UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Beyond the retina: Evidence for a face inversion effect in the environmental frame of reference

Permalink
https://escholarship.org/uc/item/170723j0

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 33(33)

ISSN
1069-7977

Authors
Davidenko, Nicolas
Flusberg, Stephen

Publication Date
2011

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Beyond the retina:
Evidence for a face inversion effect in the environmental frame of reference

Nicolas Davidenko (ndaviden@stanford.edu)
Stephen J. Flusberg (sflus@stanford.edu)

Stanford University, Department of Psychology
Jordan Hall, 450 Serra Mall, Building 420, Stanford, CA 94305 USA

Abstract

Across a wide range of face perception tasks, observers show drastically worse performance when faces are oriented upside-down versus upright. However, the meaning of *orientation* must be established in relation to a particular frame of reference. In relation to which reference frame(s) does the face inversion effect occur? Here we describe a simple, novel method for investigating potentially independent effects of retinal and environmental reference frames on face processing. Participants performed one of two face-processing tasks (emotional expression classification and recognition memory) as they lay horizontally, which served to disassociate the retinal and environmental reference frames. In both experiments we found a large effect of retinal orientation on performance and a small but reliable effect of environmental orientation. In a follow-up control study, we consider an alternative explanation based on our experimental setup. We argue that environmental orientation influences face processing, which is revealed when retinal orientation is kept constant.

Keywords: face perception, reference frames, face inversion effect, embodiment

Background

Over 50 years of research has demonstrated that orientation dramatically affects the visual processing of faces; across a wide range of perception and memory tasks, observers show markedly worse performance when faces are presented upside-down compared to upright (Yin, 1969). This *face inversion effect* is perhaps most famously illustrated by the classic Thatcher Illusion (Thompson, 1980); a picture of the former prime minister that appears normal when presented upside-down is revealed to be a disturbing grotesque of inverted features when rotated upright. The dramatic effect of inversion on face perception sets faces apart from other objects and has led many researchers to consider faces a special visual category (Rhodes et al., 1993; Farah et al., 1998).

Notice, though, that the meaning of orientation (i.e. what counts as upright or upside-down) must be established in relation to a particular frame of reference. The vast majority of experiments examining orientation effects in visual perception have participants seated in front of a computer screen. A face image displayed on the monitor might be upright with respect to the participant's retina (*retinal* frame), but it would also be upright with respect to the computer screen itself, the room the computer is situated in, and even the directional pull of gravity (*environmental* frames). In a typical laboratory study, therefore, several reference frames are conflated. This begs the question: in relation to which reference frame(s) does the face inversion effect occur?

There is reason to believe that both retinal and environmental reference frames matter in visual information processing. Irvin Rock (1973) conducted a series of pioneering experiments on object form perception, demonstrating that both retinal and environmental reference frames impact participants' memory for and interpretation of novel objects under certain circumstances. More recently, researchers have shown that our ability to perceive the stability of human body postures as well as the direction of bodily motion depends on both retinal and environmental cues (Chang, Harris & Troje, 2010; Lopez et al., 2009). Finally, research on spatial cognition has found that people are sensitive to a variety of spatial reference frames and flexibly adopt multiple different frames for representing the environment as they perform different tasks and communicate with others (Tversky, Lee, & Mainwaring, 1999).

However, to date there is no evidence that multiple frames of reference play a role in face processing. In fact, Rock (1973) himself suggested that because faces are familiar objects that have an intrinsic spatial structure (i.e. a top and bottom), their environmental orientation should not affect how we perceive them. Recent research seems to support this view that face orientation effects are restricted to the retinal frame of reference (Troje, 2003; Chang et al., 2010). In one experiment (Troje, 2003), participants had to indicate whether or not a face image displayed on a computer screen was the same face they had seen moments before. The images could be oriented upright on the screen or rotated by 90º. The participant's head was orientated upright in the room for half of the trials and rotated by 90º in the other half of the trials. The results indicated that performance was fastest and most accurate when the orientation of the face image matched the orientation of the participant's retinal frame, regardless of the environmental orientation of the image and the participant.

Troje's (2003) experiment (and related work; see Rock, 1973) was designed to pit the retinal frame *against* the environmental frame to see which one mattered more for face processing. However, it is possible that *both* reference frames affect face processing, but by pitting the two against one another the relatively large retinal frame effect masked

the ability to detect any effect of the environmental frame. Here we describe a novel, simple method for investigating potentially *independent* effects of retinal and environmental reference frames in face perception. Participants performed one of two face-processing tasks as they lay horizontally, thereby disassociating the retinal and environmental orientation of the stimuli (see figure 1). In this position, faces presented upright and upside-down in the retinal frame are both rotated by 90º in the environmental frame, while faces presented upright and upside-down in the environmental frame are both rotated by 90º in the retinal frame. This allows us to measure face inversion effects in each reference frame while keeping the orientation in the alternative frame constant.

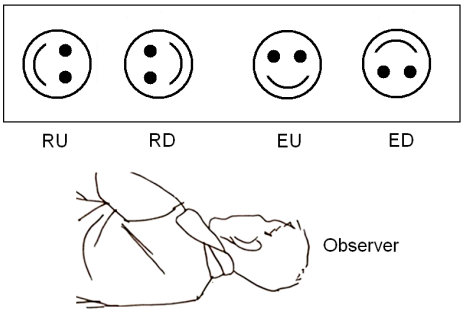


Figure 1: Schematic diagram of the possible image orientations when the observer is lying on his right side. Face images that are retinally up (RU) and retinally down (RD) are both rotated by 90º in the environmental frame, while images that are environmentally up (EU) and environmentally down (ED) are both rotated by 90º in the retinal frame.

Experiment 1

In relation to which reference frame(s) does the face inversion effect occur? In Experiment 1, we investigated this question by having participants classify the emotional expression of Mooney faces (Mooney & Ferguson, 1951) while they lay on their sides. Mooney faces are two-toned images used in many experimental studies of face processing because they are difficult to perceive when upside-down and elicit a notoriously large inversion effect (see Figure 3).

Methods

Participants. 56 individuals from the Stanford community were recruited to participate in this study in exchange for payment or class credit.

Stimuli. Mooney faces were generated by blurring gray-scale photographs and reducing them to two tones (see Figure 3A). We selected 48 faces that could be easily identified as happy (16), sad (16), or angry (16) when upright. We then mirror-reversed each of these images to create two sets of 48 faces for a total of 96 face images.

Procedure. The classification task was programmed in Matlab using Psychophysics Toolbox 3. Data was collected on a 15" Macbook Pro. The background of the display was black, and on each trial a single face image was presented at the center of the display, preceded by a 100ms inter-trial interval (ITI). Participants were asked to judge the emotional expression of each presented face as "happy," "sad," or "angry" as quickly and as accurately as possible by pressing a number on a keypad corresponding to each expression. Images remained on the screen until participants provided a response. The keypad was held in the left hand while three digits of the right hand were used to make the response. There was no fixation cross and participants were free to move their eyes to inspect the stimuli before responding. Participants first completed several practice trials while seated upright at a desk in order to familiarize them with the task.

They were then randomly assigned to lie down on their right or left side on a padded bench to begin the experimental task. A pole-mounted head and chin rest with head-strap was constructed in-house to maintain participants' heads fixed horizontally in the room as they completed the task (see figure 2). The computer was placed on a flat horizontal surface next to the bench and the screen was positioned approximately 33 cm from the face of the participant. At this distance, the face stimuli subtended approximately 5 º by 7º of visual angle. The experimental room was brightly lit.

We presented participants with Mooney faces in random order with each face in one of 4 possible orientations – up, down, right, and left on the screen (see Figure 1). After completing a block of 48 trials while lying on one side, the participant switched to lying on their other side and completed another block of 48 trials. The first block consisted of the 48 original Mooney images while the second block consisted of the 48 mirror-reversed images, each appearing in a randomly selected orientation. No feedback was given to participants.



Figure 2: Pole-mounted horizontal chin-rest used in all experiments. Response pad used in Experiment 1.

Results

Data from 6 of 56 participants were excluded in analysis because they either failed to perform above chance levels (n=1) or their reaction times were more than 3 standard deviations above the group median (n=5). There were 8 distinct types of trials (2 body positions X 4 image orientations). Because there was neither a main effect of body position ($F(1,49)= 1.40$, $p>0.2$) nor an interaction

between body position and image orientation ($F(1,49) = 1.59$, $p>0.2$), we collapsed each participant's data across the two body positions and refer to the 4 image orientation conditions as "retinally up" (RU), "retinally down" (RD), "environmentally up" (EU) and "environmentally down" (ED). A retinal face inversion effect would manifest as better performance in RU versus RD trials. An environmental face inversion effect would manifest as better performance in EU trials as compared to ED trials.

Performance in all image orientations was significantly above 33% chance ($ts(49)>16$, $p<10^{-20}$). As expected, we observed a large retinal inversion effect, with more accurate performance (0.94 vs. 0.70) and faster reaction times on correct trials (917ms vs. 1160ms) for RU versus RD faces ($t(49)= 15.6$, $p<10^{-19}$ for accuracy; $t(49)=-9.3$, $p<10^{-11}$ for reaction time; see Figure 3B,C). Intriguingly, we also found a reliable effect of environmental orientation. Responses were consistently more accurate (0.843 vs. 0.796) and faster (1039ms vs. 1084ms) for EU vs. ED faces ($t(49)= -2.4$, $p=0.02$ for both measures).

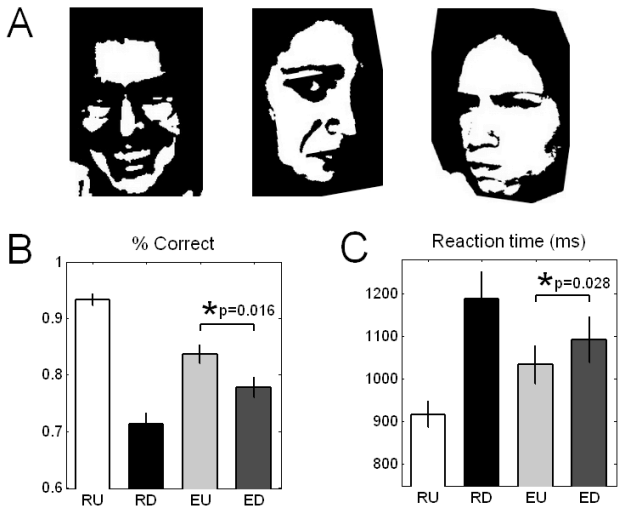


Figure 3: Sample stimuli and performance on the emotional expression classification task in Experiment 1 for retinally up (RU), down (RD), environmentally up (EU) and down (ED) faces. Error bars denote between-subjects SEM.

Discussion

The results of Experiment 1 support previous work demonstrating a large face inversion effect in the retinal frame of reference. However, the results also support the existence of a novel *environmental inversion effect* in the perception of faces. Despite the fact that faces in the EU and ED orientations were both rotated by 90º in the retinal frame, there was a small but significant advantage for classifying the faces' expressions when they were EU (upright in the environmental reference frame) versus ED (upside-down in the environmental reference frame). It is possible that earlier research on this topic (e.g. Troje, 2003) found no evidence of an environmental inversion effect in faces because pitting the two reference frames against one another causes the large retinal effect to mask the environmental effect.

It is also possible that the environmental inversion effect observed in Experiment 1 may be a result of the specific task we used: a difficult, emotional expression classification task of degraded face stimuli that recruits online face processing mechanisms. The literature on face processing has established inversion effects in a wide range of perception and memory tasks. Can environmental orientation also influence how we store faces in memory, or are these effects limited to specific online perceptual tasks?

Experiment 2

The results of Experiment 1 demonstrated that both retinal and environmental reference frames affect online processing in an emotional expression classification task of Mooney faces. In Experiment 2, we asked if these effects would extend to a very different type of task with a very different set of face stimuli: recognition memory for gray-scale face images.

Methods

Participants 26 individuals from the Stanford community were recruited to participate in this study in exchange for payment or class credit.

Stimuli & Procedure 192 front-view, gray-scale photographs of Caucasian males were selected from the FERET database (Phillips et al., 1998). Stimuli were cropped using an oval shape to remove hair and clothing around each face and normalized for size, brightness, and contrast (see Figure 4A).

We used the same apparatus and setup as in Experiment 1. The experiment was programmed in Matlab using the Psychophysics Toolbox. Participants completed study/test blocks while lying on either side, for a total of 8 study/test blocks (2 body positions X 4 image orientations).

During each study block, a sequence of 12 different faces was presented 4 times (each sequence in a new random order). All faces in each study block were displayed at one given orientation (all up, all down, all left, or all right on the screen). Each face image remained on the screen for 900ms and there was a 100ms ISI between faces. The background of the display was black.

A test block immediately followed each study block. A face was presented centrally in the same orientation as the faces in the preceding study block. Participants held a computer mouse in their right hand and indicated whether the face on the screen was one they had just studied ("old"; left mouse click) or one they had never seen before ("new"; right mouse click). Each test block thus consisted of 24 trials: the 12 old faces interspersed with 12 new faces. Participants completed 4 study/test blocks while lying in one body position (one block at each orientation), and another 4 study/test blocks while lying in the other position. The order of positions and blocks was randomized across participants.

Results

Data from 1 participant was excluded in analysis because she failed to perform above chance level on the recognition task. As in Experiment 1, there was neither a main effect of body position ($F(1,24)=0.35$, $p>0.5$) nor an interaction between body position and image orientation ($F(1,24)=1.34$, $p>0.1$). We therefore collapsed each participant's data across the two body sides and refer to 4 image orientations as "retinally up" (RU), "retinally down" (RD), "environmentally up" (EU) and "environmentally down" (ED).

We measured participants' ability to discriminate old from new faces (d') as well as their reaction time on correct trials. As expected, we found a large retinal inversion effect, with better discrimination (d' = 1.47 vs. 0.64) and faster reaction times (976ms vs. 1112ms) for RU compared to RD faces ($t(24)=5.44$, $p<0.0001$ for d', and $t(24)=-3.49$, $p=0.002$ for reaction time). In addition, we found a reliable environmental inversion effect in recognition memory of faces: EU faces were recognized better than ED faces (d' = 1.00 vs. 0.76; $t(24)=3.0$, $p<0.007$). This effect could not be attributed to a speed-accuracy trade-off; in fact, reaction time on correct trials was (non-significantly) faster for EU (1052ms) vs. ED (1096ms) faces ($p=0.3$; see Figure 4B,C).

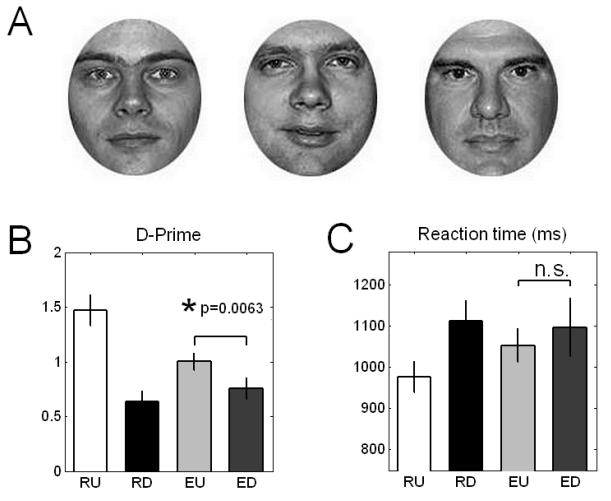


Figure 5: Sample stimuli and performance on the recognition memory task in Experiment 2 for retinally up (RU), down (RD), environmentally up (EU) and down (ED) faces. Error bars denote between-subjects SEM.

Discussion

The results of Experiment 2 suggest that environmental orientation can influence recognition memory for faces when retinal orientation is held constant. Participants had better recognition memory performance when faces were presented environmentally upright compared to environmentally upside-down. Together with Experiment 1, these results suggest that face inversion effects in the environmental frame of reference generalize across very different tasks utilizing very different face stimuli.

However, this interpretation of the results depends on the reliability of our experimental apparatus and design. In our experiments, participants were lying down horizontally in order to disassociate the retinal and environmental frames of reference, but this disassociation is only valid if both EU and ED images were precisely at 90º in the retinal frame. If participants' heads or eyes were slightly rotated towards environmentally up, this asymmetry may have contributed to the effects we have found.

In fact, there are at least two reasons an asymmetry might exist in our experimental setup. First, physiologists have identified a phenomenon known as ocular counter-roll (OCR; see Sares et al., 2007), in which people's eyes rotate slightly in the opposite direction of their head tilt. For example, when a person tilts their head clockwise, the eyes respond by exerting a small counter-roll of several degrees counter-clockwise. Second, while participants were strapped into our horizontally leveled head and chin rest, they were still able to shift their heads a few degrees and may have unwittingly tilted their heads when performing the task. Could these factors have contributed to better performance for EU versus ED faces? To address this possibility, in Experiment 3 we measured the exact position of participants' eyes as they sat upright or lay in our experimental apparatus, and constructed stimuli to counteract any resulting asymmetries in retinal orientation.

Experiment 3

Correcting for asymmetries in retinal orientation. In 13 separate participants, we measured the eyes' orientation when they lay in our experimental apparatus. We used a leveled high-resolution digital SLR camera to photograph participants' irises while they were sitting upright and lying horizontally. Using Photoshop, two independent coders measured the angular disparity between the two pictures (i.e. how many degrees the sitting-up picture needed to be rotated so that it would be aligned with the lying-horizontally picture). These measurements produced an average OCR of approximately 4.2º (SD = 1.8º) in the direction opposite of head tilt. In other words, when subjects lay on their right, their eyes were rotated left by an average of 4.2º.

Although this was a relatively small disparity consistent with previously published measurements (Sares et al., 2007), the bias results in EU faces being on average more aligned with subjects' retinal frame than ED faces. Specifically, when participants lay horizontally, EU faces were rotated away from retinal upright by an average of 85.8º, whereas ED faces were rotated by an average of 94.2º. This asymmetry could potentially drive the differences in performance and reaction time observed in Experiments 1 and 2. To correct for this asymmetry, in Experiment 3 we rotated the face images by 5º in the direction of OCR observed in our sample. This over-correction would ensure that any observed advantage in processing EU faces could not be attributed to EU faces

being more aligned than ED faces with participants' retinal frame.

Methods

Participants. 43 individuals from the Stanford community were recruited to participate in this study in exchange for payment or class credit. 39 of these completed Experiment 3a (an abbreviated version of Experiment 1) and all 43 completed Experiment 3b (an abbreviated version of Experiment 2).

Stimuli & Procedure. Experiments 3a and 3b were identical to Experiments 1 and 2, except (1) faces appeared in only EU and ED orientations (that is, we did not include RU or RD trials), and images were rotated by 5º in the direction opposite the participants' body position, to correct for the asymmetry described above. Participants first lay on one randomly selected side and performed two study/test blocks of the recognition memory task. They then switched sides and performed two more such blocks. They then completed one block of 48 emotional expression ratings of Mooney faces, switched sides once more, and completed a final block of expression ratings.

Results

Data from 7 participants in Experiment 3a were excluded from analysis because their reaction times were more than 3 standard deviations above the median, and data from 6 participants in Experiment 3b were excluded because they failed to perform above chance level on the recognition task. Despite over-correcting for retinal asymmetry across EU and ED image orientations, we found significant or trending effects in performance and reaction time similar to those of Experiments 1 and 2. In Experiment 3a (expression classification of Mooney faces), proportion correct was 0.826 on EU vs. 0.809 on ED trials ($t(31)=1.18$; $p=0.12$, 1-tail). Reaction time was 958ms vs. 978ms, respectively ($t(31)= -1.43$, $p=0.08$, 1-tail; see Figure 6A).

In Experiment 3b, we found a significant advantage in recognition memory for EU versus ED faces (Figure 6B). The average d' across subjects was 1.01 for EU vs. 0.77 for ED faces ($t(36)=2.79$; $p=0.004$, 1-tail). Reaction time was also faster for EU (1108ms) vs. ED (1181ms) faces ($t(36)= -2.42$, $p=0.01$, 1-tail).

Discussion

The results of Experiment 3 suggest that OCR does not fully account for the results observed in Experiments 1 and 2. Even after over-correcting for a 4.2º average asymmetry in the retinal orientation of images, we again found reliable advantage in processing EU faces compared to ED faces, although these effects in the expression classification task were only marginally significant. We attribute the noisier results in Experiment 3a to the fact that we utilized a single correction for OCR across all participants even though our measurements found this varied considerably from person to person. In addition, rotating the image to correct for OCR results in EU faces that are not truly upright in the environmental frame, but rather rotated by 5º. The misalignment between our EU stimuli and the environmental frame may have further weakened the efficacy of our control experiment. In an ongoing study, we are examining a different method of correcting for OCR that involves adjusting participants' orientation (rather than the images) according to individual estimates of their OCR.

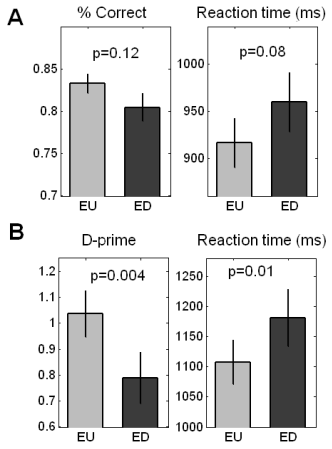


Figure 6: Results of Experiments 3a and b. A: percent correct and reaction time on the emotional expression classification task (Experiment 3a) for environmentally up (EU) and down (ED) faces. B: discrimination performance and reaction time on the recognition memory task (Experiment 3b). Error bars denote between-subjects SEM.

General Discussion

In relation to which reference frame(s) does the face inversion effect occur? We have presented several experiments suggesting that independent face inversion effects occur in both the retinal and environmental reference frames.

In Experiment 1, participants classified the emotional expression of Mooney faces; in Experiment 2, participants performed an old/new recognition task on novel face images. In both studies, we found a large effect of retinal orientation on performance and a reliable (though smaller) effect of environmental orientation. Specifically, participants in Experiment 1 were more accurate and faster at judging emotional expressions of environmentally upright versus environmentally upside-down faces, even though faces in both conditions were rotated by 90º in the retinal frame. Similarly, recognition performance in Experiment 2 was better for environmentally upright versus environmentally upside-down faces. In Experiment 3, we ruled out an alternative explanation of our findings based on the possibility that OCR or other experimental artifacts were causing environmentally upright face images to be slightly more aligned with participants' retinal frame. We conclude that there exists a reliable effect of environmental

orientation on face processing that is revealed when retinal orientation is held constant.

However, as noted in the introduction, there are actually many different environmental frames of reference, several of which were conflated in our present set of experiments. For example, a face image on a computer may be upright on the screen, upright in the room the computer is in, and upright with respect to the directional pull of gravity. In our setup, the room was brightly lit and participants had visual access to objects and other features of the experiment room, which provided cues to environmental orientation in addition to the pull of gravity. Previous work has suggested that it may be the gravitational orientation that matters most of all of the environmental cues (e.g. Chang et al., 2010), but this may or may not be the case with face processing. Future work will try to disentangle these different environmental reference frames in order to determine which one(s) play a role in face processing.

What mechanisms might underlie an environmental inversion effect in face processing? One possibility is that we learn through experience that faces tend to be upright in the world regardless of our own body's position when we observe them. For example, we may lie sideways when watching television, but nevertheless faces and other images on the screen remain upright in relation to the television, the room, and gravity. We propose that this contextual information present in our everyday experience affects our ability to process faces when we see their regular upright context (EU) versus an upside-down context (ED). Indeed, Rock's (1973) early work on novel shape perception suggests that how we experience and categorize objects during learning affects which reference frames matter later for recognition performance. Further research is currently underway in the lab to explore this possibility.

Acknowledgments

The authors would like to Jason Loftus for help with data collection and Harry Bahlman for engineering our experimental apparatus. We would like to thank Nathan Witthoft for bringing to our attention the potential effect of OCR. We would also like to thank Alexia Toskos Dils, Jon Winawer, Daniel Sternberg and Lera Boroditsky for helpful discussions throughout the design and execution of these studies. This work was supported by NRSA Award 018279 to ND.

References

Chang, D. H. F., Harris, L. R., & Troje, N. F. (2010). Frames of reference for biological motion and face perception. *Journal of Vision, 10* (Jun 28).

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? *Psychological Review, 105*(3), 482-498.

Lopez, C., Bachofner, C., Mercier, M., Blanke, O. (2009). Gravity and observer's body orientation influence the visual perception of human body postures. Journal of Vision. 9(5):1, 1-14

Mooney CM, Ferguson GA (1951) A new closure test. Can J Psychol 5:129–133.

Phillips, P. J., Wechsler, H., Huang, J., & Rauss, P. (1998). The FERET database and evaluation procedure for face recognition algorithms. *Image and Vision Computing Journal*. 16, 295–306.

Rock, I. (1973). *Orientation and form*. New York, New York. Academic Press Inc.

Rhodes, G., Brake, S., & Atkinson, A. P. (1993). What's lost in inverted faces? *Cognition, 47*(1), 25-57.

Sares, F., Granjon, L., Abdelrhani, B., & Boulinguez, P. (2007). Analyzing head roll and eye torsion by means of offline image processing. Behavior Research Methods. 39(3), 590-599.

Thompson, P.(1980) Margaret Thatcher: a new illusion. Perception 9 483-4.

Troje, N. F. (2003). Reference frames for orientation anisotropies in face recognition and biological-motion perception. Perception, 32(2), 201-210.

Tversky, B. Lee, P. U., and Mainwaring, S. (1999). Why speakers mix perspectives. *Journal of Spatial Cognition and Computation, 1*, 399-412.

Yin RK. (1969). Looking at upside-down faces. Journal of Experimental Psychology. 81:141–145.