Gender perception from Facial Structure: Categorization with and without Skin Texture and Color

Julio González-Alvarez[1] & Rosa Sos-Peña[1]

[1]Department of Basic and Clinical Psychology and Psychobiology.
University Jaume I. Castellón (Spain).

short title: Facial structure and skin texture and color.

Acknowledgements: This work was completed with resources provided by the University Jaume I of Castellon (Spain).

Correspondence should be addressed at:
Julio González-Alvarez Ph.D.
Dpt. Psicología Básica, Clínica y Psicobiología.
Av.Vicent Sos Baynat, s/n.
University Jaume I. 12071-Castellón de la Plana, Spain.

Email: gonzalez@psb.uji.es
Fax: +34 964 729267 Phone: +34 964 729670
orcid.org/0000-0002-0389-5263.

Gender Perception from Facial Structure: Categorization with and without Skin Texture and Color.

ABSTRACT

Gender/sex identification of faces without any cultural or conventional gender cue is primarily based on two independent components: a) shape or facial structure, and b) surface reflectance (skin texture and color). The present work studied the relative contribution of each component by means of two experiments based on 3D face models created with different degrees of masculinity-femininity within a gender continuum. The first experiment utilized totally artificial faces created *ex novo* by computer. The second employed face models created from photographs of real people.

The results of both experiments were consistent. As expected, when both components were present in a face, gender was correctly classified in almost all the cases. More interestingly, the contribution of the "pure" facial structure to the gender perception (with no surface reflectance) was about 80%, whereas 20% of the total information was provided by the surface reflectance. Furthermore, examination of the psychometric curves suggests that the information provided by surface reflectance contributes to a categorical perception of facial gender, since when it is removed the gender is perceived in a more continuous / less categorical way. On the other hand, our stimuli presented a certain "male" bias, repeatedly found in the literature on facial gender perception.

KEYWORDS: Gender perception, Facial structure, Face Models, Skin, Perception.

Research has demonstrated that human faces are processed differently compared with other objects and are perceived holistically (for reviews, see Richler & Gauthier, 2014; Rossion, 2013; Zhao, Bülthoff, & Bülthoff, 2016). This implies that faces are perceived as indecomposable wholes, or "gestalts", rather than an assembly of independent facial features. Furthermore, human faces are a significant source of personal information. Each face is unique and is the most important body part in identity determination (Bruce, 1990). Experimental evidence has revealed that people form first impressions about someone's personality after viewing their face for a minimal time exposure (100 ms) (Willis & Todorov, 2006), although this does not necessarily mean that the impression is correct (Todorov & Porter, 2014). Nevertheless, among the most reliable information that a human face provides is the sex or gender of its owner. In a sociocultural context, there are several conventional signals that indicate a person's gender through their faces, such as hairstyle, facial hair, clothing (hat, bonne), accessories, makeup, etc. However, research shows that even when all these traditional cues are removed, people are still able to identify gender with 96–98% accuracy from facial stimuli (Bruce et al., 1993; Bruce, Ellis, Gibling, & Young, 1987; Saether, Van Belle, Laeng, Brennen, & Øvervoll, 2009).

Gender/sex classification of "bare" faces—without any cultural or conventional gender cue—is primarily based on two independent components: a) shape or facial structure, and b) surface reflectance (skin texture and color) (Meinhardt-Injac, Persike, & Meinhardt, 2013; Russell, Sinha, Biederman, & Nederhouser, 2006). For a long time, many anthropometric studies have highlighted the differences between male vs. female skulls and facial structures. Human craniofacial dimorphism is well documented, for example, by Iscan and Steyn (2013), White, Black, and Folkens (2011), or Tanikawa, Zere, and Takada (2016). Thus, male skulls/faces tend to have stronger muscle attachments and larger mastoid processes, brow ridges, glabellar regions, thicker supraorbital margins, a more sloping forehead, more rectangular orbits, and heavier zygomatic bones. Seemingly, these differences between male vs. female skulls/faces are related to endogenous testosterone and develop during puberty in males (Enlow & Hans, 1996). Interestingly, some of the differences have been observed in faces of female-to-male transsexual people, being associated with the administration of exogenous testosterone and bilateral oophorectomy (removal of ovaries) (Mackenzie, & Wilkinson, 2017). On the other hand, females tend to show a greater vertical height of the eye fissure, shorter postero-anterior height of the nasal tip, vertically greater supraorbital ridge, shorter lower face height, a smaller nasal hump, more prominent cheeks in the infraorbital region and less so in the buccal region (Tanikawa, Zere, & Takada, 2016).

The other face component important for gender classification is the surface reflectance or pigmentation, which provides information about skin texture and color. Sexual dimorphism also exists for surface reflectance. For instance, female skin is generally lighter than male skin within the same ethnic group—an observation that dates back to Darwin (1871) and was later measured objectively (Edwards & Duntley 1939; Jablonski & Chaplin, 2000: Nestor & Tarr, 2008). More specifically, female faces tended to present greater luminance contrast between the eyes, lips, and the surrounding skin compared with male faces (Russell, 2009). Indeed, an androgynous face is perceived as female by increasing the facial contrast, or perceived as male by decreasing the facial contrast (Russell, 2009).

Taking both components—facial structure and skin (texture and color)—together for one face, even non-experts can correctly classify sex/gender of the person without the help of any cultural gender cues. Our question is whether the facial structure alone, once all the information on the skin (including eyebrows, eyelashes, lip color, …) has been removed, is sufficient to classify a person's gender, and if so with what degree of accuracy. In a pioneering effort, Bruce et al. (1993) used three-dimensional (3-D) representations of faces obtained by laser-scanning people wearing swimming caps to conceal hair. Almost three decades later, computer technology now allows experimental manipulations unknown at that time. In a first experiment, we used 3D computer-modified artificial faces to create stimuli with different degrees of masculinity-femininity within a gender continuum. We produced three versions of each stimulus, one version with the shape of the face and all the skin information, and two without skin information, one for the forehead and the other foreshortening. In a second experiment, we replicated the procedure using 3D face models created from photographs of real people.

EXPERIMENT 1

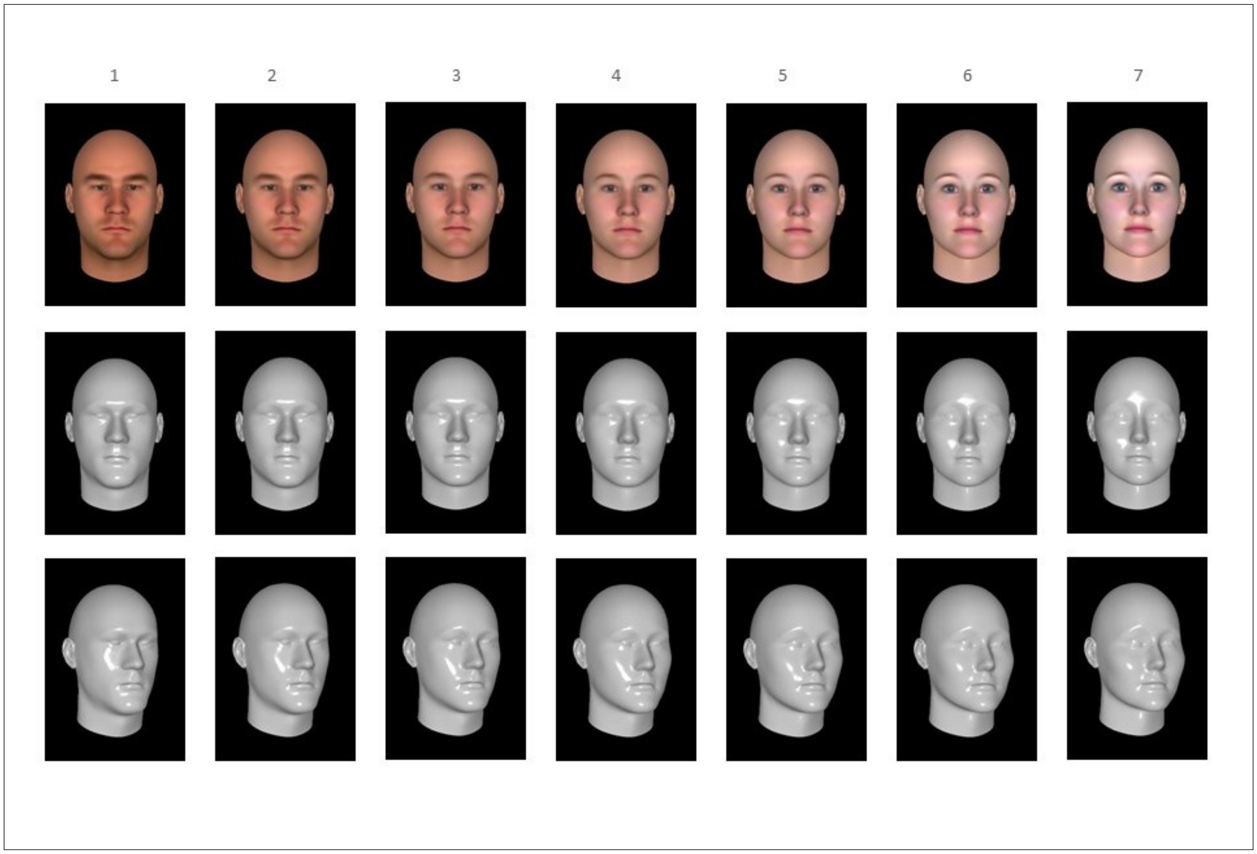
The aim of the first experiment was to study how people categorized the sex/gender of a set of 3D computer-modified artificial faces created with different degrees of masculinity-femininity within a gender continuum.

METHOD

Participants

Participants were 43 adults of both sexes (31 females) whose age range was 18–39 years (M = 23.14; SD = 5.86). All were undergraduates at the University Jaume I (Spain), who participated voluntarily and were compensated with course credit. Of those who indicated ethnicity (n=41), 78% were White/Caucasian, 17% were Hispanic/Latin American, and 5% were another ethnicity.

Figure 1. Experiment 1. Examples of stimuli derived from a single artificial face (not displayed) created with the FaceGen Modeller software. First row: front faces with full skin texture/color (F_SK); second row: front skinless faces (F_nSK); and third row: skinless faces presented at a 20º head angle (A_nSK). The seven versions show a gender continuum ranging from a "hyper-male" pole (version 1) to a "hyper-female" pole (version 7), in steps of 15 points in the general gender control of FaceGen. More details in the text.



Materials

The stimuli comprised 210 realistic 3D artificial face models created with the FaceGen Modeller Core 3.25 software (https://facegen.com/; Singular Inversions, 2010, 2020). There were 70 front faces with skin texture and color (henceforth referred to as F_SK), 70 front faces without skin texture/color (F_nSK), and 70 faces without skin texture/color, all displayed at a 20º head angle (A_nSK). FaceGen is a software that generates realistic three-dimensional faces whose shape (facial structure), skin texture/color, and facial expression can be adjusted on multiples dimensions. According to the designers, FaceGen was based on a face space created by PCA (Principal Components Analysis) from a dataset comprising 273 human faces that were laser-scanned in high-resolution 3D. This software includes age, gender and racial group controls based on linear regressions on the dataset in that face space, and it has been widely used in research on face perception (e.g., Cenac, Biotti, Gray, & Cook, 2019; González-Alvarez, 2017; Kihara & Takeda, 2019; Oh, Dotsch, & Todorov, 2019; Oosterhof & Todorov, 2008; Todorov, Pakrashi, & Oosterhof, 2009).

The stimuli were created in five steps as follows. First, FaceGen randomly generated 10 different artificial faces of apparently different ethnicities (not included as experimental stimuli) with full skin texture and color. Second, each face was modified by adjusting the general gender control of FaceGen (which ranges 100 steps, from a hyper-male to a hyper-female pole) to a central position generating a “neutral” or ambiguous face from a gender perspective. Third, this “neutral” or androgynous face constituted an experimental stimulus corresponding to the central version 4 (see example in Figure 1, first row). It served as seed to a set of 3 x 7 = 21 faces. To this end, we generated the “female” versions 5, 6, and 7 (+15, +30, and +45 points respectively) by adjusting the general gender control in steps of +15 points from the central position. Further, we generated the “male” versions 3, 2, and 1 (-15, -30, and -45 points respectively), by adjusting the general gender control in steps of -15 points from the central position. In this manner (Fig. 1, first row), we created a gender continuum of seven versions ranging from version 1 (hyper-male pole) to version 7 (hyper-female pole), going through the neutral or ambiguous version (4). The fourth step entailed generating a skinless white face (Fig. 1, second row) of each of the previous versions, thus preserving the same facial structure, but without skin texture/color. The fifth and final step was to rotate each of the previous skinless faces by an angle of 20º in order to display a foreshortened image (Fig. 1, third row). Thus, from a single original face we obtained seven versions of each of the three face modalities (F_SK, F_nSK, A_nSK) with identical facial structure between them. The modalities only varied in terms of the existence or non-existence of skin texture/color and in the angle of presentation. This process was repeated on each of the 10 original faces, leading to a set of 210 stimuli.

Procedure.

The introductory instructions stated [in Spanish]: “Artificial faces created by computer of adults of both genders (man, woman) will be presented, and your task will be to decide in each case if it is a Man or a Woman.”

This was followed by: “In each trial, we will present to you an adult artificial face, without hair or accessories. You must decide if it is a Man or a Woman. You must let yourself be carried away by the first impression. The face can have a normal appearance (with skin, eyebrows, eyes, etc.) …” [an example not included in the experimental set was presented], “or appearance without skin and other features (only facial structure). It can appear head-on…” [example presented], “or foreshortened (slightly turned)” [example presented].

After a few practice trials which did not include stimuli within the experimental set, each participant was shows all 271 experimental stimuli in random order over three sessions separated by two breaks. The task was completed individually online through the university intranet (virtual classroom). Previous research on face perception has demonstrated that laboratory and online studies yield equivalent results (e.g., DeBruine, Jones, Unger, Little, & Feinberg, 2007; Lefevre, Ewbank, Calder, von dem Hagen, & Perrett, 2013).

During each trial, the computer displayed a face in a size 9 x 13 cm, and the participant had to choose between two options (Man or Woman) by ticking one of them with a mouse. There was no time restriction.

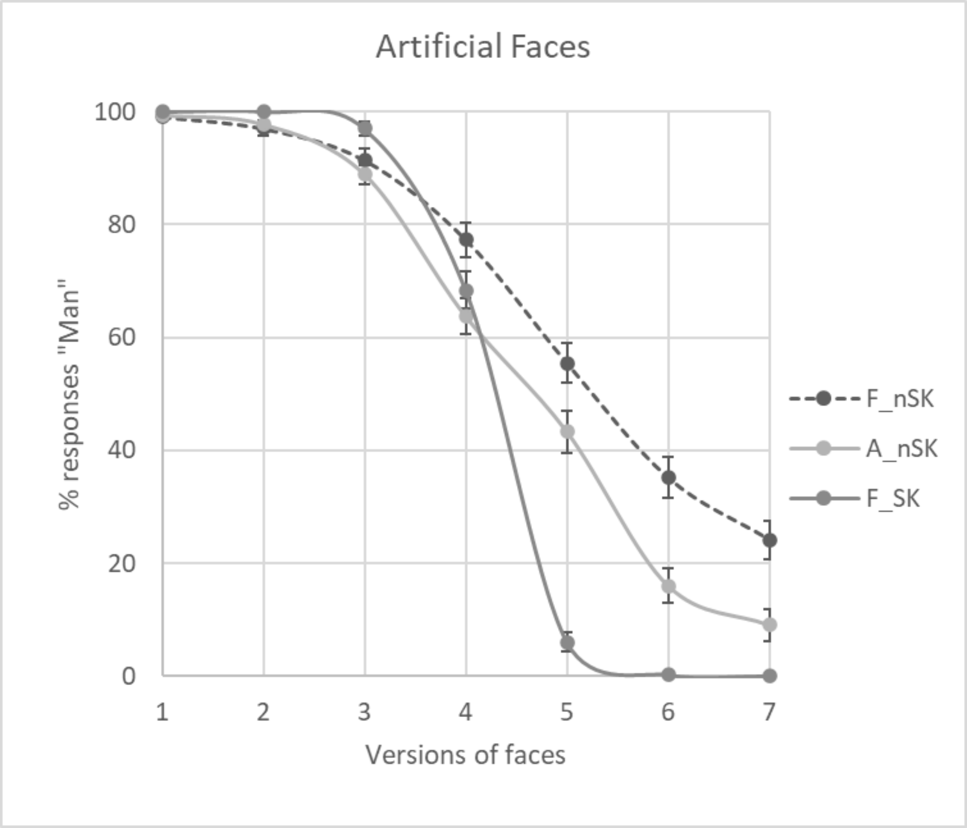
RESULTS AND DISCUSSION

We obtained the percentages of "Man" responses for each condition (the selection was arbitrary; we could also have chosen the "Woman" responses). Table 1 presents the means (and SD, standard deviations between parentheses) of the percentages of "Man" responses to the seven face versions for each of the three types of faces: front faces with full skin texture/color (F_SK); front skinless faces (F_nSK); and skinless faces presented at a 20º head angle (A_nSK). Figure 2 displays the three psychometric curves, one for each type of face, showing the typical sigmoid shape (in our case, reversed) of functions based on an identification task upon a stimulus continuum. Bar errors indicate the standard error means (± SEM) and allow the observation of the significant differences between versions and face types. As expected, the psychometric curves present different slopes, with those belonging to the F_SK faces being steeper, approximating a categorical function with extreme values on both sides and a very sharp transition ($b = - 45.5$, from versions 3 to 5) between the two gender categories. The other two types of faces (F_nSK and A_nSK) present less pronounced curves ($b = -18.0$; $b = -22.8$, respectively), suggesting less categorical/more continuous functions, with a higher overall degree of gender ambiguity as a result of loss of information provided by skin texture and color. Interestingly, the curve of A_nSK faces is closer to the curve of the full skin texture/color (F_SK) compared with F_nSK. The fact that the skinless faces are foreshortened and displayed at an angle of 20º provide more information about the depth and the third dimension (front-back axis) of the facial structure.

Table 1. Data from Experiments 1 and 2. Means (SD in parentheses) of the percentages of "Man" responses to the seven versions of faces. Data separated by face types; F_SK: front faces with full skin texture/color; F_nSK: front skinless faces; A_nSK: skinless faces presented at a 20º head angle.

	Face Versions						
	1	2	3	4	5	6	7
Experiment 1:							
F_SK:	100 (0.0)	100 (0.0)	97.0 (7.7)	68.4 (21.4)	6.0 (10.9)	0.2 (1.5)	0.0 (0.0)
F_nSK:	99.1 (3.7)	97.0 (8.3)	91.4 (12.8)	77.3 (20.0)	55.5 (23.4)	35.2 (23.5)	24.2 (22.3)
A_nSK:	99.3 (2.6)	97.7 (4.8)	88.8 (11.4)	63.7 (20.8)	43.3 (24.6)	16.0 (19.7)	9.1 (18.2)
Experiment 2:							
F_SK:	99.5 (2.2)	100 (0.0)	97.5 (4.9)	77.3 (14.3)	26.0 (16.9)	3.0 (6.0)	0.3 (1.7)
F_nSK:	99.0 (3.0)	98.0 (4.6)	87.3 (12.2)	72.5 (15.5)	45.3 (19.0)	28.0 (16.3)	19.4 (17.1)
A_nSK:	100 (0.0)	97.8 (5.2)	91.5 (10.4)	68.0 (20.8)	39.5 (20.7)	22.1 (15.7)	8.5 (13.7)

Figure 2. Experiment 1: Psychometric curves for “Man” responses for each of the three types of faces; F_SK: front faces with full skin texture/color; F_nSK: front skinless faces; A_nSK: skinless faces presented at a 20º head angle. Error bars indicate ± SEM (standard error of the mean).



A 3 (Face Type, F_SK, F_nSK, A_nSK) x 7 (Face Versions) x 2 (Sex of Participants) analysis of variance (ANOVA) was conducted with the data. As expected, the ANOVA revealed a strong Face Type effect, $F(2, 82) = 35.80$, $MS_e = 389.68$, $p < .001$, $\eta^2_p = .466$[1]. Additionally, as expected, a robust Face Version effect was obtained, $F(6, 246) = 747.96$, $MS_e = 202.71$, $p < .001$, $\eta^2_p = .948$. Interestingly, a minor but significant between-subject effect was obtained from the Sex of Participants factor, $F(1, 41) = 4.59$, $MS_e = 167.69$, $p = .038$, $\eta^2_p = .101$ due to the fact that women tended to identify faces as masculine in slightly higher proportion than men did (overall percentages 62.0% vs. 56.5%, respectively). The Face Versions x Sex of Participants interaction was also significant because the between-sex discrepancy of participants tended to be higher in the more ambiguous versions, $F(6, 246) = 8.69$, $MS_e = 202.71$, $p < .001$, $\eta^2_p = .175$. As expected, the Face Type x Face Versions interaction was significant because the three psychometric curves were not parallel through the seven versions, diverging mostly in the "female" versions (see Figure 2), $F(12, 492) = 41.41$, $MS_e = 108.34$, $p < .001$, $\eta^2_p = .502$.

Accuracy. Considering that the face versions 1, 2, and 3 belong to the "male" side of the gender continuum, we scored the "Man" responses to these versions as correct. Conversely, considering that the face versions 5, 6, and 7 belong to the "female" side of the gender continuum, we scored the "Woman" responses to these versions as correct (responses to the ambiguous version 4 were not considered). Altogether, the front faces with full skin texture/color (F_SK) achieved an accuracy level of 98.4% [95% CI (97.7, 99.10)] correct responses (see Figure 5), whereas the skinless front faces (F_nSK) obtained 78.8% (75.9, 81.7) correct responses, losing almost 20% of accuracy because the gender judgments had to be made on "pure" facial structure devoid of any superficial information on the texture or color of the skin. However, when the skinless faces were displayed obliquely (A_nSK), giving more information about the spatial depth of facial structure, they gained approximately 8–9% accuracy, at 86.2% (83.6, 88.8).

Interestingly, the loss of skin information affected gender identification of “male” versions considerably less than in the case of “female” versions. Therefore, male versions with full skin texture/color (F_SK, versions 1, 2, 3) reached a very high 99% [95% CI (98.2, 99.8)] correct responses, and the front skinless faces (F_nSK, versions 1, 2, 3) only lost about 3%, being up to 95.8% (93.6, 98.0) correctly identified. The female versions with full skin texture/color (F_SK, versions 5, 6, 7) also achieved—in a categorical perception—a very high accuracy of 97.9% (96.8, 99.0) but lost a significant 36% of accuracy, going down to 61.7% (55.6, 67.8) when the faces did not provide skin information (F_nSK, versions 5, 6, 7). In this last instance, the angular presentation (A_nSK, versions 5, 6, 7) gained 10.5% accuracy, the value increasing to 77.2% (71.7, 82.7).

We submitted the accuracy data to a 3 (Face Type, SK, FW, AW) x 6 (Face Versions, version 4 not included) x 2 (Sex of Participants) ANOVA. As expected, we obtained a strong Face Type effect, $F\ (2, 82) = 107.44$, $MS_e = 206.28$, $p < .001$, $\eta^2_p = .724$, as well as a robust Face Version effect, $F\ (5, 205) = 59.41$, $MS_e = 242.80$, $p < .001$, $\eta^2_p = .592$. However, we did not find any effect of the Sex of Participants factor, $F\ (1,41) < 1$, because men and women did not differ in their overall level of accuracy on the gender-identification task. Of all the possible interactions, we only found an effect from the Face Type x Face Versions interaction because, as seen above, the accuracy differences between the types of faces did not remain constant through the seven versions, $F\ (10, 410) = 27.93$, $MS_e = 119.28$, $p < .001$, $\eta^2_p = .405$.

This study utilized entirelly artificial faces created *ex novo* by computer. Recently, some publications have reported some evidence suggesting possible processing differences between artificial faces and photos of real faces, both at the cognitive (Balas & Pacella, 2015; Carlson, Gronlund, Weatherford, & Carlson, 2012; Kätsyri & Sams, 2008; see also González-Álvarez &

Cervera-Crespo, 2019) and neural levels (Mühlberger et al., 2009; Wheatley, Weinberg, Looser, Moran, & Hajcak, 2011). It seems that artificial faces, overall, are more difficult to recognize and remember than natural faces, and some basic emotions provoked by artificial faces could be recognized and processed differently (Ehrlich, Schiano, & Sheridan, 2000; Kätsyri & Sams, 2008). In addition, it is noteworthy that contemporary psychological science, particularly within the social field, is encountering a certain crisis of reproducibility (Open Science Collaboration, 2015; see also Baucal, Gillespie, Krstić, & Zittoun, 2020). There is a general concern about the robustness and replication success of experimental results. Consequently, we wanted to test the robustness of our results by carrying out another experiment using face models created from natural photographs of real people.

EXPERIMENT 2

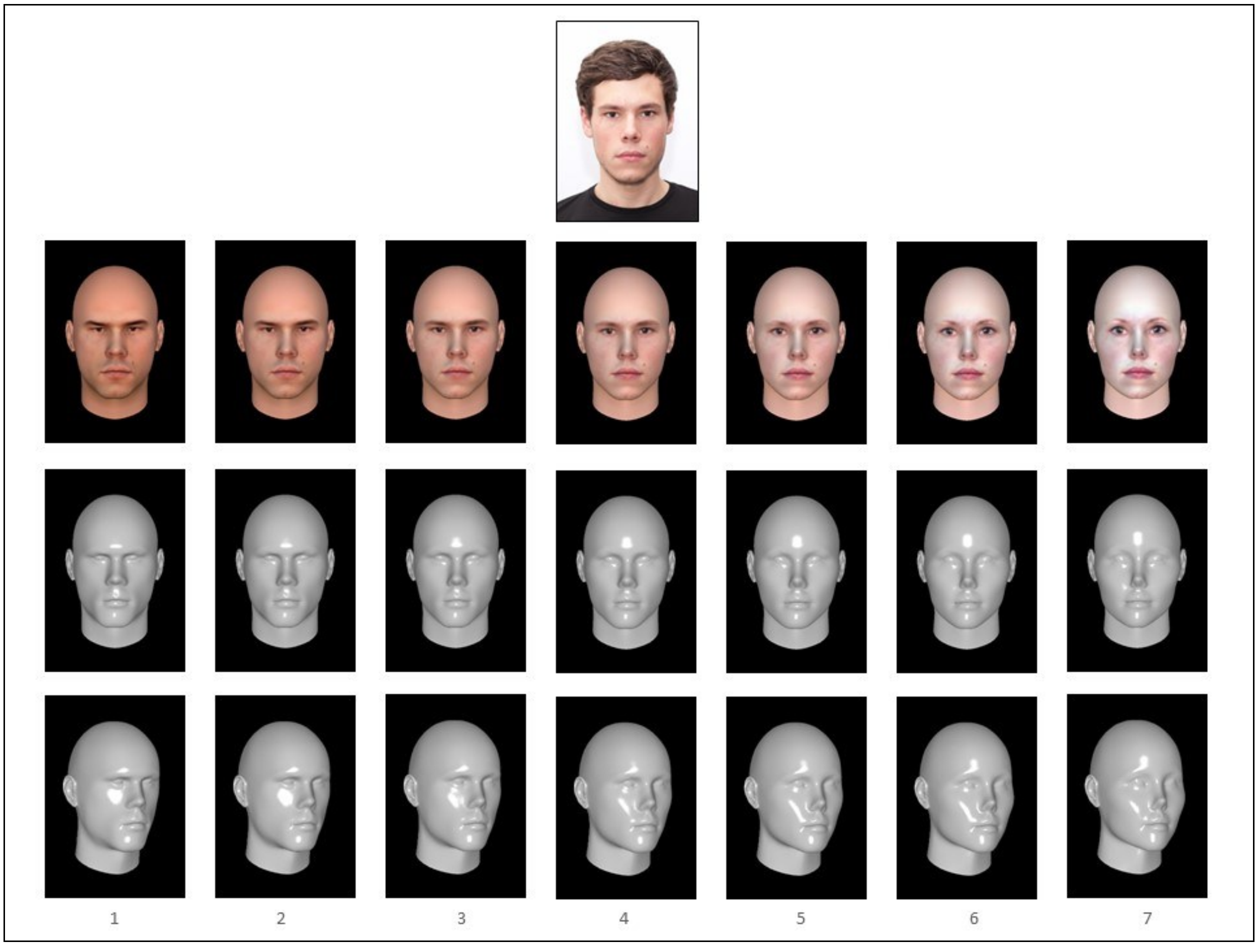
This experiment tested the robustness of the findings obtained in Experiment 1, extending the research using another type of stimuli that involved 3D face models created from photographs of real people.

METHODS

Participants.

Participants were 40 adults of both sexes (31 females) whose age range was 18-35 years (M = 20.95; SD = 3.01). None had participated in Experiment 1. All were undergraduates at the University Jaume I (Spain), who voluntarily participated in exchange for course credit. All indicated ethnicity, with 75% being White/Caucasian, 18% Hispanic/Latin American, and 7% of another ethnicity.

Figure 3. Examples of stimuli presented in the Experiment 2 and derived from natural photograph (top) of a real person selected from the Basel Face Database (Walker, Schönborn, Greifeneder, & Vetter, 2018). The face models of each row were generated and modified with the FaceGen software following the same procedures as in Experiment 1. More details in the text.



Materials.

The stimuli consisted of 210 face models created from natural photographs of real people: 70 front faces with full skin texture and color (henceforth referred to as F_SK), 70 front faces without skin texture/color (F_nSK), and 70 white faces without skin texture/color were created and displayed at a 20º head angle (A_nSK). The procedure was the same as in Experiment 1, except for the first step. In the first step, ten photographs, five of men and five of women, all Caucasian except two people, were selected from the Basel Face Database (Walker, Schönborn, Greifeneder, & Vetter, 2018) from the University of Basel, Switzerland, (photographs of individuals 1, 3, 5, 11, and 20 for men and 6, 18, 19, 26, and 35 for women). They were inputted in FaceGen software and transformed into three-dimensional face models. In a second step, as in Exp. 1, each face model was modified by adjusting the general gender control of FaceGen to a central position generating a "neutral" or ambiguous face from a gender perspective. The subsequent steps were the same as in Exp. 1. Thus, from each photograph we obtained seven versions x three face modalities (F_SK, F_nSK, A_nSK) with identical facial structure for each version, the modalities only varying in terms of the existence or non-existence of skin texture/color and in the angle of presentation (see an example in Figure 3).

Procedure.

The procedure was the same as in Experiment 1.

RESULTS AND DISCUSSION

As was done in Experiment 1, we obtained the percentages of "Man" responses from each condition. Table 1 shows the means (and SD, standard deviations between parentheses) of the percentages of "Man" responses to the seven face versions for each of the three types of faces: front faces with full skin texture/color (F_SK); front skinless faces (F_nSK); and skinless faces presented at a 20º head angle (A_nSK). Figure 4 presents the three psychometric curves, one for each type of face, showing also the typical sigmoid shape (reversed in our case) of functions based on an identification task upon a stimulus continuum. Bar errors also specify the standard error means (± SEM) and allow the observation of the significant differences between versions and face types. Additionally, as expected, the psychometric curves present different slopes, the one belonging to the F_SK faces being steepest, approximating a categorical function with extreme values on both sides and a sharp transition ($b = - 35.8$, from versions 3 to 5) between the two gender categories. It is noteworthy that this transition is a little less sharp than in the case of the artificial faces in Experiment 1 ($b = - 45.5$). As in Experiment 1, the other two types of faces (F_nSK and A_nSK) present less pronounced curves ($b = -21.0$ and $b = -26.0$, respectively), suggesting less categorical/more continuous functions. They also present a higher overall degree of gender ambiguity as a result of the loss of information provided by skin texture and color.

Figure 4. Experiment 2: Psychometric curves for “Man” responses for each of the three types of faces; F_SK: front faces with full skin texture/color; F_nSK: front skinless faces; A_nSK: skinless faces presented at a 20º head angle. Error bars indicate ± SEM (standard error of the mean).

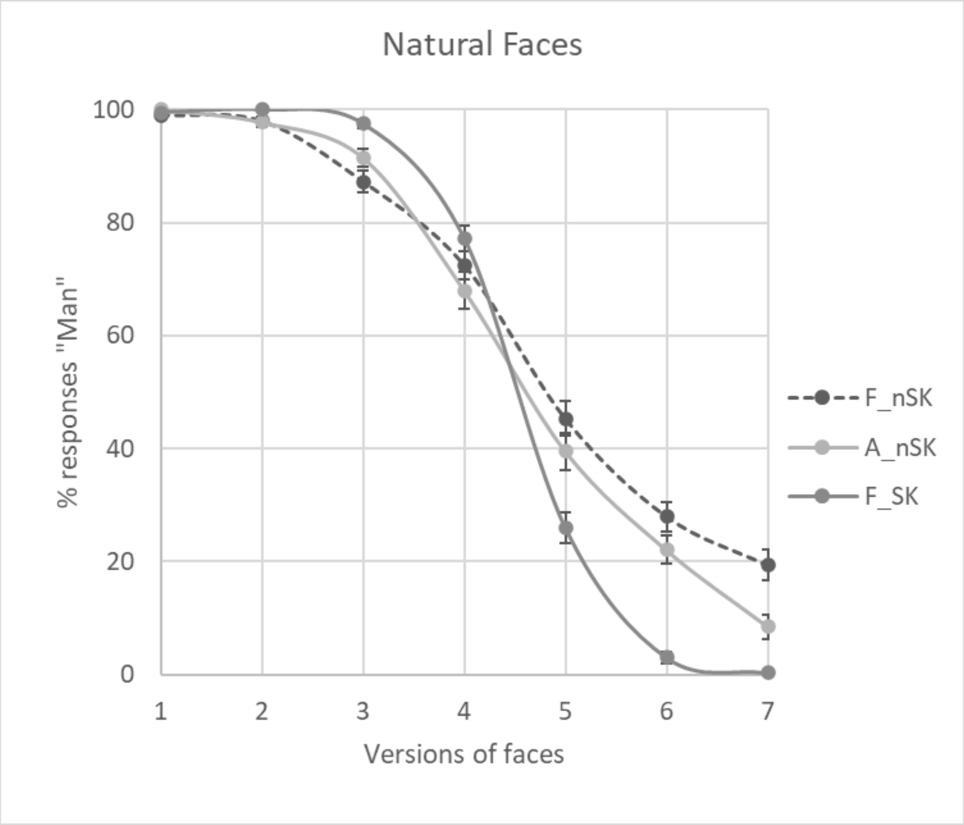
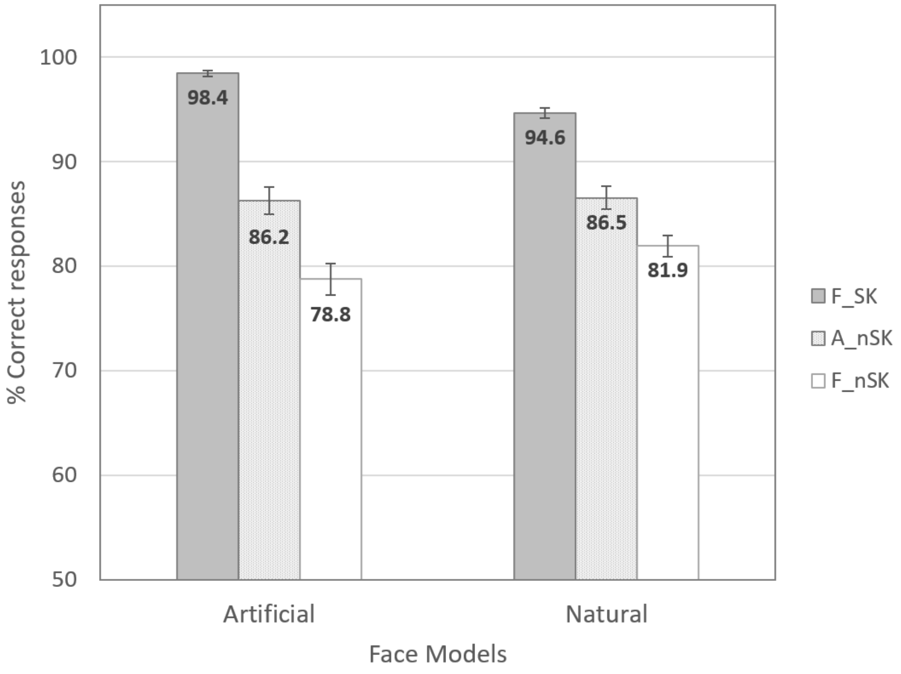


Figure 5. Percentages of correct responses for Artificial (Experiment 1) and Natural (based) face models (Experiment 2) for each of the three types of faces; F_SK: front faces with full skin texture/color; F_nSK: front skinless faces; A_nSK: skinless faces presented at a 20º head angle. Error bars indicate ± SEM (standard error of the mean).



A 2 (Photo Gender) x 3 (Face Type, F_SK, F_nSK, A_nSK) x 7 (Face Versions) x 2 (Sex of Participants) ANOVA was performed. This time the ANOVA included an additional factor not used in the Exp. 1: Photo Gender, or the gender of the people portrayed in the photographs from which the facial models were derived. The analysis revealed that this Photo Gender factor showed a significant effect, $F(1, 38) = 44.67$, $MS_e = 168.09$, $p < .001$, $\eta^2_p = .540$. Although each photo selected from the Basel Face Database was transformed into a “gender-neutral" face model, the original gender of the photo had a certain influence such that the male photos resulted, on average, in a score 4.5% more masculine than the female photos (63.2% vs. 58.7%, respectively). As expected, the ANOVA also revealed a Face Type effect, F (2, 76) = 6.27, MS_e = 570.19, p = .003, $\eta^2{}_p$ = .142, although with a smaller effect size than in the case of artificial faces (Exp. 1, $\eta^2{}_p$ = .466). Additionally, as expected, a robust Face Version effect was obtained, F (6, 228) = 944.73, MS_e = 279.29, p < .001, $\eta^2{}_p$ = .961. The Sex of Participants factor yielded a marginal significance, F (1, 38) = 3.86, MS_e = 202.37, p = .057, $\eta^2{}_p$ = .092 (and the significance would probably be high if the number of men was greater than nine, observed power: 0.482) due to the fact that, again as in Exp. 1, women tended to identify faces as masculine in a slightly higher proportion compared with men (overall percentages 61.9% vs. 57.6%, respectively). As in Exp. 1, the Face Type x Face Versions interaction was significant as indicated by the three psychometric curves not being parallel through the seven versions (see Figure 4), F (12, 456) = 13.74, MS_e = 201.82, p < .001, $\eta^2{}_p$ = .266.

Analyzing the data obtained from Exp. 1 using artificial faces (artificial) and Exp. 2 using stimuli derived from natural faces (natural), an ANOVA did not reveal any significant effect of the class of stimuli (artificial vs. natural), F (1,79) < 1. However, a Sex of Participants effect was clearly significant, attesting to the fact that in both experiments, women tended to give to the stimuli a higher masculinity score than men did; F (1, 79) = 8.15, MS_e = 135.67, p < .01, $\eta^2{}_p$ = .094.

Accuracy. As in the first experiment, we scored the “Man” responses to versions 1, 2, and 3, belonging to the “male” side of the gender continuum as correct. Conversely, we scored the “Woman” responses to the versions 5, 6, and 7, belonging to the “female” side as correct.

Altogether, the front faces with full skin texture/color (F_SK) achieved an accuracy level of 94.6% [95% CI (93.5, 95.7)] correct responses (see Figure 5). The skinless front faces (F_nSK) obtained a success of 81.9% (80.0, 83.8), losing almost 13% of accuracy because, again as in Exp. 1, the gender judgments had to be made on the "pure" facial structure devoid of any superficial information on skin texture or color. When the skinless faces were displayed obliquely (A_nSK), presenting more information about the spatial depth (third dimension) of facial structure, they gained approximately 4–5% of accuracy at 86.5% (84.4, 88.6).

Interestingly, again as in Exp.1, the loss of skin information affected the gender identification of "male" versions considerably less than it did of "female" versions. Thus, male versions with full skin texture/color (F_SK, versions 1, 2, 3) achieved a very high 99% [95% CI (98.4, 99.6)] of correct responses, and the front skinless versions (F_nSK, versions 1, 2, 3) only lost 4.7% to reach 94.7% (93.1, 96.3). The female versions with full skin texture/color (F_SK, versions 5, 6, 7) reached an accuracy of 90.2% (88.0, 92.4), but they lost 21% accuracy, the value decreasing to 69.1% (64.7, 73.5) when the faces did not provide skin information (F_nSK, versions 5, 6, 7). In this last case, the angular presentation (A_nSK, versions 5, 6, 7) gained 7.5% accuracy, the value increasing to 76.6% (72.1, 81.1). Accordingly, it is noteworthy that both in Exp. 1 and in Exp. 2, some participants informally reported that some skinless faces of the female versions were perceived as young men and therefore classified as masculine; that is, when the information on the skin texture/color is missing and only the information provided by the "pure" facial structure remains, it seems that a certain interaction between gender x age is more likely to emerge. Thus, the female facial structure is more likely to be perceived as belonging to a younger human.

We submitted the accuracy data to a 2 (Photo Gender) x 3 (Face Type, SK, FW, AW) x 6 (Face Versions, version 4 not included) x 2 (Sex of Participants) ANOVA. As above, the ANOVA included the additional Photo Gender factor, not applicable in Exp. 1. Analysis revealed that the Photo Gender factor showed a significant effect, F (1, 38) = 23.53, MS_e = 140.96, p < .001, η^2_p = .382, because stimuli derived from photos of women were slightly better classified (89.6%) than stimuli derived from photos of men (85.8%). As expected, we obtained a Face Type effect, F (2, 76) = 46.71, MS_e = 278.46, p < .001, η^2_p = .551, because, as we saw above, the stimuli with skin information (F_SK) yielded higher accuracy than skinless (F_nSK, A_nSK) stimuli (Figure 5). Also as expected, we obtained a Face Version effect, F (5, 190) = 75.98, MS_e = 339.41, p < .001, η^2_p = .667, because less ambiguous (more extreme) versions achieved higher accuracy. Unlike in Exp.1, the Sex of Participants factor showed a marginal significance, F (1, 38) = 3.88, MS_e = 95.87, p = .056, η^2_p = .093, with men yielding an overall level of accuracy (90.0%) 3% higher than women (87.0%). We also found an effect from the Face Type x Face Versions interaction because, as in Exp. 1, the accuracy differences between the types of faces did not persist consistently through the seven versions, F (10, 380) = 6.67, MS_e = 207.00, p < .001, η^2_p = .149.

Grouping the accuracy data from Exp. 1 (using artificial faces) and Exp. 2 (using stimuli derived from natural faces), an ANOVA showed no significant effect of the class of stimuli (artificial vs. natural), F (1,79) < 1. It also did not show a Sex of Participants effect, F (1,79) < 1, because men and women did not differ in their overall level of accuracy.

GENERAL DISCUSSION

Our results indicate that facial structures with full information on the texture and color of the skin are correctly classified as to their gender by most of the participants (98.4% for Exp. 1 and 94.6% for Exp. 2). If we do not consider versions 3 and 5 (close to the androgyne version 4), which contain a certain degree of gender ambiguity and only consider the less ambiguous versions (1 and 2 for male faces, and 6 and 7 for female faces), the accuracy approaches the ceiling (99.9% for Exp.1, and 99.1% for Exp.2). This is in line with previous research which observed that natural faces, devoid of any cultural signs of gender, are generally correctly categorized into their sex/gender (Bruce, et al., 1987; Bruce et al., 1993; Saether, et al., 2009).

The really interesting aspect is observed when all the information provided by the texture and color of the skin (including eyebrows, eyelashes, lip color, freckles or skin blemishes, etc.) is experimentally removed and the participants must judge the sex/gender of the faces based only on facial structure. In this case, our data shows that accuracy decreased to 78.8–81.9% for both experiments, respectively. This implies that the skin or surface reflectance provides 13–20% information about the sex/gender of the face. Bruce et al. (1993) studied sex categorization for natural photographs of people, without conventional gender cues and wearing swimming caps to conceal hair as well as for 3D representation of the same faces obtained by laser-scanning, devoid of their normal surface markings and texture. Accuracy dropped from 96% for natural photographs to 75% for the laser-scanned faces, suggesting that skin provided 21% information about the face sex/gender. The authors observed that most errors were concentrated on the laser scans of the female heads, with many participants perceiving them as male heads. Bruce et al., (1993) stated the following: “the full-face laser heads look male, and this bias disadvantages identification of the female faces while marginally benefitting identification of the male heads (performance on the male heads was slightly more accurate with laser than with natural format)”

(p. 141). They also wondered if this effect was an artifact of the laser-scanning technique: "perhaps this bias is itself the result of superficial cues present in the laser scans but not the natural heads […], the reconstruction of the images from laser-scan data introduces some wrinkling which could be interpreted as the rougher skin texture of a male rather than a female head." (p. 141). The technique that we used in the present study to create the facial models is more sophisticated and does not introduce such artifacts (see figures 1 and 3), but nevertheless our data showed a similar pattern. Accuracy for skinless front faces in the male versions (F_SK, versions 1, 2, 3) was very high: 96–95% for both experiments, whereas in the female versions (F_SK, versions 5, 6, 7) it was mediocre: 62–69% for both experiments.

Similarly, our data reveals that the "neutral" or androgynous face corresponding to version 4 is categorized as "Man" by more than 50% of judgments in all the facial types (Table 1, fourth column); that is, our stimuli present a "male" bias, as in Bruce et al.'s (1993) work. This bias has also been found by other authors. Haseltine (2007) observed that ambiguous faces were rated as males significantly more often and faster than they were rated as females. Sobey (2014) in her study used a morphing procedure to create stimuli of varying degrees of sexual ambiguity and found a predisposition to categorize ambiguous faces as males, with this effect being independent of the sex of the participants. Armann and Bülthoff (2012) ended the title of their paper with the revealing expression "… and when in doubt, he is a male." They utilized a morphing technique to create stimulus continua of faces (with skin texture and color) and observed a bias revealing that participants were more likely to classify a face as male than as female when in doubt, even when the stimuli belonged to the female side of the stimulus continuum. The authors suggest that there is an overall perceptive-cognitive bias to answer "male" when in doubt about a person's sex/gender, and that in the case of faces this predisposition would result from an anatomical lack of distinctly female features (e.g., Enlow, 1990) in faces in general. Answering ‘female’ while classifying a face’s sex would thus be a ‘no male traits’ response” (p. 79). In line with this explanation, some of our participants informally stated that some skinless female faces were perceived as young men, and therefore were (mis)classified as masculine. It seems that when the information from the surface reflectance is missing and only the information from the facial structure remains, a certain interaction between gender x age has more opportunity to emerge. From a psychobiological point of view, Armann and Bülthoff (2012) speculate (and they emphasize that it is simply speculation) that, in accordance with the history of humans, misclassifying a man as woman would potentially be more dangerous than misclassifying a woman as a man.

The psychometric curves of our experiments (figures 2 and 4) suggest that the gender/sex of full skin faces (F-SK) is perceived almost categorically, especially for the artificial faces of Exp. 1 (Figure 2). The literature shows that experimental studies using continua of images created by computer (morphing) present controversial data regarding the categorical perception of the face gender (Armann & Bülthoff, 2012; Bülthoff & Newell, 2004; Campanella, Chrysochoos, & Bruyer, 2001). Our data from the first experiment (Figure 2, Table 1) point to a categorical gender perception of the full skin faces, because the responses to the male version 3 (the closest to the ambiguous version 4) are 97% “Man”, whereas the responses to the female version 5 (the closest to the ambiguous version 4) drop to just 6%, that is 94% “Woman”, marking a very sharp transition (regression line with b = - 45.5). A similar pattern, although less pronounced, can be observed in the full color faces of the second experiment (Figure 4, Table 1). However, the psychometric curves show that the gender of the skinless faces is perceived in a more continuous

/ less categorical way, suggesting that the information provided by the pigmentation or surface reflectance contributes to the categorical perception of faces as male or female.

On the other hand, comparing the responses to skinless faces displayed in foreshortened position versus those displayed at an angle of 20º, these last stimuli gained about 5–8% of correct categorization of face gender in both experiments and their psychometric curves slightly approximated the curves of full-color faces in both experiments (figures 2 and 4). This gain was expected and found in the literature (Bruce et al., 1993; Sobey, 2014), because an angled view provides stronger shading patterns and emphasizes the 3D contribution of facial shape.

In summary, our two experiments revealed the importance of the information provided by the color and skin texture, or surface reflectance, in the perception of gender/sex of a face. When this information is experimentally removed from a continuous set of 3D face models that gradually vary in their masculinity - femininity, the accuracy of gender categorization decreases by about 20% overall. On the other hand, this information contributes to a categorical perception of the gender/sex of a face, because its non-existence, and the perception of the "pure" facial structures, gives rise to more continuous psychometric curves along the stimulus continuum. These effects have been consistently found both using totally artificial faces created *ex novo* by computer, or using 3D face models created from photographs of real people.

Footnotes

[1] Partial eta-squared (η^2_p) refers to the proportion of variability in the dependent measure, which is attributable to a factor. The effect size interpretations for η^2_p values are as follows: .01 = small, .06 = medium, and .14 = large (Cohen 2013; Gravetter & Wallnau, 2012).

References

Armann, R., & Bülthoff, I. (2012). Male and female faces are only perceived categorically when linked to familiar identities–and when in doubt, he is a male. *Vision research*, *63*, 69-80.

Bruce, V. (1990). Perceiving and recognizing faces. *Mind & Language*, *5*(4), 342-364.

Bruce, V., Burton, A. M., Hanna, E., Healey, P., Mason, O., Coombes, A., ... & Linney, A. (1993). Sex discrimination: how do we tell the difference between male and female faces? *Perception*, *22*(2), 131-152.

Bruce, V., Ellis, H. D., Gibling, F., & Young, A. (1987). Parallel processing of the sex and familiarity of faces. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *41*(4), 510.

Bülthoff, I., & Newell, F. (2004). Categorical perception of sex occurs in familiar but not unfamiliar faces. *Visual Cognition*, *11*(7), 823-855.

Campanella, S., Chrysochoos, A., & Bruyer, R. (2001). Categorical perception of facial gender information: Behavioural evidence and the face-space metaphor. *Visual cognition*, *8*(2), 237-262.

Cenac, Z., Biotti, F., Gray, K. L., & Cook, R. (2019). Does developmental prosopagnosia impair identification of other-ethnicity faces? *Cortex*, *119*, 12-19.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press. doi:10.4324/9780203771587.

Edwards, E. A., & Duntley, S. Q. (1939). The pigments and color of living human skin. *American Journal of Anatomy*, *65*(1), 1-33.

Enlow, D. H., & Hans, M. G. (1996). *Essentials of facial growth*. WB Saunders Company.

Gravetter, F. J. & Wallnau, L. B. (2012). *Statistics for the behavioral health sciences* (9th ed). Belmont, CA: Wadsworth.

González-Alvarez, J. (2017). Perception of sexual orientation from facial structure: A study with artificial face models. *Archives of Sexual Behavior*, *46*(5), 1251-1260.

Haseltine, S. *The Perception of Gender in Human Faces*. URL: https://gustavus. edu/psychology/documents/Haseltine_S. pdf.

Iscan, M. Y., & Steyn, M. (2013). *The human skeleton in forensic medicine*. Charles C. Thomas Publisher.

Jablonski, N. G., & Chaplin, G. (2000). The evolution of human skin coloration. *Journal of Human Evolution*, *39*(1), 57-106.

Kihara, K., & Takeda, Y. (2019). The role of low-spatial frequency components in the processing of deceptive faces: A study using artificial face models. *Frontiers in psychology*, *10*, 1468.

Mackenzie, S., & Wilkinson, C. (2017). Morphological and morphometric changes in the faces of female-to-male (FtM) transsexual people. *International Journal of Transgenderism*, *18*(2), 172-181.

Meinhardt-Injac, B., Persike, M., & Meinhardt, G. (2013). Holistic face processing is induced by shape and texture. *Perception*, *42*(7), 716-732.

Nestor, A., & Tarr, M. J. (2008). Gender recognition of human faces using color. *Psychological Science*, *19*(12), 1242-1246.

Oh, D., Dotsch, R., & Todorov, A. (2019). Contributions of shape and reflectance information to social judgments from faces. *Vision research*, *165*, 131-142.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, *105*(32), 11087-11092.

Richler, J. J., & Gauthier, I. (2014). A meta-analysis and review of holistic face processing. *Psychological Bulletin*, *140*, 1281-1302.

Rossion, B. (2013). The composite face illusion: A whole window into our understanding of holistic face perception. *Visual Cognition*, *21*, 139-253.

Russell, R. (2009). A sex difference in facial contrast and its exaggeration by cosmetics. *Perception*, *38*(8), 1211-1219.

Russell, R., Sinha, P., Biederman, I., & Nederhouser, M. (2006). Is pigmentation important for face recognition? Evidence from contrast negation. *Perception*, *35*(6), 749-759.

Sæther, L., Van Belle, W., Laeng, B., Brennen, T., & Øvervoll, M. (2009). Anchoring gaze when categorizing faces' sex: evidence from eye-tracking data. *Vision research*, *49*(23), 2870-2880.

Sobey, S. (2014). *Face Processing and Sex Categorization: A Behavioural and Eye-Tracking Study*. Master Science Thesis.

Tanikawa, C., Zere, E., & Takada, K. (2016). Sexual dimorphism in the facial morphology of adult humans: a three-dimensional analysis. *Homo*, *67*(1), 23-49.

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition*, *27*(6), 813-833.

Todorov, A., & Porter, J. M. (2014). Misleading first impressions: Different for different facial images of the same person. *Psychological Science*, *25*(7), 1404-1417.

White, T. D., Black, M. T., & Folkens, P. A. (2011). *Human Osteology*. Academic press.

Zhao, M., Bülthoff, H. H., & Bülthoff, I. (2016). A shape-based account for holistic face processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(4), 584.

Acknowledgements:

This work was completed with resources provided by the University Jaume I of Castellon (Spain).

Ethical Compliance Section

All procedures performed in studies involving human participants were in accordance with the Deontological commission and of the Ethical Committee of Animal Welfare of the University Jaume I (Spain) and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

The authors declare they have no conflict of interest.

Written informed consent was obtained from all individual adult participants included in the study.