavanagh, P., Labianca, A., & Thornton, I. M. (2001). Attention-based visual routines: Sprites. *Cognition, 80*, 47–60.

Chandrasekaran, C., Turner, L., Bülthoff, H. H., & Thornton, I. M. (2010). Attentional networks and biological motion. *Psihologija, 43*(1), 5–20.

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

Chuang, L., Vuong, Q. C., Thornton, I. M., & Bülthoff, H. H. (2006). Recognising novel deforming objects. *Visual Cognition, 14*, 85–88.

Chun, M. M., & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science, 10*, 360–365.

Clark, A. (1997). *Being there: Putting brain, body, and world together again*. Cambridge, MA: MIT Press.

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in background and scene perception. *Psychological Science, 15*(8), 559–564.

de Gelder, B. (2006). Towards the neurobiology of emotional body language. *Nature Reviews, Neuroscience, 7*, 242–249.

de Gelder, B. (2009). Why bodies? Twelve reasons for including bodily expressions in affective neuroscience. *Philosophical Transactions of the Royal Society B, 364*, 3475–3484.

Decety, J., Grézes, J., Costes, N., Perani, D., Jeannerod, M., Procyk, E., et al. (1997). Brain activity during observation of actions. Influence of action content and subject's strategy. *Brain, 120*, 1763–1777.

Derrington, A. M., Allen, H. A., & Delicato, L. S. (2004). Visual mechanisms of motion analysis and motion perception. *Annual Review of Psychology, 55*, 181–205.

Driver, J., & Baylis, G. C. (1989). Movement and visual attention: The spotlight metaphor breaks down. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 448–456.

Franconeri, S. L., & Simons, D. J. (2003). Moving and looming stimuli capture attention. *Perception and Psychophysics, 65*, 1–12.

Freyd, J. J. (1987). Dynamic mental representations. *Psychological Review, 94*, 427–438.

Freyd, J. J., & Johnson, J. Q. (1987). Probing the time course of representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 259–268.

Gazzola, V., & Keysers, C. (2009). The observation and execution of actions share motor and somatosensory voxels in all tested subjects: Single-subject analyses of unsmoothed fMRI data. *Cerebral Cortex, 19*, 1239–1255.

Green, C., & Hummel, J. E. (2006). Familiar interacting object pairs are perceptually grouped. *Journal of Experimental Psychology: Human Perception and Performance, 32*(5), 1107–1119.

Grossman, E. D., & Blake, R. (2002). Brain areas active during visual perception of biological motion. *Neuron, 35*(6), 1167–1175.

Hassenstein, B., & Reichardt, W. (1956). Systemtheoretische analyse der Zeit-, Reihenfolgen- und Vorzeichenauswertung bei der Bewegungsperzeption des Rüsselkäfers Chlorophanus. *Zeitschrift für Naturforschung, 11b*, 513–524.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends Cognitive Science, 4*, 223–233.

Henderson, Z., Bruce, V., & Burton, A. M. (2001). Matching the faces of robbers captured on video. *Applied Cognitive Psychology, 15*, 445–464.

Hill, H., & Johnston, A. (2001). Recognizing sex and identity from the biological motion of faces. *Current Biology, 5*(11), 880–885.

Hillstrom, A. P., & Yantis, S. (1994). Visual motion and attentional capture. *Perception and Psychophysics, 55*, 399–411.

Hoffman, E. A., & Haxby, J. V. (2000). Distinct representation of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience, 3*(1), 80–84.

Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 113–136.

Hubbard, T. L. (2005). Representational momentum and related displacements in spatial memory: A review of the findings. *Psychonomic Bulletin & Review, 12*, 822–851.

Humphreys, G. W., & Riddoch, M. J. (2001). Detection by action: Neuropsychological evidence for action-defined templates in search. *Nature Neuroscience, 4*, 84–88.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics, 14*, 201–211.

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

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24*(2), 175–219.

Kilner, J. M., Neal, A., Weiskopf, N., Friston, K. J., & Frith, C. D. (2009). Evidence of mirror neurons in human inferior frontal gyrus. *The Journal of Neuroscience, 29*(32), 10153–10159.

Knappmeyer, B., Thornton, I. M., & Bülthoff, H. H. (2003). Facial motion biases the perception of facial form. *Vision Research, 43*, 1921–1936.

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

Knoblich, G., Thornton, I. M., Grosjean, M., & Shiffrar, M. (2006). *The human body. Perception from the inside out*. New York, NY: Oxford University Press.

Kolers, P. A., Duchnicky, R. L., & Sundstroem, G. (1985). Size in the visual processing of faces and words. *Journal of Experimental Psychology: Human Perception and Performance, 11*(6), 726–751.

Kourtzi, Z., & Nakayama, K. (2001). Dissociable signatures of processing for moving and static objects. *Visual Cognition, 9*, 248–264.

Krekelberg, B., & Albright, T. D. (2005). Motion mechanisms in macaque MT. *Journal of Neurophysiology, 93*(5), 2908–2921.

Kurata, K., & Tanji, J. (1986). Premotor cortex neurons in macaques: Activity before distal and proximal forelimb movements. *Journal of Neuroscience, 6*(2), 403–411.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought*. New York: Basic Books.

Lander, K., & Bruce, V. (2000). Recognizing famous faces: Exploring the benefits of facial motion. *Ecological Psychology, 12*, 259–272.

Lander, K., & Bruce, V. (2003). The role of motion in learning new faces. *Visual Cognition, 10*, 897–921.

Lawson, R. P., Clifford, C. W. G., & Calder, A. J. (2009). About turn: The visual representation of human body orientation revealed by adaptation. *Psychological Science, 20*(3), 363–371.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin and Review, 1*, 476–490.

Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology, 5*, 552–563.

Maier, J. X., & Ghazanfar, A. (2007). Looming biases in monkey auditory cortex. *Journal of Neuroscience, 27*(15), 4093–4100.

Maier, J. X., Neuhoff, J. G., Logothetis, N. K., & Ghazanfar, A. A. (2004). Multisensory integration of looming signals by rhesus monkeys. *Neuron, 43*, 177–181.

Matthews, W. J., Benjamin, C., & Osborne, C. (2007). Memory for moving and static images. *Psychonomic Bulletin and Review, 14*(5), 989–993.

Nummenmaa, L., & Calder, A. J. (2009). Neural mechanisms of social attention. *Trends in Cognitive Science, 13*(3), 135–143.

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences, 11*, 520–527.

O'Toole, A. J., Phillips, P. J., Weimer, S., Roark, D. A., Ayyad, J., Barwick, R., et al. (in press). Recognizing people from dynamic and static faces and bodies: Dissecting identity with a fusion approach. *Vision Research*.

O'Toole, A. J., Roark, D., & Abdi, H. (2002). Recognizing moving faces: A psychological and neural synthesis. *Trends in Cognitive Science, 6*, 261–266.

Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory and Cognition, 3*, 519–526.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442.

Pelphrey, K. A., Mitchell, T. V., McKeown, M. J., Goldstein, J., Allison, T., & McCarthy, G. (2003). Brain activity evoked by the perception of human walking: Controlling for meaningful coherent motion. *Journal of Neuroscience, 23*(17), 6819–6825.

Perrett, D. I., Oram, M. W., & Ashbridge, E. (1998). Evidence accumulation in cell populations responsive to faces: An account of generalisation of recognition without mental transformations. *Cognition, 67*(1–2), 111–145.

Perrett, D. I., Xiao, D., Barroclough, N. E., Keysers, K., & Oram, M. W. (2009). Seeing the future: Natural image sequences produce 'anticipatory' neuronal activity and bias perceptual report. *Quarterly Journal of Experimental Psychology, 61*, 2081–2104.

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.

Pilz, K. S., Bülthoff, H. H., & Vuong, Q. C. (2009). Learning influences the encoding of static and dynamic faces and their recognition across different spatial frequencies. *Visual Cognition, 17*(5), 716–735.

Pilz, K. S., Thornton, I. M., & Bülthoff, H. H. (2006). A search advantage for faces learned in motion. *Experimental Brain Research, 171*, 436–447.

Poggio, T., & Reichardt, W. (1973). Considerations on models of movement detection. *Kybernetik, 13*, 223–227.

Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology, 9*, 129–154.

Puce, A., Allison, T., Bentin, S., Gore, J. C., & McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *Journal of Neuroscience, 18*(6), 2188–2199.

Puce, A., & Perrett, D. (2003). Electrophysiology and brain imaging of biological motion. *Philosophical Transactions of The Royal Society of London Series B: Biological Sciences, 358*(1431), 435–445.

Ratcliff, R. (1979a). Group reaction time distribution and an analysis of distribution statistics. *Psychological Bulletin, 86*(3), 446–461.

Ratcliff, R. (1979b). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114*(3), 510–523.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114*(3), 510–532.

Riddoch, M. J., Humphreys, G. W., Edwards, S., Baker, T., & Willson, K. (2003). Seeing the action: Neuropsychological evidence for action-based effects on object selection. *Nature Neuroscience, 6*, 82–89.

Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G., & Matelli, M. (1988). Functional organization of inferior area 6 in the macaque monkey. II. Area F5 and the control of distal movements. *Experimental Brain Research, 71*(3), 491–507.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*, 169–192.

Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Paulesu, E., Perani, D., et al. (1996). Localization of grasp representations in humans by PET: 1. Observation versus execution. *Experimental Brain Research, 111*, 246–252.

Roark, D., O'Toole, A. J., & Abdi, H. (2003). Human recognition of familiar and unfamiliar people in naturalistic video. In *Proceedings of the IEEE international workshop on analysis and modeling of faces and gestures*, pp. 36–43.

Saygin, A. P., Wilson, S. M., Hagler Jr., D. J., Bates, E., & Sereno, M. I. (2004). Poin-light biological motion perception activates human premotor cortex. *Journal of Neuroscience, 24*(27), 6181–6188.

Schiff, W. (1965). Perception of impending collision—A study of visually directed avoidant behavior. *Psychological Monographs, 79*, 1–26.

Schiff, W., Banka, L., & Galdi, G. D. (1986). Recognizing people seen in events via dynamic "mug shots". *American Journal of Psychology, 99*, 219–231.

Schiff, W., Caviness, J. A., & Gibson, J. J. (1962). Persistent fear responses in rhesus monkeys to the optical stimulus of "looming". *Science, 136*, 982–983.

Schultz, J., & Pilz, K. S. (2009). Natural facial motion enhances the cortical responses to faces. *Experimental Brain Research, 194*(3), 465–475.

Schutz-Bosbach, S., & Prinz, W. (2007). Perceptual resonance: Action-induced modulation of perception. *Trends in Cognitive Sciences, 11*(8), 349–355.

Shen, Y. C., & Franz, E. A. (2005). Hemispheric competition in left-handers on bimanual reaction-time tasks. *Journal of Motor Behaviour, 37*, 3–9.

Shimizu, H. (2002). Measuring keyboard response delays by comparing keyboard and joystick inputs. *Behaviour Research Methods, Instruments and Computers, 34*(2), 250–256.

Stone, J. V. (1998). Object recognition using spatiotemporal signatures. *Vision Research, 38*, 947–951.

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

Thornton, I. M., & Kourtzi, Z. (2002). A matching advantage for dynamic faces. *Perception, 3*, 1113–1132.

Thornton, I. M., Pinto, J., & Shiffrar, M. (1998). The visual perception of human locomotion. *Cognitive Neuropsychology, 15*, 535–552.

Thornton, I. M., Rensink, R. A., & Shiffrar, M. (2002). Active versus passive processing of biological motion. *Perception, 31*, 837–853.

Thornton, I. M., & Vuong, Q. C. (2004). Incidental processing of biological motion. *Current Biology, 14*(12), 1084–1089.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*, 520–522.

Torralba, A. (2009). How many pixels make an image? *Visual Neuroscience, 26*, 123–131.

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

Tronick, E. (1967). Approach response of domestic chicks to an optical display. *Journal of Comparative Physiology and Psychology, 64*, 529–531.

Vuong, Q. C., Hof, A. F., Bülthoff, H. H., & Thornton, I. M. (2006). An advantage for detecting dynamic targets in natural scenes. *Journal of Vision, 6*, 87–96.

Vuong, Q. C., & Tarr, M. J. (2004). Rotation direction affects object recognition. *Vision Research, 44*, 1717–1730.

Vuong, Q. C., & Tarr, M. J. (2006). Structural similarity and spatiotemporal noise effects on learning dynamic novel objects. *Perception, 35*, 497–510.

Wachsmuth, E., Oram, M. W., & Perret, D. I. (1994). Recognition of objects and their component parts – Response of single units in the temporal cortex of macaque. *Cerebral Cortex, 4*, 509–522.

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

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review, 9*, 625–636.