Faces in Context: Does Face Perception Depend on the Orientation of the Visual Scene?

Jessica Taubert and Celine van Golde
The School of Psychology, The University of Sydney, Australia

Frans A. J. Verstraten
The School of Psychology, The University of Sydney, Australia
Helmholtz Institute, The University of Utrecht, The Netherlands

Abstract

The mechanisms held responsible for familiar face recognition are thought to be orientation dependent; inverted faces are more difficult to recognize than their upright counterparts. Although this effect of inversion has been investigated extensively, researchers have typically sliced faces from photographs and presented them in isolation. As such, it is not known whether the perceived orientation of a face is inherited from the visual scene in which it appears. Here, we address this question by measuring performance in a simultaneous same–different task while manipulating both the orientation of the faces and the scene. We found that the face inversion effect survived scene inversion. Nonetheless, an improvement in performance when the scene was upside down suggests that sensitivity to identity increased when the faces were more easily segmented from the scene. Thus, while these data identify congruency with the visual environment as a contributing factor in recognition performance, they imply different mechanisms operate on upright and inverted faces.

Keywords

face perception, scene perception, segmentation, shapes/objects

Introduction

Inverted faces are difficult to recognize. This classic observation (Hochberg & Galper, 1967; Yin, 1969) has been widely replicated (for a review, see McKone & Yovel, 2009) and has converged with many different neural markers (Kanwisher, Tong, & Nakayama, 1998; Rossion, Prieto, Boremanse, Kuefner, & Van Belle, 2012; Taubert, Van Belle, Vanduffel, Rossion, & Vogels, 2015; Yovel & Kanwisher, 2005), suggesting that faces are processed by

Corresponding author:
Jessica Taubert, The School of Psychology, The University of Sydney, Sydney, New South Wales 2006, Australia.
Email: jesstaubert@gmail.com

orientation-dependent brain mechanisms in participants that have developed normally. Indeed, a number of studies have confirmed that equivocal performance with upright and inverted faces is associated with impaired or inefficient face recognition in people living with Prosopagnosia, a specific inability to recognize faces (de Gelder & Rouw, 2000; Farah, Wilson, Drain, & Tanaka, 1995; Righart & de Gelder, 2007).

In parallel to these observations, behavioral studies have shown that the perception of metric distances between features, like the interocular distance, is more affected by inversion than the shape of local features, like the shape of the mouth (Goffaux & Rossion, 2006; Le Grand, Mondloch, Maurer, & Brent, 2001; Rhodes, Brake, & Atkinson, 1993). Evidence that inversion does not have a uniform affect on all aspects of face perception has been used to argue that inversion affects face processing qualitatively rather than quantitatively (Rossion & Boremanse, 2008).

A strict qualitative account of the face inversion effect presumes there is a one- or two-dimensional code that confers a computational advantage by initiating face processing in the brain (Bartlett & Searcy, 1993; Dakin & Watt, 2009; Morton & Johnson, 1991; Taubert & Alais, 2009; Xu, Biederman, & Shah, 2014). Under this view, a face code (or template) acts like a filter, and inverted faces are disqualified on the basis of their feature sequence. The increasingly popularity of the qualitative account is driven in part by its success in improving computer algorithms for face recognition (see Tsao & Livingstone, 2008). It predicts that a face presented upside down, relative to the observer, will be harder to recognize and more difficult to discriminate from other upside down faces irrespective of external factors such as the visual scene.

However, several other studies have cast some doubts on a strict qualitative interpretation of the face inversion effect (Collishaw & Hole, 2002; McKone & Yovel, 2009; Riesenhuber, Jarudi, Gilad, & Sinha, 2004; Riesenhuber & Wolff, 2009). A quantitative view of this phenomenon was first asserted by Valentine and Bruce (1988). They posit that an inverted face represents a more costly signal due to the processing associated with normalization, something like mental-rotation (Schwaninger & Mast, 2005; Valentine & Bruce, 1988). Under this view, a loss of sensitivity to inverted faces could potentially be explained by the incongruency between the orientation of a face and the surrounding background scene.

In 1957, Mooney introduced two-tone faces as a means of evaluating mental completion (i.e., perceptual closure). Since then, the so-called *Mooney Faces* have been a valuable tool for understanding face detection in Prosopagnosia patients (Busigny, Joubert, Felician, Ceccaldi, & Rossion, 2010) and nonhuman primates (Taubert & Parr, 2012). Inverting Mooney faces makes them more difficult to detect (Kanwisher et al., 1998). In Figure 1, we demonstrate whether the visual environment can facilitate the detection of an inverted Mooney face. A quantitative view that explains the face inversion effect as the result of the incongruency between the face and the background scene would predict recovery of perception in the bottom right panel where the inverted face appears in an upside down scene (the inverted congruent condition). Instead, it seems scene inversion facilitates a kind of visual pop-out in the bottom left panel.

Here, we empirically test whether the effect, hinted at in Figure 1, is evident when observers are asked to match photographs of familiar famous people. As with the demonstration in Figure 1, we created a plausible scenario for our participants tasking them to recognize whether two photographs hanging in a lounge room were taken of the same person or two different people. This enabled us to independently manipulate the orientation of the local face stimuli and surrounding background scene across four conditions to determine how congruency with the scene influences face recognition performance.
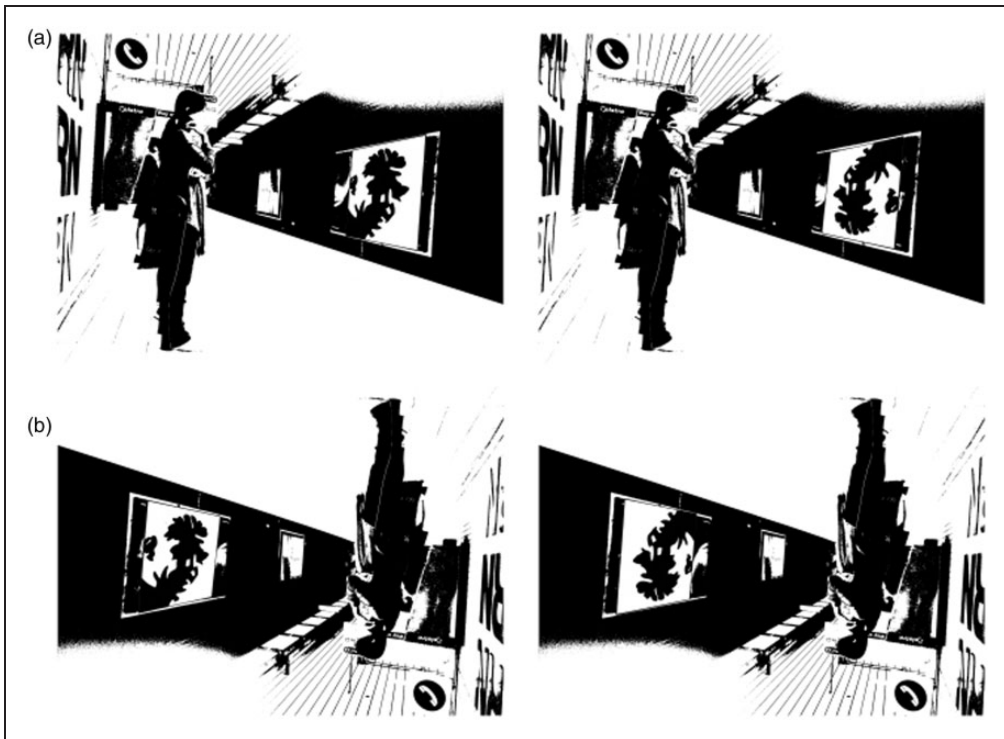
Figure 1. The Mooney face demonstration. (a) On the left, a Mooney face was superimposed on a visual scene with its perspective and orientation cues preserved (image was converted to black and white to ease integration). On the right is virtually the same image with the Mooney face turned upside down (b). The images in this panel are identical to those in Figure 1(a) except that they are rotated 180°.

Although there is plenty of behavioral evidence to suggest that matching across instances is a trivial task when faces are familiar (Burton, Jenkins, Hancock, & White, 2005; Burton, Kramer, Ritchie, & Jenkins, 2016; Hancock, Bruce, & Burton, 2000), this matching ability has not been tested when faces are inverted. If the face inversion effect holds for recognition judgements made across instances, then we predict performance will decrease when face images are upside down while the background scene remains upright (i.e., the face inversion effect; Yin, 1969). In this article, we will also extend what is known about the face inversion effect by testing whether congruency between a face and the background scene can explain the face inversion effect; a quantitative account expects an interaction between face orientation and scene orientation whereby performance in the congruent conditions ("Faces Upright/Scene Upright" and "Face Inverted/Scene Inverted") is better than performance in the corresponding incongruent conditions ("Faces Upright/Scene Inverted" and "Face Inverted/Scene Upright"). Whereas a qualitative view would expect a main effect of Face Orientation in the absence of an interaction.

Results and Discussion

Participant Sensitivity

A 2×2 analysis of variance (ANOVA) with Face Orientation (faces upright vs. faces inverted) and Scene Orientation (scene upright vs. scene inverted) both included in the
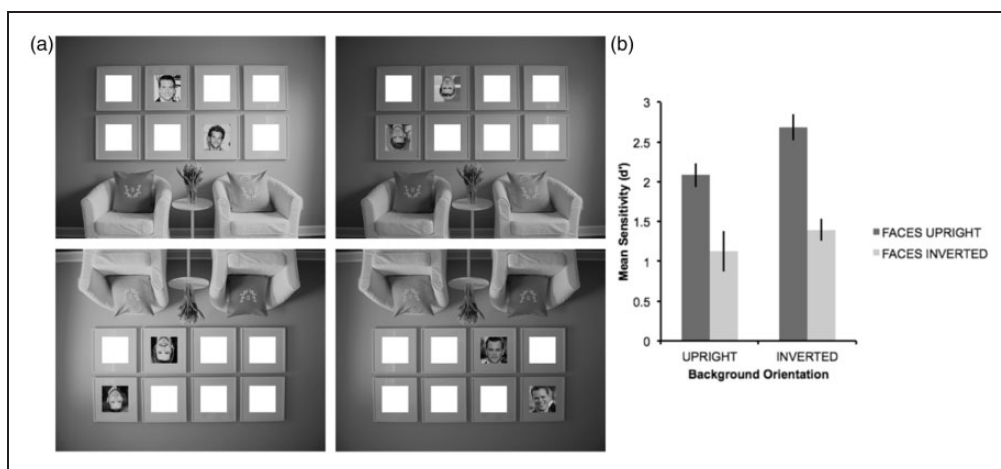
Figure 2. Experimental stimuli and main results: (a) Examples of the four unique conditions (all same trials) and (b) average subject sensitivity scores across the four unique conditions.

model as repeat factors. This analysis yielded the expected main effect of Face Orientation ($F(1, 13) = 47.04$, $p < .001$, $\eta_p^2 = 0.78$; see Figure 2b) which was further investigated with a set of a priori pairwise contrasts designed to test for evidence of the predicted local inversion effect (i.e., sensitivity in upright face trials > sensitivity in inverted face trials) in both Scene Orientation conditions (corrected p values < .01; under the Bonferroni rule $\alpha = 0.05/2$).

The ANOVA analysis also revealed a main effect of Scene Orientation ($F(1, 13) = 9.61$, $p = .008$, $\eta_p^2 = 0.42$; see Figure 2b) indicating that the participants were able to complete the matching task with greater accuracy when the scene was inverted. However, there was no evidence of an interaction between Face Orientation and Scene Orientation ($F(1, 13) = 0.85$, $p = .37$, $\eta_p^2 = 0.06$) implying that turning the scene upside down had the same effect on upright face matching and inverted face matching. Nonetheless, two a priori contrasts were run to determine whether turning the scene upside down improved sensitivity for upright faces ($p = .002$, $\alpha = 0.05/2$) but not inverted faces ($p = .352$, $\alpha = 0.05/2$). The results of these follow-up tests suggest that inverting the orientation of the scene increased sensitivity to matching identity pairs when faces were upright but not when faces were inverted.

Inverse Efficiency Scores

Figure 3 illustrates each observer's performance (Average Correct Reaction Time × Average Percent Correct) as a function of (a) condition and (b) block order. There is some indication in Figure 3(a) that the participants were responding faster in some conditions than others. To take into account any speed accuracy trade-off, we computed inverse efficiency (IE) scores for each observer, by dividing the average response time by the corresponding accuracy within each condition (Townsend & Ashby, 1983). A 2×2 repeated measures ANOVA (Face Orientation × Scene Orientation) on the IE data yielded the same pattern of overall results as in the analysis of the d' data (*main effect of Face Orientation, $F(1, 13) = 44.81$, $p < .001$, $\eta_p^2 = 0.77$; main effect of Scene Orientation, $F(1, 13) = 5.25$, $p = .04$, $\eta_p^2 = 0.29$; Face Orientation × Scene Orientation, $F(1, 13) = 2.05$, $p = .18$, $\eta_p^2 = 0.14$*). Meanwhile, the follow-up contrasts revealed new information; while there was evidence that the two local inversion effects on faces were still present in both the upright scene and inverted scene conditions
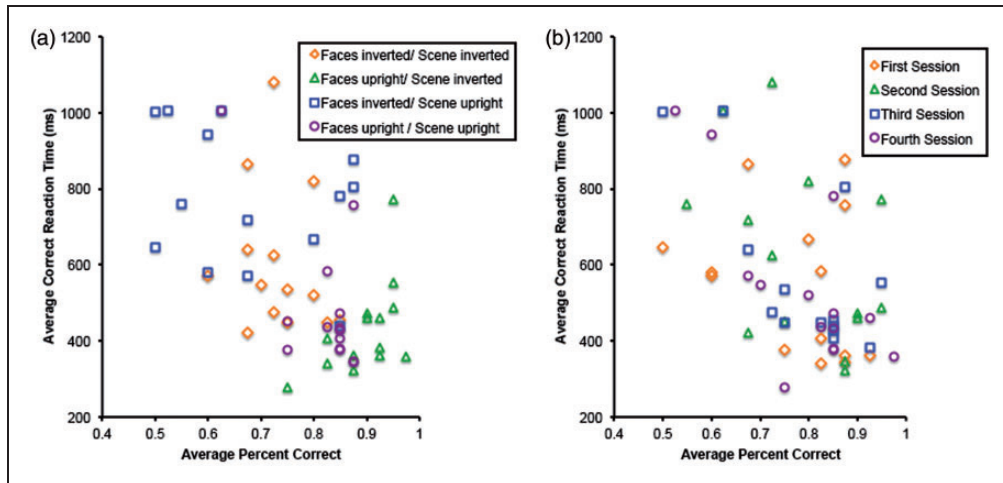
Figure 3. Individual participant data in each of the four unique conditions (average percentage correct as a function of average correct reaction time). (a) Data presented as a function of the four conditions. The distribution of individual data in this plot suggests participants may have been trading speed for accuracy in the "faces upright/scene inverted" condition compared with the other conditions. (b) Condition order was counterbalanced across participants; here, the data have been sorted by the "block order" rather than "condition" to check for effects of repetition (i.e., practice).

(p values $< .001$; $\alpha = 0.05/2$), there was an initial indication that, when faces were presented upside down, inverting the scene improved performance (i.e., the IE scores, on average, were smaller; $p = .04$; NB not significant after Bonferroni correction) whereas there was no evidence that when faces were upright, scene orientation made any difference to performance ($p = .16$).

Participant IE scores were also used to evaluate the effect (if any) of block order on performance (see Figure 3(b)). A one-way ANOVA was performed on Block Order, a repeated measures factor with four levels (first, second, third, and fourth). The resulting test statistic was not significant ($F(39) = 0.315$, $p > .8$, $\eta_p^2 = 0.24$) ruling out the order of the conditions as a contributing factor to participant performance.

General Discussion

We report a robust face inversion effect indicating that a simultaneous same different task is performed both better and faster when the photographs are upright compared with when photographs are inverted. By asking participants to match identity across two separate instances, these data extend what is known about within-person variation (Burton et al., 2005); not only does inversion reduce our sensitivity to differences between novel individuals in a simultaneous same or different task (Le Grand et al., 2001; see a review McKone & Yovel, 2009), but our data indicate that it also reduces our sensitivity to familiar individuals when you use the same paradigm. We did not need to train our participants with static images of novel identities before collecting data, nor did we rely on their recognition memory and naming skills. Instead, we increased ecological validity by relying on robust representations of highly familiar faces (in this case, celebrities). Moreover, by using a number of different photographs to represent each identity, we have eliminated any concern that participants were able to determine matching pairs based on image properties, rather than faces per se.

Therefore, this method of using multiple photographs of highly familiar faces may prove useful in the future because it avoids both the problems normally associated with simultaneous matching tasks and those associated traditional recognition memory paradigms (for recent reviews, see McKone & Yovel, 2009; Rossion, 2009).

Importantly, the main effect of Face Orientation is largely consistent with the qualitative account of face inversion effect (Le Grand et al., 2001; McKone & Yovel, 2009; Rhodes et al., 1993; Rossion, 2009; Rossion & Boremanse, 2008), suggesting that upright faces are processed as such irrespective of the background scene orientation. However, when we took correct reaction time into account, the IE data suggested that congruency with the scene had a different impact on upright faces than it did on inverted faces; congruency facilitated performance when faces were upside down. Participants performed better in the incongruent condition (Faces Inverted/Scene Inverted) compared with the congruent condition (Faces Upright/Scene Inverted). Meanwhile, this effect of congruency was not present at all when faces were presented upright. One explanation for this result is that the mechanism responsible for processing upright faces is largely insensitive to the orientation of the scene. In contrast, the mechanism responsible for processing inverted faces seems to benefit from orienting faces in the same direction as the background scene. Once again, this is largely consistent with the qualitative account of the face inversion effect where different mechanisms operate on upright and inverted stimuli.

Turning back to the initial analysis of participant sensitivity, an outstanding question is why we, counter to expectation, found that participants were more sensitive to upright faces when the scene was inverted compared with when the scene was upright. This is also evident in Figure 1 where we use Mooney faces in a back and white scene. We think it is tempting to assert that this might be an example of efficient figure/ground segmentation resulting from the incongruency between a salient stimulus, the upright face, and an unusual, inverted, background scene. In the past, studies of face perception have presented participants with isolated faces, floating on medium gray backgrounds. This has given researchers greater control of the visual input but at the cost of ecological validity. These data, and the demonstration in Figure 1, collectively represent an attempt to understand how faces are processed when they appear in visual scenes (i.e., their natural environment). The implication is that under normal circumstances processing visual information in an upright scene may distract from, or interact with, the processes that underlie face perception. Future research could extend this line of research by testing whether the face inversion effect would survive if a participant's entire field of view was turned upside down in a virtual environment or if the participant themselves were upside down.

Method

Observers

We collected data from 14 undergraduate students enrolled in third-year Psychology at the University of Sydney (NSW, Australia). All aspects of the data collection and analysis were approved by the Human Research Ethics Committee of The University of Sydney (Project No. 2015/336).

Stimuli

The scene was an image downloaded from an online furniture catalogue that was 1024×768 pixels in size (see Figure 2a). This image of furniture against a wall with eight equally spaced and equally sized photo frames hanging above a couch and equally in two rows of four was

converted to gray-scale, and square apertures were cut into the background scene, at locations corresponding to the photographs, that were all 100×100 pixels in size. On any given trial, two of the apertures were filled with two face images (exactly 100×100 pixels in size). These face images were positioned so that one was always on the top row, while the other was on the bottom row—also the two images always occupied adjacent columns.

Face images were drawn at random from an image set comprising 10 different famous males (eight different photographs per male $= 80$ images in total). These 10 identities were confirmed as familiar to each participant before the experiment began, and participants often happily commented about seeing these identities throughout each of the blocks. After these images were downloaded from the public domain, they were also converted to gray scale and resized where necessary to 100×100 pixels. Images were not rotated or aligned in any way; however, mean luminance was equated across the set (see Figure 2a).

Procedure

All observers completed the experiment in a dimly lit curtained booth. Both experiments were programmed in MATLAB version R2010a using the Psychophysics Toolbox 3. The program was run on an Apple Mac Pro running Mac OSX Lion 10.7.5. Subjects were seated approximately 57 cm in front of a CRT monitor (18 in. viewable screen size) set at a screen resolution of 1024×768 pixels with a refresh rate of 100 Hz. Participant responses were recorded using an Apple wired USB keyboard.

A trial began with the presentation of a central white fixation cross that was superimposed on a uniform gray background for 2 s (± 400 ms). It was replaced with a stimulus array that filled the entire screen for 700 ms. Our participants were instructed to the task using the left or right arrow keys (_left arrow key_ = the two photographs were taken of different people; _right arrow key_ = the two photographs were taken of the same person). As soon as a response was recorded, the next trial began immediately.

Each stimulus array comprised three components: the background scene (always the same image, although its orientation depended on the condition) and two face images. In any given condition, each of the 10 celebrity identities appeared in two same trials and four different trials (each of the eight photographs associated with a single identity were only repeated across conditions and never in the same condition). Thus, each observer completed 40 trials per condition (160 trials in total). When they finished, they were asked whether they knew each of the 10 celebrity names and whether they believed they saw that celebrity during the experiment.

References

Bartlett, J. C., & Searcy, J. (1993). Inversion and configuration of faces. _Cognitive Psychology_, _25_, 281–316.

Burton, A. M., Jenkins, R., Hancock, P. J., & White, D. (2005). Robust representations for face recognition: The power of averages. *Cognitive Psychology*, *51*, 256–284.

Burton, A. M., Kramer, R. S., Ritchie, K. L., & Jenkins, R. (2016). Identity from variation: Representations of faces derived from multiple instances. *Cognitive Science*, *40*, 202–223.

Busigny, T., Joubert, S., Felician, O., Ceccaldi, M., & Rossion, B. (2010). Holistic perception of the individual face is specific and necessary: Evidence from an extensive case study of acquired prosopagnosia. *Neuropsychologia*, *48*, 4057–4092.

Collishaw, S. M., & Hole, G. J. (2002). Is there a linear or a nonlinear relationship between rotation and configural processing of faces? *Perception*, *31*, 287–296.

Dakin, S. C., & Watt, R. J. (2009). Biological ''bar codes'' in human faces. *Journal of Visualization*, *9*, 2.1–10.

de Gelder, B., & Rouw, R. (2000). Configural face processes in acquired and developmental prosopagnosia: Evidence for two separate face systems? *Neuroreport*, *11*, 3145–3150.

Farah, M. J., Wilson, K. D., Drain, H. M., & Tanaka, J. R. (1995). The inverted face inversion effect in prosopagnosia: Evidence for mandatory, face-specific perceptual mechanisms. *Vision Research*, *35*, 2089–2093.

Goffaux, V., & Rossion, B. (2006). Faces are ''spatial"–Holistic face perception is supported by low spatial frequencies. *Journal of Experimental Psychology:Human Perception and Performance*, *32*, 1023–1039.

Hancock, P. J., Bruce, V. V., & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends Cognitive Science*, *4*, 330–337.

Hochberg, J., & Galper, R. E. (1967). Recognition of faces: I. An exploratory study. *Psychonomic Science*, *9*, 619–620.

Kanwisher, N., Tong, F., & Nakayama, K. (1998). The effect of face inversion on the human fusiform face area. *Cognition*, *68*, B1–11.

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2001). Neuroperception. Early visual experience and face processing. *Nature*, *410*, 890.

McKone, E., & Yovel, G. (2009). Why does picture-plane inversion sometimes dissociate perception of features and spacing in faces, and sometimes not? Toward a new theory of holistic processing. *Psychonomic Bulletin Review*, *16*, 778–797.

Morton, J., & Johnson, M. H. (1991). CONSPEC and CONLERN: A two-process theory of infant face recognition. *Psychology Review*, *98*, 164–181.

Rhodes, G., Brake, S., & Atkinson, A. P. (1993). What's lost in inverted faces? *Cognition*, *47*, 25–57.

Riesenhuber, M., Jarudi, I., Gilad, S., & Sinha, P. (2004). Face processing in humans is compatible with a simple shape-based model of vision. *Proceedings of the Royal Society B: Biological Sciences*, *271*, S448–S450.

Riesenhuber, M., & Wolff, B. S. (2009). Task effects, performance levels, features, configurations, and holistic face processing: A reply to Rossion. *Acta Psychologica (Amst)*, *132*, 286–292.

Righart, R., & de Gelder, B. (2007). Impaired face and body perception in developmental prosopagnosia. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 17234–17238.

Rossion, B. (2009). Distinguishing the cause and consequence of face inversion: The perceptual field hypothesis. *Acta Psychologica (Amst)*, *132*, 300–312.

Rossion, B., & Boremanse, A. (2008). Nonlinear relationship between holistic processing of individual faces and picture-plane rotation: Evidence from the face composite illusion. *Journal of Visualization*, *8*, 3.1–13.

Rossion, B., Prieto, E. A., Boremanse, A., Kuefner, D., & Van Belle, G. (2012). A steady-state visual evoked potential approach to individual face perception: Effect of inversion, contrast-reversal and temporal dynamics. *Neuroimage*, *63*, 1585–1600.

Schwaninger, A., & Mast, F. W. (2005). The face-inversion effect can be explained by the capacity limitations of an orientation normalization mechanism1. *Japanese Psychological Research*, *47*, 216–222.

Taubert, J., & Alais, D. (2009). The composite illusion requires composite face stimuli to be biologically plausible. *Vision Research*, *49*, 1877–1885.

Taubert, J., & Parr, L. A. (2012). The perception of two-tone Mooney faces in chimpanzees (Pan troglodytes). *Cognitive Neuroscience*, *3*, 21–28.

Taubert, J., Van Belle, G., Vanduffel, W., Rossion, B., & Vogels, R. (2015). The effect of face inversion for neurons inside and outside fMRI-defined face-selective cortical regions. *Journal of Neurophysiology*, *113*, 1644–1655.

Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes*. Cambridge, MA: Cambridge University Press.

Tsao, D. Y., & Livingstone, M. S. (2008). Mechanisms of face perception. *Annual Review of Neuroscience*, *31*, 411–437.

Valentine, T., & Bruce, V. (1988). Mental rotation of faces. *Memory and Cognition*, *16*, 556–566.

Xu, X., Biederman, I., & Shah, M. P. (2014). A neurocomputational account of the face configural effect. *Journal of Visualization*, *14*, 9.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*, 141–145.

Yovel, G., & Kanwisher, N. (2005). The neural basis of the behavioral face-inversion effect. *Current Biology*, *15*, 2256–2262.