

# Learning to recognize face shapes through serial exploration

Christian Wallraven · Lisa Whittingstall ·  
Heinrich H. Bühlhoff

Received: 8 October 2012 / Accepted: 18 February 2013 / Published online: 7 March 2013  
© Springer-Verlag Berlin Heidelberg 2013

**Abstract** Human observers are experts at visual face recognition due to specialized visual mechanisms for face processing that evolve with perceptual expertise. Such expertise has long been attributed to the use of configural processing, enabled by fast, parallel information encoding of the visual information in the face. Here we tested whether participants can learn to efficiently recognize faces that are *serially* encoded—that is, when only partial visual information about the face is available at any given time. For this, ten participants were trained in gaze-restricted face recognition in which face masks were viewed through a small aperture controlled by the participant. Tests comparing trained with untrained performance revealed (1) a marked improvement in terms of speed and accuracy, (2) a gradual development of configural processing strategies, and (3) participants' ability to rapidly learn and accurately recognize novel exemplars. This performance pattern demonstrates that participants were able to learn new strategies to compensate for the serial nature of information encoding. The results are discussed in terms of expertise acquisition and relevance for other sensory modalities relying on serial encoding.

**Keywords** Face recognition · Face processing strategies · Perceptual expertise · Learning · Serial encoding

## Introduction

Human observers are experts at visual face recognition. Consequently, face processing has received a lot of attention in vision research providing evidence for specialized visual mechanisms that evolve with perceptual expertise. Among the hallmarks of expertise face processing are as follows:

- The use of configural as opposed to featural processing (processing of individual face parts). Three types of such configural processing have been defined (Maurer et al. 2002, see also Gauthier and Tarr 2002): (1) sensitivity to first-order relations (individual object parts are better recognized in the context of other parts than in isolation), (2) holistic processing (ability or tendency to consider all parts of an object simultaneously, regardless of the exact configuration of parts), and (3) sensitivity to second-order relations (perceiving inter-feature distances; individual object parts are placed in the context of other individual parts).
- Face perception is *orientation specific*, that is, faces are processed more accurately when they are presented in the normal upright position than when they are inverted (for reviews, see Searcy and Bartlett 1996; Valentine 1988). According to a long-standing and influential hypothesis of this so-called 'face inversion' effect, first demonstrated by Yin (1969), vertical inversion selectively impairs our ability to extract configural information from faces, while leaving featural processing largely intact. This impairment mostly affects holistic processing and sensitivity to second-order relations (Leder and

---

C. Wallraven (✉) · H. H. Bühlhoff  
Department of Brain and Cognitive Engineering,  
Korea University Anam-Dong 5ga, Seongbuk-gu,  
Seoul 136-713, Korea  
e-mail: wallraven@korea.ac.kr

H. H. Bühlhoff  
e-mail: Heinrich.buelthoff@tuebingen.mpg.de

L. Whittingstall · H. H. Bühlhoff  
Department of Human Perception, Cognition and Action,  
Max Planck Institute for Biological Cybernetics,  
Spemannstraße 38, 72076 Tübingen, Germany  
e-mail: Lisa.dopjans@tuebingen.mpg.de

Bruce 2000; Schwanger et al. 2006, Gold et al. 2012). Whereas an upright face is processed at the global level of the whole face, an inverted face would have to be processed at a more local level, feature by feature (but see also Sekuler et al. 2004, Gold et al. 2012).

Van Belle et al. (2010) recently showed that the face inversion effect may, indeed, be caused by an inability to perceive the individual face as a whole rather than as a collection of specific features, thus supporting the view that observers' expertise at upright face recognition is due to configural processing. Using a gaze-contingent stimulus presentation method to study the visual face inversion effect, they compared participants' face discrimination performance on (1) faces presented in full view with (2) only the central window of vision revealed by means of a gaze-contingent aperture and with (3) only the fixated feature masked by means of a gaze-contingent mask. The authors found a face inversion effect for faces presented in full view but none for the second condition, when observers had their vision constrained such that they could see only through a small aperture. These results support the notion that the inversion effect is not primarily caused by a difficulty in perceiving local detailed facial features but rather by the observers' inability to simultaneously extract diagnostic information at different locations on an inverted face, in effect showing that configural face perception is impaired for inverted faces. The authors, however, did not specifically discuss the effect of constraining the effective field of view on encoding strategies. While unrestricted vision can process all aspects of an image in parallel, so that both local facial features and their global configuration can be rapidly processed (Tanaka and Sengco 1997), constraining the effective field of view limits participants to serial exploration of an object, that is, it entails piecemeal or feature-by-feature analysis (Loomis et al. 1991; Loomis and Lederman 1986).

We previously tested the effect of encoding differences on face recognition performance in unrestricted vision, gaze-restricted vision, and haptics (the latter as a sensory modality that is 'naturally' limited to serial encoding; Dopjans et al. 2012). This was achieved by using a gaze-restricted display that promoted serial encoding in vision. The gaze-restricted display limited the effective field of view in vision such that only one feature, determined by the observer him/herself, was available at any given time on a face. In a first series of experiments, we compared haptic, gaze-restricted, and unrestricted visual face recognition of face masks. Secondly, we used the face inversion paradigm to assess how encoding differences might affect face processing strategies (featural vs. configural face information processing). By promoting serial encoding in vision, we found the same pattern of results for haptic and visual recognition performance using a gaze-restricted display. Not

only was face recognition performance across the visual and haptic sense equated by reducing the visual window to the narrowness of the effective field of view in haptics (due to a decrease in visual face recognition accuracy as compared to unrestricted visual face recognition), but we also found a strong face inversion effect for unrestricted visual but none for gaze-restricted face recognition. Taken together, our results suggest that configural processing is enabled by fast and parallel information encoding and impeded by restricted, serial encoding.

These conclusions are also supported by two recent studies by Dalrymple et al. (2010, 2011) in which a gaze-contingent display was used to compare patients with simultanagnosia (for whom visual attention is restricted to a small spatial window) with healthy controls. In one study, participants were asked to attend to different global or local aspects of Navon letters consisting of a (global) letter shape made up of (local) smaller letters (Dalrymple et al. 2010). They found that the use of the gaze-contingent stimulus display resulted in significantly higher errors for the healthy controls in the global task, suggesting limitations in working memory or failure to integrate across fixations during serial visual encoding (see also discussion in Dopjans et al. 2012). In another experiment, fixation patterns for viewing of social scenes were compared between a patient and healthy controls (Dalrymple et al. 2011). When healthy controls were using the gaze-contingent display, fixation patterns mimicked that of the patient, again showing that serial encoding results in a change of how visual stimuli are processed and integrated.

Given these observed effects of encoding differences on information processing strategies, the question arises whether participants can *learn* to efficiently recognize faces that are serially encoded. It is well established that hallmarks of expert face processing, such as orientation sensitivity and configural processing, take many years to develop (Carey and Diamond 1977; Dahl et al. 2009; Hay and Cox 2000; Maurer et al. 2002; Mondloch et al. 2003; Pellicano and Rhodes 2003; Schwarzer 2000). Inasmuch as we have little to no training in gaze-restricted visual face recognition throughout life, it is possible that participants might be able to develop strategies to compensate for processing differences introduced by serial encoding. In particular, if participants were able to learn to accurately integrate information gained through serial encoding into a more global representation (e.g., Lakatos and Marks 1999), gaze-restricted face processing might also benefit from configural processing.

Although real-world expertise occurs on the scale of years (e.g., Carey and Diamond 1977; Maurer et al. 2002; Mondloch et al. 2003), whereas typical laboratory training studies require only hours of training (Gauthier and Tarr 1997, 2002; Malpass et al. 1973), such laboratory training studies allow for the manipulation of different factors

that may contribute to the acquisition of expertise, providing better control over variables influencing this process. Compared to work on face recognition in general, relatively few experimental studies of laboratory-acquired perceptual expertise have been reported (Gauthier et al. 1998, 1999a, b; Gauthier and Tarr 1997, 2002; Scott et al. 2006, 2008; Wong et al. 2009). Perhaps the most influential work comes from the studies of Gauthier and colleagues, who examined the acquisition of perceptual expertise using novel objects called ‘Greebles’ (Gauthier et al. 1998, 1999a, b; Gauthier and Tarr 1997, 2002) or ‘Ziggerins’ (Wong et al. 2009). The first of these investigations found that training not only led to faster and more accurate responses, but training also increased the configural (and thus face-like) processing of Greebles (Gauthier and Tarr 1997): The study demonstrated increased reaction time to transformed ‘Greeble’ configurations (studied parts, in a different configuration) compared to trained ‘Greeble’ configurations (studied parts, in a studied configuration)—this was only true for upright presentation, however. Tests of generalization of learning after ‘Greeble’ training suggested that learning generalized to ‘Greebles’ that were structurally similar to the training set, but did not generalize to ‘Greebles’ that were less similar to the training set (Gauthier et al. 1998).

Taken together, the following hallmarks of the acquisition of perceptual expertise have been proposed (Palmeri and Cottrell 2010):

- A marked improvement in terms of speed and accuracy for expert recognition compared to novices.
- The gradual development of configural processing strategies, more accurately, the development of holistic and relational processing, as, for example, measured using the face inversion effect. While novices are largely unaffected by inversion, expert recognition is significantly impaired when faces are presented upside down. Note that, in general, inversion as an index of expertise is most likely only effective for objects encountered in unique orientations (see Diamond and Carey (1986) for a study with dog experts, and Reed et al. (2003) for an inversion effect for bodies).
- The ability to rapidly learn and accurately recognize new exemplars. That is, expertise allows generalization to previously unknown members of an expert object class. More specifically, generalization of learning occurs when performance improvements with a specific set of trained exemplars generalize or transfer to previously unlearned exemplars. Again, note that generalization will have limits in terms of the stimulus difference that can be tolerated (Gauthier et al. 1998).

Motivated by our earlier results that serial encoding by means of gaze restriction seems to equate haptic and visual

performance in face recognition, here we set out to investigate the degree with which training is able to redevelop some hallmarks of normal, visual face recognition. In this study, we therefore aim to trace the development of perceptual expertise (as defined by the three properties of increase in performance, emergence of configural processing strategies, and generalization) in recognition of serially encoded face masks using a gaze-restricted display.

Participants were trained in gaze-restricted recognition of face masks on five consecutive days. The pretest on day one established baseline gaze-restricted recognition performance of novices on upright and inverted faces, followed by 2 days of gaze-restricted training on upright faces. The post-test on day four again tested gaze-restricted recognition performance for upright and vertically inverted faces. Face inversion effects were used as one marker of face expertise. Intuitively, the most obvious way to assess expertise with a perceptual category is to determine how well experts learn new exemplars of the category. If our trained participants were truly gaze-restricted recognition experts, we would expect them to be able to transfer expertise derived from our training procedure to new stimuli. We therefore tested their *generalization* ability for the learned stimulus class—a critical test of ‘true’ learning. Finally, we assessed the *persistence* of any learning effect by repeating the post-test in another session 7 days after training was completed.

## Methods

Because relatively few experimental studies of perceptual expertise have been reported, little is known about the best methods for manipulating the level of expertise. It is obvious that experts are generally more experienced than novices but it is not clear exactly how much experience is necessary to produce significant ‘expertise effects’ and how to determine the reached level of expertise. Here, we chose to have each participant perform exactly the same training procedure until a group average criterion of  $d' > 3.6$  was reached ( $d'$  is a measure of sensitivity and in this case was set to an equivalent of 100 % hit rate and 10 % false alarm rate; in the experiment described later, this translates to only 2 incorrect ‘recognitions’ per participant and block). For this, all participants were tested on the same day. Moreover, we opted for a semi-supervised learning task in which participants first learned to identify three faces by name (with receiving feedback) and then performed an old/new recognition task (without receiving feedback) (Dopjans et al. 2012; a similar task was also used in Dopjans et al. 2009). This mixed learning design was chosen as previous training studies have stressed the importance of unsupervised exposure in the formation of expert perceptual abilities (Scott et al. 2006, 2008). Whereas a fully supervised procedure, in

which participants received feedback for every trial, might potentially have resulted in larger training effects, training (as will be seen later) happened very efficiently even in this semi-supervised context, in which participants did not receive feedback throughout the whole experiment.

#### Ethics statement

The experiments described in this manuscript were conducted with human volunteers. Informed written consent was obtained prior to any experiment or recording from all participants. Participants and data from participants were treated according to the Declaration of Helsinki. All data were kept and analyzed anonymously. The experiments were conducted in accordance with the Max Planck Institute's IRB and were approved by the local ethics committee of the University of Tübingen (Project number: 89/2009BO2).

#### Participants

Ten experimentally naïve participants (6 females, mean age of 24.3 years, right-handed) were paid 8 Euros an hour to perform the experiment. All participants reported normal or corrected-to-normal vision and had no sensory impairment.

#### Stimuli

Stimuli in the experiments reported here were two sets of 19 images of face masks (38 images in total—note that in the following, we refer to faces and face masks interchangeably for the purpose of describing experimental results; for a discussion of the role pigmentation, see Discussion below). The faces for the two sets were chosen from the Max Planck Institute face database (Troje and Bühlhoff 1996) on the basis of visual similarity ratings from a pilot experiment to ensure comparability. For this, frontal face images from a subset of the database were used in a standard, pair-wise similarity rating task. The similarity ratings were embedded in a three-dimensional space using multidimensional scaling. This space was used to identify two sets of faces (set A and set B) whose average distances in the reconstructed space were comparable.

Experimental stimuli were generated and presented under Matlab 7.11 using the Psychophysics Toolbox (Brainard 1997; Pelli 1997). Each gaze-restricted stimulus was created from photographs of a frontal view of the white plastic face masks previously used in haptic experiments (Dopjans et al. 2009, 2012). The faces spanned  $14.7 \pm 1.2^\circ$  visual angle in the vertical plane and  $9.1 \pm 0.5^\circ$  visual angle in the horizontal plane and were presented on a black background spanning  $25.85^\circ$  visual angle in the horizontal plane and  $19.52^\circ$  visual angle in the vertical plane. In addition to making the two face sets comparable in terms of their average distance

in similarity space, we tested whether there were any differences in recognizability of the faces as they were shown here. We used the old/new recognition task as reported in Dopjans et al. (2012) and below with 16 participants (not part of the present study) and found that average recognition performance for set A was  $d' = 2.3$ , whereas we obtained  $d' = 2.2$  average performance for set B. The difference was not significant ( $t_{15} = -0.19$ ,  $p = 0.86$ ).

A Gaussian weight mask (an aperture, whose transparency values were defined by a Gaussian function) was then used to uncover the photograph at the position of the mouse cursor, such that only the (Gaussian-weighted) image intensities in the aperture were visible. The aperture uncovered a window of  $2^\circ$  visual angle of the underlying photograph (corresponding to a standard deviation of the Gaussian of  $\sigma = 0.5^\circ$ ). This visual angle is equivalent to two fingers at arm's length, reflecting the most commonly used exploratory procedure by participants observed in previous haptic face recognition experiments (See Fig. 1 for examples of the stimuli used).

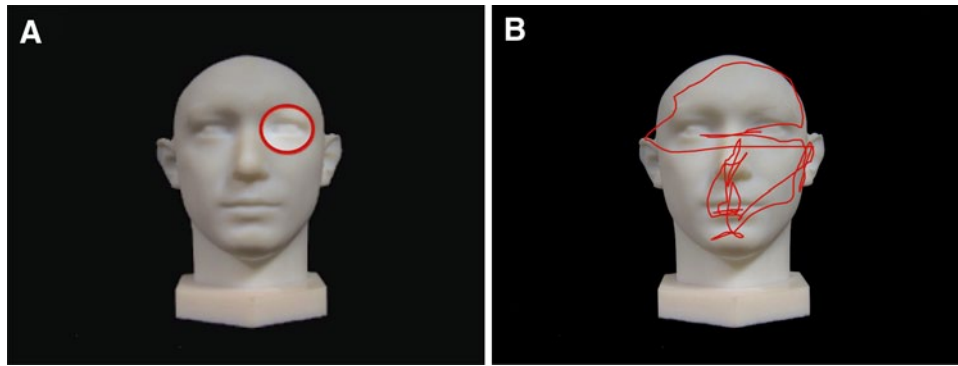
#### Experimental design

Participants were trained in gaze-restricted face recognition using an old/new recognition task on 5 consecutive days. The pretest on day 1 assessed participants' ability to recognize upright and vertically inverted faces using a gaze-restricted display. During training on days 2 and 3, participants performed the old/new recognition task on upright faces only. As participants reached our pre-defined criterion on day 3, the post-test on day 4 assessed gaze-restricted face recognition performance of upright and vertically inverted faces. Participants saw set A on days 1 to 4. On day 5, set B was used for upright and inverted gaze-restricted face recognition to rule out that potential learning effects might be solely due to familiarity with the faces. Finally, six participants returned for a final session 7 days later (on day 12) to perform the old/new recognition task with upright and inverted faces from set A.

See Fig. 2 for an overview of the procedure.

#### Procedure

Participants were seated about 60 cm away from a computer screen (21-inch CRT) resting their chin on a chin rest. They used a mouse to move a Gaussian window, which uncovered  $2^\circ$  of the photograph of the face mask for all trials such that their visual input was always limited to the small aperture. Participants were instructed not to move the mouse rapidly back and forth, as such a method would have produced a much larger effective visual field, since very rapid scanning differs little from simultaneous dull display (Ikeda and Uchikawa 1978) due to screen and visual persistence.



**Fig. 1** **a** Demonstration of the gaze-restricted display: The *red circle* indicates the size of the aperture. Only the part of the image inside the aperture was visible as indicated by the difference in brightness of the images inside and outside of the aperture. The aperture of 2°

visual angle was moved over the frontal photograph of the face mask. **b** Example of a recorded trajectory during gaze-restricted face recognition

In addition, we recorded the trajectories to check for any occurrence of these types of mouse movements, as they would need to be excluded from analysis. No trials had to be excluded for any participant.

Participants were first familiarized with three upright faces randomly chosen from set A. We labeled each face with a short first name. Participants were told to explore the face masks carefully and to learn their names because they would be asked to recognize those particular faces later. During this stage, they were allowed to explore the faces for as long as they wanted. No further information was given about the nature of the following experiment during the familiarization. As the task was fairly difficult, the set of three faces was the same during days 1, 2, 3, 4, and 12. On day 5, faces from set B were used to test for generalization.

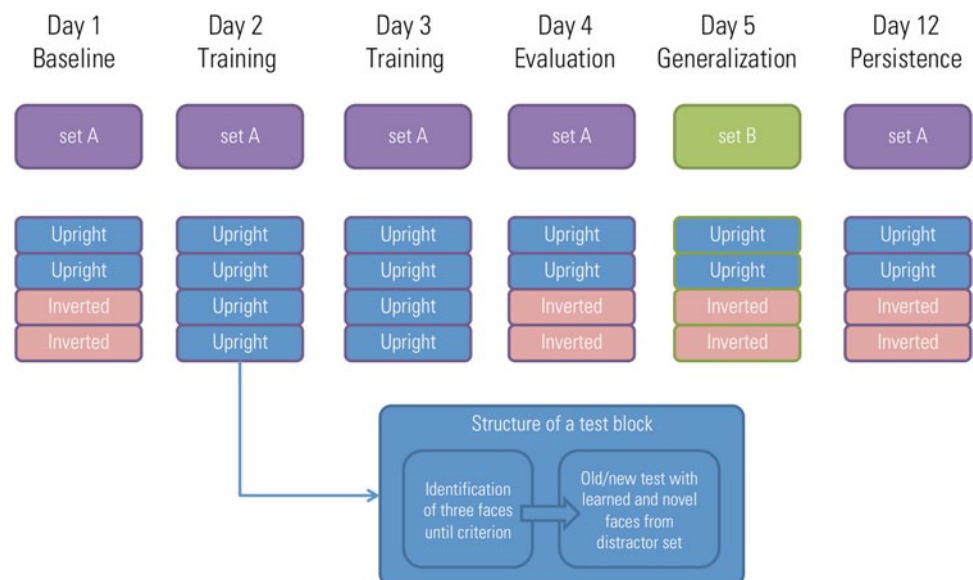
Each of the following test blocks had two parts: In the first part of the block, participants had to identify the three

familiarized faces in a forced choice test. After having passed this criterion, participants moved to the second part of the block, in which they had to perform a standard old/new recognition task with several faces.

In the identification task, participants were randomly presented with the three learned (upright) faces and had to name each face after exploration. Feedback was provided in that participants were told whether the face was recognized correctly or not. Faces were repeated in blocks of three in randomized order. Each face mask had to be identified correctly twice before the experiment continued—the minimum number of test blocks was therefore 2. This identification task with feedback was repeated before each test block to refresh memory of the training faces (see Dopjans et al. 2012 for an in-depth discussion of memory effects).

The old/new recognition task immediately followed the identification task and consisted of 19 trials, corresponding

**Fig. 2** Experimental design for assessing expertise development in gaze-restricted face recognition. A total of 6 days were used to test upright and inverted recognition of face masks. For a detailed description of the procedure, see text





to 3 old faces (i.e., faces that had been learned in the previous familiarization stage) and 16 new faces. This asymmetric design was chosen because of time constraints for gaze-restricted learning. Face masks were shown one at a time in random order. Participants were asked to explore each face mask and to report whether it was one of the three faces they had learned (old) or not (new). Although exploration time was unrestricted, they were instructed to respond as quickly and accurately as possible by pressing an ‘old’ or ‘new’ labeled key on a keyboard with their left hand. No feedback was provided for the old/new recognition task.

The training sessions on days 2 and 3 consisted of 4 blocks with upright faces only (one familiarization phase, followed by four test blocks of identification/recognition of upright faces). On days 1, 4, and 12, the experiment comprised two test blocks with upright and two test blocks with vertically inverted (upside-down) faces. During the familiarization phase and during the identification tasks, only upright faces were used.

On day 5, participants were tested with the same task; however, this time, faces from set B were used for training and testing—this was done to test for generalization to a different set of faces. Finally, on day 12, participants returned for another test of faces from set A to test how well training effects would persist.

## Analysis

The dependent variables for this study consisted of the old/new response as well as the response time. From the old/new responses, we determined the hit rates and false alarm rates, which were then converted to standard  $d'$  scores.  $D'$  was determined for each test block separately and averaged across test blocks. Response times were averaged across test blocks as well. We conducted separate analyses for responses to correct trials only, or for all trials—the results did not change such that we report here responses to all trials.

Both  $d'$  scores and response times were then subjected to statistical tests looking for effects of ‘day’ (that is, how performance changed across training and/or face sets) and ‘orientation’ (that is, how performance was affected by inversion).

## Results

### Behavioral data

The means and standard error for day and orientation are shown in Fig. 3. One-tailed  $t$  tests corrected for multiple comparisons showed that performance was above chance for each face orientation on each day (all  $p < 0.01$  for all

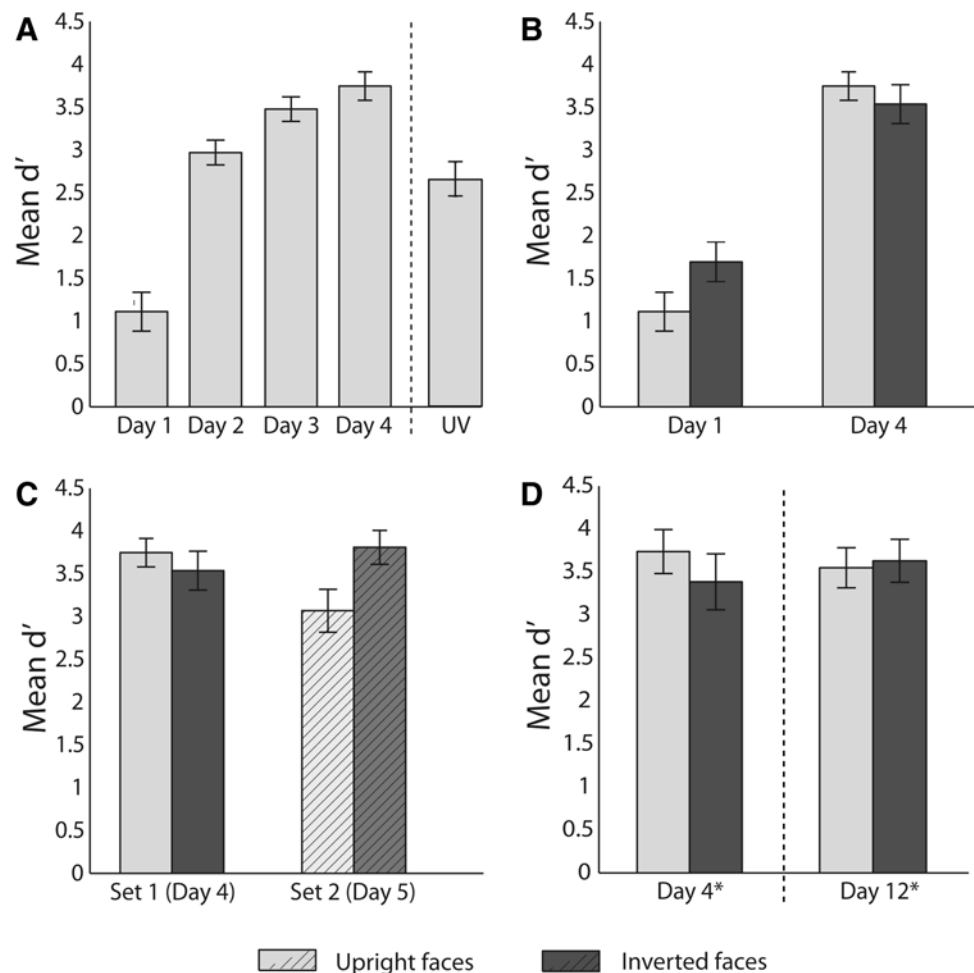
conditions). Our pre-defined criterion of  $d' \geq 3.6$  for the whole group of participants was reached after only 3 days of training (Fig. 3a, we include the pretest in the count of training days; the range of  $d'$  values on this day was from  $d' = 2.7$  to  $d' = 4.6$  indicating that all participants had learned the task well).

We further analyzed the results for the upright face orientation on days 1–4 using a repeated-measures ANOVA to test for a learning effect. We found a significant main effect for ‘day’ ( $F_{1,9} = 167.01, p < 0.001$ ) indicating that face recognition performance significantly improved through training. This performance increase was such that post-training performance seemed even better than for unrestricted *visual* face recognition of the same face masks (see Fig. 2a; data taken from a previous experiment using the same task and stimuli testing *unrestricted visual face recognition*, Dopjans et al. 2009; both this and the present study used fully comparable participant populations in terms of age, gender, and handedness—if one allows for statistical comparison of these populations, the resulting  $t$  tests become significant with  $t_{26} = 3.44, p < 0.01$ ).

Secondly, we tested for inversion effects (different recognition performance for inverted than upright faces) on days 1, 4, 5, and 12 (Fig. 3b–d). Importantly, a repeated-measures ANOVA with factors ‘day’ and ‘orientation’ failed to find a significant interaction ( $F_{1,9} = 2.11, p = 0.18$ ). Hence the observed improvement in face recognition performance (main effect of ‘day’  $F_{1,9} = 52.06, p < 0.001$ ) was not accompanied by a change in information processing strategies (replicating and extending Dopjans et al. 2012). Indeed, conducting planned two-tailed  $t$  tests failed to find evidence for inversion effects on any day (day 1:  $t_9 = -1.36, p = 0.21$ ; day 4:  $t_9 = 0.70, p = 0.50$ ; day 5:  $t_9 = -1.68, p = 0.11$ ; day 12:  $t_5 = -0.87, p = 0.42$ ). A sensitivity analysis shows that our sample size of 10 participants would have been enough to detect a large effect size of  $d_z = 1$  (equivalent to a difference in  $d'$  scores of 1 between upright and inverted faces, divided by a standard deviation of 1), given standard values of the Type I error probability of  $\alpha = 0.05$ , and associated power of the test of  $1 - \beta = 0.8$ . Observed effect sizes for paired  $t$  tests comparing performances on each single day, however, average only around 0.18 for all days. This indicates that if there were any group differences, they would be very small—indeed, much smaller than usually found for inversion effects in the face literature (e.g.,  $d_z > 0.9$  for studies in Jiang et al. 2011).

Thirdly, we used two-tailed paired  $t$  tests to compare face recognition performance on upright faces for stimulus set 1 (on day 4) and set 2 (on day 5) to test generalizability of the observed learning effect to a new set of stimuli (Fig. 3c). We found a marginally significant difference between performance on days 4 and 5 ( $t_9 = 2.23, p = 0.06$ , mean  $d'$  scores for set 1 of  $d' = 3.7$  and for set 2 of  $d' = 3.2$ ,

**Fig. 3** Plots comparing face recognition performance **a** for upright faces on days 1–4 and compared to unrestricted visual face recognition, **b** for upright and inverted faces for gaze-restricted face recognition novices (day 1) and ‘experts’ (day 4), **c** for upright and inverted faces for experts using the training set (set 1, day 4) and a new set of faces (set 2, day 5), and **d** for upright and inverted faces for experts at the end of the training (day 4\*) and after 7 days (day 12\*; \* indicates data from only 6 subjects). Data are measured in mean  $d'$   $\pm$  1 Standard Error of the Mean (SEM)



respectively). Despite the slightly lower performance, however, recognition results for set 2 are still far greater than those for set 1 on day 1 ( $t_9 = 6.97, p < 0.001$ ). As the selection of the two sets of faces followed the same criteria (see ‘Methods’ above), the solid recognition performance for the new face set shows that participants were indeed able to generalize newly learned strategies for efficient gaze-restricted face recognition.

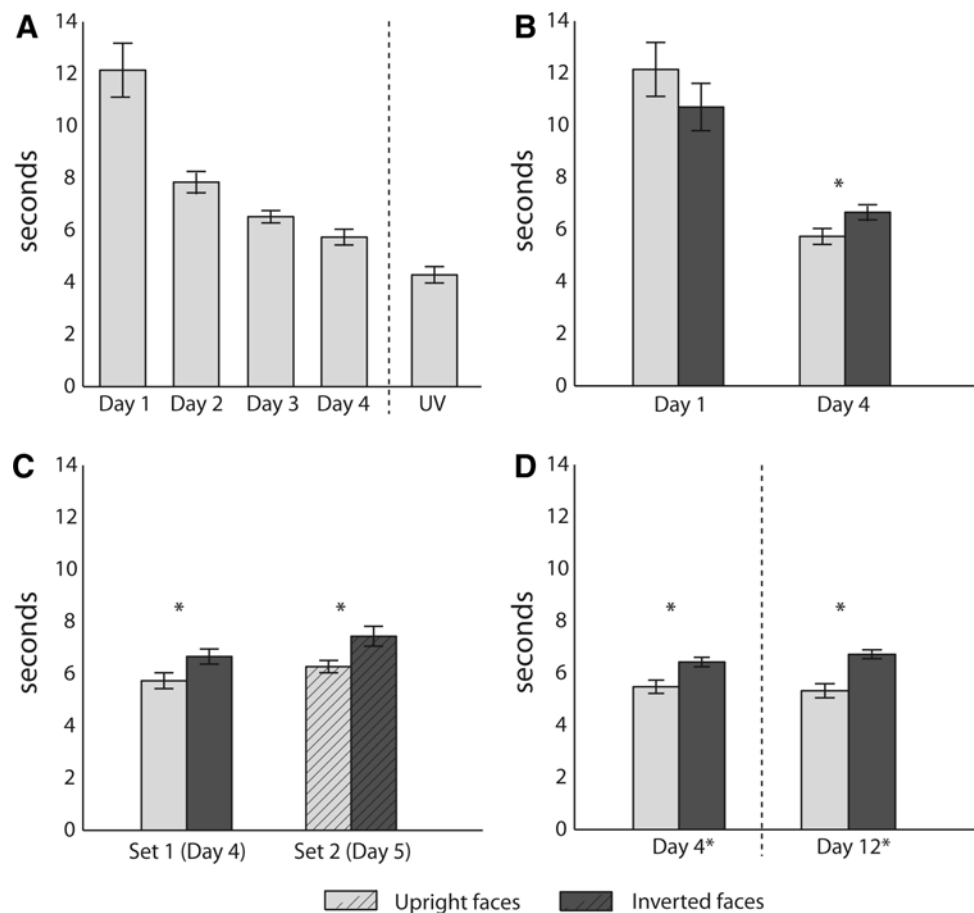
Finally, we compared recognition performance on upright and inverted faces on days 4 and 12 to test persistence of the observed learning effect (Fig. 3d). Since we were only able to test six participants for the experiment on day 12, we re-analyzed performance on day 4 for those six participants only and used a paired  $t$  test to compare performances. We found no significant difference between results on day 4 and 12 (upright faces:  $t_5 = -0.11, p = 0.92$ , inverted faces:  $t_5 = 0.28, p = 0.79$ ). In addition, performance on day 12 was still significantly higher than that of day 1 (upright faces:  $t_5 = 7.81, p < 0.001$ , inverted faces:  $t_5 = 5.44, p < 0.01$ ). Taken together, this indicates that the newly acquired perceptual skill persisted for at least 1 week.

#### Response times

Response times of all trials were averaged across test blocks. The means and standard errors for each day and orientation are shown in Fig. 4. We analyzed response times for the upright face orientation on days 1–4 using a repeated-measures ANOVA to test for learning effects. We found a significant main effect for days ( $F_{1,9} = 144.44, p < 0.001$ ) indicating that participants became faster at gaze-restricted face recognition performance through training (Fig. 4a). More interestingly, we compared response times on day 4 (after training) to those from a previous experiment using the same task and stimuli testing *unrestricted visual* face recognition (Dopjans et al. 2009) using two-tailed  $t$  tests. We found only a marginally significant advantage for unrestricted visual recognition ( $t_{26} = 1.80, p = 0.08$ , Fig. 4a), indicating that our training improved response times almost to ‘normal’ face recognition levels.

Secondly, we tested for inversion effects (longer response times for inverted than upright faces) on days 1, 4, 5, and 12 (Fig. 4b–d). A repeated-measures ANOVA revealed a

**Fig. 4** Plots comparing response times for gaze-restricted face recognition performance **a** for upright faces on days 1–4 and compared to unrestricted visual face recognition, **b** for upright and inverted faces for gaze-restricted face recognition novices (day 1) and ‘experts’ (day 4), **c** for upright and inverted faces for experts using the training set (set 1, day 4) and a new set of faces (set 2, day 5), and **d** for upright and inverted faces for experts at the end of the training (day 4\*) and after 7 days (day 12\*; \* indicates data from only 6 subjects). Data are measured in mean seconds  $\pm$  1 Standard Error of the Mean (SEM). Asterisks indicate significant differences between conditions



significant interaction for factors Training and Orientation ( $F_{1,9} = 14.92, p < 0.01$ ). While we failed to find an inversion effect on day 1 with participants almost being faster for *inverted* than upright faces ( $t_9 = -2.26, p = 0.07$ ), post-training participants were significantly faster at recognizing *upright* than inverted faces on day 4 ( $t_9 = 2.42, p < 0.05$ ). Furthermore, one-tailed, paired  $t$  tests revealed that participants were significantly faster for the second set of faces when recognizing upright than inverted faces on day 5 and on day 12 (day 5:  $t_9 = 2.34, p < 0.05$ ; day 12:  $t_5 = 2.26, p < 0.05$ ).

Thirdly, we used two-tailed paired  $t$  tests to compare response times for upright faces of stimulus set A on day 4 and set B on day 5 to test generalizability of the observed learning effect to a new set of stimuli (Fig. 4c). We found no significant difference between response times on days 4 and 5 ( $t_9 = -1.23, p = 0.25$ ). In addition, differences between response times on days 1 and 5 were still highly significant ( $t_9 = 8.59, p < 0.001$ ). Taken together, these results show that participants were indeed able to generalize newly learned strategies for efficient gaze-restricted face recognition to a new set of stimuli.

Finally, we compared response times on upright faces on days 4 and 12 to test persistence of the observed learning effect (Fig. 4d). We re-analyzed performance on day 4 for the six participants, who did the experiment on day 12, and used a paired  $t$  test to compare performance. Again, we found no significant difference between results on day 4 and 12 ( $t_5 = 0.47, p = 0.66$ ), and again, performance remained higher on day 12 than on day 1 ( $t_5 = 6.03, p < 0.01$ ) demonstrating that whatever strategies participants had picked up still enabled them to solve the difficult recognition task after 1 week.

## Discussion

We have previously shown that serial encoding similarly impedes visual and haptic face recognition (Dopjans et al. 2012). In both cases, we failed to find a face inversion effect, which could be due to promoting featural versus configural processing of facial information, the latter being widely considered as a hallmark of expert face processing. This finding raised the question whether modalities that rely on



serial encoding of information actually allow for expert face processing.

Consequently, here we trained participants in gaze-restricted face recognition to assess whether they can learn to efficiently recognize face masks that are serially encoded. In summary, we found (1) that participants became significantly faster (almost to levels of unrestricted visual face recognition) and better (up to and exceeding levels of unrestricted visual face recognition) at gaze-restricted face recognition through short training, as our pre-defined performance criterion was met after 3 days of training, (2) an inversion effect for response times and fixations (but not accuracy) for trained but not untrained participants, (3) that improvement in performance did not arise from familiarity to the faces but transferred to novel faces, and (4) that the newly learned skills in gaze-restricted recognition persisted for at least a week at similar performance levels. Our results therefore suggest that participants in this training did indeed learn how to learn and recognize faces during the training procedure.

First, the significant improvement in performance and response times that we observed during expertize acquisition represent a clear learning effect. Although participants received little feedback during training (only during the identification task), performance improved quickly and passed our pre-defined criterion after only 3 days of training (at least 36 trials of the identification task and 228 trials of the old/new recognition task). Interestingly, post-training recognition of serially encoded faces even exceeded levels of unrestricted visual recognition.

It seems clear that expertize should be more than simply a practice effect involving a qualitative shift in processing strategies. As mentioned above, we used the inversion effect as a measure to evaluate the nature of the abilities acquired by experts in processing faces (Diamond and Carey 1986; Sergent 1984; Yin 1969): For novices, we replicated previous results as we failed to find an inversion effect in terms of accuracy as well as response times, indicating the use of featural processing strategies (Dopjans et al. 2012). Interestingly, novices even showed a ‘paradoxical’ inversion effect for response times as they were almost *slower* on upright than on inverted faces. While this effect has previously been shown in prosopagnosic patients, its cause remains unclear (Farah et al. 1995; de Gelder and Rouw 2000). For trained participants, however, response times were significantly faster on upright than inverted faces. Crucially, this difference indicates that the expertize manipulation produced a speed advantage for upright over inverted faces.

What is not entirely clear is why our participants showed this sensitivity in response times and not in accuracy. We previously found a strong face inversion effect for *unrestricted* visual but none for haptic or gaze-restricted face recognition novices using our 3D face masks (Dopjans et al.

2012). This inversion effect was found for recognition accuracy, as well as response times (for similar results, see Van Belle et al. 2010). Of course, psychophysical models rarely allow one to predict a priori whether a difference between conditions will manifest itself in one dependent measure or another (Gauthier and Tarr 1997; Tanaka and Farah 1993; Tanaka and Sengco 1997). Alternatively, novices might not abruptly switch from one type of processing to another during expertize acquisition but rather undergo a more continuous shift of the type of processing. Studying the acquisition of perceptual expertize with ‘Greebles’, Gauthier and Tarr (2002), for example, found that holistic processing and second-order processing—both types of configural processing that are affected by inversion—develop on different time scales and appear to be very strongly related to the amount of expertize. This gradual shift in strategies might have manifested itself in the response time measure first, whereas an advantage of configural processing for recognition accuracy might only be established over a longer period of time with more training. After all, the inversion effect takes many years to first develop in children (Carey and Diamond 1977; Dahl et al. 2009; Hay and Cox 2000; Maurer et al. 2002; Mondloch et al. 2003; Pellicano and Rhodes 2003; Schwarzer 2000). Future studies with a more extensive training phase are needed to assess the evolution of processing strategies during acquisition of expertize for serially encoded faces and their manifestation in terms of response times and accuracy, in more detail.

Alternatively, the observed decrease in response time might reflect the optimization of exploratory strategies. While the face inversion effect has been shown to be a robust marker of face expertize (Diamond and Carey 1986; Sergent 1984; McKone et al. 2006), further studies with our setup using other experimental paradigms to test configural processing such as the composite effect (e.g., Hole 1994), configural changes in features (e.g., Freire et al. 2000), the Thatcher Illusion (e.g., Boutsen and Humphreys 2003), or scrambling facial features (e.g., Collishaw and Hole 2000) are necessary to fully investigate the use of processing strategies in gaze-restricted face recognition. Moreover, as we trained participants on upright faces, one might also argue that the observed inversion effect could be attributed to a mere exposure effect. It would therefore be interesting to train participants on inverted faces. Robbins and McKone (2003) have shown that orientation specificity of unrestricted visual face processing is highly stable against practice when participants failed to learn holistic processing for inverted faces (in contrast to the situation for objects; Tarr and Pinker 1989). Whether or not participants would be able to more efficiently learn processing for inverted faces using gaze-restricted vision would shed further light on the question of orientation specificity of gaze-restricted face recognition.

Secondly, the most obvious way to assess expertize with a perceptual category is to determine how well experts learn new exemplars of the category (Gauthier et al. 1998). After their training had been completed, our participants performed the task on a new set of faces. Since we found no significant difference in accuracy or response times between recognition of face masks from the training set and novel faces, the expertize derived from our training procedure appeared to have transferred well to a novel stimulus set. This finding, again, suggests that through training participants learned what kind of information is helpful in recognizing serially encoded faces and that this knowledge is not based on familiarity with the training stimuli but can be generalized to novel stimuli.

Thirdly, participants were as accurate and fast at recognizing serially encoded face masks after 7 days without practice. It would be interesting to rerun the study with longer time intervals to test for retention of expertize beyond this time period. In this case, a more detailed study of the exploration patterns—that is, how information is sampled from the face—would also be interesting to check whether typical strategies remain intact after a longer time, or if specific sampling patterns would need to be ‘re-discovered’.

Finally, it is important to note that the present study tested acquisition of expertize using face *masks*, that is, faces lacking natural pigmentation cues. This was mainly done to enable comparison with our earlier studies on haptic and (unrestricted) visual face recognition on the *same* stimulus set (Dopjans et al. 2009, 2012). Beyond this, several facts make us believe that the results may be able to be generalizable to recognition of ‘natural’ faces. First, participants in the unrestricted visual condition (Dopjans et al. 2012) actually showed a strong and reliable inversion effect—hence, our face masks support this important hallmark of expert face processing. Previous studies on the role of pigmentation and shape cues have been shown to contribute equally to face recognition (O’Toole et al. 1999, Yip and Sinha 2002), such that a lack of either cue might worsen recognition performance. Russell et al. (2007) found that both faces different only in shape and only in pigmentation carried inversion effects showing that both types of information support configural processing. In addition, a recent study (Jiang et al. 2011) used a morphable face model to test recognition effects in faces differing in pigmentation, shape, or both. They found that inversion had a slightly stronger impact on faces differing in shape than in pigmentation. In addition, the composite effect—another type of effect related to configural processing—was more evident for faces defined through shape differences. Finally, a prosopagnosic patient was significantly more impaired in matching faces differing in shape than in pigmentation. Taken together, these results suggest that processing of shape can be taken as representative of normal (expert) face processing.

In future studies, we will investigate the exploration pattern (that is, the information contained in the movement of the apertures over the face) in more detail. This includes comparison with fixation strategies from eye-tracking studies, as well as computational modeling of the scan path. Despite the similarities between the hand-controlled aperture task and eye movements, one difference is that the hand-controlled apertures may still reveal information between ‘fixation’ locations (these are locations for which participants remain relatively still, taking in the visual information) as opposed to the saccadic suppression experienced during eye movements. Note also that in this context, exploring the face with eye movements is a kind of ‘serial’ processing—the differences between unrestricted visual exploration and gaze-restricted as well as haptic exploration are that the latter two do not have access to peripheral information around fixations and that integration across fixations may be impaired. In order to address these issues, we are currently planning eye-tracking studies to analyze and compare the different exploration patterns.

**Acknowledgments** This research was supported by a Ph.D. stipend from the Max Planck Society, by the World Class University (WCU) program through the National Research Foundation of Korea funded by the Ministry of Education, Science, and Technology (R31-1008-000-10008-0), and through the National Research Foundation of Korea funded by the Ministry of Education, Science, and Technology (2010-0011569).

## References

- Boutsen L, Humphreys GW (2003) The effect of inversion on the encoding of normal and ‘thatcherized’ faces. *Q J Exp Psychol* 56A(6):955–975
- Brainard DH (1997) The psychophysics toolbox. *Spat Vis* 10:433–436
- Carey S, Diamond R (1977) From piecemeal to configurational representation of faces. *Science* 195:312–314
- Collishaw SM, Hole GJ (2000) Featural and configurational processes in the recognition of faces of different familiarity. *Perception* 29(8):893–909
- Dahl CD, Wallraven C, Bühlhoff HH, Logothetis NK (2009) Humans and macaques employ similar face-processing strategies. *Curr Biol* 19:509–513
- Dalrymple KA, Bischof WF, Cameron D, Barton JJ, Kingstone A (2010) Simulating simultanagnosia: spatially constricted vision mimics local capture and the global processing deficit. *Exp Brain Res* 202(2):445–455
- Dalrymple KA, Birmingham E, Bischof WF, Barton JJ, Kingstone A (2011) Experiencing simultanagnosia through windowed viewing of complex social scenes. *Brain Res* 1367:265–277
- De Gelder B, Rouw R (2000) Paradoxical configuration effects for faces and objects in prosopagnosia. *Neuropsychologia* 38:1271–1279
- Diamond R, Carey S (1986) Why faces are and are not special: an effect of expertize. *J Exp Psychol Gen* 115:107–117
- Dopjans L, Wallraven C, Bühlhoff HH (2009) Cross-modal transfer in visual and haptic face recognition. *IEEE Trans Haptics* 2: 236–240
- Dopjans L, Bühlhoff HH, Wallraven C (2012) Serial exploration of faces: comparing vision and touch. *J Vis* 12(1):6

- Farah MJ, Wilson KD, Drain HM, Tanaka JR (1995) The inverted face inversion effect in prosopagnosia: evidence for mandatory, face-specific perceptual mechanisms. *Vis Res* 35(14):2089–2093
- Freire A, Lee K, Symons LA (2000) The face-inversion effect as a deficit in the encoding of configural information: direct evidence. *Perception* 29:159–170
- Gauthier I, Tarr MJ (1997) Becoming a ‘Greeble’ expert: exploring the face recognition mechanism. *Vis Res* 37:1673–1682
- Gauthier I, Tarr MJ (2002) Unraveling mechanisms for expert object recognition: bridging brain activity and behavior. *J Exp Psychol Human Learn Memory* 28:432–446
- Gauthier I, Williams P, Tarr MJ, Tanaka JW (1998) Training ‘greeble’ experts: a framework for studying expert object recognition processes. *Vis Res* 38:2401–2428
- Gauthier I, Behrmann M, Tarr MJ (1999a) Can face recognition really be dissociated from object recognition? *J Cognit Neurosci* 11:349–370
- Gauthier I, Tarr MJ, Anderson A, Skudlarski P, Gore JC (1999b) Activation of the middle fusiform ‘face area’ increases with expertise in recognizing novel objects. *Nat Neurosci* 2:568–573
- Gold JM, Mundy PJ, Tjan BS (2012) The perception of a face is no more than the sum of its parts. *Psychol Sci*. doi:10.1177/0956797611427407
- Hay DC, Cox R (2000) Developmental changes in the recognition of faces and facial features. *Infant Child Dev* 9:199–212
- Hole GJ (1994) Configurational factors in the perception of unfamiliar faces. *Perception* 23(1):64–74
- Ikeda M, Uchikawa K (1978) Integrating time for visual pattern perception and a comparison with the tactile mode. *Vis Res* 18:1565–1571
- Jiang F, Blanz V, Rossion B (2011) Holistic processing of shape cues in face identification: evidence from face inversion, composite faces, and acquired prosopagnosia. *Visual Cognit* 19(8):1003–1034
- Lakatos S, Marks L (1999) Haptic form perception: relative salience of local and global features. *Percept Psychophys* 61(5):895–908
- Leder H, Bruce V (2000) When inverted faces are recognized: the role of configural information in face recognition. *Q J Exp Psychol* 53A:513–536
- Loomis JM, Lederman SJ (1986) Tactual perception. In: Boff KR, Kaufman L, Thomas JP (eds) *Handbook of perception and human performances*, vol 2, cognitive processes and performance. Wiley, New York, pp 31/1–31/41
- Loomis JM, Klatzky RL, Lederman SJ (1991) Similarity of tactual and visual picture recognition with limited field of view. *Perception* 20:167–177
- Malpass RS, Laviguer H, Weldon DE (1973) Verbal and visual training in face recognition. *Percept Psychophys* 14:283–292
- Maurer D, LeGrand R, Mondloch CJ (2002) The many faces of configural processing. *Trends Cognit Sci* 6:255–260
- McKone E, Kanwisher N, Duchaine BC (2006) Can generic expertise explain special processing for faces? *Trends Cognit Sci* 11(1): 8–15
- Mondloch CJ, Geldart S, Maurer D, LeGrand R (2003) Developmental changes in face processing skills. *J Exp Child Psychol* 86: 67–84
- O’Toole AJ, Vetter T, Blanz V (1999) Three-dimensional shape and two-dimensional surface reflectance contributions to face recognition: an application of three-dimensional morphing. *Vis Res* 39:3145–3155
- Palmeri TJ, Cottrell GW (2010) Modeling perceptual expertise. In: Gauthier I, Tarr MJ, Bub D (eds) *Perceptual expertise: bridging brain activity and behavior*. Oxford University Press, Oxford
- Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis* 10:437–442
- Pellicano E, Rhodes G (2003) Holistic processing of faces in preschool children and adults. *Psychol Sci* 14:618–622
- Reed CL, Stone VE, Bozova S, Tanaka J (2003) The body-inversion effect. *Psychol Sci* 14(4):302–308
- Robbins R, McKone E (2003) Can holistic processing be learned for inverted faces? *Cognition* 88:79–107
- Russell R, Biederman I, Nederhouser M, Sinha P (2007) The utility of surface reflectance for the recognition of upright and inverted faces. *Vis Res* 47:157–165
- Schwaninger A, Wallraven C, Cunningham DW, Chiller-Glaus S (2006) Processing of identity and emotion in faces: a psychophysical, physiological and computational perspective. *Prog Brain Res* 156:321–343
- Schwarzer G (2000) Development of face processing: the effect of face inversion. *Child Dev* 71:391–401
- Scott LS, Tanaka J, Sheinberg DL, Curran T (2006) A reevaluation of the electrophysiological correlates of expert object processing. *J Cognit Neurosci* 18:1453–1465
- Scott LS, Tanaka JW, Sheinberg DL, Curran T (2008) The role of category learning in the acquisition and retention of perceptual expertise: a behavioral and neurophysiological study. *Brain Res* 1210:204–215
- Searcy JH, Bartlett JC (1996) Inversion and processing of component and spatial-relational information of faces. *J Exp Psychol Hum Percept Perform* 22:43–47
- Sekuler AB, Gaspar CM, Gold JM, Bennett PJ (2004) Inversion leads to quantitative not qualitative, changes in face processing. *Curr Biol* 14(5):391–396
- Sergent J (1984) An investigation into component and configural processes underlying face perception. *Br J Psychol* 75(2):221–242
- Tanaka JW, Farah MJ (1993) Parts and wholes in face recognition. *Q J Exp Psychol* 12:242–248
- Tanaka JW, Sengco J (1997) Features and their configuration in face recognition. *Memory Cognit* 25:583–592
- Tarr MJ, Pinker S (1989) Mental rotation and orientation-dependence in shape recognition. *Cogn Psychol* 21(2):233–282
- Troje NF, Bühlhoff HH (1996) Face recognition under varying pose: the role of texture and shape. *Vis Res* 36:1761–1771
- Valentine T (1988) Upside-down faces: a review of the effects of inversion upon face recognition. *Br J Psychol* 79:471–491
- Van Belle G, De Graef P, Verfaillie K, Rossion B, Lefèvre P (2010) Face inversion impairs holistic perception: evidence from gaze-contingent stimulation. *J Vis* 10:1–13
- Wong ACN, Palmeri T, Gauthier I (2009) Conditions for face-like expertise with objects: becoming a Ziggerin expert—but which type? *Psychol Sci* 20(9):1109–1117
- Yin RK (1969) Looking at upside-down faces. *J Exp Psychol* 81:141–145
- Yip A, Sinha P (2002) Role of color in face recognition. *Perception* 31:995–1003