Human Perception of Visual Realism for Photo and Computer-generated Face Images

Shaojing Fan and Rangding Wang*, Ningbo University
Tian-Tsong Ng and Cheston Y. -C. Tan, Institute for Infocomm Research
Jonathan S. Herberg, Institute of High Performance Computing; National University of Singapore
Bryan L. Koenig, Lindenwood University

Computer generated (CG) face images are common in video games, advertisements, and other media. CG faces vary in their degree of realism, a factor which impacts viewer reactions. Therefore, efficient control of visual realism of face images is important. Efficient control is enabled by a deep understanding of visual realism perception: the extent to which viewers judge an image as a real photograph rather than a computer-generated image. Across two experiments we explored the processes involved in visual realism perception of face images. In Experiment 1, participants made visual realism judgments on original face images, inverted face images, and images of faces that had the top and bottom halves misaligned. In Experiment 2, participants made visual realism judgments on original face images, scrambled faces, and images that showed different parts of faces. Our findings indicate that both holistic and piecemeal processing are involved in visual realism perception of faces, with holistic processing becoming more dominant when resolution is lower. Our results also suggest that shading information is more important than color for holistic processing, and inversion makes visual realism judgments harder for realistic images but not for unrealistic images. Furthermore, we found that eyes are the most influential face part for visual realism, and face context is critical for evaluating realism of face parts. To the best of our knowledge, this work is a first realism-centric study that attempted to bridge the human perception of visual realism on face images with general face perception tasks.

Categories and Subject Descriptors: I.3.7 [**Computer Graphics**]: Three-Dimensional Graphics and Realism—*Color, shading, shadowing, and texture*; I.4.1 [**Image Processing and Computer Vision**]: Digitization and Image Capture—*Reflectance*; J.4 [**Psychology**] Social band Behavioral Sciences

General Terms: Design, Experimentation

Additional Key Words and Phrases: Visual realism, Human perception, Holistic face perception, Piecemeal face perception

ACM Reference Format:

S. Fan et al. 2014. Human Perception of Visual Realism for Photo and Computer-generated Face Images. ACM Trans. Appl. Percept. 2, 3, Article 1 (January YY), 20 pages.
DOI = 10.1145/0000000.0000000 http://doi.acm.org/10.1145/0000000.0000000

* Corresponding author. This work was supported in part by the National Natural Science Foundation of China #61301247, Ningbo Natural Science Foundation of China #2013A610059, Open Fund of Zhejiang Provincial Top key Discipline of Information and Communication Engineering #XKXL1313.
Author's address: S. Fan, R. Wang, College of Information and Computer Science, Ningbo University, Ningbo, China. 315000; email: fanshaojing@gmail.com;wangrangding@nbu.edu.cn; T. Ng, C. Tan, 1 Fusionopolis Way, 21-01 Connexis, Singapore 138632. email: {ttng, cheston-tan}@i2r.a-star.edu.sg; J. Herberg, 1 Fusionopolis Way, 16-16 Connexis North, Singapore 138632. email: herbergjs@ihpc.a-star.edu.sg; B. Koenig, Lindenwood University, 209 S Kingshighway St St Charles, MO 63301. email: bkoenig@lindenwood.edu.

DOI 10.1145/0000000.0000000 http://doi.acm.org/10.1145/0000000.0000000

1. INTRODUCTION

Virtual characters have become common in digital media. The most important part of virtual characters – their faces – receives much emphasis from CG artists. Proper assessment and control of the realism of CG faces is important because people's reactions when observing CG faces depends on their visual realism. For example, images that are close to but not quite realistic may cause discomfort for human observers [Mori 1970]. Surprisingly, little research has focused on the underlying processes for realism perception of CG faces.

In this paper, we aim to gain insight into how humans perceive the visual realism of face images. Specifically, we investigated the roles of holistic versus piecemeal processing in visual realism perception for static face images. "Holistic" is a loosely defined concept in face perception, with various definitions [Richler et al. 2012; Piepers and Robbins 2012]. In the current context, "holistic" refers to "the emergent features of a face that only become apparent when two or more of its basic features (e.g., the eyes, nose, or mouth) are processed at the same time" [Piepers and Robbins 2012], whereas "piecemeal" represents the processing of separate face components. We thus use the term "holistic" differently from [Tanaka and Farah 1993], who use the term to mean that the image is processed without any deconstruction into smaller units. Based on prior research in holistic face perception [Moscovitch et al. 1997; Farah et al. 1998b; Wang et al. 2012; Laguesse and Rossion 2013; Rossion 2013], one possibility is that perception of visual realism of faces relies primarily on holistic perceptual processing. However, previous studies also show the importance of piecemeal processing on separate face parts [Schwaninger et al. 2003a; Amishav and Kimchi 2010]. A recent study by [Gold et al. 2012] suggested the human ability to process face features viewed together is no better than their ability to use each individual face part when viewed in isolation. Thus another possibility is that visual realism of face images might rely heavily on piecemeal processing. That is, people might notice telltale signs that an image is CG when looking at parts of faces, which could have local artifacts from the rendering process.

We conducted two experiments to test the two opposing hypotheses. In Experiment 1 we tested whether holism effects exist in realism perception of face images, and whether the factors influencing image realism interact with holistic or piecemeal processing. In Experiment 2 we directly explored the influence of separate face parts on realism perception, as well as underlying perception processes.

2. RELATED WORK

Our study builds on the previous research of multiple disciplines in computer graphics, multimedia forensics, cognitive psychology, and neuroscience.

Computer Graphics: Since the early 1980's, the CG community has been interested in how true CG images are to reality, *i.e.*, their CG fidelity [Meyer et al. 1986; Rademacher et al. 2001; McNamara 2005; Ramanarayanan et al. 2007]. CG fidelity is similar to visual realism, but whereas CG fidelity evaluates perceived similarity of CG images to reality, visual realism focuses on what characteristics observers use to detect whether an image is a photo or CG. Both topics involve identifying visual cues that influence how realistic CG images appear and developing better CG rendering algorithms. In CG fidelity research, a common approach is to render a CG scene that is visually identical to reality, then to adjust some rendering parameters, such as illumination, shadow, and surface roughness, and have people judge whether the scene was real or CG. Research under this approach has suggested that image properties important for CG fidelity include illumination, shadow, and surface roughness [Meyer et al. 1986; Rademacher et al. 2001; McNamara 2005].

Difficulty producing photorealistic human faces is exacerbated by the *uncanny valley* – the finding that as something looks more human-like it also looks more agreeable, until it looks nearly-but-not-quite-human and its nonhuman imperfections are unsettling [Mori 1970; MacDorman and Ishiguro

2006]. A series of experiments have explored such phenomena using androids, mechanical-looking robots, and images of human faces [MacDorman and Ishiguro 2006; MacDorman et al. 2009]. Generally, more human-like CG faces have more details and realistic textures, and when stylized (and therefore less realistic) observers tolerate more facial distortion [MacDorman et al. 2009].

Other research has focused on the perception of visual realism for face images, including facial expressions [Deng and Ma 2008; Wallraven et al. 2008]. A recent study showed that intrinsic image components such as shading and reflectance, and factors influencing perceptual judgment, such as expertise and ethnicity, are critical for visual realism perception of faces [Fan et al. 2012]. Another study investigated the effect of render style on virtual human appeal, and the results showed that both highly realistic and highly abstract styles were appealing [McDonnell et al. 2012]. For composite images, other research found that the most important factors for perceived realism were illumination, color, and saturation [Xue et al. 2012]. The current research included some visual factors considered by the CG community and incorporated other factors identified in research on human cognition. The current focus is more on the nature of visual realism perceptual processing as a whole, rather than evaluating the effect of specific visual properties in isolation.

Multimedia Processing: Visual realism perception is closely related to multimedia communications and multimedia forensics. Multimedia communications are often lossy representations enabling efficient transmission of multimedia [Ohm 2004]. Although they are generally designed to preserve image fidelity, preserving image realism is also important. Multimedia forensics is concerned with multimedia authenticity. Authenticity of digital evidence is important in forensics investigation and courts of law. The authenticity of a photo can be challenged as being fully or partially CG. Various computational methods have been developed to distinguish CG images from photos [Ng and Chang 2013]. These data-driven methods provide little insight into what makes an image appear real or CG. Only recently have image-forensics studies looked at factors such as image resolution, JPEG compression, and color as influences on visual realism perception [Farid and Bravo 2012]. The current study differs in that we evaluated hypotheses regarding the underlying human perceptual processing, as well as a broader range of influencing factors.

Face Perception – Holistic or Piecemeal: Configural and holistic coding are generally believed to be the hallmarks of face perception. The fact that faces are processed holistically was first put on record by [Galton 1883], who suggested that a face stimulus was perceived as whole, without any deconstruction into smaller units. The holistic concept was later developed most notably by [Young et al. 1987] and [Sergent 1984]. [Young et al. 1987] demonstrated the importance of configural information in face perception. [Sergent 1984] showed that faces have both component and configural properties and lend themselves to different processing strategies that are not mutually exclusive and can unfold simultaneously. Numerous follow-up studies support the holistic and configural processing of faces [Tanaka and Farah 1993; Farah et al. 1998a; Hancock et al. 2000; Leder and Bruce 2000; Itier and Taylor 2002; Maurer et al. 2002; Le Grand et al. 2004; Sinha et al. 2006; Goffaux and Rossion 2006; Durand et al. 2007; Goffaux et al. 2011; Laguesse and Rossion 2013; Rossion 2013; Omigbodun and Cottrell 2013; Watson 2013].

Many studies suggest that both featural (piecemeal) and configural (holistic) processing make important contributions to face recognition [Sergent 1984; Cabeza and Kato 2000; Collishaw and Hole 2000; Ingvalson and Wenger 2005; Rhodes et al. 2006; Schwaninger et al. 2009; Amishav and Kimchi 2010; Kimchi and Amishav 2010]. Moreover, these two processes are dissociable [Cabeza and Kato 2000; Schwaninger et al. 2009; Amishav and Kimchi 2010], suggesting a dual-model hypothesis [Ingvalson and Wenger 2005; Schwaninger et al. 2009; Amishav and Kimchi 2010]. Research has also shown that recognition of novel faces, compared to learned faces, relies relatively more on the processing of featural information. In the course of familiarisation the importance of configural information increases [Lobmaier and Mast 2007]. [Amishav and Kimchi 2010] suggested that the essential and interactive processing of both componential and configural information, rather than the relative dominance of either, is the hallmark of upright face perception.

In opposition to such ideas, [Gold et al. 2012] argued that observers perceptually integrate features of a whole face with no improvement over their processing of each separate part in isolation. That is, a whole face is perceived no better than the sum of its individual parts.

Studies on configural and featural processing of faces have commonly used three manipulations: *inversion, part-whole*, and *face composite*. Researchers have long known that turning a face upside down makes it harder to recognize [Yin 1969; Valentine 1988], and findings suggest that inversion affects the perception of face configuration more than the perception of local features [Van Belle et al. 2010]. The part-whole method demonstrated that subjects are better at identifying face parts in the context of a whole face than in isolation [Tanaka and Farah 1993]. The face composite method involves composite stimuli created by joining the top half of a familiar face with the bottom half of another familiar face. Observers are generally slower to name the top half of a composite face when the top and bottom parts are vertically aligned than when they are offset laterally (*i.e.*, misaligned) [Young et al. 1987]. A recent study by [Laguesse and Rossion 2013] indicated that the composite face effect results from breaking the whole configuration rather than the increase in relative distance between the face parts. [Rossion 2013] concluded the composite face paradigm is a fantastic tool for face perception research. The three manipulations could be used together. Indeed, [McKone et al. 2013] suggested that the inverted face control should always be tested in all composite and part-whole tasks.

As this research suggests occurs for face recognition, we hypothesized that realism perception for face images might be holistic, piecemeal, or a combination of both. In this paper, we used the inversion method, part-whole method, and face-composite method (misalignment of each image's top and bottom) and collapsed across these manipulations to make inferences regarding holistic versus piecemeal processes underlying the perception of visual realism for face images.

Perception of Eye Region: A range of research highlights the importance of eyes for face perception. The detrimental effects of viewing photos in negative color are largely eliminated if the eye regions alone are rendered positive (in normal color) [Gilad et al. 2009]. Face recognition is better when observers are cued to fixate in the eye region rather than mouth region. The face-inversion effect is reduced with cues to gaze at the eyes [Hills et al. 2011]. [Itier et al. 2007] has shown that the face-specific scheme of human brain is reflected mainly by the contribution of the eye region. [Ryan and Schwartz 2013] found that eyes are more important than noses during face recognition in emotional scenes. These findings suggest that face representations might significantly favour information around the eyes. The social importance of the eyes in a face-context is indicated by the finding that the eyes are specifically informative when comprehending the nonverbal intentions of others, but only if inserted in an intact face [Cecchini et al. 2013]. To preview our findings, eyes were more influential than any other face part for visual realism.

3. IMAGE PREPARATION AND ANALYTICAL TOOLS

In this section, we first describe how stimulus images were produced through intrinsic image decomposition. We then describe our statistical analysis methods.

3.1 Image Collection

We collected a database of 1260 pairs of images, each image depicting one unique face. Each pair was one CG image and one photo with faces matched for age, gender, race, and pose. Most CG images were

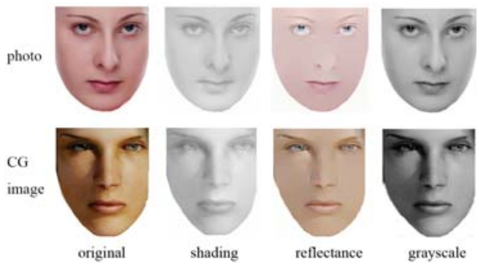


Fig. 1. Images from one pair decomposed into shading and reflectance components. We used grayscale images with high spatial frequency content well-preserved as a replacement for the blurrier shading components.

from a popular computer graphics forum (www.forums.cgsociety.org). The photos were from a photo ranking website (www.photosig.com) and face recognition datasets [Hayward et al. 2008; Kasinski et al. 2008]. We included only realistic CG images, but had no strict criteria for photos. Thus the photos in our dataset varied in lightness and color, as do photos seen in daily life. We cropped the images such that only the face was visible, with minimum hair and no clothes visible (see Fig. 1).

3.2 Intrinsic Image Decomposition

Modified shading and reflectance components of face images elicit different visual realism judgments from human subjects [Fan et al. 2012]. Reflectance refers to the surface color when 3D geometry effects are absent. Shading refers to the brightness change due to the object geometry. We manipulated these intrinsic components in Experiment 1, using the same methods as [Fan et al. 2012]. Specifically, we applied the widely-adopted Retinex-based approach [Grosse et al. 2009] to each R, G, and B color channel of an image to obtain its reflectance component. The shading component from image decomposition is unusable as it loses too much high frequency information, which is important for visual realism perception. Therefore, as a replacement we simply used grayscale versions of images, which are similar to their decomposed shading component (see Fig. 1). Besides the classical Retinex algorithm, additional methods have recently been proposed for intrinsic image decomposition (*e.g.,* [Garces et al. 2012]). We leave it to future work to test their suitability on our face images.

3.3 Analytical Methods

We applied *Signal Detection Theory* (SDT) [Wickens 2001] in our data analysis. SDT is a common analysis tool in psychophysics and biology. It models the decision making process of determining whether items are members of one class or another (e.g., CG or photo). We defined photo as the *signal* and CG as *noise*. In SDT, an observer's sensitivity, indexed by d', is a measure of how well the observer can correctly detect a "signal," when the observer's incorrect identification of non-signals as signals (false alarms) is also taken into account. Thus it represent how well an observer identifies a "signal" relative to "noise". Higher values of d' represent higher signal discrimination. d' values near zero indicate chance performance.

We conducted inferential statistical tests for differences in d'. These tests are standard in behavioral and other sciences. See, for example, [Bailey 2008] for an introduction to inferential statistics. Our analysis strategy was as follows. We evaluated whether d' values differed by image type and manipulation (inversion, misalignment, and isolated parts). For each research question, we first performed an ANOVA (univariate analyses of variance) to check for general effects on d'. If the ANOVA was significant, we conducted followup t-tests or post hoc tests to isolate the various manipulation effects.

4. EXPERIMENT 1: HOLISM

Research in psychology has demonstrated that face perception utilizes holistic and configural information [Bruce and Young 2012]. Yet according to some accounts piecemeal information is processed simultaneously and in parallel with the holistic and configural information [Schwaninger et al. 2003a]. In Experiment 1, we tested the importance of holistic processing for face realism perception. That is, is holistic information important over and above piecemeal information, or do observers predominately look for telltale features (in isolation) indicating that an image is CG versus photo? Across conditions, we manipulated whether an image was aligned versus misaligned, and upright versus inverted, to test in different ways the impact of disrupting holistic processing. [Hancock et al. 2000] showed that representations derived from unfamiliar faces are based on relatively low-level image descriptions, like color, lighting, and local face features. In line with this idea, [Fan et al. 2012] investigated the perceived realism of CG images and demonstrated that shading and color are important cues for image realism perception. Therefore, we also evaluated whether effects of shading and color manipulations interacted with those of the holism-disrupting manipulations.

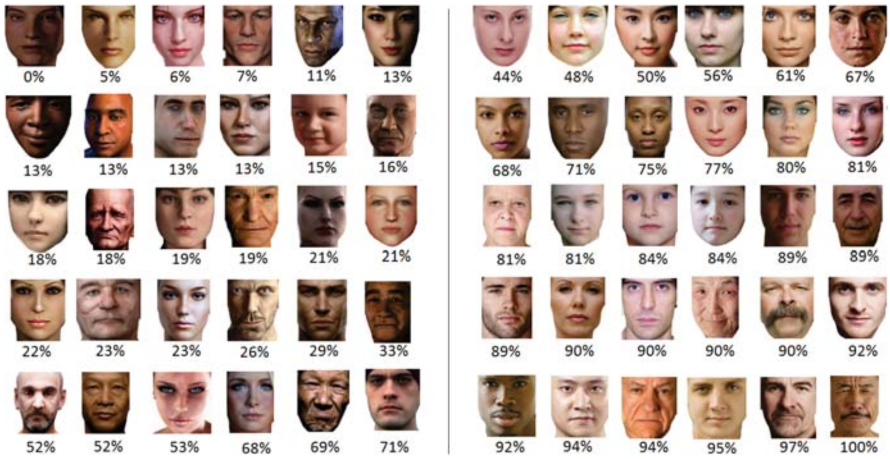
4.1 Method

Participants: 51 Asian undergraduate students (ages between 19-23, 9 females) from the Department of Electronic and Information Engineering (Ningbo University of Technology, China) received course credit for completing the experiment. They had normal or corrected-to-normal visual acuity. To encourage effort, participants were informed that those who were in the top 5% in terms of accuracy on the experimental task would receive an extra prize (a USB hub suite).

Fig. 2. An example of a face image stimulus across experimental conditions (upright versus inverted, crossed with aligned versus misaligned)

Stimuli: From our dataset we selected 30 pairs of images in which the CG images were photo-realistic and such that images were diverse across race, age, and gender (see Fig. 3). The person in each image faced the viewer. All images were down-sampled such that the area depicting the face was about 280 pixels in width and 300 pixels in height. We first performed intrinsic decomposition on the images (see Section 3.2 and Fig. 1). We then created manipulated versions of each image for the upright-misaligned, inverted-aligned, and inverted-misaligned conditions (with the original images comprising the stimuli for the upright-aligned condition) (see Fig. 2). Thus for each face image there were three types: original, grayscale, and reflectance. For each image type there were four conditions: upright versus inverted (orientation) crossed with aligned versus misaligned (alignment). Images manipulations were made using Matlab. The restricted size of the stimulus set was intended to keep the experiment session durations within two hours to avoid participant fatigue and impatience.

Design: There were 12 conditions (3 image-type $\times$ 4 manipulation conditions; image type: original, grayscale, and reflectance; manipulation: upright-aligned, inverted-aligned, upright-misaligned, and inverted-misaligned). All conditions were within-subjects. Therefore participants saw a total of 720 images (60 images $\times$ 12 conditions) presented in a different random order to each participant.



(a) CG images (b) photos

Fig. 3. CG images (a) and photos (b) from all thirty stimulus pairs used in Experiment 1 (with image dimensions shrunk relative to experimental viewing dimensions). The images are arranged lowest to highest in terms of realism ratings for non-manipulated images (*i.e.*, original image-types in the upright-aligned condition). Realism ratings are defined as the percent of participants (displayed under each image) who judged the image as a photo.

Definitions: In both Experiment 1 and 2, CG images were defined as those solely generated with computer software. Photos with post-processing were still defined as photos. In each experiment, these definitions were provided to participants on the introduction page.

Procedure: The experimenter first explained the image manipulations to participants and then told them that the task was judging whether each image was CG or photo. All participants took the experiment at the same time. They sat in a quiet room, each viewing their own 17-inch PC monitor (HP LE1711, 60 Hz refresh rate; 1280 $\times$ 1024 pixel resolution) from about 100cm. Stimuli subtended proximate visual angles of 4.2°horizontally and 4.5°vertically. The log-in page provided an introduction to the experiment, displayed examples of manipulated images, defined CG images and photos, and described the task. Participants entered their student ID number. The second page asked participants their gender, nationality, and prior exposure to CG images. Then the main experimental task began. For each trial, the screen displayed a single image, with text above the image asking, "Is it computer generated (CG), or taken by a digital camera (photo)?". Above this question was text indicating any image manipulations (intrinsic decomposition, inversion, or misalignment). Each image was presented on a white background. Underneath each image was a reminder instructing participants not to judge the image according to the outline, as the background of the image had been manually removed. Participants completed a two-alternative forced-choice decision task. They judged each image as CG or photo by clicking one of two buttons located under the image. One button was labelled "CG", the other "photo". Viewing times were unconstrained. A one second interval followed clicking the button before next image was displayed. A page at the end of the experiment asked participants to indicate the main cue they used to distinguish photos from CG images, from the following options: eyes, skin, color, shape, glossiness, lip, nose, or other. If they chose "other", they were allowed to type in their own specific cue. Each participant saw their overall judgment accuracy for the experiment on the last page, after all the trials were concluded.

4.2 Results

4.2.1 *Realism Rating* . We defined the *realism rating* for an image as the percent of trials in which participants classified it as a photo. Fig. 3 illustrates the realism rating of the 60 non-manipulated images (*i.e.* corresponding to the original image-type condition and the upright-aligned manipulation). As would be expected, CG images had lower realism ratings than photos, $t(58) = -11.97$, $p < .001$. Notably, six CG images had realism ratings higher than 50%, suggesting that they appeared more like photos than CG. Conversely, two photos had ratings lower than 50%, suggesting they appeared more like CG images than photos.

4.2.2 *Inversion Effect on d'*. Fig. 4 shows the mean d' values for inverted versus upright images, collapsed over aligned and misaligned. There was a significant decrease in d' for inverted relative to upright images, $F(1, 50)$ = 194.85, $p < .001$, η_p^2 = .796. The orientation $\times$ alignment and the image-type $\times$ orientation $\times$ alignment interactions were non-significant. There was however an image-type $\times$ orientation interaction, $F(2, 49)$ = 3.49, $p < .001$, η_p^2 = .125. This means that the inversion effect differed in magnitude across image types. We next unpack this interaction with follow-up analyses that indicated a moderating role of the loss of shading information on orientation effects (see also Section 4.2.4).

Follow-up t-tests indicated that for all three image types inversion reduced d'. Thus the t-tests for upright versus inverted images were significant for original images, $t(50)$ = 9.90, $p < .001$, for grayscale images, $t(50)$ = 9.63, $p < .001$, and for reflectance images, $t(50)$ = 8.67, $p < .001$. Furthermore, the effect sizes were similar for original and grayscale images (Cohen's d = .88 and .83, respectively), both of which were larger than the effect size for reflectance images (Cohen's d = .64). **Thus the inversion effect of decreased discrimination performance (d' values) was larger for original and grayscale images than for reflectance images**.

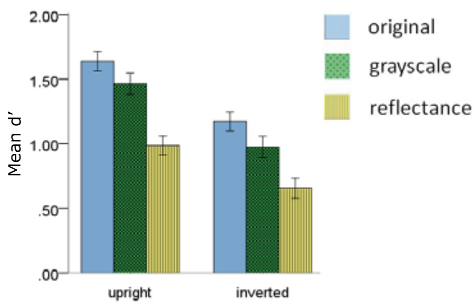


Fig. 4. Mean d' values $(+/-SD)$ for original, grayscale, and reflectance images in the upright versus inverted conditions.

Regression Analysis on Inversion Effect: A regression analysis indicated that most of the variance in inverted-image realism ratings is accounted for by the upright image realism ratings (unstandardized β = .77, SE = .03, $p < .001$, adjusted R^2 = .90). Furthermore, with the reference line $y = x$ drawn on the scatterplot, it is apparent that for most images with upright realism scores greater than 0.5, the inverted realism score was smaller than the upright realism score (see Fig 5). To further evaluate this observation, we split the data into the upright images with realism ratings above 0.5 (upright-realistic group) and those below 0.5 (upright-unrealistic group), then calculated the inversion effect for each image, defined as the upright-image realism score minus the inverted-image realism score. The inversion effect did not significantly differ from zero for the upright-unrealistic group, $t(26) = -0.87$,

$p = .392$. However, for the the upright-realistic group, the inversion effect was significant, $t(32) = 8.55$, $p < .001$. Thus inversion caused realism ratings for the more realistic images to regress towards chance (*i.e.*, $y = 0.5$). We refer to this finding as "*reversion effect*". In short, **inversion made visual realism judgments harder for realistic images but not for unrealistic images**.

An ANCOVA showed that realism scores of upright images significantly predict the inversion effect, $F(1, 57) = 5.41$, $p < .05$, $\eta_p^2 = .087$, but image category (whether it is CG or photo) was unrelated to the inversion effect, $F(1, 57) = .23$, $p = .633$, $\eta_p^2 = .004$. A regression with all datapoints indicated a consistent overall increase in the inversion effect as realism increased, unstandardized β = .23, SE = .03, $p < .001$, adjusted R^2 = .44, suggesting that **the larger the realism score the larger the absolute value of the inversion effect**.

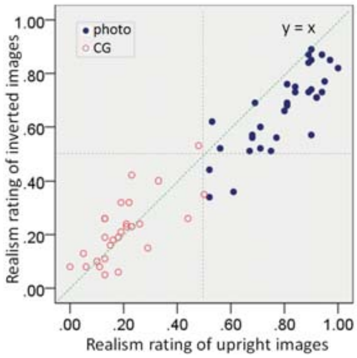


Fig. 5. Mean realism ratings for inverted versus upright images.

4.2.3 *Alignment Effect.* There was a main effect of misalignment (collapsed over upright and inverted) on d', $F(1, 50) = 11.03$, $p = .002$, $\eta_p^2 = .181$. However, misaligned images did not have significantly lower d' values than aligned images for original and reflectance images, $ts(50) < 1.86$, $ps > .068$, but they did for grayscale images, $t(50) = 2.47$, $ps < .05$. **Thus the misalignment effect was mainly carried by grayscale images. Overall, the disruptive effect of misalignment on realism perception was small compared to that of inversion**.

4.2.4 *Image Type Analysis.* There was a substantial effect of image-type on d', $F(2, 118) = 73.56$, $p < .001$, $\eta_p^2 = .555$. As reported above, however, image-type had a significant interaction with orientation, $F(2, 49)$ = 3.49, $p < .001$, η_p^2 = .125. For upright images, paired-sample t-tests indicated higher d' values on original images than grayscale images, $t(50) = 3.58$, $p < .001$, and higher d' values on grayscale images than reflectance images, $t(50) = 8.93$, $p < .001$. A similar overall pattern emerged for inverted images: discrimination performance was better on original images than grayscale images, $t(50) = 4.07$, $p < .001$, and on grayscale images than reflectance images, $t(50) = 4.93$, $p < .001$. The above differences were all significant after Bonferroni correction.

To further analyze the image-type $\times$ orientation interaction, especially the impacts of the removal of shading versus color, we calculated the effect size for each pairwise-comparison among image-type conditions separately for upright images and inverted images (see Fig. 6). Inversion generally resulted in smaller image-type effects. Inversion appears to decrease the impact of presence versus absence of image-components on visual realism perception. As the inversion manipulation was used to measure the impact of disrupting holistic processing, the results of our image-type analysis taken together suggest that **shading is more important than color**.

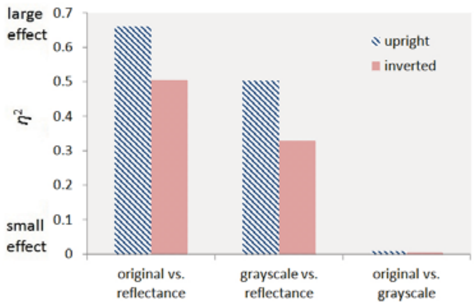


Fig. 6. Image-type effect sizes (η^2) on d' as a function of orientation.

4.2.5 *Self-reported Cues.* When indicating which cue was most important for judgments of whether an image was CG or a photo, most participants selected eyes, followed by glossiness, then face parts and face shape (see Fig. 7).

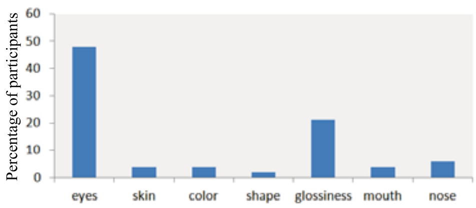


Fig. 7. Main judgment cues claimed by different percentage of participants in Experiment 1.

4.2.6 *Result Summary.* The Experiment 1 results [1] suggest that holistic processing is important for visual realism perception of face images, similar to its general importance in face perception.

We found, however, that misalignment had marginal effects compared to those of inversion. Misalignment allows normal use of parts like eyes during realism judgment but disrupts holism perceptual processes. We also found the realism ratings of upright-unrealistic images were not affected by inversion (Fig. 5). Based on the Anna Karenina principle [Gorban et al. 2010], people might judge using a strategy of, "it is CG if it has something weird and photo if not." Inverting realistic images might make image parts look weird (*e.g.*, inverted eyes), but inverting weird (CG) images may make the weirdness less apparent, yet still identifiable. Combining the above findings, we postulate that piecemeal processing may play a role in realism perception.

[1]Online labour source websites like Amazon Mechanical Turk [Paolacci et al. 2010] are becoming increasingly popular. It is therefore important to compare online experiments and controlled lab experiments. Reported in supplementary materials, we did a pilot study online that otherwise replicated Experiment 1, but used a different 10 pairs of CG images and photos. The results were very similar (see supplementary material). The different set of images did not alter human perception characteristics. We concluded that for this type of psychophysics experiment, with sufficient participants, online and lab experiments are comparable. We performed Experiment 2 online.

As recent research suggests that inversion can affect not only holistic processing, but piecemeal processing also [Piepers and Robbins 2012], we conducted an additional experiment to explore the role of holistic processing with different manipulations. In Experiment 2, we test the role of holistic versus piecemeal processing in realism judgments using a part-whole paradigm and resolution manipulation, two complementary approaches that induce opposite effects on holistic and piecemeal processing. Experiment 2 also explores more directly whether the sensitivity of realism perception differed for different face parts, as an extension of the self-reported cues in Section 4.2.5.

5. EXPERIMENT 2: FACE PARTS

Research has shown that component information and configural information are both important for face perception [Schwaninger et al. 2009; Laguesse and Rossion 2013]. In Experiment 2, we test the contribution of both component and configural information to visual realism perception of face images by separating faces into parts and by scrambling faces. We notice that viewing distance influences the perception of CG artifacts [Larkin and O'Sullivan 2011], and image resolution affects realism perception [Farid and Bravo 2012]. As resolution could impact participants' utilization of component or configural information, we presented images of faces, face parts, and scrambled faces at two resolutions to test how resolution influences holistic or piecemeal processing.

5.1 Method

Participants: Participants were 670 students aged 19-25 (325 females), recruited by advertisements at Ningbo University of Technology, China. All participants were Asian and had normal or corrected-to-normal visual acuity. Students who participated in previous experiments were excluded. Every participant was given a souvenir (a key chain) for their participation. The top 5% most accurate were rewarded with an extra prize (USB hub suite).

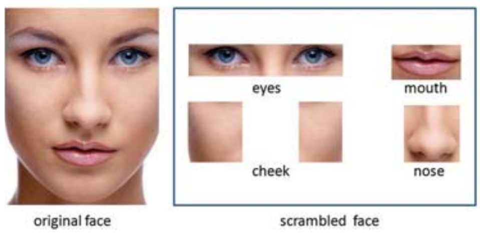


Fig. 8. Face parts used in our experiment. Scrambled face parts were presented in one horizontal row to participants.

Stimuli: From our dataset, we selected 30 pairs of photo-realistic CG and photo faces, which differed from the Experiment 1 stimuli. Images were diverse in race, age, and gender. We separated each image into face parts and reorganized them as a scrambled face (see Fig. 8) using Adobe Photoshop. We further synthesized two resolutions for all resulting images while keeping their aspect ratio fixed, with high resolution face images having approximately 450×520 pixels, and low resolution ones about 270×312 pixels.

Design: The experiment was a 6 (face-part: unscrambled face, scrambled face, eyes, nose, mouth, cheek) $\times$ 2 (resolution: high, low) between-subjects design. The stimulus set had a total of $6 \times 2 \times 30 \times 2 = 720$ images. Each participant was randomly assigned to one face-part condition at one resolution level. The number of test stimuli for each participant was $30 \times 2 = 60$ images. The number of participants in each condition varied from 53 to 61. For each participant, image presentation order was random.

Procedure: The procedure was similar to that of Experiment 1, except participants in Experiment 2 completed the experiment online, accessing the experiment via a link to its homepage. Instructions were on the homepage. Participants were required to input their student ID as a unique identifier. Participants entered their gender, nationality, and prior exposure to CG images. The layout of the judgment pages was the same as in Experiment 1. Participants saw 60 single images sequentially displayed on a white background. They judged each image as CG or photo, in response to a two-alternative forced-choice question by clicking the corresponding button below the image. A one second interval followed clicking the button before the next image was displayed. Participants saw their overall judgment accuracy at the end of experiment.

5.2 Results

5.2.1 *Face Part Analysis.* There was a significant main effect of face-part on d', $F(5, 658) = 31.58$, $p < .001$, $\eta_p^2 = .411$. The d' values were significantly above zero on all face parts except cheek, indicating participants had considerable discrimination ability on isolated face parts (except cheek). The face-part $\times$ resolution interaction was non-significant, $F(5, 658) = 1.92$, $p = .089$, $\eta_p^2 = .014$. Post hoc Turkey tests showed significant differences among all possible face-part condition combinations on d', $ps < .05$, except for the eyes versus scrambled-face, and nose versus mouth conditions. This indicates that **seeing all face parts when scrambled provided no benefit for visual realism beyond that of seeing only the eyes, and nose and mouth have similar effect on realism of static faces**. Participants performed best on unscrambled faces, followed by eyes and scrambled faces, then mouth and nose, and worst on cheek (see Fig. 9). The large difference between the unscrambled face and scrambled face conditions suggests a strong holism effect on realism perception on face images.

Among isolated face parts, eyes most enabled participants to distinguish between CG images and photos. The cheek was the least informative isolated face-part for visual realism perception. The cheek, merely a skin surface, may have been least informative for image realism, due to its lack of geometry information or variability in color. As we found in Experiment 1 that shading information was important for realism perception, the lack of shading in cheeks might also be a cause.

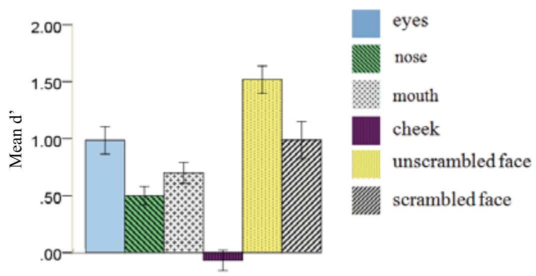


Fig. 9. Mean d' for each face-part condition ($+/-SD$), collapsed over high and low resolutions.

5.2.2 *Resolution Effect.* There was a moderate but statistically significant effect of resolution on d', $F(1, 658) = 22.59$, $p < .001$, $\eta_p^2 = .033$. Pairwise contrasts showed that d' values were higher for high resolution than low resolution for images of eyes, a nose and a scrambled face, $|t|s(108) > 2.82$, $ps < .05$, but not for images of a mouth, cheeks, or an unscrambled face, $|t|s(109) < 1.26$, $ps > .211$ (see Fig. 10). This indicates that **resolution is a meaningful factor for realism perception when viewing all face parts in scrambled arrangement, the eyes in isolation, or the nose in isolation.**

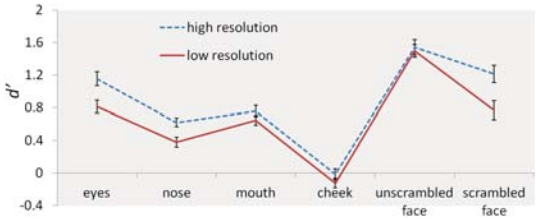


Fig. 10. Mean d' $(+/-SD)$ across face-part conditions for high and low resolution images.

We were also interested in comparing more specifically the influence of resolution across two scrambling conditions (unscrambled versus scrambled). The two scrambling conditions both presented images with the same amount of pixel information but with the parts arranged either to facilitate or to disrupt holistic processing. There was a moderate resolution by scrambling interaction for d', $F(1,218) = 4.13$, $p = .043$, $\eta_p^2 = .019$. The effect of scrambling was larger for lower resolution than higher resolution images. **This suggests that holistic processing becomes increasingly dominant when image resolution is lower.** To further analyze this finding, for each resolution condition, we created a scatterplot of realism ratings for scrambled face images versus realism ratings for unscrambled face images (See Fig. 11). On low resolution images, there is a notable ”reversion effect” of scrambling causing judgments for the more realistic images to regress towards chance (In Fig. 11(a), above $y = x$ line for unrealistic images, and below $y = x$ line for realistic images). This was larger than the reversion effect from inversion in Experiment 1 discussed in Section 4.2.2. Defining the ”scrambling” effect as the unscrambled minus scrambled mean realism-rating, “realistic” images as those with an unscrambled realism mean rating above 0.5, and “unrealistic” images as those with an unscrambled realism mean rating below 0.5, we furthered our analysis by examining the scrambling effect on realistic versus unrealistic images. Scrambling effects tended to be negative for unrealistic images, indicating that scrambling tended to raise realism ratings for unrealistic images, $t(29) = -6.11$, $p < .001$. However scrambling effects tended to be positive for realistic images (and with larger magnitudes), indicating that realistic images had their realism ratings lowered by scrambling, in line with the reversion effect, $t(26) = 9.15$, $p < .001$. In short, **scrambling of low resolution images made visual realism judgments harder for both realistic and unrealistic images**. Turning to high resolution images, we did not find these differential patterns for how scrambling impacted the realism ratings on realistic versus unrealistic images, in line with decreased reliance of holistic processing for high resolution images (see Fig. 11(b)).

5.2.3 *Result Summary.* These findings confirm and delineate the importance of holistic information for visual realism perception of faces. From the larger-than-zero d' on all face parts (except cheeks) and scrambled faces, we found that piecemeal processing also plays a role. We also found that eyes provide essential piecemeal-information for visual realism perception. Higher reliance on holistic information occurred for lower-resolution images.

6. DISCUSSION

6.1 Comparison with Face Perception Studies

Prior face perception research has mainly focused on face identification and face recognition. Findings consistently indicate that holistic and configural processing are critical for those tasks [Tanaka and Farah 1993; Farah et al. 1998a; Hancock et al. 2000; Itier and Taylor 2002; Maurer et al. 2002; Le Grand et al. 2004; Sinha et al. 2006; Goffaux and Rossion 2006; Durand et al. 2007; Goffaux et al.

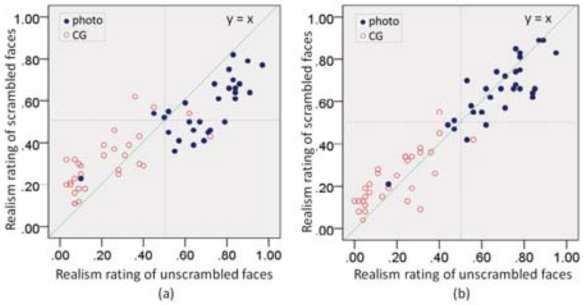


Fig. 11. Mean realism ratings for scrambled versus unscrambled face images at low resolution (a) and high resolution (b). The $y = x$ line is drawn to highlight reversion effects.

2011; Laguesse and Rossion 2013; Rossion 2013; Omigbodun and Cottrell 2013; Watson 2013]. In line with this idea, we found that holistic processing plays a critical role in visual realism perception, but piecemeal processing is important too. Our results showed that the discrimination performance on isolated eyes, mouth, nose, and scrambled faces were above chance, indicating that people can successfully utilize piecemeal information during realism perception of faces.

Still controversial is the argument that face processing uses or is even dominated by component (piecemeal) information. [Schwaninger et al. 2003a; Schwaninger et al. 2009; Amishav and Kimchi 2010] argue that both component and configural information are extracted and processed in face perception. [Kimchi and Amishav 2010] claim that both component and configural properties contribute to the processing of upright faces and neither property necessarily dominates the other. Instead, the interactive processing of component and configural properties may be the dominant form of face processing in everyday life [Kimchi and Amishav 2010]. The study in [Favelle and Palmisano 2012] suggests that holistic processing does not occur for all views of upright faces (*e.g.*, not for uncommon pitch rotated views), only those that can be matched to a generic global representation of a face. In our experiment, we also identified some evidence of piecemeal processing. What's more, we found that holistic processing was increasingly dominant when resolution was lower. This observation is in line with previous findings on spatial face perception – they suggest that whereas holistic face processing depends mostly on low spatial frequency, the extraction of face parts depends mostly on high spatial frequency [Goffaux et al. 2011; Gao and Bentin 2011].

We also situate our findings to prior research regarding the contribution of individual face parts during perception tasks. In [Brown and Perrett 1993], participants did a binary perception task of gender recognition from face parts. The contributions of face parts was ordered in descending importance from the eye brows & eyes, eye brows alone, eyes alone, the whole jaw, the chin, the nose & mouth, to the mouth alone, with the nose the only part insufficient to distinguish between the genders. We used slightly different face parts, but our visual realism judgment was also a binary perception task. Our face parts ordered in descending importance from the eyes, the mouth and nose, and the cheek (see Section 5.2.1). Interestingly, in both tasks, eyes were found to be most important. Moreover, the classification accuracy rates were higher for our visual realism judgment task than the gender recognition task, and the nose is still sufficient to distinguish the image category. This suggests either that our task was easier, or that more information is usable for piecemeal processing during visual realism judgment than during gender recognition. The latter argues for the relative importance of piecemeal processing for visual realism judgment, which was also seen in the regression effect on inversion (see Section 4.2.2). However, recent research has shown that the eyes were the most salient feature (in terms of first fixation, number of fixations, and duration of fixation) for upright faces. For inverted faces, in contrast, other features were sampled first [Hills et al. 2012]. This suggests that the order of importance of face parts might vary for different face presentations. Finally, our face parts excluded eyebrows, which many studies have demonstrated as important during face perception [Sadrô et al. 2003]. We leave their effect on visual realism to future work.

It is important to note that face perception includes a number of dedicated processes and structures in the human brain. Which is used varies as a function of task: determining face identity is processed differently from identifying emotion or expression. [Schwaninger et al. 2003b] suggested that configural processing does not obey the same rules during perceptual tasks that it does during detection and recognition tasks – the perception of configural information is less orientation-sensitive. This might be one reason for why holistic processing is less generally dominant in visual realism perception.

Finally, we found that resolution affected realism perception. Holistic processing became increasingly dominant when image resolution was lower. It would be interesting to explore the perception of visual realism for face images of even higher and lower resolutions, which we leave for future work.

6.2 Generalization to Objects

Classic behavioral work has shown that faces are processed in a distinctive holistic manner that is unlike the processing of objects (*i.e.*, using relatively less part decomposition than other types of objects) [Farah 1996; Moscovitch et al. 1997; Farah et al. 1998b; Leder and Bruce 2000; Nelson 2001; Bruce and Young 2012]. Moreover, face perception appears to be domain specific – it uses specific brain areas dedicated to that purpose [McCarthy et al. 1997; Farah et al. 1998b; Kanwisher et al. 2000; Haxby et al. 2000; Kanwisher and Yovel 2006; McKone et al. 2007].

However, studies indicate, first, that experts trained to identify non-face objects exhibit similar perception characteristics for such objects as for faces [Gauthier et al. 1997; Rossion et al. 2002; Bukach et al. 2006; McKeeff et al. 2010], and, second, that experience shapes holistic object representations in the visual system [Wong et al. 2012b]. Holistic processing may be a general marker of expertise across a wide domain of visual discrimination [Wong et al. 2012a]. In neuroscience, some experimental evidence supports the idea that face perception shares brain areas with object perception [Moscovitch et al. 1997; Rossion et al. 2002; Xu 2005; McKeeff et al. 2010; Rossion et al. 2012; Goffaux et al. 2013], and the functional overlap between face and object perception is increased by expertise [McKeeff et al. 2010]. [Tan and Poggio 2013] raised the possibility that face perception uses the same template matching mechanisms as object perception. If these hypotheses are true, our findings are likely to be generalizable to understanding visual realism perception of non-face objects for trained experts.

In object recognition tasks, visual processing in cortex is classically modeled as a hierarchy of increasingly sophisticated representations [Riesenhuber and Poggio 1999]. Recent research indicated that the human visual system uses a hierarchical representation scheme to process multiple objects in natural scenes: an "average mechanism" in posterior brain regions helps retain information of individual objects in cluttered scenes, whereas a "nonaverage mechanism" in anterior regions uses contextual information to optimize the representation of multiple objects [Song et al. 2013]. Faces being a special object type, it still undergoes the similar brain mechanisms. We predict that realism perception for general objects might have similarities with realism perception for faces. This means that, as we observed for face images, lower resolution images of general scenes may not be as easily and accurately judged as real versus CG.

Other research suggests a reduced role of some kinds of information for non-face objects [Gauthier et al. 1997; Wong et al. 2012a; Tan and Poggio 2013]. We predict that configural information such as the distances between object parts may be relatively less important for realism judgments of non-face objects, as compared to realism judgments of faces. Such hypotheses about general objects provide interesting directions for future research.

6.3 Implications for CG Rendering

As visual realism of face images was related to holistic as well as piecemeal processing, a potential implication is that visual realism perception can be systematically controlled by adjusting local face parts as well as their relative placements. Refined control of visual realism could then help to achieve other effects, such as liking, trust [McDonnell et al. 2012] and negative responses [Nowak et al. 2008]. Although such fine-grained control of visual realism is a topic for future work, we can derive a few general principles for CG artists from our studies. The importance of configural processing in visual realism perception of low-resolution faces suggests that CG games targeted for small-size displays on mobile devices can lessen the degradation of visual quality from reduced resolution by preserving the configural information of CG characters. In this case, a low-dimensional global-basis-based synthesis approach may be appropriate for efficiently generating realistic faces [Blanz and Vetter 1999].

Our part-whole analysis has implications for rendering CG face parts. In the case of limited computing resources, a renderer may want to begin by dedicating equal rendering resources to all face parts until all CG face parts attain sufficient realism. Beyond this point, the renderer may benefit by devoting surplus resources to specific parts according to their relative importance to visual realism: first to eyes, then the mouth, and finally the nose.

The reversion effect observed in Fig. 4 and 5 has several implications if sufficiently dominant. First, a highly realistic CG face in a movie may carry the risk of a large drop in apparent realism when undergoing movements that result in occlusions, which may interrupt holistic perception. Such an effect is reminiscent of the *uncanny valley* [Mori 1970]. Second, a poorly rendered CG face or a CG face with ambiguous realism may maintain a stable visual realism perception amidst such movements.

Finally, perception-based rendering has become an increasingly important paradigm [Vangorp et al. 2013]. Our studies on perceptual processes could motivate new perception-based rendering techniques.

6.4 Limitations and Future Work

This research focused on static face images. Visual realism for motion pictures and general objects might function differently. It seems likely that visual realism perception for moving faces is more complex than for static faces. The application of 3D representations needs to be considered in this case, since 3D information substantially contributes to face recognition tasks [Liu and Ward 2006; Schwaninger and Yang 2011]. [Piepers and Robbins 2012] suggested important advances can be made by studying moving faces, helping us better understand whether "relationships between parts" means between the edges of nameable features of key elements.

Research indicates that face recognition depends on prior exposure to racial groups: performance is better for familiar than unfamiliar races. [Michel et al. 2006] suggested that holistic processing is finely tuned for faces of one's own race and same-race faces are perceived more holistically than other-race faces. The own-race advantage is a well-established phenomenon that occurs across countries and racial groups [Meissner and Brigham 2001; Michel et al. 2006; Bernstein et al. 2007; Natu et al. 2011]. [Golby et al. 2001] has shown that differential activation in fusiform regions in human brain contributes to same-race memory superiority. Research has also shown that Asians may make more general use of face-specific mechanisms than Caucasians [Crookes et al. 2013]. During facial expression recognition, east Asian observers tend to focus heavily on the eye (and sometimes nose) region whereas Westerners pay more attention to the mouth (with some attention on the nose and eyes) [Jack et al. 2009]. This demonstrates that cultural background can potentially influence the perception of realism for viewed face parts. For realism perception, [Fan et al. 2012] found that during realism perception Chinese and Caucasian participants had higher sensitivity to faces of their own races. Future work can more fully investigate the influence of cognitive factors on realism perception for face images.

We also hope to build perceptual models and computational models for visual realism perception of face images, like those created for face perception [Schwaninger et al. 2003a; Schwaninger et al. 2009]. More experiments are needed for this goal, such as the investigation of participants' response time.

7. CONCLUSIONS

We conducted psychophysics experiments to investigate perceptual processing factors underlying visual realism. The research strategy we developed and the resulting findings provide novel contributions towards understanding how people distinguish between CG faces versus photo faces. Our findings have implications for optimizing visual realism through effective face image rendering, resource allocation across face parts, and holism disruption. Future work can build on our findings to develop specific guidelines for enhancing or controlling visual realism of face images.

Unlike other studies on perception-based rendering, our research was psychologically-oriented. Our goal was not to identify rendering parameters that influence visual realism, but to investigate at a perceptual processing level some causes of why parameters influence visual realism. In the future, we plan to utilize these findings in developing a computational model for automatic estimation of the visual realism of face images.

REFERENCES

AMISHAV, R. AND KIMCHI, R. 2010. Perceptual integrality of componential and configural information in faces. *Psychonomic bulletin & review 17,* 5, 743–748.

BAILEY, R. 2008. *Design of comparative experiments*. Vol. 25. Cambridge University Press.

BERNSTEIN, M. J., YOUNG, S. G., AND HUGENBERG, K. 2007. The cross-category effect mere social categorization is sufficient to elicit an own-group bias in face recognition. *Psychological Science 18,* 8, 706–712.

BLANZ, V. AND VETTER, T. 1999. A morphable model for the synthesis of 3d faces. In *Proceedings of SIGGRAPH*. 187–194.

BROWN, E. AND PERRETT, D. 1993. What gives a face its gender. *Perception 22*, 829–840.

BRUCE, V. AND YOUNG, A. 2012. *Face Perception*. Psychology Press.

BUKACH, C. M., GAUTHIER, I., AND TARR, M. J. 2006. Beyond faces and modularity: The power of an expertise framework. *Trends in cognitive sciences 10,* 4, 159–166.

CABEZA, R. AND KATO, T. 2000. Features are also important: Contributions of featural and configural processing to face recognition. *Psychological Science 11,* 5, 429–433.

CECCHINI, M., ACETO, P., ALTAVILLA, D., PALUMBO, L., AND LAI, C. 2013. The role of the eyes in processing an intact face and its scrambled image: A dense array erp and low-resolution electromagnetic tomography (sloreta) study. *Social Neuroscience* ahead-of-print, 1–12.

COLLISHAW, S. M. AND HOLE, G. J. 2000. Featural and configurational processes in the recognition of faces of different familiarity. *PERCEPTION-LONDON- 29,* 8, 893–910.

CROOKES, K., FAVELLE, S., AND HAYWARD, W. G. 2013. Holistic processing for other-race faces in chinese participants occurs for upright but not inverted faces. *Frontiers in psychology 4*.

DENG, Z. AND MA, X. 2008. Perceptually guided expressive facial animation. In *Proceedings of the 2008 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. Eurographics Association, 67–76.

DURAND, K., GALLAY, M., SEIGNEURIC, A., ROBICHON, F., AND BAUDOUIN, J.-Y. 2007. The development of facial emotion recognition: The role of configural information. *Journal of experimental child psychology 97,* 1, 14–27.

FAN, S., NG, T.-T., HERBERG, J., KOENIG, B., AND XIN, S. 2012. Real or fake?: human judgments about photographs and computer-generated images of faces. In *Technical Briefs, ACM SIGGRAPH Asia*. 17.

FARAH, M., WILSON, K., DRAIN, M., AND TANAKA, J. 1998a. What is" special" about face perception? *Psychological Review; Psychological Review 105,* 3, 482.

FARAH, M. J. 1996. Is face recognition special? evidence from neuropsychology. *Behavioural brain research 76,* 1, 181–189.

FARAH, M. J., WILSON, K. D., DRAIN, M., AND TANAKA, J. N. 1998b. What is" special" about face perception? *Psychological review 105,* 3, 482.

FARID, H. AND BRAVO, M. 2012. Perceptual discrimination of computer generated and photographic faces. *Digital Investigation 8,* 3, 226–235.

FAVELLE, S. AND PALMISANO, S. 2012. The face inversion effect following pitch and yaw rotations: investigating the boundaries of holistic processing. *Frontiers in Psychology 3,* 563.

GALTON, F. 1883. *Inquiries into human faculty and its development*. Macmillan and Company.

GAO, Z. AND BENTIN, S. 2011. Coarse-to-fine encoding of spatial frequency information into visual short-term memory for faces but impartial decay. *Journal of Experimental Psychology: Human Perception and Performance 37,* 4, 1051.

GARCES, E., MUNOZ, A., LOPEZ-MORENO, J., AND GUTIERREZ, D. 2012. Intrinsic images by clustering. In *Computer Graphics Forum*. Vol. 31. Wiley Online Library, 1415–1424.

GAUTHIER, I., TARR, M., ET AL. 1997. Becoming a “greeble” expert: exploring mechanisms for face recognition. *Vision research 37,* 12, 1673–1682.

GILAD, S., MENG, M., AND SINHA, P. 2009. Role of ordinal contrast relationships in face encoding. *Proceedings of the National Academy of Sciences 106,* 13, 5353–5358.

GOFFAUX, V., PETERS, J., HAUBRECHTS, J., SCHILTZ, C., JANSMA, B., AND GOEBEL, R. 2011. From coarse to fine? spatial and temporal dynamics of cortical face processing. *Cerebral Cortex 21,* 2, 467–476.

GOFFAUX, V. AND ROSSION, B. 2006. Faces are” spatial”–holistic face perception is supported by low spatial frequencies. *Journal of Experimental Psychology: Human Perception and Performance 32,* 4, 1023.

GOFFAUX, V., SCHILTZ, C., MUR, M., AND GOEBEL, R. 2013. Local discriminability determines the strength of holistic processing for faces in the fusiform face area. *Frontiers in Psychology 3,* 604.

GOLBY, A. J., GABRIELI, J. D., CHIAO, J. Y., AND EBERHARDT, J. L. 2001. Differential responses in the fusiform region to same-race and other-race faces. *Nature neuroscience 4,* 8, 845–850.

GOLD, J. M., MUNDY, P. J., AND TJAN, B. S. 2012. The perception of a face is no more than the sum of its parts. *Psychological science 23,* 4, 427–434.

GORBAN, A., SMIRNOVA, E., AND TYUKINA, T. 2010. Correlations, risk and crisis: From physiology to finance. *Physica A: Statistical Mechanics and its Applications 389,* 16, 3193–3217.

GROSSE, R., JOHNSON, M., ADELSON, E., AND FREEMAN, W. 2009. Ground truth dataset and baseline evaluations for intrinsic image algorithms. In *IEEE International Conference on Computer Vision*. 2335–2342.

HANCOCK, P. J., BRUCE, V., AND BURTON, A. M. 2000. Recognition of unfamiliar faces. *Trends in cognitive sciences 4,* 9.

HAXBY, J. V., HOFFMAN, E. A., AND GOBBINI, M. I. 2000. The distributed human neural system for face perception. *Trends in cognitive sciences 4,* 6, 223–233.

HAYWARD, W. G., RHODES, G., AND SCHWANINGER, A. 2008. An own-race advantage for components as well as configurations in face recognition. *Cognition 106,* 2, 1017–1027.

HILLS, P. J., ROSS, D. A., AND LEWIS, M. B. 2011. Attention misplaced: The role of diagnostic features in the face-inversion effect. *Journal of Experimental Psychology: Human Perception and Performance 37,* 5, 1396.

HILLS, P. J., SULLIVAN, A. J., AND PAKE, J. M. 2012. Aberrant first fixations when looking at inverted faces in various poses: The result of the centre-of-gravity effect? *British Journal of Psychology 103,* 4, 520–538.

INGVALSON, E. M. AND WENGER, M. J. 2005. A strong test of the dual-mode hypothesis. *Perception & Psychophysics 67,* 1.

ITIER, R. J., ALAIN, C., SEDORE, K., AND MCINTOSH, A. R. 2007. Early face processing specificity: It’s in the eyes! *Journal of Cognitive Neuroscience 19,* 11, 1815–1826.

ITIER, R. J. AND TAYLOR, M. J. 2002. Inversion and contrast polarity reversal affect both encoding and recognition processes of unfamiliar faces: a repetition study using erps. *Neuroimage 15,* 2, 353–372.

JACK, R. E., BLAIS, C., SCHEEPERS, C., SCHYNS, P. G., AND CALDARA, R. 2009. Cultural confusions show that facial expressions are not universal. *Current Biology 19,* 18, 1543–1548.

KANWISHER, N. ET AL. 2000. Domain specificity in face perception. *Nature neuroscience 3*, 759–763.

KANWISHER, N. AND YOVEL, G. 2006. The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences 361,* 1476, 2109–2128.

KASINSKI, A., FLOREK, A., AND SCHMIDT, A. 2008. The put face database. *Image Processing and Communications 13,* 3-4.

KIMCHI, R. AND AMISHAV, R. 2010. Faces as perceptual wholes: The interplay between component and configural properties in face processing. *Visual Cognition 18,* 7, 1034–1062.

LAGUESSE, R. AND ROSSION, B. 2013. Face perception is whole or none: disentangling the role of spatial contiguity and interfeature distances in the composite face illusion. *Perception 42*, 1013–1026.

LARKIN, M. AND O’SULLIVAN, C. 2011. Perception of simplification artifacts for animated characters. In *ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization*. 93–100.

LE GRAND, R., MONDLOCH, C. J., MAURER, D., AND BRENT, H. P. 2004. Impairment in holistic face processing following early visual deprivation. *Psychological Science 15,* 11, 762–768.

LEDER, H. AND BRUCE, V. 2000. When inverted faces are recognized: The role of configural information in face recognition. *The Quarterly Journal of Experimental Psychology: Section A 53,* 2, 513–536.

LIU, C. H. AND WARD, J. 2006. The use of 3d information in face recognition. *Vision research 46,* 6, 768–773.

LOBMAIER, J. S. AND MAST, F. W. 2007. Perception of novel faces: The parts have it! *Perception 36,* 11, 1660.

MACDORMAN, K. F., GREEN, R. D., HO, C.-C., AND KOCH, C. T. 2009. Too real for comfort? uncanny responses to computer generated faces. *Computers in human behavior 25,* 3, 695–710.

MACDORMAN, K. F. AND ISHIGURO, H. 2006. The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies 7,* 3, 297–337.

MAURER, D., GRAND, R. L., AND MONDLOCH, C. J. 2002. The many faces of configural processing. *Trends in cognitive sciences 6,* 6, 255–260.

MCCARTHY, G., PUCE, A., GORE, J. C., AND ALLISON, T. 1997. Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience 9,* 5, 605–610.

MCDONNELL, R., BREIDT, M., AND BÜLTHOFF, H. 2012. Render me real?: investigating the effect of render style on the perception of animated virtual humans. *ACM Transactions on Graphics 31,* 4, 91:1–91:11.

MCKEEFF, T. J., MCGUGIN, R. W., TONG, F., AND GAUTHIER, I. 2010. Expertise increases the functional overlap between face and object perception. *Cognition 117,* 3, 355–360.

MCKONE, E., DAVIES, A. A., DARKE, H., CROOKES, K., WICKRAMARIYARATNE, T., ZAPPIA, S., FIORENTINI, C., FAVELLE, S., BROUGHTON, M., AND FERNANDO, D. 2013. Importance of the inverted control in measuring holistic face processing with the composite effect and part-whole effect. *Frontiers in psychology 4*.

MCKONE, E., KANWISHER, N., AND DUCHAINE, B. C. 2007. Can generic expertise explain special processing for faces? *Trends in cognitive sciences 11,* 1, 8–15.

MCNAMARA, A. 2005. Exploring perceptual equivalence between real and simulated imagery. In *ACM symposium on Applied perception in graphics and visualization*. 123–128.

MEISSNER, C. A. AND BRIGHAM, J. C. 2001. Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law 7,* 1, 3.

MEYER, G., RUSHMEIER, H., COHEN, M., GREENBERG, D., AND TORRANCE, K. 1986. An experimental evaluation of computer graphics imagery. *ACM Transactions on Graphics 5,* 1, 30–50.

MICHEL, C., ROSSION, B., HAN, J., CHUNG, C.-S., AND CALDARA, R. 2006. Holistic processing is finely tuned for faces of one's own race. *Psychological Science 17,* 7, 608–615.

MORI, M. 1970. The uncanny valley. *Energy 7,* 4, 33–35.

MOSCOVITCH, M., WINOCUR, G., AND BEHRMANN, M. 1997. What is special about face recognition? nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *Journal of Cognitive Neuroscience 9,* 5, 555–604.

NATU, V., RABOY, D., AND O'TOOLE, A. J. 2011. Neural correlates of own-and other-race face perception: Spatial and temporal response differences. *NeuroImage 54,* 3, 2547–2555.

NELSON, C. A. 2001. The development and neural bases of face recognition. *Infant and child development 10,* 1-2, 3–18.

NG, T.-T. AND CHANG, S.-F. 2013. Discrimination of computer synthesized or recaptured images from real images. In *Digital Image Forensics*. Springer, 275–309.

NOWAK, K., KRCMAR, M., AND FARRAR, K. 2008. The causes and consequences of presence: Considering the influence of violent video games on presence and aggression. *Presence: Teleoperators and Virtual Environments 17,* 3, 256–268.

OHM, J.-R. 2004. *Multimedia communication technology: Representation, transmission and identification of multimedia signals*. Springer Verlag.

OMIGBODUN, A. AND COTTRELL, G. 2013. Evidence for holistic facial expression processing with a neurocomputational model. *Journal of Vision 13,* 9, 99–99.

PAOLACCI, G., CHANDLER, J., AND IPEIROTIS, P. 2010. Running experiments on amazon mechanical turk. *Judgment and Decision Making 5,* 5, 411–419.

PIEPERS, D. AND ROBBINS, R. 2012. A review and clarification of the terms "holistic," "configural," and "relational" in the face perception literature. *Frontiers in Psychology 3,* 559.

RADEMACHER, P., LENGYEL, J., CUTRELL, E., AND WHITTED, T. 2001. Measuring the perception of visual realism in images. In *Eurographics Workshop on Rendering Techniques*. 235–248.

RAMANARAYANAN, G., FERWERDA, J., WALTER, B., AND BALA, K. 2007. Visual equivalence: towards a new standard for image fidelity. *ACM Transactions on Graphics 26,* 3, 76.

RHODES, G., HAYWARD, W. G., AND WINKLER, C. 2006. Expert face coding: Configural and component coding of own-race and other-race faces. *Psychonomic Bulletin & Review 13,* 3, 499–505.

RICHLER, J. J., PALMERI, T. J., AND GAUTHIER, I. 2012. Meanings, mechanisms, and measures of holistic processing. *Frontiers in psychology 3*.

RIESENHUBER, M. AND POGGIO, T. 1999. Hierarchical models of object recognition in cortex. *Nature neuroscience 2,* 11, 1019–1025.

ROSSION, B. 2013. The composite face illusion: A whole window into our understanding of holistic face perception. *Visual Cognition* ahead-of-print, 1–115.

ROSSION, B., GAUTHIER, I., GOFFAUX, V., TARR, M., AND CROMMELINCK, M. 2002. Expertise training with novel objects leads to left-lateralized facelike electrophysiological responses. *Psychological Science 13,* 3, 250–257.

ROSSION, B., HANSEEUW, B., AND DRICOT, L. 2012. Defining face perception areas in the human brain: a large-scale factorial fmri face localizer analysis. *Brain and cognition 79,* 2, 138–157.

RYAN, K. F. AND SCHWARTZ, N. Z. 2013. Face recognition in emotional scenes: observers remember the eye shape but forget the nose. *Perception 42*, 330–340.

SADRÔ, J., JARUDI, I., AND SINHAÔ, P. 2003. The role of eyebrows in face recognition. *Perception 32*, 285–293.

SCHWANINGER, A., CARBON, C.-C., AND LEDER, H. 2003a. Expert face processing: Specialization and constraints. *Development of face processing*, 81–97.

SCHWANINGER, A., LOBMAIER, J. S., WALLRAVEN, C., AND COLLISHAW, S. 2009. Two routes to face perception: evidence from psychophysics and computational modeling. *Cognitive science 33,* 8, 1413–1440.

SCHWANINGER, A., RYF, S., AND HOFER, F. 2003b. Configural information is processed differently in perception and recognition of faces. *Vision Research 43,* 14, 1501–1505.

SCHWANINGER, A. AND YANG, J. 2011. The application of 3d representations in face recognition. *Vision research 51,* 9, 969–977.

SERGENT, J. 1984. An investigation into component and configural processes underlying face perception. *British Journal of Psychology 75,* 2, 221–242.

SINHA, P., BALAS, B., OSTROVSKY, Y., AND RUSSELL, R. 2006. Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE 94,* 11, 1948–1962.

SONG, Y., LUO, Y. L., LI, X., XU, M., AND LIU, J. 2013. Representation of contextually related multiple objects in the human ventral visual pathway.

TAN, C. AND POGGIO, T. 2013. Faces as a" model category" for visual object recognition. *MIT CSAIL Technical Report*.

TANAKA, J. AND FARAH, M. 1993. Parts and wholes in face recognition. *Journal of Experimental Psychology 46,* 2, 225–245.

VALENTINE, T. 1988. Upside-down faces: A review of the effect of inversion upon face recognition. *British journal of psychology 79,* 4, 471–491.

VAN BELLE, G., DE GRAEF, P., VERFAILLIE, K., ROSSION, B., AND LEFÈVRE, P. 2010. Face inversion impairs holistic perception: Evidence from gaze-contingent stimulation. *Journal of Vision 10,* 5.

VANGORP, P., RICHARDT, C., COOPER, E. A., CHAURASIA, G., BANKS, M. S., AND DRETTAKIS, G. 2013. Perception of perspective distortions in image-based rendering. *ACM Transactions on Graphics 32,* 4.

WALLRAVEN, C., BREIDT, M., CUNNINGHAM, D. W., AND BÜLTHOFF, H. H. 2008. Evaluating the perceptual realism of animated facial expressions. *ACM Transactions on Applied Perception (TAP) 4,* 4, 4.

WANG, R., LI, J., FANG, H., TIAN, M., AND LIU, J. 2012. Individual differences in holistic processing predict face recognition ability. *Psychological Science 23,* 2, 169–177.

WATSON, T. L. 2013. Implications of holistic face processing in autism and schizophrenia. *Frontiers in Psychology 4*, 414.

WICKENS, T. 2001. *Elementary signal detection theory*. Oxford University Press, USA.

WONG, A. C.-N., BUKACH, C. M., HSIAO, J., GREENSPON, E., AHERN, E., DUAN, Y., AND LUI, K. F. 2012a. Holistic processing as a hallmark of perceptual expertise for nonface categories including chinese characters. *Journal of vision 12,* 13.

WONG, Y. K., FOLSTEIN, J. R., AND GAUTHIER, I. 2012b. The nature of experience determines object representations in the visual system. *Journal of Experimental Psychology: General 141,* 4, 682.

XU, Y. 2005. Revisiting the role of the fusiform face area in visual expertise. *Cerebral Cortex 15,* 8, 1234–1242.

XUE, S., AGARWALA, A., DORSEY, J., AND RUSHMEIER, H. 2012. Understanding and improving the realism of image composites. *ACM Transactions on Graphics 31,* 4, 84:1–84:10.

YIN, R. K. 1969. Looking at upside-down faces. *Journal of experimental psychology 81,* 1, 141.

YOUNG, A., HELLAWELL, D., AND HAY, D. 1987. Configurational information in face perception. *Perception 16,* 6, 747–759.

Received July 2013; revised February 2014; accepted April 2014

Online Appendix to: Human Perception of Visual Realism for Photo and Computer-generated Face Images

Shaojing Fan and Rangding Wang*, Ningbo University
Tian-Tsong Ng and Cheston Y. -C. Tan, Institute for Infocomm Research
Jonathan S. Herberg, Institute of High Performance Computing; National University of Singapore
Bryan L. Koenig, Lindenwood University

A. ONLINE PILOT EXPERIMENT FOR EXPERIMENT 1

We conducted an online pilot for Experiment 1. It investigated the difference between an online experiment and a controlled lab experiment. The design and results follow.

A.1 Method

Participants: Participants were 583 laypersons and 35 experts (290 females), aged between 19-35, recruited by campus advertisements on Ningbo University of Technology, China. All participants were Asian and had normal or corrected-to-normal visual acuity. Every participant was given a souvenir (a key chain) for their participation. To encourage effort, participants were informed that those who were in the top 5% in terms of accuracy on the experimental task would receive an extra prize (a USB hub suite).

Stimuli: The preparation of stimuli was same to that in Experiment 1, except that we select only 10 pairs of CG images and photos that were not used in Experiment 1.

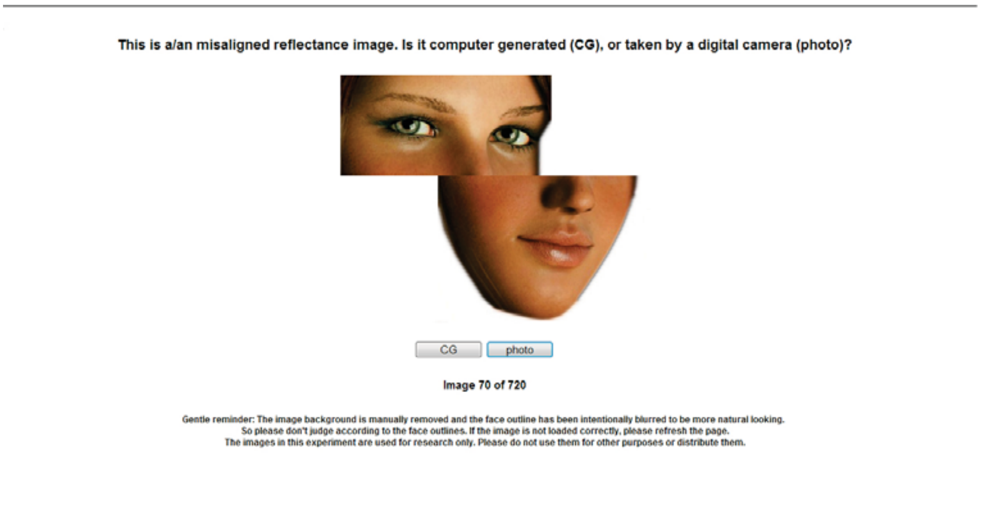
Design: There were 3 within-subjects conditions (image-type: original, grayscale, and reflectance) and four between-subjects conditions (manipulation: upright, inverted, aligned, and misaligned). Thus all subjects saw all image-types, and different groups of subjects were randomly exposed to different manipulation conditions. Participants were presented with the different image-type conditions according to one of six random orders. A near equal number of participants (25 to 27) were randomly assigned to each combination of image type, manipulation, and order. The stimulus set contained a total of 60 images (*i.e.*, 20 images $\times$ 3 types). Within each image type, image presentation order was random for each participant. The differences in design compared to Experiment 1 were intended to shorten the experiment session for online purposes.

Procedure: The procedures were similar to Experiment 1, except that participants completed the experiment online. They accessed it via a link to the study homepage. Instructions and illustrations were on the homepage (12). Participants were required to input their student ID as a unique identifier. Participants entered their gender, nationality and prior exposure to CG images. The layout of the judgment pages was the same as in Experiment 1. Participants saw 60 single images displayed sequentially on a white background. They judged each image as CG or photo in response to a two-alternative forced choice question by clicking the corresponding button below the image. There was one second interval after the button was clicked before the next image was displayed. Participants saw their overall judg-

DOI 10.1145/0000000.0000000 http://doi.acm.org/10.1145/0000000.0000000

ment accuracy at the end of experiment. Fig. 12 shows the screenshots of the introduction page and a judgment page.

(a)



(b)

Fig. 12. Screenshot for Experiment 1. (a) Introduction page; (b) Judgment page.

A.2 Results

A.2.1 *Realism Rating* . Fig.13 illustrates the realism ratings of the 10 pairs of original-upright test images. CG images were located more on the low end of realism scale, whereas photos concentrated near the high-realism end of the scale.

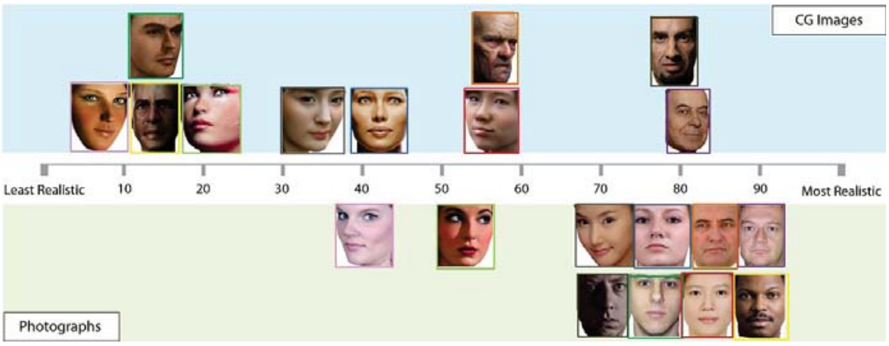


Fig. 13. Realism ratings for 10 photorealistic computer-generated faces (top) and 10 photos of faces (bottom). Paired CG/photo images are outlined with the same color.

A.2.2 *Inversion Effect* . **SDT Analysis on Inversion Effect:** Fig. 14 shows the effects of inverted (versus upright) images on d'. A repeated-measures ANOVA showed a main effect of orientation, $F(1, 614)$ = 11.37, $p < .001$, η_p^2 = .018. The 2-way interaction between orientation and alignment was non-significant, as was the 3-way interaction between image type, orientation and alignment. Interestingly, the 2-way interaction between image type and orientation was moderate but significant, $F(1, 650)$ = 3.96, $p < .05$, η_p^2 = .006. This means that the inversion effect differed across image types. We next unpack this interaction with follow-up analyses that indicated a moderating role of the loss of shading information on orientation effects.

Follow-up t-tests indicated that for both original and grayscale images inversion reduced values of d' (indicating lower performance), $|t|s(616) > 3.01$, $ps < .003$, but inversion had little if any effect on reflectance images, $|t|(616) = 1.69$, $p = .092$. **Thus the inversion effect was stronger on original and grayscale images than on reflectance images**.

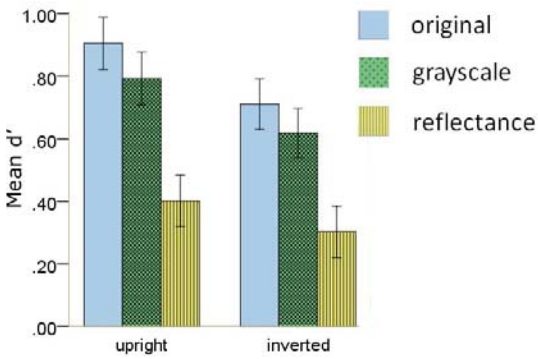


Fig. 14. Mean d' values $(+/-SD)$ for original, grayscale, and reflectance images in the upright versus inverted conditions.

Regression Analysis on Inversion Effect: A regression analysis indicated that most of the variance in inverted-image realism ratings is accounted for by the upright image realism ratings (unstandardized β = .83, SE = .03, $p < .001$, adjusted R^2 = .93). Furthermore, with the reference line $y = x$ drawn on the scatterplot, it is apparent that on most points with upright realism scores above 0.5, the inverted realism score decreases relative to the upright realism score (see Fig 15). To further evaluate this observation, we split the data into the upright images with realism ratings above 0.5 (upright-realistic group) and those below 0.5 (upright-unrealistic group), then calculated the inversion effect for each image, defined as the upright-image realism score minus the inverted-image realism score. The inversion effect did not significantly differ from zero for the upright-unrealistic group, $t(26) = -1.51$, $p = .142$. However, for the the upright-realistic group, the inversion effect was significant, $t(26) = 5.16$, $p < .001$. Thus inversion caused realism ratings for the more realistic images to regress towards chance (*i.e.*, $y = 0.5$). We refer to this finding as "*reversion effect*". In short, **inversion made visual realism judgments harder for realistic images but not for unrealistic images**. An ANCOVA showed that realism scores of upright images significantly predict the inversion effect, $F(1, 53) = 16.70$, $p < .001$, $\eta_p^2 = .242$, but image category (whether it is CG or photo) was unrelated to the inversion effect, $F(1, 53) = 1.15$, $p = .233$, $\eta_p^2 = .027$. A regression with all datapoints indicated a consistent overall increase in the inversion effect as realism increased, unstandardized β = .17, SE = .03, $p < .001$, adjusted R^2 = .37, suggesting that **the larger the realism score the larger the absolute value of the inversion effect**.

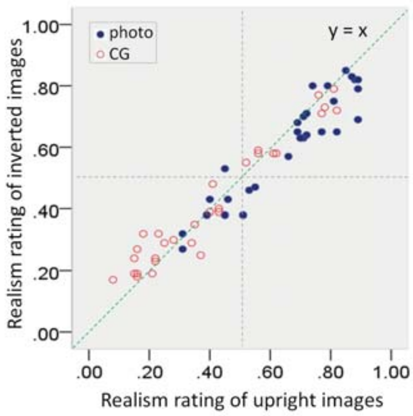


Fig. 15. (Mean realism ratings for inverted versus upright images.

A.2.3 *Alignment Effect.* For alignment, an ANOVA showed main effect on d', $F(1, 614) = 5.65$, $p < .05$, $\eta_p^2 = .009$. Interestingly, follow-up t-tests showed that misalignment caused lower performance (indicated by lower values of d') only on grayscale images, $|t|(616) = 2.57, p < .05$. Misalignment did not significantly effect original or reflectance images, $|t|s(616) < 1.82$, $ps > .070$. **Thus the misalignment effect was carried by grayscale images. As misalignment leaves the entire top and bottom half of an image intact, resulting in only the nose among all face parts being affected, misalignment had a weaker effect on both configural and piecemeal processing for realism perception than did inversion (indicated by smaller effect size)**.

A.2.4 *Image Type Analysis.* An ANOVA indicated a substantial effect of image-type on d', $F(2, 1300) = 145.22$, $p < .001$, $\eta_p^2 = .183$. For upright images, paired-sample t-tests suggested better performance (larger d') on original images than grayscale images ($t(330) = 2.11$, $p < .05$), which had better performance than reflectance images ($t(330) = 10.72$, $p < .001$). A similar overall pattern emerged for inverted images: original images had better performance than grayscale images ($t(322) = 2.57$, $p < .05$), which had better performance than reflectance images ($t(322) = 7.41$, $p < .001$).

To further analyze the image-type $\times$ orientation interaction, and whether the removal of shading versus color impacts, we calculated the effect size for each pairwise-comparison among image-type conditions separately for upright images and inverted images (see Fig. 16). Inversion generally resulted in smaller image-type effects. Inversion appears to decrease the impact of presence versus absence of image-components on visual realism perception. As the inversion manipulation was used to measure to impact of disrupting holistic processing, the results of our image-type analysis taken together suggest that **shading is more important than color**.

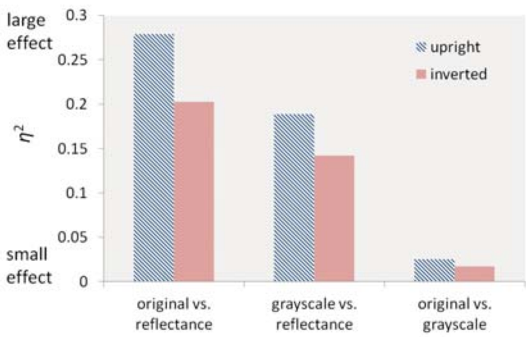


Fig. 16. Effect size η^2 on d' across image-type conditions.

A.2.5 *Self-reported Cues.* When indicating which cue was most important for judgments of whether an image was CG or a photo, most participants selected eyes, followed by glossiness, skin, color, and then shape (see Fig. 17).

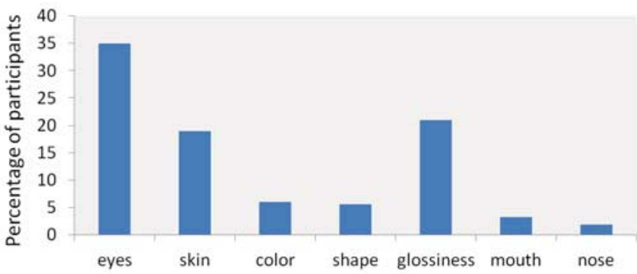


Fig. 17. Main judgment cues claimed by different percentage of participants in Experiment 1.

A.2.6 *Discussion.* To summarize, the findings from the online experiments closely replicated those from the controlled experiment. One difference was that the effect sizes for orientation and misalignment were smaller online than for the lab experiment, but this might due to different experiment design. For the online experiment, to keep the experiment to a brief duration, we used both within-subject and between-subject conditions to reduce the number of images each participant saw. Orientation and misalignment were between-subject conditions in the online experiment, but they were within-subject conditions in the controlled lab experiment. This different might explain the reduced effect size of orientation and alignment in the online experiment. We conclude that with sufficient participant size, online experiments on this topic are reliable and comparable to lab experiments.