Vision Research xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Face recognition of full-bodied avatars by active observers in a virtual environment

Isabelle Bülthoff[a,*], Betty J. Mohler[a,1], Ian M. Thornton[b]

[a] Human Perception, Cognition and Action, Max Planck Institute for Biological Cybernetics, Germany
[b] Department of Cognitive Science, University of Malta, Malta

ARTICLE INFO

No of reviewers = 2

Keywords:
Faces
Active vision
Virtual reality
Avatars
View dependency
Motion

ABSTRACT

Viewing faces in motion or attached to a body instead of isolated static faces improves their subsequent recognition. Here we enhanced the ecological validity of face encoding by having observers physically moving in a virtual room populated by life-size avatars. We compared the recognition performance of this *active* group to two control groups. The first control group watched a *passive* reenactment of the visual experience of the active group. The second control group saw *static* screenshots of the avatars. All groups performed the same old/new recognition task after learning. Half of the learned faces were shown at test in an orientation close to that experienced during learning while the others were viewed from a new viewing angle. All observers found novel views more difficult to recognize than familiar ones. Overall, the active group performed better than both other groups. Furthermore, the group learning faces from static images was the only one to be at chance level in the novel-view condition. These findings suggest that active exploration combined with a dynamic experience of the faces to learn allow for more robust face recognition and point out the value of such techniques for integrating facial visual information and enhancing recognition from novel viewpoints.

1. Introduction

Identifying the people around us is a crucial social task in everyday life. Defining the circumstances under which we are best at this task will allow us to better understand face recognition mechanisms. In the current study we created a "virtual museum" – a virtual reality (VR) cave, populated with life-size, static avatars – to examine whether variations in perceptual experience during initial encounters with a new face influences subsequent identity encoding. Our VR environment allowed us to examined three specific learning components that might influence identity processing during day-to-day encounters with other people: 1) dynamic rather than static exposure to target identities as the participants moved around the avatars and thereby were exposed to rigid facial motion; 2) faces shown attached to a body rather than as isolated heads, and; 3) active rather than passive exploration of the stimuli by the participants. In the following sections we first review what is already known about these three aspects of face encoding. We then introduce the notion of viewpoint change as a useful measure of facial expertise, before finally giving a short overview of our study and its goals.

1.1. Dynamic advantage: facial motion facilitates recognition

There is now considerable evidence to suggest that exposure to moving faces can facilitate identity processing (for review see Lander & Butcher, 2015; O'Toole, Roark, & Abdi, 2002; Xiao et al., 2014). Dynamic face advantages have been reported both for rigid movements of the head, for example when a person shakes her head (e.g., Hill & Johnston, 2001; Pike, Kemp, Towell, & Phillips, 1997; Xiao, Quinn, Ge, & Lee, 2012), and the non-rigid deformations of the face that occur during speech or display of expression (e.g., Butcher, Lander, Fang, & Costen, 2011; Knappmeyer, Thornton, & Bülthoff, 2003; Knight & Johnston, 1997; Lander & Bruce, 2000). Such effects appear to generalize across types of task, for example standard old/new recognition tests of memory (e.g., Butcher et al., 2011; Lander & Chuang, 2005; Lander, Chuang, & Wickham, 2006), matching tasks (e.g., Girges, Spencer, & O'Brien, 2015; Knappmeyer et al., 2003; Thornton & Kourtzi, 2002), composite paradigms (Xiao et al., 2012; Xiao, Quinn, Ge, & Lee, 2013) and visual search (Pilz, Thornton, & Bülthoff, 2006), as well as types of stimuli, including video clips (e.g., Bruce & Valentine, 1988; Knight & Johnston, 1997; Lander & Bruce, 2000), point-light displays (Bruce & Valentine, 1988; Kozlowski & Cutting,

* Corresponding author at: Department of Human Perception, Cognition and Action, Max Planck Institute for Biological Cybernetics, Max-Planck-Ring 8, 72076 Tübingen, Germany.
E-mail address: isabelle.buelthoff@tuebingen.mpg.de (I. Bülthoff).
[1] Present address: Institut für Sportwissenschaft, Technical University Darmstadt, Germany.

https://doi.org/10.1016/j.visres.2017.12.001
Received 27 July 2017; Received in revised form 1 December 2017; Accepted 13 December 2017

1977) and computer-generated stimuli (e.g., Girges, Spencer, & O'Brien, 2015; Hill & Johnston, 2001; Knappmeyer et al., 2003). Importantly, such dynamic advantages can be found even when the quantity of static information is equated in appropriate control conditions (e.g., Lander & Bruce, 2000; Pilz et al., 2006).

Although it would be wrong to suggest that facial motion always gives rise to performance advantages in face recognition tasks (see for example, Bennetts et al., 2013, Experiment 1; Bruce & Valentine, 1988; Christie & Bruce, 1998; O'Toole et al., 2011), a dynamic learning context clearly has the potential to improve identity encoding. The studies reviewed above usually involved neurotypical young adults. However, it may also be the case that such dynamic contexts can be of particular benefit to individuals who might otherwise perform poorly on standard, static assessments of face recognition, either due to age or face-specific processing difficulties (e.g., Albonico, Malaspina, & Daini, 2012; Bennetts, Butcher, Lander, Udale, & Bate, 2015; Bulf & Turati, 2010; Longmore & Tree, 2013; Maguinness & Newell, 2014; Steede, Tree, & Hole, 2007). Finally, we should note that most of these studies on face recognition have only looked at moving stimuli presented in front of static observers. In contrast, the observers in our study were active; they moved around the faces and thus gained dynamic experience of otherwise static faces.

1.2. Body advantage: presence of the body with the face can facilitate face recognition

As with the majority of face research, the above findings have almost exclusively been based on studies using isolated heads. However, work from several domains of face processing now make it clear that it is also important to consider the face within the context of the body (e.g., Cox, Meyers, & Sinha, 2004; de Gelder, 2009; Ghuman, McDaniel, & Martin, 2010; Kiiski, Hoyet, Woods, O'Sullivan, & Newell, 2015; Robbins & Coltheart, 2012). With specific reference to identity processing, a number of laboratories have previously examined the influence of *movement* when a face is presented in the natural context of a body. For example, Burton, Bruce and colleagues (e.g., Bruce et al., 1999; Burton, Wilson, Cowan, & Bruce, 1999; Henderson, Bruce, & Burton, 2001) explored person recognition from CCTV footage. Although familiar individuals could be recognized quite successfully, the ability to match or recognize previously unfamiliar people from such videos was generally found to be quite poor. However, although it was found that participants typically rely mostly on the face when determining identity, when the body or gait information was removed from videos of familiar individual, recognition performance dropped slightly, but significantly, indicating some influence of body context (Burton et al., 1999).

O'Toole and colleagues (e.g., Hahn, O'Toole, & Phillips, 2015; O'Toole et al., 2011; O'Toole, Roark, Jiang, & Abdi, 2005; Rice, Phillips, & O'Toole, 2013; Roark, O'Toole, Abdi, & Barrett, 2006) have extensively studied the interaction between face and bodies using videos of natural walking movements and dyadic interaction. Findings from their work suggest that the body can make an independent contribution to person recognition, for example when facial detail is reduced by distance (Hahn, O'Toole, & Phillips, 2015). However, a more important role may be to provide a context for improving whole person (i.e., face and body) recognition, a role that is amplified when people are in motion (i.e., a *dynamic* body context).

Another approach that examined the idea of such a dynamic body context, is work that used avatars to simulate an actor walking towards an observer (Pilz, Vuong, Bülthoff, & Thornton, 2011). In one series of experiments, a range of 3D heads were placed on a single, identical body model. Although body shape and walking animation contained no distinguishing features, this familiar, dynamic body context allowed participants to respond more rapidly to a subsequent probe image than when the same identities were seen walking backwards away from the observer, moving without a body, or standing still. In a second series of experiments, participants were familiarised with two identities, one using the dynamic body context just described, and the other as either a single still image or multiple still images taken from the approaching animation sequence. Performance was assessed using a delayed visual search paradigm (Pilz et al., 2006). Dynamically learned target identities were found more rapidly than those learnt in either of the static conditions.

In a similar vein, Simhi and Yovel (2016) used whole-body walking videos taken from the Database of Moving Faces and People (O'Toole et al., 2005) to examine matching accuracy. Participants saw clips of unfamiliar individuals approaching the camera (dynamic condition) or multiple still images taken from the same video sequence (static condition). The task was to immediately match (Thornton & Kourtzi, 2002; Pilz et al., 2011) those identities to static images of either full bodies or isolated heads. In contrast to Pilz et al. (2011), they only found a dynamic advantage when testing with a full body image, not when testing with an isolated head. A comparison between dynamic and static body images during testing (Exp. 2) revealed no additional performance gains, suggesting that for unfamiliar individuals, the influence of a body context does not rely on characteristic movement, but rather on structural or representational advantages of seeing a familiar dynamic event (i.e. a walking approach sequence) during initial exposure.

As with the dynamic face advantage mentioned above, it is not always the case that exposure to a moving body necessarily leads to clearly improved performance. For example, Robbins & Coltheart, 2012 exposed adults and children to target identities using full head-body, head only, or body only stimuli, under both static and moving conditions. Although performance in both a match-to-sample and a learning task were best when test items contained the head and body, showing the influence of body context, motion did not lead to unequivocal performance advantages, although it did modulate patterns of results in a number of ways.

In summary, although there have been far fewer studies on the influence of body context on identity processing – compared to the literature on faces – there is clearly a suggestion that exposure to whole-body stimuli during learning can facilitate later processing. There is also some evidence that the addition of motion can further improve performance, although this dynamic body advantage would appear to require further verification.

1.3. Active advantage: dynamic observers may facilitate face recognition

The distinction between active and passive learning has been extensively studied in a number of research areas, including spatial navigation (e.g., Appleyard, 1970; for a recent review see Chrastil & Warren, 2012); scene recognition (e.g., Christou & Bülthoff, 1999; Motes, Finlay, & Kozhevnikov, 2006; Wang & Simons, 1999) and objects or the spatial layout of objects (e.g., James et al., 2002; Meijer & Van der Lubbe, 2011; Teramoto & Riecke, 2010; Wang & Simons, 1999). Typically in these studies – most of which have used either desktop or immersive VR environments – a crucial question is whether the sensori-motor and cognitive control signals which are unique to active conditions provide any form of performance advantage relative to passive conditions that share the same rich visual input. Generally, evidence for such an active advantage has been quite mixed (Chrastil & Warren, 2012). In the context of object recognition, for example, although an active advantage was found by Simons, Wang, and Roddenberry (2002), when the quality of passive visual information was equated, neither Christou and Bülthoff (1999), nor Teramoto and Riecke (2010) found performance advantages for active conditions.

Returning to the topic of face recognition, the study most relevant to the current work is a paper by Liu and colleagues (Liu, Ward, & Markall, 2007). In a series of experiments, using both old/new recognition and matching tasks, these authors compared the performance of participants who could actively rotate views of 3D laser-scanned faces to those who could only passively view playback of the active condition. Once passive participants had been given a comparable manual task to equate overall task difficulty, the authors found very consistent advantages for active exploration in five separate experiments. Interestingly, using the same procedures, they did not find an advantage in another experiment using non-biological category of objects, namely chairs (see also Meijer & Van der Lubbe, 2011).

Although, as noted above, the evidence for an active advantage from scene and object studies has been rather mixed, the study of Liu et al. (2007) clearly suggests that the context of faces has the potential to reveal differences between passively viewing and actively exploring a target stimulus. As described in more detail below, in the current study we examine the more naturalistic situation in which the active component involves a face-to-face encounter with life-size statues.

1.4. View dependency

Many important questions have arisen in relation to object recognition that also pertain to face recognition. One of the most heavily investigated questions in object recognition concerns our ability to recognize a given object under various orientations and illuminations that give rise to very different 2D images on the retina. Most familiar objects can be easily recognized from any viewpoint, although novel views typically incur a cost (Bülthoff & Edelman, 1992). For newly learned objects, a clear viewpoint effect occurs such that recognition is easier for views close to those under which the new object has already been seen (Braje, Kersten, Tarr, & Troje, 1998; Bülthoff & Edelman, 1992; Logothetis & Pauls, 1995; Tarr, 1995; Tarr & Pinker, 1989). It has also been suggested that the visual system is able to adopt a flexible approach when representing objects, combining view-based and view-independent properties depending on the precise nature of the stimuli and current task demands (Foster & Gilson, 2002; Vanrie, Willems, & Wagemans, 2001).

How does this view-based recognition model apply to faces considering their unique status among objects and the level of expertise that they require for proper subordinate level of recognition? Bülthoff & Edelman had used bent paperclips for their experiments testing view-based hypotheses (Bülthoff & Edelman, 1992). Wallraven, Schwaninger, Schuhmacher, and Bülthoff (2002) repeated these experiments using faces instead of paperclips, finding that the view-dependent model of object recognition could be extended to the domain of face recognition. Specifically, faces were better recognized from views that were near to learned views rather than further away. Such view dependency for faces has also been shown in neurophysiological studies (e.g., Oram & Perrett, 1992).

Faces can be tilted upward or downward around the horizontal axis with respect to their body or with respect to the observer ("pitch" see Fig. 1), turned around the vertical axis to the left or to the right ("yaw") or tilted from one side (roll). Faces seen under a slight yaw rotation during learning are better recognized at test (Troje & Bülthoff, 1996), but see Liu (2002). Unsurprisingly, changing viewpoint between learning and test is in general detrimental for recognition performance, but it depends on how viewpoint is changed. Viewpoint changes implemented by yaw rotations affect recognition more than by roll rotations (Van der Linde & Watson, 2010) but less than by pitch rotations (Favelle, Palmisano, & Maloney, 2007). Further, downward pitch rotations of the faces are more disruptive than upward pitch rotation (Favelle, Palmisano, & Avery, 2011; Favelle et al., 2007). In the current study, we use variations in pitch that naturally occur when looking at up at (tall) standing figures and down at sitting figures to modulate the difficulty of old/new face recognition.

1.5. Outline of the current study

A quick overview of the current study is given here, with more details of each learning condition given in the Methods section. We tested three different groups of participants, with each group learning faces in a different way. The participants of the *active-dynamic* group explored a full-size virtual room filled with life-size, static avatars (statues). They wore a head-mounted display which rendered a 3D view of the room and were required to physically walk around the museum and to naturally move their heads in order to view the faces of the avatars. Each participant in the *passive-dynamic* group was seated and used the same head-mounted display to watch a reenactment of the visual exploration of one participant from the active-dynamic group. The *passive-static* group viewed static views of the faces to learn while sitting in front of a computer screen. The recognition performance of all groups was evaluated afterwards with an identical same/different recognition test. Half of the learned faces were shown at test in an orientation close to that experienced during learning while the others were viewed from a new viewing angle.

1.6. Goals, mechanisms and predictions

There were three main goals to the current work. First, we wanted to establish whether the dynamic *visual experience* associated with walking up to and exploring a life-sized, full-body character leads to a performance advantage compared to more traditional mug-shot learning involving isolated heads. To do this, we contrasted old/new recognition performance in groups of participants who had dynamic exposure to the avatars (*active-dynamic* group; *passive-dynamic group*), to those who only saw static mug-shots of the faces to learn (*passive-static* group). This comparison thus relates both to the facial motion and the body context aspects of identity processing reviewed above.

It is important to note that as our avatars were realized as statues, "motion" in the current study refers to the smooth transition of views that occurs as participants approach and explore the avatars. Such dynamic exposure thus combines the type of motion explored in previous person recognition studies, where video or animation is used to present the situation where a moving actor approaches a static observer (e.g., O'Toole et al., 2011; Pilz et al., 2011; Simhi & Yovel, 2016) with more traditional rigid facial motion studies, which typically involve head tilt and rotation (e.g., Hill & Johnston, 2001; Pike et al., 1997; Xiao et al., 2012).

It is believed that such dynamic exposure can give rise to richer representations of identity through a variety of mechanisms, such as structure-from-motion (O'Toole et al., 2002), the involvement of dynamic mental representations (Freyd, 1987; Pilz et al., 2011; Thornton & Kourtzi, 2002) and/or shifts in the flexibility of face processing (Xiao et al., 2014). The operation of such mechanisms together with the presence of a body context, leads to the prediction that recognition performance should be better in the active-dynamic and passive-dynamic groups compared to the passive-static group.

Our second goal was to determine whether any performance advantages related to learning in our virtual reality (VR) environment came for the active (i.e. self-generated) aspects of movement and cognition or were simply a consequence of having a richer visual input. To do this, we contrasted old/new recognition performance in those observers who physically moved through to the space (*active-dynamic* group) with the group who simply watched playback of someone else moving through the space (*passive-dynamic* group).

In the literature on active processing, it is typically assumed that the additional demands/process involved in volitional control somehow enhance the associated visual representations (Chrastil & Warren, 2012). More specifically, given the results of Liu and colleagues (Liu et al., 2007), we can clearly predict that *active-dynamic* participants would perform more accurately in the later memory test than those participants who only watch passively a playback (*passive-dynamic* group). Nevertheless, the somewhat mixed findings alluded to above in terms of active/passive advantages make the outcome of the current work a little less certain. Clearly, our immersive VR environment more closely approximates natural viewing conditions, so our findings could have important implications for the generalizability of effects due to

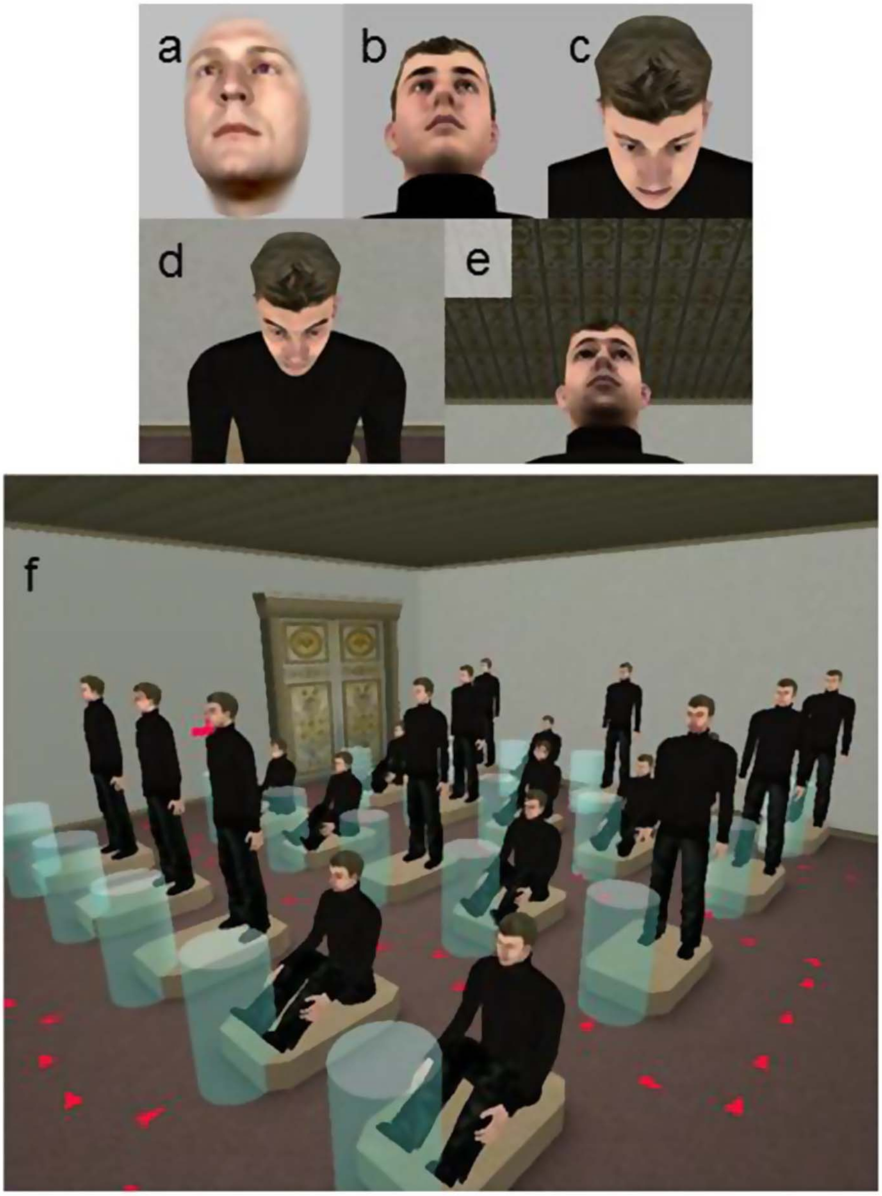

Fig. 1. a. Ambiguous view of a face: The person displayed might be looking up or being viewed from below. b, c, d, e. The presence of a body disambiguates face orientation. b. and c. Upper bodies of avatars at test, the museum background has been replaced by a homogeneous background. d and e. Partial views of the museum with sitting and standing avatars viewed from above and from below. f General view of the museum room. Viewing cylinders are placed in front of each avatar. The red arrows visible in the color version of the figure served as walking guides for the active observers. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

active face exploration.

Third and finally, we wanted to contribute to the literature on viewpoint effects with faces. In particular, we wanted to know whether learning faces in the current conditions benefits face recognition across viewpoints. To date, most studies have used isolated faces to test viewpoint dependency. Here we used whole heads attached to standing (faces seen from below) or sitting bodies (faces seen from above). To study viewpoint effects, we ensured that half of all studied faces in each learning condition were seen from above, and half from below, and we manipulated the congruency of these views at test. If more robust representations of the faces were gained during any one of our learning conditions, we expect recognition to be less affected by viewpoint variations at test.

2. Methods

2.1. Participants

Seventy two volunteers from our subject-database as well as campus employees aged between 18 and 38 (37 women) participated in this study. All were naïve as to the purpose of the experiment. They were assigned to three groups of 24 participants with differing learning procedures. All participants were given written and verbal instructions before the start of the experiment. Participants were paid 8 euro per hour for their time. The procedures were approved by the Ethical Review Board of the Max Planck Society. The work was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) and signed consent was obtained from each participant before the experiment.

2.2. Virtual environment

A virtual *museum room* environment was set up to display avatars with test faces using the software package Virtools 4.1 (Dassault Systèmes) on a Dell Inspiron M6400 laptop with an Intel Core2 Duo 2.53 GHz processor. The graphics were rendered on an NVidia Quadro FX3700M with 1024 MB of RAM. The virtual room (10 m length, 10 m width, 4 m height) included 20, 20 cm high pedestals on which ten “statues” (immobile avatars) were standing and ten were sitting (Fig. 1d and e show details and Fig. 1f shows a general view of the museum room). The avatars were placed in four rows of 5 statues, regularly spaced. The placement of the avatars on the pedestals was randomized for each participant to avoid order bias and the body position (sitting or standing) was balanced across face stimuli.

The statues in the virtual room were made out of four meshes: the body (which included the head and the hair), a sweater, ears, and a face. Forty male 3D face models with low distinctiveness ratings (Schieche, Bülthoff & Wallraven, unpublished data) were chosen from the face database of the Max Planck Institute for Biological Cybernetics (http://faces.kyb.tuebingen.mpg.de, Blanz & Vetter, 1999; Troje & Bülthoff, 1996). The resolution of the faces was chosen based on the detail level that could be displayed on the HMD in the virtual environment (around 17000 vertices). The ears were one set of ears chosen from one face of the same face database. Each face was added to the same body for all statues (a male avatar named *casual03* from Rocketbox Studios GmbH: Complete Characters Library) with a sweater mesh (from the male avatar, *casual1)* used to hide any suture artifacts between face and body. All statues had the same ears and hairline. Ten statues were in a sitting position and the others were in a standing position. Each statue was placed on a 20 cm high pedestal. The avatar body was 200 cm high when standing; body height of a sitting avatar was 100 cm. With the pedestal, the maximal height was 220 cm for the standing avatars and 120 cm for the sitting avatars. Two screenshots of each avatar's face were taken (as seen when the avatar was sitting and standing). These screenshots were used for the learning session of the passive-static group and for all subsequent test phases. Diffuse illumination was used for the scene and the stimuli.

Depending on the learning condition, participants either experienced this room in virtual reality as active observers, or they saw a rendering of the visual exploration of the active group, or they saw images extracted from this virtual environment. More details about the learning conditions for each group are given below.

2.3. Learning conditions

2.3.1. Active-dynamic observers in a virtual environment

During the learning phase, participant in this group viewed faces under conditions quite close to a natural experience, that is: (1) they viewed faces attached to bodies, (2) they had to bend their neck to look down at sitting avatars or to look up at the standing ones, and (3) they needed to walk around in the virtual museum to see all faces and by moving around they could see all faces from various viewing angles and distances.

The learning session was carried out in a fully tracked free-walking space, 11.13 m width × 12.60 m length × 8.4 m height that allowed participants to walk freely without physical limitations within the displayed virtual environment. Participants' motion in real space was tracked via a motion capture system with 16 Vicon MX 13 cameras and Vicon IQ 2.5 software. Each participant viewed the virtual room with the statues through an nVisor SX head-mounted display (HMD) which displayed a stereoscopic image of the virtual world with a resolution of 1280 × 1024 pixels, a frame rate of 60 Hz for each eye, and diagonal field of view of 60°. The end-to-end latency was less than 41 ms for the active-dynamic group walking through this environment (Di Luca, 2010). Participants were allowed to navigate freely although guidance was given by a path marked on the virtual floor that guided them from one avatar to the next to make sure they saw every avatar. A blueish transparent observation cylinder was visible in front of each avatar (see Fig. 1f). Participants were required to enter each observation cylinder and look at the corresponding avatar's face for at least 5 s. The position of the observation cylinders meant that participants were standing between 50 and 60 cm away from the target avatars. This ensured a common basic viewing experience for all participants. Each participant had the same starting point in the room and followed the same path along the pedestals. Participants were not allowed to bend down or to jump and were given the same virtual viewing height of 170 cm in the virtual environment. This viewing height was equidistant between the standing avatar face and the sitting avatar face. Their virtual height allowed them to have exactly the same basic visual experience at viewing the faces of sitting and standing avatars in terms of vertical viewing angles (pitch). In the viewing cylinders, vertical viewing angle upon the faces reached a maximum of 45°; this viewing angle was reduced at other more distant locations. Each participant walked this virtual environment three times. Each walk lasted approximately 10–13 min. There was a short time between runs whilst the program was loading, of around 2–3 min.

2.3.1.1. Pre-training in the virtual environment. Participants were given a short training phase to get them used to wearing the equipment as they had to wear the laptop providing the visuals to the head-mounted display in a backpack. They learned to navigate a virtual museum populated by 10 practice avatars not used in the experiment, and to remain for 5 s inside each viewing cylinder. This pre-training lasted approximately 5 min.

2.3.1.2. Instructions. Active-dynamic participants (1) were told that they would view statues in a museum, all with different faces but with the same height, build and black uniform. They were asked to learn their faces, as they would be tested on them afterwards in a desktop experiment. They knew that they would have three runs through the museum with the avatars all in the same position and order for each of the three runs. (2) They were informed that they could not bend down or stand up in order to view the face of standing or sitting avatars better. They knew that they could take as long as they needed to learn the face. They were informed that the scene would be blacked-out if they attempted to come too close to the avatar (closer than 50 cm) and that they had to enter each observation cylinder for at least 5 s, a warning sign would admonish them not to leave the cylinder for stays under 5 s.

2.3.2. Passive-dynamic participants viewing a video of the virtual museum

Sitting participants experienced the rendition of the visual experience of the active-dynamic participants during their learning phase on the same head-mounted display as for the active-dynamic group. Each participant was paired to one participant of the active-dynamic group and saw a reenactment of his/her three runs through the museum.

2.3.2.1. Instructions

The instructions replicated those given under point (1) for the active-dynamic group. Point (2) of the instructions was not needed

2.3.3. Passive-static participants viewing static images extracted from the virtual museum

Sitting participants saw static representations of the faces to learn on a computer screen. These face stimuli were screenshots of the avatar's faces and upper bodies extracted from the virtual museum. These screenshots were close to the viewing experience of the participants of the other group in the viewing cylinders. The passive-static participants saw two screenshots displaying the face to learn slightly turned away from a frontal view to the left and the right by 5°. These screenshots were shown side by side on the screen, for 5 s followed by a blank screen for a minimum of 500 ms between each face set. Participants were required to press a button to display the next face set. Each image covered an angular size of approximately 15° × 15°, a size that approximates the size of the faces viewed by the other two groups from the observation cylinder. Each block of 20 face sets was repeated 3 times to match the learning experience of the other two groups.

2.3.3.1 Instructions. Written instruction replicated the instructions given under point (1) for the active-dynamic group. Point (2) of the instructions was not needed.

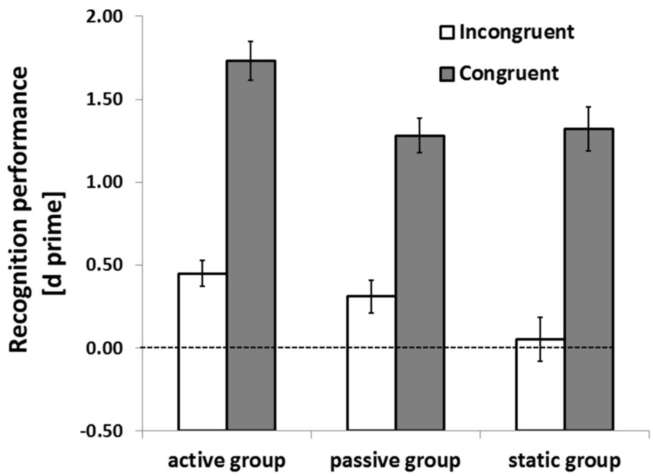

Fig. 2. Recognition performance in terms of d′ for the three learning conditions. Solid gray bars show performance when test views were congruent with learning views. Open white bars show performance when test views were incongruent with learning views. Error bars represent one standard error of the mean.

2.4. Test phase

The test phase was identical for all groups. As test images, we created screenshots of the top parts of all avatars as seen in both their sitting and standing positions from the observation cylinder. The participants were standing around 50–60 cm away from the avatar at that location in the virtual room. The test images were identical for all observers, because they had all been given the same virtual height in the scene (see description in 2.3.1). The vertical viewing angle was thus approximately 45° (up or down) for each test image. A view of 0° would happen when avatars' and participants' eyes are at the same height. These test images could be congruent, that is they showed a face from a viewpoint compatible with the viewing experience of the observer during the learning phase, incongruent, the learned faces were pictured from a viewpoint not experienced by the observers (i. e. the view on the avatar seen in its other (sitting or standing) position) or new, the faces had not been seen during learning. All test images were edited to remove the museum background that was replaced by a uniform grey background and were shown in full color on a computer monitor. They covered a visual angle of approximately 18° × 18°. They showed the faces and upper body so that the orientation of the head was not ambiguous, but the position of the body (sitting or standing) was not indicated by the visible body in the pictures (see examples in Fig. 1b and c.

Each face was shown until a response was entered and it was immediately followed by the next face. There was no feedback. The screenshots of the target faces (old) were intermixed with a similar number of screenshots of distracter faces (new). Of the 20 learned target faces, 10 were pictured in screenshots showing them in the same learnt orientation (congruent) and 10 were in an orientation not experienced during learning (incongruent). The distracter faces were also shown in a similar fashion; half of them were seen from below and the others from above. Images were presented sequentially in random order and participants were to decide whether they had seen the face in the learning phase (old) or if it was a new distracter face (new). They were asked to complete the test as quickly and as accurately as possible using a response box with two buttons labeled "old" and "new". They performed four practice trials before the test proper.

2.5. Design

The 40 faces were separated into two groups of 20 – counterbalanced across participants – that served as targets and distractors. There were 8 different configurations of the experiment so that each of the 40 faces was seen as a target in all four conditions at test across participants: sitting congruent, standing congruent, sitting incongruent, standing incongruent. A given face was only seen by a participant in one condition. In each group, three participants were assigned to each of the 8 configurations.

2.6. Procedure

The experiment consisted of a learning and a testing phase. Although there were three different learning conditions, the test phase was the same for all participants. In the test phase, participants of all groups performed a desktop picture recognition task with 40 trials shown in one block. Accuracy and response times were recorded. The whole experiment lasted between 1 and 1.5 h.

3. Results

We calculated recognition performance in terms of response sensitivity (d′) which allowed for comparisons between all groups. Recognition performance was assessed separately for each learning group and test congruency condition. One-sample two-tailed t-tests assured that in all conditions but one (passive-static in incongruent trials, $p = .694$, $d = 0.081$), participants were better than chance (all other $ps < .001$, all $ds > 0.634$). Fig. 2 displays response sensitivity for each test condition. A visual inspection of the responses reveals that congruent target faces were always better recognized than incongruent ones.

The d′ data were submitted to a 3 (learning group: active-dynamic, passive-dynamic, passive-static) × 2 (test congruency: congruent test views, incongruent test views) mixed-design repeated-measures ANOVA, with congruency as a within-participant factor and learning group as between-participant factor. As suggested by Fig. 2, we found a significant main effect of test congruency ($F(1, 69) = 199.17$, $p < .001$, $\eta_p^2 = 0.743$) with better performance for congruent than incongruent test faces (1.45 vs. 0.27, Bonferroni-corrected paired t-tests for each learning condition were significant, all $ps < .001$, all $ds > 1.423$). There was also a significant main effect of learning condition ($F(2, 69) = 5.75$, $p = .005$, $\eta_p^2 = 0.143$) with the active-dynamic group performing best (1.09), and the passive-dynamic group performing better (0.80) than the passive-static group (0.69). Bonferroni corrected post-hoc pairwise comparisons showed that the active-dynamic group performed significantly better than the passive-static group ($p = .005$) while the difference was close to significant for the comparison with the passive-dynamic group ($p = .058$). The two passive groups did not differ from each other ($p = 1.0$). The interaction between factors (congruency x learning group) was not significant ($F(2, 69) = 1.53$, $p = .225$, $\eta_p^2 = 0.042$).

In addition to response sensitivity, we also calculated *response criterion* c to assess whether participants changed their decision process depending on learning and testing conditions. Fig. 3 shows that the averaged criteria c calculated across all participants for each condition separately are all more liberal in congruent than in incongruent trials. Correspondingly, a 3 (learning group: active-dynamic, passive-dynamic, passive-static) × 2 (test congruency: congruent test views, incongruent test views) mixed-design repeated-measures ANOVA on the criterion data demonstrated a main effect of congruency. Participants were more biased to respond New in the old/new task when the viewpoint changed (in incongruent trials) than when it did not (congruent trials), $F(1,69) = 197.62$, $p < .001$, $\eta_p^2 = 0.741$). There was no main effect of learning condition and no interaction (both $Fs \leq 1.53$, $ps \geq .223$, $\eta_p^2 \leq 0.043$), indicating that participants' decision process did not vary across learning conditions. The most obvious interpretation of the congruency main effect is that when participants were uncertain, they adopted a more conservative criteria and were less likely to confirm having seen an identity before. Analysis of RT data showed qualitatively the same pattern as the d′ data. In particular, there was no

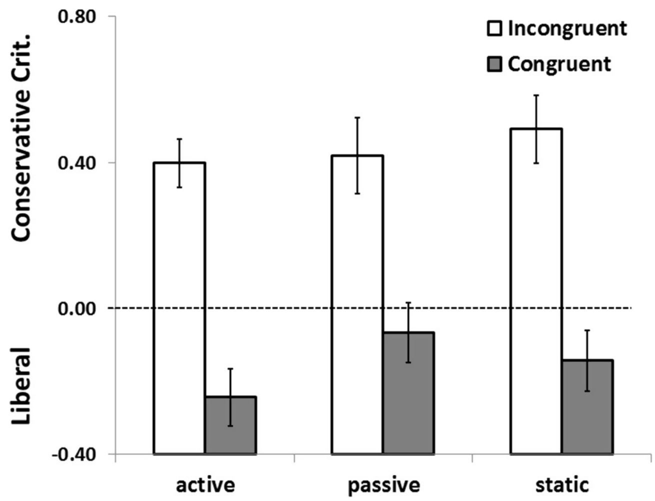

Fig. 3. Criterion averaged across all participants for each condition. Solid gray bars show criterion when test views were congruent with learning views. Open white bars show criterion when test views were incongruent with learning views. Error bars depict one standard error of the mean.

Table 1
Reaction times for correct responses averaged across all participants for each condition.

Test condition	Mean	Std. Error
Active Congruent	2272 ms	168 ms
Active Incongruent	3610 ms	294 ms
Passive Congruent	2326 ms	245 ms
Passive Incongruent	2815 ms	329 ms
Static Congruent	1893 ms	148 ms
Static Incongruent	2564 ms	301 ms

evidence of speed-accuracy trade-offs that might account for group differences. For the sake of brevity, we have not included details of the analysis, but the data are summarized in Table 1.

4. Discussion

In the current paper, we used life-sized avatars to investigate the impact of three different learning conditions on person recognition. A first group of participants (active-dynamic) walked through a virtual museum to inspect and learn the faces of statues. A second group (dynamic-passive) watched the visual input received by the first group during its exploration of the museum. Finally, the third group saw two different static snapshots of each face to learn (static-passive). All groups were tested for their memory of the faces encountered during their learning phase with an old/new recognition task where we also manipulated viewpoint.

There were three main findings that we believe make a useful contribution to the person recognition literature. First, our results show an active-dynamic advantage, such that recognition performance was best when participants directly interacted with target identities. Second, across all conditions we find very robust viewpoint costs when the pitch (up/down orientation) of the target faces was changed between study and test. Not only is this view manipulation far less studied than typical yaw (side-to-side) rotations, but here we show that memory costs may naturally occur when we interact with seated (looking down) or standing (looking up) individuals. Third, we have shown how a virtual museum scenario can be used to study person recognition. As noted in the Introduction, there has been a move away from using static snapshots of faces towards a more integrated study of dynamic faces and bodies, mostly using video clips. Here we have shown proof-in-principle of how VR technology – which is becoming more available in terms of technology and costs – can be used to further increase the ecological relevance of lab-based studies. In the remainder of this discussion we explore each of these contributions in turn, before considering other aspects of our data and possible future directions.

Our first finding adds to the somewhat mixed literature on active exploration during visual learning. Although there appears to be a general assumption that active exploration provides richer experience that can enhance memory representations, such advantages have not always been found (e.g., Christou & Bülthoff, 1999; Teramoto & Riecke, 2010). Our finding thus provides another example where active exploration appears to have a positive outcome. Of course, we need to add a small caveat to this claim, as our post-hoc comparison between active-dynamic and passive-dynamic was only marginally significant. Nevertheless, the overall pattern of data shown in Fig. 2 is certainly consistent with the claim that active-dynamic participants generally perform better. Moreover, had we analyzed congruent and incongruent trials separately, then the post-hoc difference between these two conditions would have been significant in the former, but not the latter case. It is somewhat surprising that the clearest difference between active-dynamic and passive-dynamic is seen in the congruent condition. One of our expectations was that active experience might help most under conditions of uncertainty, but this does not seem to be the case. We return to this issue shortly.

More generally, it is interesting that the additional task demands placed on our active-dynamic participants – they were immersed in an unusual virtual environment, had to physically navigate around the space, judge their approach to the statues, all the while carrying a backpack and wearing a head-mounted display – did not distract them from the task of learning the faces. Clearly, our other two groups were more able to focus resources directly on the face learning, so in terms of resource allocation, they might have been expected to do somewhat better. In fact, Liu et al. (2007) were only able to find an active advantage when passive observers were given an additional task to equate overall levels of difficulty. As we did not do this in the current scenario, this may go some way to explain why the difference between active-dynamic and passive-dynamic observers was not more robust.

Our second finding relates to the question of viewpoint dependency during perception. Wallraven et al. (2002) had shown convincingly that methods previously applied to novel objects, also yielded viewpoint costs in the context of faces. Here, we were interested in exploring whether these costs would be reduced or even eliminated when additional context and volitional control accompanied learning. Specifically, our active-dynamic group were not only exposed to life-sized faces on bodies, rather than photographs or isolated heads – but had additional sensori-motor signals while they walked toward and around the avatars and while they moved their heads to view the statues. Despite this additional input and the volitional aspect of their learning that might have helped encoding, test images showing the learned faces from the viewpoint most familiar to the participants were much better and faster recognized than when these faces were shown from unfamiliar views. Hence face representation gained during a naturalistic learning experience is not robust against viewpoint changes.

One could argue that the height difference (50 cm) implemented between observer and standing or sitting statues is much larger than what is experienced between adults generally in real life and might have impeded generalization over unfamiliar views. To test this possibility, we repeated the active-dynamic learning condition with a new group of 24 participants after modifying the virtual museum so as to have a lower height difference between observers and avatars. We used a height difference of 30 cm rather than 50 cm, which is closer to what might be typically experienced in everyday life (6.3% of German men are 190 cm tall or taller, 44% of all German women are 164 cm or smaller; SOEP 2006, Körpergröβe der Deutschen. Statistik des Sozio-oekonomischen Panels (SOEP), presented by statista.org). In this control experiment, the 24 participants were still better in congruent than incongruent trials even though the viewing angles under which the faces were seen during learning and testing are commonly experienced in real life. Thus, view generalization still did not emerge under those more natural viewing conditions.

The third novel aspect of this study was our attempt to design a learning situation as close as possible to the same task in real life for the active-dynamic group. That is we tried to simulate a visit to a real museum containing life-sized statues. We believe that as VR technology becomes more available and affordable, this is a good time to encourage researchers to move away from traditional photographic, isolated face studies to explore the potential of more complex and immersive environments. Similar efforts to increase the ecological validity and realism of stimuli using virtual reality and virtual humans have been made by many other research groups (Fox, Arena, & Bailenson, 2009; Slater & Sanchez-Vives, 2016).

Our participants were required to follow a path and enter each viewing cylinder. These requirements assured that they all had a similar basic viewing experience. This could be understood as a reduced freedom in looking at the faces to learn as one wished. We argue that this is not quite true as the videos of their runs (they were used for the passive-dynamic group) confirmed that they looked at the faces not only from within the cylinders but also from other locations while they accomplished their three runs in the museum. Our study is the first to use such a setup for investigating face recognition under more ecological conditions. Following this trend, similar paradigms will allow a large variety of new studies, for example one could investigate viewing behavior and concomitant face recognition performance when there are less viewing constraints or when some modifications like changes in illumination conditions occur.

Our current results show that despite potential additional costs resulting from the demands of the active learning condition, it is possible to study person recognition in VR. Indeed the performance of our active participants suggests that such learning environments can give rise to more robust face representations, representations that support improved recognition compared to traditional photographic, and even possibly, video-based conditions. Clearly, our current active environment could be greatly improved. For example, including avatars with dynamic faces – to explore how characteristic expressions are interpreted in VR – and/or using diagnostic body information – to more fully explore realistic encounters with new people. Nevertheless, we believe this study takes an important first step in showing the potential of VR in the context of face and person perception.

Before concluding, one other aspect of our data should be mentioned. This concerns the lack of a general performance difference between the passive-dynamic and passive-static groups. As mentioned in the Introduction, there is now considerable evidence that dynamic exposure to faces can facilitate performance using a variety of tasks and types of stimuli (Lander & Butcher, 2015; O'Toole et al., 2002; Xiao et al., 2014). It is important to ask why this form of dynamic advantage was not seen in the current study. Two likely causes suggest themselves. First, as noted in several places in this paper the statues in our museum did not actually move. That is the dynamic exposure we explored was generated by the movement of the observer, rather than movement of the actor. Although such observer-generated dynamic input should still provide richer spatio-temporal input than simple static views, it may differ in quantity and/or quality from previously studied types of motion. For example, we did not enforce detailed side-to-side viewing of the target heads, although this was encouraged. In approaching the statues coherent "looming" events would have occurred, similar to those used in other studies. Possibly the duration of these looming events, or even the relative short range over which they occurred may not have been enough alone (as opposed to in tandem with active participation) to afford our passive-dynamic participants with a learning advantage.

More generally, all of the observer movement in our study gave rise to rigid changes in viewpoint rather than non-rigid characteristic motion, such as those involved in smiling or talking. In the literature on dynamic faces rigid head movement advantages have previously proven difficult to find. A motion advantage seems to appear mostly in specific conditions, for example when participants are very young (Otsuka et al., 2009) or older (Maguinness & Newell, 2014), but does not necessarily appear for young adults (e.g., Christie & Bruce, 1998). It is thus possible that the face representation improvement due to rigid motion is subtle and our current viewing conditions with young adult participants simply not did bring it to light.

To end on a more positive note, although there were no general performance differences between passive-dynamic and passive-static groups, there is a hint of a difference if we return to the view dependence question. We had expected better generalization abilities for both dynamic groups over the passive-static group, because of the enhanced possibility given by structure-from-motion to those participants to create a strong representation of the shape of the faces. In line with this expectation, we found that performance in the incongruent condition was at chance level for the static group but remained above chance for both dynamic groups. In this sense, then, we can say that in terms of generalisation performance, the passive-dynamic group did outperform the passive-static group, giving rise to at least one form of dynamic advantage.

5. Conclusion

To our knowledge, this is the first study to show that the visual input generated by active participants exploring life-sized avatars can provide a performance advantage during face recognition. Additionally, our data show that even when learning faces under naturalistic conditions there is still strong viewpoint dependency for a class of objects that is extremely homogeneous and for which we have considerable expertise. Finally, our use of immersive VR with a standard old/new recognition paradigm provides a useful example of how experimental control can be retained while using technology to improve the ecological validity of lab-based studies.

Author contributions

I.B. and I.M.T. conceived the experiments, I.B. developed the experimental design and B.J.M. programmed the experiments in virtual reality, I.B. and I.M.T. analyzed the data; I.B. and B.J.M. contributed reagents/materials/analysis tools; I.B. and I.M.T. wrote the paper; B.J.M. discussed the results and commented on the manuscript.

Competing financial interests

The authors declare no competing financial interests.

Acknowledgments

We wish to thank the internship students Sarah Shrimpton and Laura McDowall and our technical assistant Karin Bierig for their help preparing the stimuli and running the experiments. We are grateful for the help of Paolina Teneva, Ivelina Piryankova and Markus Leyrer for installing and running the virtual setup and obtaining the learning sequences for the passive-dynamic observers. This research was supported by the Max Planck Society.

References

Albonico, A., Malaspina, M., & Daini, R. (2012). Congenital Prosopagnosia: The role of changeable and invariant aspects in famous face identification. In ECVP 2012.

Appleyard, D. (1970). Styles and methods of structuring a city. *Environment and Behavior, 2*, 100–118.

Bennetts, R. J., Butcher, N., Lander, K., Udale, R., & Bate, S. (2015). Movement cues aid face recognition in developmental prosopagnosia. *Neuropsychology, 29*(6), 855–860.

Bennetts, R. J., Kim, J., Burke, D., Brooks, K. R., Lucey, S., Saragih, J., & Robbins, R. A. (2013). The movement advantage in famous and unfamiliar faces: A comparison of point-light displays and shape-normalised avatar stimuli. *Perception, 42*(9), 950–970. http://dx.doi.org/10.1068/p7446.

Blanz, V., & Vetter, T. (1999). A morphable model for the synthesis of 3D faces. In: Proceedings of the 26th annual conference on Computer graphics and interactive techniques – SIGGRAPH '99 (pp. 187–194). New York, New York, USA: ACM Press. doi: 10.1145/311535.311556.

Braje, W. L., Kersten, D., Tarr, M. J., & Troje, N. F. (1998). Illumination effects in face recognition. *Psychobiology, 26*(4), 371–380.

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. http://dx.doi.org/10.1037//1076-898X.5.4.339.

Bruce, V., & Valentine, T. (1988). When a nod's as good as a wink: The role of dynamic information in facial recognition. In P. E. Morris, M. M. Gruneberg, & R. N. Sykes (Eds.). *Practical aspects of memory: Current research and issues. Vol. 1. Memory in everyday life*(pp. 164–174). Oxford, England: John Wiley & Sons Inc.

Bulf, H., & Turati, C. (2010). The role of rigid motion in newborns' face recognition. *Visual Cognition, 18*(4), 504–512. http://dx.doi.org/10.1080/13506280903272037.

Bülthoff, H. H., & Edelman, S. Y. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences of the United States of America, 89*(1), 60–64. Available from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=48175&tool=pmcentrez&rendertype=abstract.

Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science, 10*(3), 243–248. http://dx.doi.org/10.1111/1467-9280.00144.

Butcher, N., Lander, K., Fang, H., & Costen, N. (2011). The effect of motion at encoding and retrieval for same- and other-race face recognition, 931–942. doi: 10.1111/j.2044-8295.2011.02060.x.

Chrastil, E. R., & Warren, W. H. (2012). Active and passive contributions to spatial learning. *Psychonomic Bulletin and Review, 19*(1), 1–23. http://dx.doi.org/10.3758/s13423-011-0182-x.

Christie, F., & Bruce, V. (1998). The role of dynamic information in the recognition of unfamiliar faces. *Memory and Cognition, 26*(4), 780–790. http://dx.doi.org/10.3758/BF03211397.

Christou, C. G., & Bülthoff, H. H. (1999). View dependence in scene recognition after active learning. *Memory and Cognition, 27*(6), 996–1007. Available from http://www.ncbi.nlm.nih.gov/pubmed/10586576.

Cox, D. D., Meyers, E., & Sinha, P. (2004). Contextually evoked object-specific responses in human visual cortex. *Science (New York, N.Y.), 304*(5667), 115–117. http://dx.doi.org/10.1126/science.1093110.

de Gelder, B. (2009). Why bodies? Twelve reasons for including bodily expressions in affective neuroscience. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1535), 3475–3484. http://dx.doi.org/10.1098/rstb.2009.0190.

Di Luca, M. (2010). New method to measure end-to-end delay of virtual reality. *Presence, 19*(6), 569–584.

Favelle, S. K., Palmisano, S., & Avery, G. (2011). Face viewpoint effects about three axes: The role of configural and featural processing. *Perception, 40*(7), 761–784. http://dx.doi.org/10.1068/p6878.

Favelle, S. K., Palmisano, S., & Maloney, R. T. (2007). Things are looking up: Differential decline in face recognition following pitch and yaw rotation. *Perception, 36*(9), 1334–1352. http://dx.doi.org/10.1068/p5637.

Foster, D. H., & Gilson, S. J. (2002). Recognizing novel three-dimensional objects by summing signals from parts and views. In: *Proceedings of the Royal Society of London. Series B, Biological Sciences*, (pp. 1939–1947). 269(1503), doi: 10.1098/rspb.2002.2119.

Fox, J., Arena, D., & Bailenson, J. N. (2009). Virtual reality: A survival guide for the social scientist. *Journal of Media Psychology, 21*(3), 95–113. http://dx.doi.org/10.1027/1864-1105.21.3.95.

Freyd, J. (1987). Dynamic mental representations. *Psychological Review, 94*(4), 427–438.

Ghuman, A. S., McDaniel, J. R., & Martin, A. (2010). Face adaptation without a face. *Current Biology, 20*(1), 32–36. http://dx.doi.org/10.1016/j.cub.2009.10.077.

Girges, C., Spencer, J., & O'Brien, J. (2015). Categorizing identity from facial motion. *Quarterly Journal of Experimental Psychology (2006),* 1–12. http://dx.doi.org/10.1080/17470218.2014.993664.

Hahn, C. A., O'Toole, A. J., & Phillips, P. J. (2015). Dissecting the time course of person recognition in natural viewing environments. *British Journal of Psychology (London, England: 1953),* 1–18. http://dx.doi.org/10.1111/bjop.12125.

Henderson, Z., Bruce, V., & Burton, A. M. (2001). Matching the faces of robbers captured on video. *Applied Cognitive Psychology, 15*(4), 445–464. http://dx.doi.org/10.1002/acp.718.

Hill, H., & Johnston, A. (2001). Categorizing sex and identity from the biological motion of faces. *Current Biology, 11*(11), 880–885. http://dx.doi.org/10.1016/S0960-9822(01)00243-3.

James, K. H., Humphreys, G. W., Vilis, T., Corrie, B., Baddour, R., & Goodale, M. A. (2002). "Active" and "passive" learning of three-dimensional object structure within an immersive virtual reality environment. *Behavior Research Methods, Instruments, and Computers, 34*(3), 383–390. http://dx.doi.org/10.3758/BF03195466.

Kiiski, H., Hoyet, L., Woods, A. T., O'Sullivan, C., & Newell, F. N. (2015). Strutting hero, sneaking villain: Utilizing body motion cues to predict the intentions of others. *ACM Transactions on Applied Perception (TAP), 13*(1), 1. http://dx.doi.org/10.1145/2791293.

Knappmeyer, B., Thornton, I. M., & Bülthoff, H. H. (2003). The use of facial motion and facial form during the processing of identity. *Vision Research, 43*(18), 1921–1936. http://dx.doi.org/10.1016/S0042-6989(03)00236-0.

Knight, B., & Johnston, A. (1997). The role of movement in face recognition. *Visual Cognition, 4*(3), 265–273. http://dx.doi.org/10.1080/713756764.

Kozlowski, L. T., & Cutting, J. E. (1977). Recognizing the sex of a walker from a dynamic point-light display. *Perception Psychophysics, 21*(6), 575–580. http://dx.doi.org/10.3758/BF03198740.

Lander, K., & Bruce, V. (2000). Recognizing famous faces: Exploring the benefits of facial motion. *Ecological Psychology, 12*(4), 259–272. http://dx.doi.org/10.1207/S15326969ECO1204.

Lander, K., & Butcher, N. (2015). Independence of face identity and expression processing: Exploring the role of motion. *Frontiers in Psychology, 6*(March), 1–6. http://dx.doi.org/10.3389/fpsyg.2015.00255.

Lander, K., & Chuang, L. (2005). Why are moving faces easier to recognize? *Visual Cognition, 12*(3), 429–442. http://dx.doi.org/10.1080/13506280444000382.

Lander, K., Chuang, L., & Wickham, L. H. V. (2006). Recognizing face identity from natural and morphed smiles. *Quarterly Journal of Experimental Psychology, 59*(5), 801–808. http://dx.doi.org/10.1080/17470210600576136.

Liu, C. H. (2002). Reassessing the 3/4 view effect in face recognition. *Cognition, 83*(1), 31–48. http://dx.doi.org/10.1016/S0010-0277(01)00164-0.

Liu, C. H., Ward, J., & Markall, H. (2007). The role of active exploration of 3D face stimuli on recognition memory of facial information. *Journal of Experimental Psychology: Human Perception and Performance, 33*(4), 895–904. http://dx.doi.org/10.1037/0096-1523.33.4.895.

Logothetis, N. K., & Pauls, J. (1995). Psychophysical and physiological evidence for viewer-centered object representations in the primate. *Cerebral Cortex (New York, N.Y. : 1991), 5*(3), 270–288. Available from http://www.ncbi.nlm.nih.gov/pubmed/7613082.

Longmore, C. A., & Tree, J. J. (2013). Motion as a cue to face recognition: Evidence from congenital prosopagnosia. *Neuropsychologia, 51*(5), 864–875. http://dx.doi.org/10.1016/j.neuropsychologia.2013.01.022.

Maguinness, C., & Newell, F. N. (2014). Motion facilitates face perception across changes in viewpoint and expression in older adults. *Journal of Experimental Psychology: Human Perception and Performance, 40*(6), 2266–2280. http://dx.doi.org/10.1037/a0038124.

Meijer, F., & Van der Lubbe, R. H. J. (2011). Active exploration improves perceptual sensitivity for virtual 3D objects in visual recognition tasks. *Vision Research, 51*(23–24), 2431–2439. http://dx.doi.org/10.1016/j.visres.2011.09.013.

Motes, M. A., Finlay, C. A., & Kozhevnikov, M. (2006). Scene recognition following locomotion around a scene. *Perception, 35*(11), 1507–1520. http://dx.doi.org/10.1068/p5459.

O'Toole, A. J., Harms, J., Snow, S. L., Hurst, D. R., Pappas, M. R., Ayyad, J. H., & Abdi, H. (2005). A video database of moving faces and people. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27*(5), 812–816. http://dx.doi.org/10.1109/TPAMI.2005.90.

O'Toole, A. J., Phillips, P. J., Weimer, S., Roark, D. A., Ayyad, J., Barwick, R., & Dunlop, J. (2011). Recognizing people from dynamic and static faces and bodies: Dissecting identity with a fusion approach. *Vision Research, 51*(1), 74–83. http://dx.doi.org/10.1016/j.visres.2010.09.035.

O'Toole, A. J., Roark, D. A., & Abdi, H. (2002). Recognizing moving faces: A psychological and neural synthesis. *Trends in Cognitive Sciences, 6*(6), 261–266. http://dx.doi.org/10.1016/S1364-6613(02)01908-3.

O'Toole, A. J., Roark, D. A., Jiang, F., & Abdi, H. (2005). Predicting human performance for face recognition. In W. Zhao, & R. Chellappa (Eds.). *Face processing: Advanced modeling and methods*(pp. 293–319). Academic Press.

Oram, M. W., & Perrett, D. I. (1992). Time course of neural responses discriminating different views of the face and head. *Journal of Neurophysiology, 68*(1), 70–84.

Otsuka, Y., Konishi, Y., Kanazawa, S., Yamaguchi, M. K., Abdi, H., & O'Toole, A. J. (2009). Recognition of moving and static faces by young infants. *Child Development, 80*(4), 1259–1271. http://dx.doi.org/10.1111/j.1467-8624.2009.01330.x.

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*(4), 409–438. http://dx.doi.org/10.1080/713756769.

Pilz, K. S., Thornton, I. M., & Bülthoff, H. H. (2006). A search advantage for faces learned in motion. *Experimental Brain Research, 171*(4), 436–447. http://dx.doi.org/10.1007/s00221-005-0283-8.

Pilz, K. S., Vuong, Q. C., Bülthoff, H. H., & Thornton, I. M. (2011). Walk this way: Approaching bodies can influence the processing of faces. *Cognition, 118*(1), 17–31. http://dx.doi.org/10.1016/j.cognition.2010.09.004.

Rice, A., Phillips, P. J., & O'Toole, A. J. (2013). Variable use of the face and body in person identification. *Journal of Vision, 13*(9), 977.

Roark, D. A., O'Toole, A. J., Abdi, H., & Barrett, S. E. (2006). Learning the moves: The effect of familiarity and facial motion on person recognition across large changes in viewing format. *Perception, 35*(6), 761–773. http://dx.doi.org/10.1068/p5503.

Robbins, R. A., & Coltheart, M. (2012). The effects of inversion and familiarity on face versus body cues to person recognition. *Journal of Experimental Psychology: Human Perception and Performance, 38*. http://dx.doi.org/10.1037/a0028584.

Simhi, N., & Yovel, G. (2016). The contribution of the body and motion to whole person recognition. *Vision Research, 122*, 12–20. http://dx.doi.org/10.1016/j.visres.2016.02.003.

Simons, D. J., Wang, R. F., & Roddenberry, D. (2002). Object recognition is mediated by extraretinal information. *Perception and Psychophysics, 64*(4), 521–530. http://dx.doi.org/10.3758/BF03194723.

Slater, M., & Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Frontiers in Robotics and AI, 3*, 1–47. http://dx.doi.org/10.3389/frobt.2016.00074.

Steede, L. L., Tree, J. J., & Hole, G. J. (2007). I can't recognize your face but I can recognize its movement. *Cognitive Neuropsychology, 24*(4), 451–466. http://dx.doi.org/10.1080/02643290701381879.

Tarr, M. J. (1995). Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic*

Bulletin and Review, 2(1), 55–82. http://dx.doi.org/10.3758/BF03214412.

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology, 21*(2), 233–282. http://dx.doi.org/10.1016/0010-0285(89)90009-1.

Teramoto, W., & Riecke, B. E. (2010). Dynamic visual information facilitates object recognition from novel viewpoints. *Journal of Vision, 10*(13), 11. http://dx.doi.org/10.1167/10.13.11 1–13.

Thornton, I. M., & Kourtzi, Z. (2002). A matching advantage for dynamic human faces. *Perception, 31*(1), 113–132. http://dx.doi.org/10.1068/p3300.

Troje, N. F., & Bülthoff, H. H. (1996). Face recognition under varying poses: The role of texture and shape. *Vision Research, 36*(12), 1761–1771. http://dx.doi.org/10.1016/0042-6989(95)00230-8.

Van der Linde, I., & Watson, T. L. (2010). A combinatorial study of pose effects in unfamiliar face recognition. *Vision Research, 50*(5), 522–533. http://dx.doi.org/10.1016/j.visres.2009.12.012.

Vanrie, J., Willems, B., & Wagemans, J. (2001). Multiple routes to object matching from different viewpoints: Mental rotation versus invariant features. *Perception, 30*(9), 1047–1056. http://dx.doi.org/10.1068/p3200.

Wallraven, C., Schwaninger, A., Schuhmacher, S., & Bülthoff, H. H. (2002). View-based recognition of faces in man and machine : Re-visiting inter-extra-ortho. *Lecture Notes in Computer Science, 2525*, 651–660.

Wang, R. F., & Simons, D. J. (1999). Active and passive scene recognition across views. *Cognition, 70*(2), 191–210. Available from http://www.ncbi.nlm.nih.gov/pubmed/10349763.

Xiao, N. G., Perrotta, S., Quinn, P. C., Wang, Z., Sun, Y. H. P., & Lee, K. (2014). On the facilitative effects of face motion on face recognition and its development. *Frontiers in Psychology, 5*, 1–16. http://dx.doi.org/10.3389/fpsyg.2014.00633.

Xiao, N. G., Quinn, P. C., Ge, L., & Lee, K. (2012). Rigid facial motion influences featural, but not holistic, face processing. *Vision Research, 57*, 26–34. http://dx.doi.org/10.1016/j.visres.2012.01.015.

Xiao, N. G., Quinn, P. C., Ge, L., & Lee, K. (2013). Elastic facial movement influences part-based but not holistic processing. *Journal of Experimental Psychology: Human Perception and Performance, 39*(5), 1457–1467. http://dx.doi.org/10.1037/a0031631.