

Attractiveness Differentially Affects Direct Versus Indirect Face Evaluations in Two Cultures

Evolutionary Psychology
July-September 2023: 1–13
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/14747049231186119
journals.sagepub.com/home/evp



Brittany S. Cassidy¹ , S. Adil Saribay² , Hüseyin Yüksel³
and Karel Kleisner⁴

Abstract

Although decades of research have identified facial features relating to people's evaluations of faces, specific features have largely been examined in isolation from each other. Recent work shows that considering the relative importance of these features in face evaluations is important to test theoretical assumptions of impression formation. Here, we examined how two facial features of evolutionary interest, facial attractiveness and facial-width-to-height ratio (FWHR), relate to evaluations of faces across two cultures. Because face evaluations are typically directly measured via self-reports, we also examined whether these features exert differential effects on both direct and indirect face evaluations. Evaluations of standardized photos naturally varying in facial attractiveness and FWHR were collected using the Affect Misattribution Procedure in the United States and Turkey. When their relative contributions were considered in the same model, facial attractiveness, but not FWHR, related to face evaluations across cultures. This positive attractiveness effect was stronger for direct versus indirect evaluations across cultures. These findings highlight the importance of considering the relative contributions of facial features to evaluations across cultures and suggest a culturally invariant role of attractiveness when intentionally evaluating faces.

Keywords

facial attractiveness, facial width-to-height ratio, face perception, affect misattribution procedure, cross-cultural effects

Received 7 June 2023; accepted 19 June 2023

Evaluating faces is a highly consequential part of social perception (e.g., Wilson & Rule, 2015). The ecological approach to social perception asserts that these evaluations stem from an evolutionary motive to thrive within one's environment (McArthur & Baron, 1983). This approach suggests that face evaluations are self-protective in that they provide adaptive cues about potential interactions (Slepian et al., 2017; Zebrowitz & Collins, 1997). Indeed, face evaluations affect outcomes ranging from mate choice (Fisher & Cox, 2009) to hiring decisions (Chiu & Babcock, 2002). Given their consequential nature, decades of research situated in the ecological approach have focused on what facial features contribute to relative impression valence (Todorov et al., 2014; Zebrowitz & Collins, 1997). Of particular interest in this approach is facial attractiveness. From an evolutionary lens (for a review, see Little et al., 2011), facial attractiveness reveals an adaptation reflecting health that guides mate selection (Scheib et al., 1999; Thornhill & Gangestad, 1999). Another facial feature

studied from an evolutionary approach is the facial width-to-height ratio (FWHR). FWHR is theorized to reflect an evolved cue in that people evaluate faces (and particularly men's faces) with larger FWHRs as more threatening and dominant (Geniole et al., 2015). Supporting that evaluations of these features are evolutionarily adaptive, some research has found cross-cultural generalizability in evaluations of facial attractiveness (Langlois et al., 2000) and FWHRs (Tsujimura & Banissy, 2013).

¹University of North Carolina at Greensboro, Greensboro, NC, USA

²Kadir Has University, Istanbul, Turkey

³UniQus, Istanbul, Turkey

⁴Charles University, Prague, Czech Republic

Corresponding author:

Brittany S. Cassidy, Department of Psychology, University of North Carolina at Greensboro, 296 Eberhart, PO Box 26180, Greensboro, NC 27412, USA.
Email: bccassid@uncg.edu



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Although many facial features are related by the evaluations they elicit (Jones & Jaeger, 2019), most research has examined these features in isolation (e.g., Stirrat & Perrett, 2010). Examining these features in isolation has been important to identify what people instinctively use from faces to form adaptive impressions. Yet, it is also a limitation because people do not evaluate singular facial features when they encounter others. Even though the valenced and adaptive impressions that facial attractiveness and FWHR elicit have been widely studied from an evolutionary lens, we do not have a clear understanding of the *relative* contributions of these features. This understanding is important because evincing *relative* contributions to evaluations identifies the features actually relied upon when forming impressions (Jaeger & Jones, 2022). Whereas theoretical foundations supporting the importance of facial attractiveness and FWHR to evaluations might both seem plausible, examining their relative contributions to evaluations can thus reveal the strength of these foundations. To this end, we examined the relative contributions of facial attractiveness and FWHR on two types of valenced evaluations across American (Sample 1) and Turkish (Sample 2) samples of participants and face identities.

Facial attractiveness research situated in an evolutionary perspective is driven by the hypothesis that emergent evaluations reflect information about a target's health (Thornhill & Gangestad, 1999). As such, people seem to instinctively evaluate others on their facial attractiveness. Consensually positive evaluations of attractive faces are typified by *what-is-beautiful-is-good* effect (Dion et al., 1972), a widely replicable phenomenon (Eagly et al., 1991). Given the replicability of this effect, social psychological research has identified facial attractiveness to be a unique dimension of face evaluation that can be interpreted within an evolutionary framework (Sutherland et al., 2013) and that seems to be culturally universal (Sutherland et al., 2018). Indeed, although some face evaluation models suggest that attractiveness is somewhat related to trustworthiness in reflecting a broader valence dimension of face perception (Oosterhof & Todorov, 2008), other work suggests that attractiveness is not especially related to morality-related evaluations like trustworthiness (Eagly et al., 1991). Similarly, work on romantic partner preferences suggests that attractiveness is separable from warmth and status dimensions (Fletcher et al., 2000). Unlike such trait-based dimensions of face evaluation, facial attractiveness is likely determined by the featural content of faces that mark genetic quality (e.g., symmetry; Perrett et al., 1999) rather than structural resemblance to facial emotions (Todorov, 2008). Evaluations based on facial attractiveness indeed seem to reflect evolutionary motives. For example, masculine characteristics of men's faces correspond with their reproductive potential (Rhodes et al., 2005), and women more strongly prefer masculine characteristics of men's faces when they are most fertile (Johnston et al., 2001).

Also suggesting the adaptive nature of evaluations based on facial attractiveness, people within and across cultures strongly agree on what faces are and are not attractive, with higher facial

attractiveness advantaging both child and adult targets in ways that align with an evolutionary perspective (Langlois et al., 2000). For example, connectionist modeling work suggests that evaluations of facial attractiveness are overgeneralized responses to adaptively significant facial qualities (Zebrowitz et al., 2003). Such support is also evident in cross-cultural work. For example, evaluations from facial attractiveness among Americans and the Tsimane', a culturally isolated group in the Bolivian rainforest showed within- and between-group agreement for American and Tsimane' faces, as well as an attractiveness halo (Zebrowitz et al., 2012). This work suggests that facial attractiveness is an adaptive predictor of impressions. An open question regards its strength when considered in tandem with other features of interest from an evolutionary lens. Indeed, despite a robust *what-is-beautiful-is-good* effect (Dion et al., 1972), attractiveness was only the second-most informative predictor of trustworthiness evaluations in work using machine learning to test how dimensions of face evaluation predict impressions (Jaeger & Jones, 2022).

In contrast to facial attractiveness being an adaptive cue to fitness (Little et al., 2011), work on FWHR from an evolutionary lens focuses on FWHR as an evolved threat cue (Geniole et al., 2015). FWHR is defined as the distance between the zygomatic bones divided by the distance between the upper lip and mid-brow (Weston et al., 2007). Interest in FWHR originated from the proposal that it may be a sexually dimorphic facial trait potentially related to testosterone (Carré & McCormick, 2008). Although some work has supported this possibility by showing shown a relation between FWHR and the threat corresponding with increased testosterone (Stirrat & Perrett, 2010), other work has found no evidence for this relation (Kosinski, 2017; Whitehouse et al., 2015). Regardless of the accuracy of these relations, however, there is consistent evidence that FWHR is associated with people's evaluation of faces. Here, we examined FWHR because of its general simplicity and the ability to compare it with the longstanding and broad literature on FWHR. Moreover, FWHR represents a one-dimensional quantity that can be expressed by a single variable, which is practical for our purpose as it can be readily included in any statistical model. We are aware, however, that although FWHR provides valuable insights, it should be used with caution. Facial features and their associations are complex and can vary across populations, cultures, and individual differences.

Research examining the relationship between the FWHRs of target faces and perceptions of naïve perceivers has consistently found that as FWHR increases, evaluations become more negative (Geniole et al., 2015). Indeed, people generally evaluate faces with greater relative to lower FWHR to be more threatening, dominant, aggressive, and less trustworthy (Eisenbruch et al., 2016; Kleisner et al., 2013; Stirrat & Perrett, 2010; Trębický et al., 2015). Greater FWHR does elicit positive evaluations in certain contexts. Faces with greater FWHR, for instance, are sometimes evaluated as more successful (Alrajih & Ward, 2014).

Evidence for the adaptive nature of FWHR-based evaluations can be greatly strengthened by demonstrations of cross-cultural generalizability. Although some work has shown these patterns to emerge across cultures (Tsujimura & Banissy, 2013), however, the FWHR literature is biased toward Western samples (Haselhuhn et al., 2015). There are an even smaller number of studies examining faces or perceivers from different cultures (see Saribay et al., 2018). Beyond the utility of examining relative FWHR effects on valenced impressions when accounting for other theoretically relevant features like attractiveness, examining FWHR effects on a diversity of perceivers and target faces may contribute to a broader understanding of the cultural universality of such effects. Thus, the adaptive nature of relative FWHR-based contributions to evaluations could be greatly strengthened by the cross-cultural assessment presented in the current work.

Although face evaluations are characterized as being largely spontaneous (Willis & Todorov, 2006), a great deal of work examining facial attractiveness and FWHR has used standard *direct* ratings paradigms. In these paradigms, perceivers often have long or unlimited exposures to individual faces and are asked transparent questions about their evaluations (e.g., by making a rating via a Likert-type scale). Although some studies address a link between, for example, FWHR and interpersonal behavior (Stirrat & Perrett, 2010) or real-world judgments (Wilson & Rule, 2015), most work regarding perceptual effects involves relatively intentional judgments. By contrast, *indirect* measures tap relatively automatic evaluations and prohibit, to some extent, perceivers from editing their responses based on deliberative considerations. Thus, under some circumstances, such as when people feel pressure to respond in socially desirable ways, direct and indirect measures may diverge to potentially reveal different aspects of mental functioning (Nosek, 2007).

The reliance on face evaluation research on direct ratings and intentional behavioral reactions does not allow for a comprehensive assessment of how facial features contribute to them. For example, as far as we know, no evidence exists on the role of FWHR in relatively automatic or less intentional face evaluations. This gap in the literature, however, allows for the possibility that perceptual effects on face evaluations are inconsistent across direct and indirect measures. For instance, perceptual effects of attractiveness and FWHR may rely on lay theories about facial features. In this case, people may expect targets with more positive traits (e.g., likability) to have more attractive faces; just as they expect babyfaced targets to be more trustworthy (Zebrowitz et al., 2012). Less intentional reactions to faces, which may be colored less by such lay theories, may therefore tell a different story about the role of specific facial features in impressions.

The converse is also possible. For example, it could be that an effect of FWHR on face evaluations may be disrupted by relatively intentional attention to faces. Supporting this possibility is research showing that FWHR affects impressions in the expected direction under suboptimal conditions such as when faces are cropped or blurred (Carré et al., 2010). Thus, effortless

exposure to faces may be sufficient for at least some facial features to affect subsequent judgments, with additional exposure or intentional attention potentially diluting or even disrupting effects. A comparison of conditions under which controlled attention to faces is manipulated is necessary to examine these possibilities. The current research was designed as an initial step to fill this gap in the literature by considering the relative effects of facial attractiveness and FWHR on both direct and indirect evaluations.

One challenge in comparing direct and indirect measures that involve different degrees of controlled attention is their typical reliance on distinct tasks with little structural commonality. It is difficult to learn much from a comparison of different measures when their structures greatly differ. To address this challenge, the current work used the affect misattribution procedure (AMP; Payne et al., 2005). The AMP capitalizes on the psychological tendency to misattribute the evaluative effects of a prime on an unrelated target. This misattribution allows for an assessment of the indirect effects of properties of primes (e.g., facial attractiveness). We relied on a modified version of the AMP to keep task structure constant while varying whether relatively intentional versus automatic evaluations were measured (Payne, Burkley et al., 2008).

Based on the above-described literature, we deem it necessary to compare the relative effects of facial attractiveness and FWHR on direct and indirect face evaluations using measures with high structural fit. This comparison serves to compare the relative effects of facial features with evolutionary interest, and tests whether any relative effects of these facial features generalize across evaluations formed in different ways. We first examined this possibility in an American sample (Sample 1). We next conducted a close replication in Turkey (Sample 2) both to assess the cross-cultural generalizability expected from an evolutionary perspective (e.g., Sutherland et al., 2018), as well as to provide one of the first tests of relative effects of facial features (e.g., Jaeger & Jones, 2022) from a non-Western culture. We expected that if they exert uniquely adaptive effects on face evaluations, facial attractiveness, and FWHR would, respectively, positively and negatively relate to evaluations across samples. We explored differences between direct and indirect measures without a guiding hypothesis given that the literature offers no systematic test of such differences.

Sample 1

Method

Participants. One hundred forty-three undergraduates provided informed consent and participated for course credit. Based on a priori criteria of not using each number on the scale at least once or recognizing a character in the below-described task, 24 participants were excluded. These exclusions yielded a sample of 119 participants ($M_{age} = 19.44$ years, $SD_{age} = 1.93$, 75 female). The experiment was approved by the Indiana University Institutional Review Board.

Stimuli. Images of 50 European-American male front-facing faces with neutral expressions were selected from the Chicago Face Database (Ma et al., 2015). This database includes attractiveness norms (1 [*very unattractive*] to 7 [*very attractive*]) as well as a validated FWHR for each face. These images are widely used in research on attractiveness (e.g., Alaei et al., 2022) and FWHR (e.g., Deska et al., 2018). The FWHR of selected faces ranged from 1.589 to 2.15 ($M=1.86$, $SD=0.129$). The attractiveness of selected faces ranged from 1.73 to 4.66 ($M=2.86$, $SD=0.55$). Attractiveness and FWHR were not significantly correlated, $r(48)=-.15$, $p=.29$, 95% CI $[-0.41, 0.13]$. Faces were displayed at a size of 256*298 pixels. This size was selected because it approximated the sizes of both the mask and Chinese ideograms (see Task).

Task. The task was a variation of the Affective Misattribution Procedure (AMP; Payne & Lundberg, 2014; Payne, Govorun et al., 2008) that maintained structural consistency across direct and indirect evaluations. Participants were told that the “task is about making judgments while avoiding distractions” and that they would see pairs of images in each trial and that they would evaluate one while ignoring the other. They were told that there was no right or wrong answer, that they should indicate their “gut reaction as quickly as possible” and were asked to “not judge all of the images as pleasant or all of them as unpleasant” but instead to “judge each image based on whether they think it is more or less pleasant than average.”

In each trial, one face was shown at the center of the display followed by a blank screen and then a randomly selected Chinese ideogram. Each of these events took 100 ms. The Chinese ideogram was masked with a noise pattern of black and white dots that remained on the screen until participants responded. A rating scale appeared at the bottom of the screen along with the mask. The scale showed the numbers

−2, −1, +1, and +2 from left to right. These numbers corresponded to “very unpleasant,” “slightly unpleasant,” “slightly pleasant,” and “very pleasant” and could be selected using, respectively, the Z, X, N, and M keys on a standard keyboard. See Figure 1 for an example trial and its timecourse.

For direct evaluations, participants were asked to “rate the photos of the people” and to try “not to be influenced by the Chinese characters.” For indirect evaluations, they were asked to “rate the Chinese characters” and to try “not to be influenced by the photographs.” To maintain target awareness, the rating scale was captioned with the phrase “rate the photo of the person” and “rate the Chinese character,” depending on the evaluation type. The direct and indirect trials were blocked. Each of the 50 target faces was evaluated once per block, resulting in 100 evaluation trials. Block order was randomized. Within-blocks, face presentation order was randomized. Before the task, participants completed three practice trials in which they evaluated faces not used in the actual trials.

Procedure. Participants were seated in individual rooms facing 19” LCD monitors set to a screen resolution of 1280*1024 pixels. The experiment was implemented using DirectRT v2012. Participants provided demographic information and then received instructions on the screen. After the task, participants indicated whether they recognized any characters in the task, but did not indicate which specific characters, if any, were recognized. Participants indicating recognition were excluded.

Analytic Strategy. Across experiments, we used the lme4 package (Bates et al., 2014) to create linear mixed effects models and the lmerTest package (Kuznetsova et al., 2017) to calculate model p -values. We used the *emtrends* function from the emmeans package (Lenth, 2018) to calculate

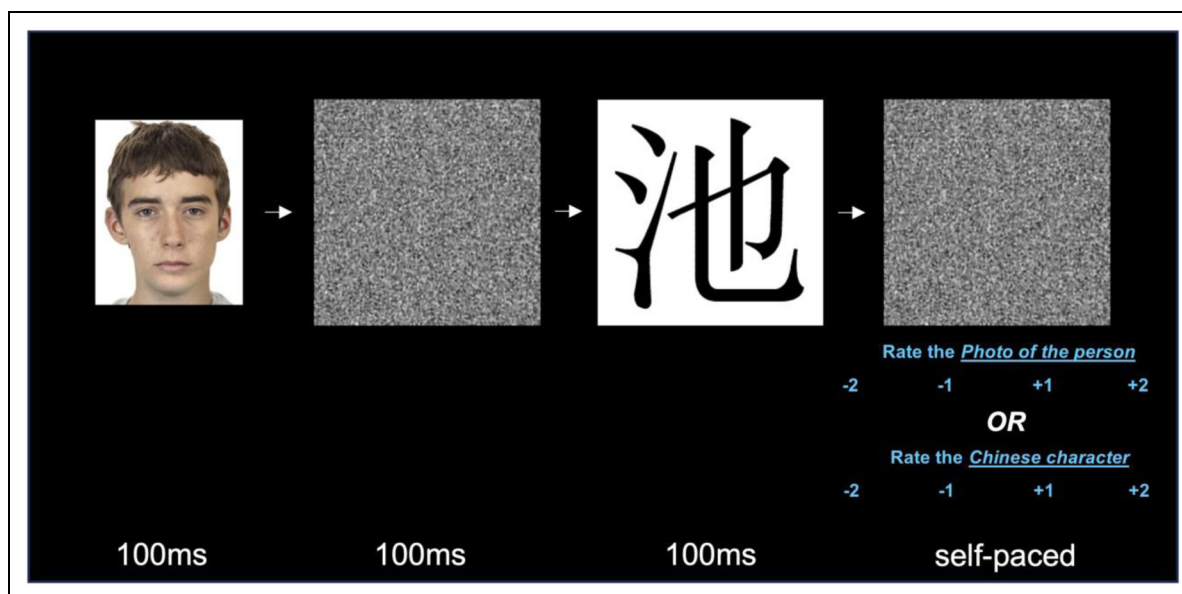


Figure 1. Example task trial.

simple slopes to characterize interaction effects. All data and code are available at https://osf.io/ehpf6/?view_only=350f01f895d441499e51f54454acc128.

Results and Discussion

A linear mixed effect model regressed pleasantness ratings on Evaluation (direct = -1, indirect = 1), FWHR (standardized around the mean database-provided value of selected faces), attractiveness (standardized around the mean database-provided norm of selected faces) and interactions between Evaluation with the continuous variables as fixed effects. The model included a random effects structure such that both intercepts and evaluation effects varied by participant and by face identity.

An Evaluation effect emerged such that indirect ($M = 0.23$, $SE = 0.04$) relative to direct ($M = -0.22$, $SE = 0.07$) evaluations were more positive, $b = 0.22$, $SE = 0.03$, $t = 6.54$, $p < .001$. An Attractiveness effect emerged, $b = 0.20$, $SE = 0.04$, $t = 5.11$, $p < .001$, consistent with the longstanding literature on the *what-is-beautiful-is good* effect (e.g., Dion et al., 1972). Contrasting previous research showing FWHR to negatively relate to evaluations (e.g., Eisenbruch et al., 2016), no FWHR effect emerged, $b = -0.02$, $SE = 0.04$, $t = 0.41$, $p = .69$.

An interaction between Evaluation and Attractiveness emerged, $b = -0.15$, $SE = 0.03$, $t = 5.51$, $p < .001$. A stronger Attractiveness effect was observed for direct, $b = 0.35$, $SE = 0.06$, $z = 5.46$, $p < .001$, than for indirect, $b = 0.04$, $SE = 0.02$, $z = 2.10$, $p = .04$, evaluations (Figure 2(a)). A sensitivity analysis for this interaction indicated a minimum detectable effect of $b = -0.075$ with power = 0.80 and alpha = 0.05. This finding suggests that facial attractiveness more strongly relates to valence evaluations when people explicitly evaluate faces themselves. No interaction between Evaluation and FWHR emerged, $b = 0.03$, $SE = 0.03$, $t = 1.10$, $p = .28$.

Exploratory Analyses Including Participant Gender. Because the face stimuli were all male and because participants spanned the genders, we conducted analyses including Participant Gender (female = -1, male = 1) and its interactions with the variables in the above-described model as fixed effects on an exploratory basis. This model explained more variance than the first model, $\chi^2 = 14.80$, $p = .02$. All effects from the first model retained direction and significance (Table 1(a)). Unique to this model, a Participant Gender effect showed that men ($M = 0.11$, $SD = 0.06$) had more positive evaluations than women ($M = -0.06$, $SD = 0.05$), $b = 0.08$, $SE = 0.03$, $t = 2.56$, $p = .01$. An interaction between Participant Gender and Attractiveness emerged, $b = -0.03$, $SE = 0.01$, $t = 2.67$, $p = .01$. The Attractiveness effect was stronger for women, $b = 0.22$, $SE = 0.04$, $z = 5.59$, $p < .001$, than men, $b = 0.15$, $SE = 0.04$, $z = 3.74$, $p < .001$ (Figure 3(a)). Greater sensitivity to facial attractiveness among women than men is consistent with work showing women to be more sensitive and accurate when perceiving and evaluating affective facial expressions (Montagne et al., 2005) and work showing that women are better than men at recognizing subtle differences in the affective

content of faces (Hoffmann et al., 2010). This pattern is also consistent with women being socialized to be more communal and interpersonally focused than are men (Eagly & Steffen, 1984; Eagly & Wood, 2012). Speculatively, differential gender socialization may leave women, in part, more sensitive to differences in some facial features than men.

Sample 2

For Sample 2, we examined whether the findings of Sample 1 replicated in a different culture. Because most social cognition research relies on samples from cultures that are Western, educated, industrialized, rich, and democratic (i.e., WEIRD; Henrich et al., 2010), we collected data from Turkey. Turkey has a predominantly Muslim culture that reflects a mix of Western and Eastern influences. Relatively few studies have examined face impressions among Turkish faces with various psychological and sociopolitical qualities (e.g., Ozener, 2012; Saribay et al., 2018; Saribay & Kleisner, 2018). Collecting data from a Turkish sample provided an opportunity to examine whether the relative effects of facial attractiveness and FWHR shown in WEIRD culture of Sample 1 would hold. It also provided the opportunity for further cross-cultural comparison (with the United States) in terms of the direct-indirect evaluation differences shown in Sample 1.

Method

Participants. One hundred forty-two Boğaziçi University students provided informed consent and participated for course credit. Twelve were excluded using the same exclusion criteria used for Sample 1. Three were excluded due to computer error in recording ratings, yielding a sample of 127 participants ($M_{age} = 21.48$ years, $SD_{age} = 1.77$, 73 female). This experiment was approved by the Boğaziçi University Institutional Review Board.

Stimuli. Images of 50 Turkish front-facing male faces of undergraduates with neutral expressions were selected from the Boğaziçi Face Database (Saribay et al., 2018) to ensure FWHR variability. This database includes attractiveness norms (1 [*very unattractive*] to 7 [*very attractive*]). We used Turkish faces to ensure evaluations of cultural ingroup faces by perceivers. We selected male faces that had little or no facial hair to keep this analysis comparable to the one using Sample 1, in which all faces were clean shaven. Using NIH's ImageJ software (<https://imagej.nih.gov/ij/>), independent raters who did not complete the described experiments took two measures of facial height and width that corresponded to the distance between the zygomatic bones divided by the distance between the upper lip and mid-brow. These values were averaged since they were highly correlated ($r_s > .99$, $p < .001$). FWHR was computed by dividing width by height and ranged from 1.806 to 2.265 ($M = 2.03$; $SD = 0.12$) (for more details on stimuli norms, see Saribay et al., 2018). The average attractiveness ratings of the selected faces ranged from 1.46 to 4.53 ($M = 2.48$, $SD = 0.82$). Attractiveness

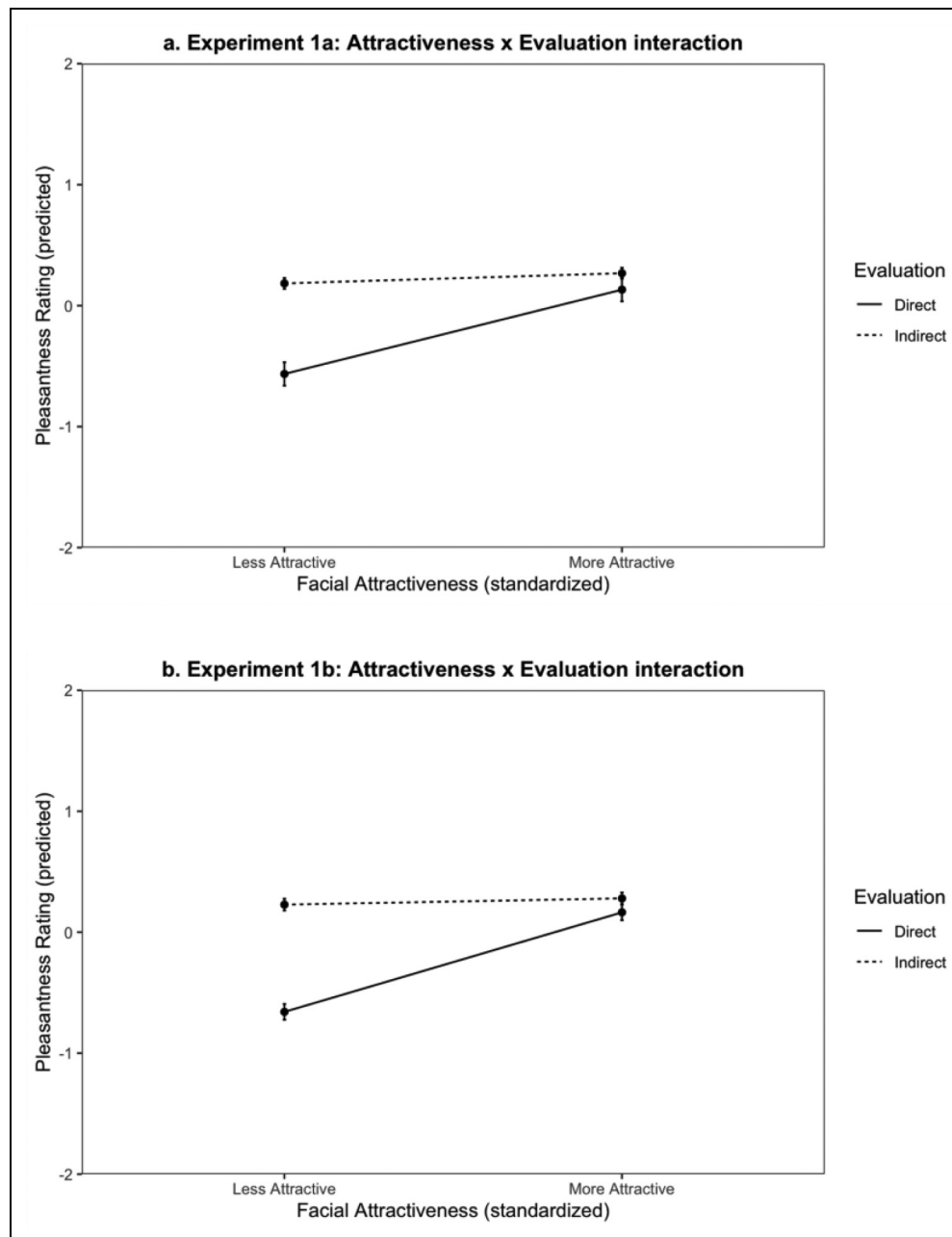


Figure 2. Stronger facial attractiveness effects on pleasantness ratings emerged for direct relative to indirect evaluations across Sample 1 (a) and Sample 2 (b).

Note. Error bars reflect standard error.

norms and FWHR were not significantly related, $r(48) = -.11$, $p = .44$, 95% CI $[-0.38, 0.17]$.

The Facial image format differed in minor ways from Sample 1. Images were cropped slightly more narrowly (from chin to hairline and ear to ear) and converted to grayscale. Image size (277*313 pixels) was like that for Sample 1.

Task and Procedure. We replicated the task used for Sample 1 with the following differences. Instructions and rating scales were translated to Turkish. Participants faced a 19" LCD monitor with a screen resolution of 1440*900 pixels. Because

participants and target faces were from the same university, participants were shown each face after the main task and indicated whether they had encountered it in real life, even if from a distance. Trials with recognized faces were excluded, reducing the number of analysed trials from 12,700 to 10,844.

Results and Discussion

Suggesting cross-cultural generalizability from American to Turkish samples, the results of this model replicated Sample 1 in both direction and significance. An Evaluation effect again emerged such that

Table 1. Regression Summaries on Pleasantness Ratings for Analyses That Included Participant Gender in Sample 1 (a) and Sample 2 (b) and Analyses in Sample 2 That Included Recognized Face Trials (c).

Predictor	a. Sample 1		b. Sample 2		c. Sample 2	
	Estimate [95% CI]	<i>p</i>	Estimate [95% CI]	<i>p</i>	Estimate [95% CI]	<i>p</i>
(Intercept)	0.03 [−0.07, 0.12]	.58	0.01 [−0.06, 0.09]	.72	0.04 [−0.04, 0.11]	.35
Evaluation	0.22 [0.15, 0.29]	<.001	0.24 [0.18, 0.30]	<.001	0.22 [0.17, 0.28]	<.001
Participant Gender	0.08 [0.02, 0.15]	.01	0.07 [0.01, 0.12]	.03		
Facial-width-to-height ratio (FWHR) (standardized)	−0.02 [−0.09, 0.06]	.69	−0.02 [−0.07, 0.03]	.53	−0.01 [−0.06, 0.04]	.64
Attractiveness (standardized)	0.19 [0.11, 0.26]	<.001	0.22 [0.17, 0.27]	<.001	0.21 [0.16, 0.27]	<.001
Evaluation * Participant Gender	−0.01 [−0.06, 0.04]	.69	−0.05 [−0.10, 0.00]	.07		
Evaluation * FWHR	0.03 [−0.02, 0.09]	.23	−0.01 [−0.04, 0.03]	.72	−0.01 [−0.04, 0.03]	.77
Evaluation * Attractiveness	−0.15 [−0.21, 0.10]	<.001	−0.20 [−0.23, −0.16]	<.001	−0.19 [−0.22, −0.15]	<.001
Participant Gender * FWHR	0.00 [−0.02, 0.02]	.99	−0.01 [−0.04, 0.01]	.24		
Participant Gender * Attractiveness	−0.03 [−0.06, −0.01]	.01	−0.01 [−0.04, 0.02]	.43		
Evaluation * Participant Gender * FWHR	0.01 [−0.01, 0.04]	.35	−0.01 [−0.03, 0.02]	.69		
Evaluation * Participant Gender * Attractiveness	0.00 [−0.02, 0.02]	.99	−0.01 [−0.04, 0.01]	.28		

indirect ($M=0.25$, $SE=0.04$) relative to direct ($M=-0.25$, $SE=0.05$) evaluations were more positive, $b=0.25$, $SE=0.03$, $t=8.43$, $p<.001$. Replicating Sample 1, an Attractiveness effect emerged, $b=0.22$, $SE=0.03$, $t=8.45$, $p<.001$, and no FWHR effect emerged, $b=-0.01$, $SE=0.03$, $t=0.53$, $p=.60$.

Partially replicating Sample 1 was an interaction between Evaluation and Attractiveness, $b=-0.19$, $SE=0.02$, $t=10.94$, $p<.001$ (Figure 2(b)). A stronger Attractiveness effect emerged for direct, $b=0.41$, $SE=0.04$, $z=11.20$, $p<.001$, than for indirect, $b=0.03$, $SE=0.02$, $z=1.04$, $p=.30$, evaluations. Although this differential pattern in relation to strength indeed replicated Sample 1, we note that the Attractiveness effect on indirect evaluations was significant in Sample 1 but not Sample 2. A sensitivity analysis for this interaction indicated a minimum detectable effect of $b=-0.047$ with power = 0.80 and alpha = 0.05. No interaction between Evaluation and FWHR emerged, $b=-0.01$, $SE=0.02$, $t=0.32$, $p=.75$.

Exploratory Analyses Including Participant Gender. Contrasting Sample 1, including Participant Gender as described above did not explain more variance than the first model, $\chi^2=10.09$, $p=.12$. All effects from the first model retained direction and significance (Table 1(b)). A Participant Gender effect again showed men ($M=0.08$, $SE=0.05$) to have more positive evaluations than women ($M=-0.05$, $SE=0.05$), $b=0.07$, $SE=0.03$, $t=2.14$, $p=.03$. The interaction between Participant Gender and Attractiveness was not significant, $b=0.02$, $SE=0.02$, $t=1.83$, $p=.43$ (Figure 3(b)). However, the Attractiveness effect was in the direction of Sample 1 in that it numerically was larger for women, $b=0.23$, $SE=0.03$, $z=8.13$, $p<.001$, than men, $b=0.21$, $SE=0.03$, $z=6.90$, $p<.001$.

Exploratory analyses including all trials in which participants recognized the target face did not change the direction or significance of the above-described results (Table 1(c)).

Exploratory Cross-Cultural Analysis. Analyses within samples showed a significant Attractiveness effect on indirect evaluations in Sample 1, but not Sample 2. This difference, however, does not necessarily reflect an interactive effect (Nieuwenhuis et al., 2011). To explore this possibility, we conducted an exploratory analysis combining the samples and including all trials for Sample 2. An exploratory linear mixed effect model regressed pleasantness ratings on Culture (USA = −1, Turkey = 1), Evaluation (direct = −1, indirect = 1), FWHR (standardized around the database-specific mean value of selected faces), attractiveness (standardized around the database-specific norm of selected faces) and interactions between Evaluation and Culture with the continuous variables as fixed effects. The random effects allowed both intercepts and evaluation effects to vary by participant and by face identity.

Showing cross-cultural generalizability, results mirrored those of Samples 1 and 2. The Evaluation effect showed that indirect ($M=0.24$, $SE=0.03$) relative to direct ($M=-0.20$, $SE=0.05$) evaluations were more positive, $b=0.22$, $SE=0.03$, $t=10.03$, $p<.001$. An Attractiveness effect emerged, $b=0.25$, $SE=0.03$, $t=8.86$, $p<.001$, and no FWHR effect emerged, $b=-0.01$, $SE=0.09$, $t=0.60$, $p=.55$. The Attractiveness effect varied by Evaluation, $b=-0.17$, $SE=0.02$, $t=10.45$, $p<.001$. A stronger Attractiveness effect emerged for direct, $b=0.38$, $SE=0.04$, $z=10.45$, $p<.001$, than for indirect, $b=0.03$, $SE=0.01$, $z=2.42$, $p=.02$, evaluations. Culture did not affect pleasantness ratings on its own or in combination with the other variables. See Table 2 for all coefficient information.

Exploratory Cross-Cultural Bayesian Analysis. Null effects of FWHR and its interactions emerged across analyses both within samples and combining across them. A frequentist

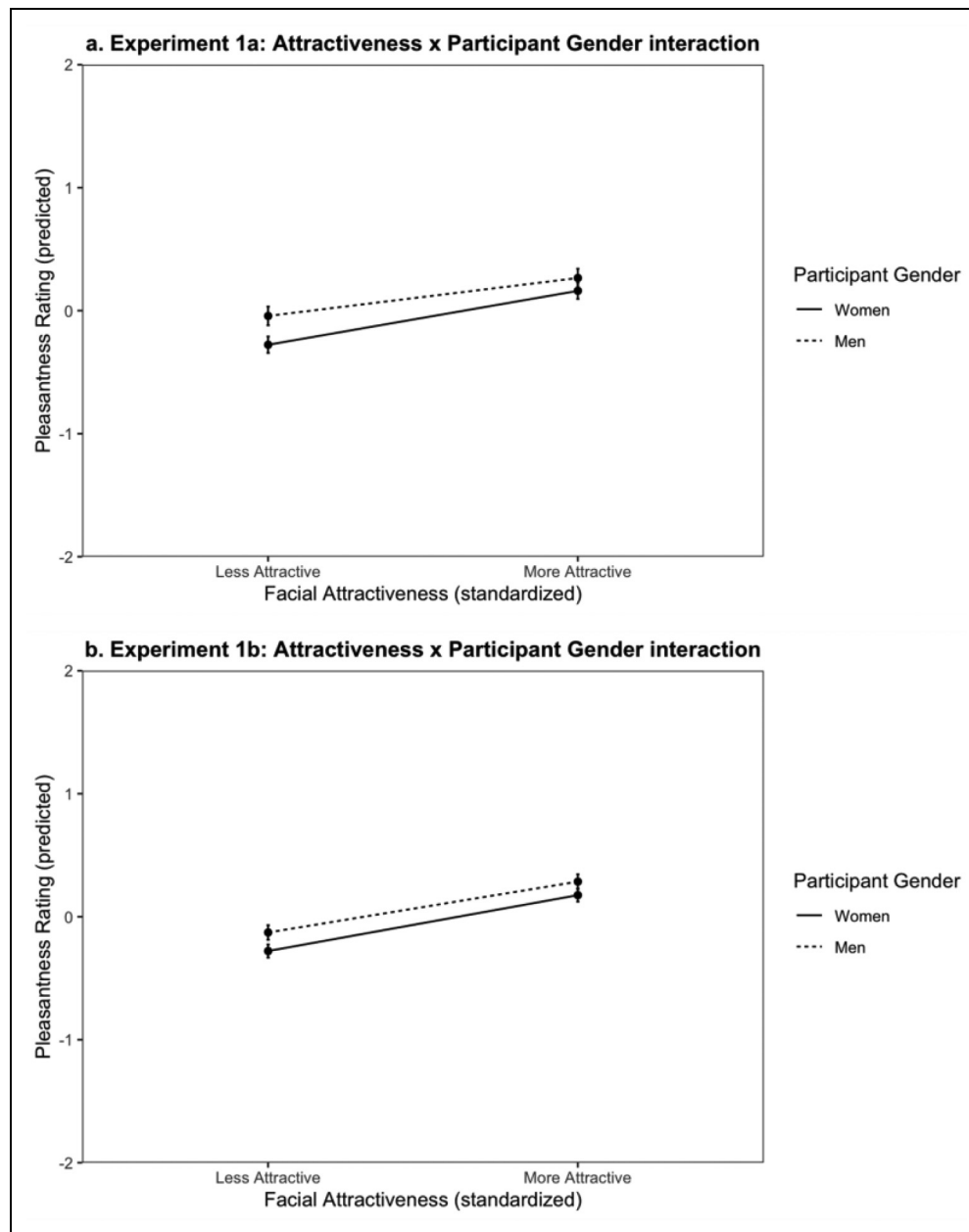


Figure 3. Stronger facial attractiveness effects on pleasantness ratings emerged among women relative to men across Sample 1 (a) and Sample 2 (b).

Note. Error bars reflect standard error.

statistical framework, however, does not distinguish between evidence for a null effect and a lack of statistical power. To this end, we conducted an exploratory Bayesian analysis to assess evidence for such null effects. We conducted this analysis by combining the two samples to increase the number of observations (see Supplemental Material on OSF for code and output). These results largely paralleled the frequentist results. Of note, the distribution of likely slope estimates for the FWHR effect and the interactive effect of Evaluation and FWHR both centered around zero, providing evidence for null effects.

General Discussion

The current research examined relative facial attractiveness and FWHR effects on automatic and intentional face evaluations across American and Turkish samples of participants and face stimuli. Across samples, facial attractiveness positively related to evaluations overall. This pattern is consistent with longstanding work on the *what-is-beautiful-is good* effect (Dion et al., 1972). The cross-cultural nature of this effect also aligns with evolutionary accounts positing that positive evaluations of more attractive faces reflect the adaptive nature of these faces

Table 2. Regression Summary for Exploratory Cross-Cultural Model Including Samples 1 and 2.

Predictors	Pleasantness ratings		
	Estimates	CI	p
(Intercept)	0.02	−0.04–0.08	0.513
Evaluation	0.22	0.18–0.27	<0.001
Culture	0.02	−0.04–0.08	0.617
Facial-width-to-height ratio (FWHR) (standardized)	−0.01	−0.06–0.03	0.548
Attractiveness (standardized)	0.21	0.16–0.25	<0.001
Evaluation × Culture	0.00	−0.04–0.04	0.967
Evaluation × FWHR	0.01	−0.02–0.04	0.434
Evaluation × Attractiveness	−0.17	−0.20–0.14	<0.001
Culture × FWHR	0.00	−0.04–0.05	0.943
Culture × Attractiveness	0.01	−0.04–0.06	0.671
Evaluation × Culture × FWHR	−0.02	−0.05–0.01	0.273
Evaluation × Culture × Attractiveness	−0.02	−0.05–0.01	0.287

(Scheib et al., 1999; Thornhill & Gangestad, 1999; Zebrowitz et al., 2012; Zebrowitz & Montepare, 2006). Because explanations for the adaptive value of evaluations based on facial attractiveness suggest that variations in several facial cues characteristic of attractiveness signal fitness (e.g., symmetry and skin texture; for a review, see Little et al., 2011), future work may systematically manipulate these cues to determine their unique relative contributions to evaluations.

Across American and Turkish participants and face stimuli, the positive attractiveness effect was stronger for direct relative to indirect evaluations. This pattern suggests that facial attractiveness effects on valenced evaluations may be stronger when they are intentional, and reflects the self-reported and intentional nature of the paradigms used in much work on face impressions (e.g., Cassidy et al., 2017; Ma et al., 2015; Saribay et al., 2018). Differential effects of facial attractiveness based on evaluation type is also consistent with work showing a replicable but variable attractiveness stereotype based on the type of evaluations people make (Eagly et al., 1991).

Despite the consistent overall pattern in relation strength across American and Turkish samples, these relations also seemed to differ across samples. Within the American sample, facial attractiveness had a significantly positive relation with evaluations regardless of whether evaluations were direct or indirect. Within the Turkish sample, this relation only significantly emerged for direct evaluations. At first blush, this potential cross-cultural difference may seem counterintuitive. Indeed, people from Western cultures tend to attend to specific objects within view whereas people in more Eastern cultures, like Turks (Schwartz et al., 2014) attend more holistically to an entire field of view (e.g., Boduroglu et al., 2009; Nisbett et al., 2001). From this lens, it would seem more likely for Americans than Turks to allow evaluations of characters to be indirectly influenced by evaluations of preceding faces.

An alternative perspective supports the emergent pattern of results. Although the *what-is-beautiful-is-good* effect seems broadly adaptive given its universality (e.g., Eagly et al., 1991; Zebrowitz et al., 2012), some work suggests that it is

stronger in Western than in Eastern cultures because physical attractiveness is an individual identity attribute (see Wheeler & Kim, 1997). Indeed, students involved in the Chinese community (who are presumably more influenced by Eastern culture) applied attractiveness stereotypes more than less involved people (who are presumably more influenced by Western culture (Dion et al., 1990). Albeit speculative, this work raises the possibility that the American and Turkish faces in the current work may have systematically differed in ways that allowed attractiveness effects to be more broadly apparent among Americans than Turks. For example, it could be that Turks value faces signaling the potential for harmonious relationships more than Americans do. If true, a broad attractiveness effect across evaluations may not emerge for Turks if faces do not signal this potential. For Americans, this potential may not matter.

It is important to note that although this difference emerged when the samples were analyzed separately, cross-cultural differences did not emerge from the exploratory combined analysis. Thus, any evidence for cross-cultural differences in the current work should be cautiously interpreted. This caution, however, should not stop assessments of potential cultural differences in future work. For example, it will be important for future work to have people from different cultures to evaluate within- and across-culture faces to better assess the cultural universality that would support an evolutionary interpretation of the findings (e.g., Sutherland et al., 2018). Indeed, whereas cross-cultural similarities in evaluations would support a universal mechanism, effects moderated by the participant and face culture (e.g., Zebrowitz et al., 2012) would be consistent with culturally specific perceptual learning beyond a broad evolutionary adaptation (McArthur & Baron, 1983). Future work may assess this possibility, among others.

Despite much research showing that FWHR negatively relates to evaluations (Eisenbruch et al., 2016; Kleisner et al., 2013; Stirrat & Perrett, 2010; Třebický et al., 2015), no relative FWHR effects emerged across experiments. This finding highlights a need to consider *relative* contributions of facial features

to evaluations. Indeed, recent work assessing the relative contributions of 28 oft-studied facial features to trustworthiness and dominance impressions found that FWHR was relatively uninformative (Jaeger & Jones, 2022). One possibility is that FWHR may better inform evaluations in contexts where perceivers are specifically drawn to examine it over other features (e.g., Hehman et al., 2013). Future work may test this possibility.

We used cultural ingroup male faces across experiments for task simplicity. Because this choice required using different sets of target faces, it hinders direct statistical comparison of data across cultures. Future work should enable direct comparison by using target faces from both cultures within the same paradigm. Such experiments will clarify the possibility of cultural differences in the scope of facial attractiveness effects on face evaluation, as suggested here. Such work could additionally extend the literature by manipulating group membership and gender of target faces.

A potential limitation of the current research is the use of relatively small face images. It could be that presenting such images could obfuscate some facial information, leading to assessments that might differ from when more information is available. For example, one could posit that perceivers could pay attention to facial features where only a coarse idea of a face is needed to assess it (e.g., FWHR) and disregard features (e.g., symmetry) more uniquely contributing to facial attractiveness. Empirical work does not support this possibility, however. The magnitude of holistic processing decreases as faces become smaller (Ross & Gauthier, 2015), for example. Other work has found that neural adaptation to repeated presentations of faces in core face processing regions is not sensitive to changes in image size (Andrews & Ewbank, 2004). This finding suggests that neural representations reflective of recognizing identifying features of faces are size invariant. That facial attractiveness, but not FWHR, affected evaluations consistently across samples suggests the faces were large enough that facial attractiveness was easily interpretable. One possibility, however, is that people might initially use information from coarse ideas of faces like FWHR to make initial approach decisions when faces are very far away and rely on facial attractiveness to make evaluations when faces become close enough for potential interaction. It will be important for future work to manipulate aspects of images such as size, contrast, and viewpoint to identify conditions where some facial features may contribute to evaluations more than others.

We note that we used the AMP because it offered a well-controlled comparison of direct and indirect evaluations. However, this choice also meant that target faces were seen for a very short time period in the direct evaluation trials. This period contrasts much work in which people make a self-reported evaluation after seeing a face for an unlimited amount of time. Previous work, however, has shown that correlations between face evaluations made under different exposure durations are generally high (Willis & Todorov, 2006). On this basis, we would expect responses gathered in the direct evaluation trials to be comparable to responses gathered under unconstrained exposure. Future research may verify this

expectation. More broadly, given the exploratory nature of this work, knowing whether the potential cross-cultural differences we observed are reliable requires replication and, eventually, testing a priori predictions about the nature of such differences. We consider our work an important initial step in beginning to rectify identified gaps in the literature (e.g., reliance on intentional face evaluations).

The current work extended the literature by providing the first systematic examination of relative facial attractiveness and FWHR effects on direct and indirect evaluations. Emergent consensus and contrast in findings across American and Turkish samples of participants and face stimuli both inform our understanding of how facial features of evolutionary interest affect face impressions. Understanding whether these cross-cultural and cross-measure effects of facial attractiveness are reliable and why they occur should be a focus of future research. For now, our findings highlight the need to examine face impressions from a variety of paradigms across cultures to better understand this core and consequential aspect of human social perception.

Acknowledgments

The authors thank Petr Tureček for assistance with Bayesian analyses and the research assistants at the Neuroscience of Mind and Behavior lab at Indiana University for help with data collection.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research was funded by a 2219 research grant from The Scientific and Technological Research Council of Turkey (TÜBİTAK) to the second author (no: 1059B191600901). Karel Kleisner was supported by the Czech Science Foundation project number 21-10527S. BSC was supported by NIA grant F32AG051304.

ORCID iDs

Brittany S. Cassidy  <https://orcid.org/0000-0002-1449-0886>
S. Adil Saribay  <https://orcid.org/0000-0001-7070-7106>

References

- Alaei, R., Deska, J. C., Hugenberg, K., & Rule, N. O. (2022). People attribute humanness to men and women differently based on their facial appearance. *Journal of Personality and Social Psychology*, 123(2), 400–422. <https://doi.org/10.1037/pspi0000364>
- Alrajih, S., & Ward, J. (2014). Increased facial width-to-height ratio and perceived dominance in the faces of the UK's leading business leaders. *British Journal of Psychology*, 105(2), 153–161. <https://doi.org/10.1111/bjop.12035>
- Andrews, T. J., & Ewbank, M. P. (2004). Distinct representations for facial identity and changeable aspects of faces in the human

- temporal lobe. *NeuroImage*, 23(3), 905–913. <https://doi.org/10.1016/j.neuroimage.2004.07.060>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). *Fitting Linear Mixed-Effects Models using lme4* (arXiv:1406.5823). arXiv. <https://doi.org/10.48550/arXiv.1406.5823>
- Boduroglu, A., Shah, P., & Nisbett, R. E. (2009). Cultural differences in allocation of attention in visual information processing. *Journal of Cross-Cultural Psychology*, 40(3), 349–360. <https://doi.org/10.1177/0022022108331005>
- Carré, J. M., & McCormick, C. M. (2008). In your face: Facial metrics predict aggressive behaviour in the laboratory and in varsity and professional hockey players. *Proceedings of the Royal Society B: Biological Sciences*, 275(1651), 2651–2656. <https://doi.org/10.1098/rspb.2008.0873>
- Carré, J. M., Morrissey, M. D., Mondloch, C. J., & McCormick, C. M. (2010). Estimating aggression from emotionally neutral faces: Which facial cues are diagnostic? *Perception*, 39(3), 356–377. <https://doi.org/10.1068/p6543>
- Cassidy, B. S., Krendl, A. C., Stanko, K. A., Rydell, R. J., Young, S. G., & Hugenberg, K. (2017). Configural face processing impacts race disparities in humanization and trust. *Journal of Experimental Social Psychology*, 73, 111–124. <https://doi.org/10.1016/j.jesp.2017.06.018>
- Chiu, R. K., & Babcock, R. D. (2002). The relative importance of facial attractiveness and gender in Hong Kong selection decisions. *The International Journal of Human Resource Management*, 13(1), 141–155. <https://doi.org/10.1080/09585190110092857>
- Deska, J. C., Lloyd, E. P., & Hugenberg, K. (2018). Facing humanness: Facial width-to-height ratio predicts ascriptions of humanity. *Journal of Personality and Social Psychology*, 114(1), 75–94. <https://doi.org/10.1037/pspi0000110>
- Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, 24(3), 285–290. <https://doi.org/10.1037/h0033731>
- Dion, K. K., Pak, A. W.-P., & Dion, K. L. (1990). Stereotyping physical attractiveness: A sociocultural perspective. *Journal of Cross-Cultural Psychology*, 21(2), 158–179. <https://doi.org/10.1177/0022022190212002>
- Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but...: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, 110(1), 109–128. <https://doi.org/10.1037/0033-2909.110.1.109>
- Eagly, A. H., & Steffen, V. J. (1984). Gender stereotypes stem from the distribution of women and men into social roles. *Journal of Personality and Social Psychology*, 46(4), 735–754. <https://doi.org/10.1037/0022-3514.46.4.735>
- Eagly, A. H., & Wood, W. (2012). Social role theory. In P. Van Lange, A. Kruglanski, & E. Higgins, *Handbook of theories of social psychology* (pp. 458–476). Sage Publications Ltd. <https://doi.org/10.4135/9781446249222.n49>
- Eisenbruch, A. B., Grillot, R. L., Maestripieri, D., & Roney, J. R. (2016). Evidence of partner choice heuristics in a one-shot bargaining game. *Evolution and Human Behavior*, 37(6), 429–439. <https://doi.org/10.1016/j.evolhumbehav.2016.04.002>
- Fisher, M., & Cox, A. (2009). The influence of male facial attractiveness on women's receptivity. *Journal of Social, Evolutionary, and Cultural Psychology*, 3(1), 49–62. <https://doi.org/10.1037/h0099332>
- Fletcher, G. J. O., Simpson, J. A., & Thomas, G. (2000). Ideals, perceptions, and evaluations in early relationship development. *Journal of Personality and Social Psychology*, 79(6), 933–940. <https://doi.org/10.1037/0022-3514.79.6.933>
- Geniole, S. N., Denson, T. F., Dixon, B. J., Carré, J. M., & McCormick, C. M. (2015). Evidence from meta-analyses of the facial width-to-height ratio as an evolved cue of threat. *PLoS One*, 10(7), e0132726. <https://doi.org/10.1371/journal.pone.0132726>
- Haselhuhn, M. P., Ormiston, M. E., & Wong, E. M. (2015). Men's facial width-to-height ratio predicts aggression: A meta-analysis. *PLoS One*, 10(4), e0122637. <https://doi.org/10.1371/journal.pone.0122637>
- Hehman, E., Leitner, J. B., & Gaertner, S. L. (2013). Enhancing static facial features increases intimidation. *Journal of Experimental Social Psychology*, 49(4), 747–754. <https://doi.org/10.1016/j.jesp.2013.02.015>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2-3), 61–83. <https://doi.org/10.1017/S0140525X0999152X>
- Hoffmann, H., Kessler, H., Eppel, T., Rukavina, S., & Traue, H. C. (2010). Expression intensity, gender and facial emotion recognition: Women recognize only subtle facial emotions better than men. *Acta Psychologica*, 135(3), 278–283. <https://doi.org/10.1016/j.actpsy.2010.07.012>
- Jaeger, B., & Jones, A. L. (2022). Which facial features are central in impression formation? *Social Psychological and Personality Science*, 13(2), 553–561. <https://doi.org/10.1177/19485506211034979>
- Johnston, V. S., Hagel, R., Franklin, M., Fink, B., & Grammer, K. (2001). Male facial attractiveness: Evidence for hormone-mediated adaptive design. *Evolution and Human Behavior*, 22(4), 251–267. [https://doi.org/10.1016/S1090-5138\(01\)00066-6](https://doi.org/10.1016/S1090-5138(01)00066-6)
- Jones, A. L., & Jaeger, B. (2019). Biological bases of beauty revisited: The effect of symmetry, averageness, and sexual dimorphism on female facial attractiveness. *Symmetry*, 11(2), 279. <https://doi.org/10.3390/sym11020279>
- Kleisner, K., Priplatova, L., Frost, P., & Flegr, J. (2013). Trustworthy-looking face meets brown eyes. *PLoS One*, 8(1), e53285. <https://doi.org/10.1371/journal.pone.0053285>
- Kosinski, M. (2017). Facial width-to-height ratio does not predict self-reported behavioral tendencies. *Psychological Science*, 28(11), 1675–1682. <https://doi.org/10.1177/0956797617716929>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin*, 126(3), 390–423. <https://doi.org/10.1037/0033-2909.126.3.390>
- Lenth, R. (2018). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (1.2.3) [R; R]. <https://cran.r-project.org/package=emmeans>
- Little, A. C., Jones, B. C., & DeBruine, L. M. (2011). Facial attractiveness: Evolutionary based research. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1568), 183–192. <https://doi.org/10.1098/rstb.2010.0287>

- the Royal Society B: Biological Sciences, 366(1571), 1638–1659. <https://doi.org/10.1098/rstb.2010.0404>
- Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, 47(4), 1122–1135. <https://doi.org/10.3758/s13428-014-0532-5>
- McArthur, L. Z., & Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychological Review*, 90(3), 215–238. <https://doi.org/10.1037/0033-295X.90.3.215>
- Montagne, B., Kessels, R. P. C., Frigerio, E., de Haan, E. H. F., & Perrett, D. I. (2005). Sex differences in the perception of affective facial expressions: Do men really lack emotional sensitivity? *Cognitive Processing*, 6(2), 136–141. <https://doi.org/10.1007/s10339-005-0050-6>
- Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E.-J. (2011). Erroneous analyses of interactions in neuroscience: A problem of significance. *Nature Neuroscience*, 14(9), 1105–1107. <https://doi.org/10.1038/nn.2886>
- Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review*, 108(2), 291–310. <https://doi.org/10.1037/0033-295X.108.2.291>
- Nosek, B. A. (2007). Implicit–explicit relations. *Current Directions in Psychological Science*, 16(2), 65–69. <https://doi.org/10.1111/j.1467-8721.2007.00477.x>
- Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, 105(32), 11087–11092. <https://doi.org/10.1073/pnas.0805664105>
- Ozener, B. (2012). Facial width-to-height ratio in a Turkish population is not sexually dimorphic and is unrelated to aggressive behavior. *Evolution and Human Behavior*, 33(3), 169–173. <https://doi.org/10.1016/j.evolhumbehav.2011.08.001>
- Payne, B. K., Burkley, M. A., & Stokes, M. B. (2008). Why do implicit and explicit attitude tests diverge? The role of structural fit. *Journal of Personality and Social Psychology*, 94(1), 16–31. <https://doi.org/10.1037/0022-3514.94.1.16>
- Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, 89(3), 277–293. <https://doi.org/10.1037/0022-3514.89.3.277>
- Payne, B. K., Govorun, O., & Arbuckle, N. L. (2008). Automatic attitudes and alcohol: Does implicit liking predict drinking? *Cognition & Emotion*, 22(2), 238–271. <https://doi.org/10.1080/02699930701357394>
- Payne, K., & Lundberg, K. (2014). The affect misattribution procedure: Ten years of evidence on reliability, validity, and mechanisms. *Social and Personality Psychology Compass*, 8(12), 672–686. <https://doi.org/10.1111/spc3.12148>
- Perrett, D. I., Burt, D. M., Penton-Voak, I. S., Lee, K. J., Rowland, D. A., & Edwards, R. (1999). Symmetry and human facial attractiveness. *Evolution and Human Behavior*, 20(5), 295–307. [https://doi.org/10.1016/S1090-5138\(99\)00014-8](https://doi.org/10.1016/S1090-5138(99)00014-8)
- Rhodes, G., Simmons, L. W., & Peters, M. (2005). Attractiveness and sexual behavior: Does attractiveness enhance mating success? *Evolution and Human Behavior*, 26(2), 186–201. <https://doi.org/10.1016/j.evolhumbehav.2004.08.014>
- Ross, D. A., & Gauthier, I. (2015). Holistic processing in the composite task depends on face size. *Visual Cognition*, 23(5), 533–545. <https://doi.org/10.1080/13506285.2015.1049678>
- Saribay, S. A., Biten, A. F., Meral, E. O., Aldan, P., Třebický, V., & Kleisner, K. (2018). The Bogazici face database: Standardized photographs of Turkish faces with supporting materials. *PLoS One*, 13(2), e0192018. <https://doi.org/10.1371/journal.pone.0192018>
- Saribay, S. A., & Kleisner, K. (2018). Are political views and religiosity related to facial morphology? Evidence from a Turkish sample. *Personality and Individual Differences*, 124, 25–30. <https://doi.org/10.1016/j.paid.2017.11.035>
- Scheib, J. E., Gangestad, S. W., & Thornhill, R. (1999). Facial attractiveness, symmetry and cues of good genes. *Proceedings of the Royal Society B: Biological Sciences*, 266(1431), 1913–1917. <https://doi.org/10.1098/rspb.1999.0866>
- Schwartz, A. J., Boduroglu, A., & Gutchess, A. H. (2014). Cross-cultural differences in categorical memory errors. *Cognitive Science*, 38(5), 997–1007. <https://doi.org/10.1111/cogs.12109>
- Slepian, M. L., Young, S. G., & Harmon-Jones, E. (2017). An approach-avoidance motivational model of trustworthiness judgments. *Motivation Science*, 3(1), 91–97. <https://doi.org/10.1037/mot0000046>
- Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychological Science*, 21(3), 349–354. <https://doi.org/10.1177/0956797610362647>
- Sutherland, C. A. M., Liu, X., Zhang, L., Chu, Y., Oldmeadow, J. A., & Young, A. W. (2018). Facial first impressions across culture: Data-driven modeling of Chinese and British perceivers' unconstrained facial impressions. *Personality and Social Psychology Bulletin*, 44(4), 521–537. <https://doi.org/10.1177/0146167217744194>
- Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Michael Burt, D., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition*, 127(1), 105–118. <https://doi.org/10.1016/j.cognition.2012.12.001>
- Thornhill, R., & Gangestad, S. W. (1999). Facial attractiveness. *Trends in Cognitive Sciences*, 3(12), 452–460. [https://doi.org/10.1016/S1364-6613\(99\)01403-5](https://doi.org/10.1016/S1364-6613(99)01403-5)
- Todorov, A. (2008). Evaluating faces on trustworthiness. *Annals of the New York Academy of Sciences*, 1124(1), 208–224. <https://doi.org/10.1196/annals.1440.012>
- Todorov, A., Olivola, C., Dotsch, R., & Mende-Siedlecki, P. (2014). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology*, 66(1), 519–545. <https://doi.org/10.1146/annurev-psych-113011-143831>
- Třebický, V., Fialová, J., Kleisner, K., Roberts, S. C., Little, A. C., & Havlíček, J. (2015). Further evidence for links between facial width-to-height ratio and fighting success: Commentary on Zilioli et al. (2014). *Aggressive Behavior*, 41(4), 331–334. <https://doi.org/10.1002/ab.21559>
- Tsujimura, H., & Banissy, M. J. (2013). Human face structure correlates with professional baseball performance: Insights from professional Japanese baseball players. *Biology Letters*, 9(3), 20130140. <https://doi.org/10.1098/rsbl.2013.0140>
- Weston, E. M., Friday, A. E., & Liò, P. (2007). Biometric evidence that sexual selection has shaped the hominin face. *PLoS One*, 2(8), e710. <https://doi.org/10.1371/journal.pone.0000710>

- Wheeler, L., & Kim, Y. (1997). What is beautiful is culturally good: The physical attractiveness stereotype has different content in collectivistic cultures. *Personality and Social Psychology Bulletin*, 23(8), 795–800. <https://doi.org/10.1177/0146167297238001>
- Whitehouse, A. J. O., Gilani, S. Z., Shafait, F., Mian, A., Tan, D. W., Maybery, M. T., Keelan, J. A., Hart, R., Handelsman, D. J., Goonawardene, M., & Eastwood, P. (2015). Prenatal testosterone exposure is related to sexually dimorphic facial morphology in adulthood. *Proceedings of the Royal Society B: Biological Sciences*, 282(1816), 20151351. <https://doi.org/10.1098/rspb.2015.1351>
- Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-Ms exposure to a face. *Psychological Science*, 17(7), 592–598. <https://doi.org/10.1111/j.1467-9280.2006.01750.x>
- Wilson, J. P., & Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal-sentencing outcomes. *Psychological Science*, 26(8), 1325–1331. <https://doi.org/10.1177/0956797615590992>
- Zebrowitz, L. A., & Collins, M. A. (1997). Accurate social perception at zero acquaintance: The affordances of a Gibsonian approach. *Personality and Social Psychology Review*, 1(3), 204–223. https://doi.org/10.1207/s15327957pspr0103_2
- Zebrowitz, L. A., Fellous, J.-M., Mignault, A., & Andreoletti, C. (2003). Trait impressions as overgeneralized responses to adaptively significant facial qualities: Evidence from connectionist modeling. *Personality and Social Psychology Review*, 7(3), 194–215. https://doi.org/10.1207/S15327957PSPR0703_01
- Zebrowitz, L. A., & Montepare, J. (2006). The ecological approach to person perception: Evolutionary roots and contemporary offshoots. In *Evolution and social psychology* (pp. 81–113). Psychosocial Press.
- Zebrowitz, L. A., Wang, R., Bronstad, P. M., Eisenberg, D., Undurraga, E., Reyes-García, V., & Godoy, R. (2012). First impressions from faces among U.S. and culturally isolated Tsimane' people in the Bolivian rainforest. *Journal of Cross-Cultural Psychology*, 43(1), 119–134. <https://doi.org/10.1177/0022022111411386>