The "Central Tendency Bias" in the assessment of facial attractiveness in group-based and individual ratings—A survey-based study in 727 volunteers

Michael Alfertshofer [a],[*],[1],[2], Joanna Kempa [b],[1],
Brian S. Biesman [c], Samuel Knoedler [d], Leonard Knoedler [e],
Carlos Bravo [f], Kristina Davidovic [g], Boguslaw Antoszewski [h],
Wojciech Timler [i], Anna Kasielska-Trojan [h],[1],
Sebastian Cotofana [j],[k],[l],[1]

[a] Department of Oromaxillofacial Surgery, Ludwig-Maximilians-University, Munich, Germany
[b] Individual Course of Study in the Department of Plastic, Reconstructive and Aesthetic Surgery, Institute of Surgery, Medical University of Łódź, Poland
[c] Private Practice, Nashville, TN, USA
[d] Division of Plastic Surgery, Department of Surgery, Yale New Haven Hospital, Yale School of Medicine, New Haven, CT, USA
[e] Division of Plastic and Reconstructive Surgery, Department of Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA
[f] Private Practice, San José, Costa Rica
[g] Department of Radiology and Magnetic Resonance Imaging, Clinical Center of Serbia, Belgrade, Serbia
[h] Department of Plastic, Reconstructive, and Aesthetic Surgery, Institute of Surgery, Medical University of Łódź, Poland
[i] Department of Family Medicine, Medical University of Łódź, Poland
[j] Department of Dermatology, Erasmus Medical Center, Rotterdam, the Netherlands
[k] Center for Cutaneous Research, Blizard Institute, Queen Mary University of London, London, UK
[l] Department of Plastic and Reconstructive Surgery, Guangdong Second Provincial General Hospital, Guangzhou, Guangdong Province, China

* Corresponding author.
  E-mail address: michaelalfertshofer@gmx.de (M. Alfertshofer).
[1] contributed equally.
[2] Socialmedia: michael.alfertshofer.

Abstract   *Background:* The increasing number of esthetic procedures emphasizes the need for effective evaluation methods of outcomes. Current practices include the individual practitioners' judgment in conjunction with standardized scales, often relying on the comparison of before and after photographs. This study investigates whether comparative evaluations influence the perception of beauty and aims to enhance the accuracy of esthetic assessments in clinical and research settings.

Objective: To compare the evaluation of attractiveness and gender characteristics of faces in group-based versus individual ratings.

Methods: A sample of 727 volunteers (average age of 29.5 years) assessed 40 facial photographs (20 male, 20 female) for attractiveness, masculinity, and femininity using a 5-point Likert scale. Each face was digitally edited to display varying ratios in four lip-related proportions: vertical lip position, lip width, upper lip esthetics, and lower lip esthetics. Participants rated these images both in an image series (group-based) and individually.

Results: Differences in the perception of the most attractive/masculine/feminine ratios for each lip proportion were found in both the group-based and individual ratings. Group ratings exhibited a significant central tendency bias, with a preference for more average outcomes compared with individual ratings, with an average difference of 0.50 versus 1.00. (p = 0.033)

Conclusion: A central tendency bias was noted in evaluations of attractiveness, masculinity, and femininity in group-based image presentation, indicating a bias toward more "average" features. Conversely, individual assessments displayed a preference for more pronounced, "non-average" appearances, thereby possibly pointing toward a malleable "intrinsic esthetic blueprint" shaped by comparative context.

© 2024 British Association of Plastic, Reconstructive and Aesthetic Surgeons. Published by Elsevier Ltd.

According to the annual procedural report released by the American Society of Plastic Surgeons, a total of 1,498,361 surgical and 23,672,269 minimally invasive cosmetic procedures were performed in 2022, underscoring the constantly growing need and acceptance of esthetic procedures in today's society.[1] Consequently, the increasing demand for such esthetic procedures underscores the necessity for reliable and effective assessment methods of the esthetic status, both before and after a procedure. Esthetic assessment is typically carried out according to the practitioners' expertise in conjunction with standardized scales tailored for clinical assessment. The development of such scales relies on semiquantitative ratings of images, which help to categorize and rank esthetic features to support decision-making in both clinical and research settings. Such images are presented to the raters either in pairs (before and after) or on an individual basis, the latter especially in scenarios necessitating blinded assessment to ensure impartiality. Blinded rating is regarded to be of advantage over paired rating because no comparison or reference is available, and a more objective and unbiased outcome could be expected. This accounts for the effect of attractiveness comparison, which is known to influence the perception of beauty, as esthetic evaluations are not absolute but rather based on comparisons.[2-4] A recent functional magnetic resonance imaging (fMRI) imaging-based study substantiated this by showing overlapping neural pathways and brain loci involved in the comparison of physical attractiveness and nonsocial magnitudes, such as the evaluation of other people's sizes.[5,6]

This creates a dilemma when it comes to the evaluation of beauty and esthetic outcomes: it could be speculated that the outcome of an esthetic evaluation is different when individual images are rated or when additional images are presented to the rater as a reference. However, at this point, it remains elusive whether such an effect exists in the evaluation of beauty and esthetic parameters based on images and whether such an effect can be quantified. Therefore, it is the objective of this study to investigate the effect of comparative evaluation of attractiveness and masculinity/femininity when performing esthetic assessments of facial images. Frontal images showing five different ratios of four lip-related proportions were presented to study volunteers both individually and in the form of an image series. The study volunteers had to evaluate each presented image twice: presented in an image series (i.e., group-based) versus presented individually. The comparison between the two different rating modalities is the subject of this study, and the results will hopefully provide insights into the effects of comparative attractiveness assessment. It is hoped that the outcome will enable practitioners and researchers to critically re-evaluate patient assessment outcomes and study designs to provide improved and unbiased patient care.

Material and methods

Study sample

This cross-sectional observational study investigated the survey results of 727 volunteers (621 females and 106 males), all with a Caucasian ethnic background of Polish

Table 1 Comparison of rankings in terms of attractiveness and masculinity/femininity as assessed in the group-based and individual rating for the different ratios of the four investigated lip proportions. Congruencies in the group-based and individual ranking are indicated in bold.

	Males				Females			
	Attractiveness Group-based rating	Attractiveness Individual rating	Masculinity Group-based rating	Masculinity Individual rating	Attractiveness Group-based rating	Attractiveness Individual rating	Femininity Group-based rating	Femininity Individual rating
Lip position (vertical)								
1:1	10 (1.2%)	3.86 ± 0.9	22 (2.7%)	3.83 ± 1.0	9 (1.1%)	2.91 (1.0)	13 (1.6%)	2.50 (0.9)
1:1.6	25 (3.0%)	3.26 ± 0.9	34 (4.1%)	3.27 ± 1.0	22 (2.7%)	2.33 (0.8)	13 (1.6%)	2.16 (0.8)
1:2	**477 (57.6%)**	2.85 ± 0.8	**392 (47.3%)**	2.83 ± 0.8	**596 (72.0%)**	**2.15 (0.7)**	**514 (62.1%)**	**2.05 (0.7)**
1:2.5	167 (20.2%)	**2.65 ± 0.7**	220 (26.6%)	2.59 ± 0.8	106 (12.8%)	2.71 (0.7)	150 (18.1%)	2.58 (0.7)
1:3	149 (18.0%)	2.69 ± 0.8	160 (19.3%)	**2.42 ± 0.7**	95 (11.5%)	3.09 (0.7)	138 (16.7%)	2.86 (0.8)
Lip width (horizontal)								
1:1.8	22 (2.7%)	4.27 (0.7)	32 (3.9%)	3.63 (1.0)	1 (0.1%)	4.24 (0.7)	5 (0.6%)	3.97 (0.9)
1:2	5 (0.6%)	3.72 (0.8)	26 (3.1%)	3.24 (0.9)	67 (8.1%)	3.57 (0.8)	63 (7.6%)	3.35 (1.0)
1:2.2	88 (10.6%)	2.94 (0.8)	153 (18.5%)	2.68 (0.8)	**424 (51.2%)**	**2.75 (0.8)**	**356 (43.0%)**	2.63 (0.9)
1:2.5	**352 (42.5%)**	**2.51 (0.8)**	**336 (40.6%)**	**2.25 (0.7)**	314 (37.9%)	2.83 (0.7)	335 (40.5%)	**2.55 (0.8)**
1:3	361 (43.6%)	2.97 (0.8)	281 (33.9%)	2.89 (0.9)	22 (2.7%)	3.54 (0.8)	69 (8.3%)	3.06 (1.0)
Upper lip esthetics								
1:1	34 (4.1%)	3.86 (0.9)	19 (2.3%)	3.83 (1.0)	157 (19.0%)	2.91 (1.0)	**254 (30.7%)**	2.50 (0.9)
1:1.6	88 (10.6%)	3.26 (0.9)	51 (6.2%)	3.27 (1.0)	283 (34.2%)	2.33 (0.8)	240 (29.0%)	2.16 (0.8)
1:2	**279 (33.7%)**	**2.65 (0.7)**	**277 (33.5%)**	2.83 (0.8)	**296 (35.7%)**	**2.15 (0.7)**	**254 (30.7%)**	**2.05 (0.7)**
1:3	247 (29.8%)	2.69 (0.8)	233 (28.1%)	2.59 (0.8)	56 (6.8%)	2.71 (0.7)	39 (4.7%)	2.58 (0.7)
1:4	180 (21.7%)	2.69 (0.8)	248 (30.0%)	**2.42 (0.7)**	36 (4.3%)	3.09 (0.7)	41 (0.5%)	2.86 (0.8)
Lower lip esthetics								
1:1	14 (1.7%)	4.59 (0.6)	13 (1.6%)	4.16 (0.9)	12 (1.4%)	4.44 (0.6)	53 (6.4%)	3.68 (1.1)
1:1.6	9 (1.1%)	3.95 (0.8)	6 (0.7%)	3.65 (1.0)	55 (6.6%)	3.25 (0.9)	86 (10.4%)	2.85 (1.0)
1:2	77 (9.3%)	3.15 (0.9)	67 (8.1%)	3.06 (1.0)	164 (19.8%)	2.95 (0.9)	234 (28.3%)	2.65 (0.9)
1:3	**373 (45.0%)**	2.27 (0.7)	323 (39.0%)	2.21 (0.7)	**503 (60.7%)**	2.70 (0.7)	**375 (45.3%)**	**2.55 (0.7)**
1:4	355 (42.9%)	**2.14 (0.7)**	**419 (50.6%)**	**2.12 (0.6)**	94 (11.4%)	**2.69 (0.7)**	80 (9.7%)	2.56 (0.8)

Bold values indicate the most commonly selected ratio for each proportion.
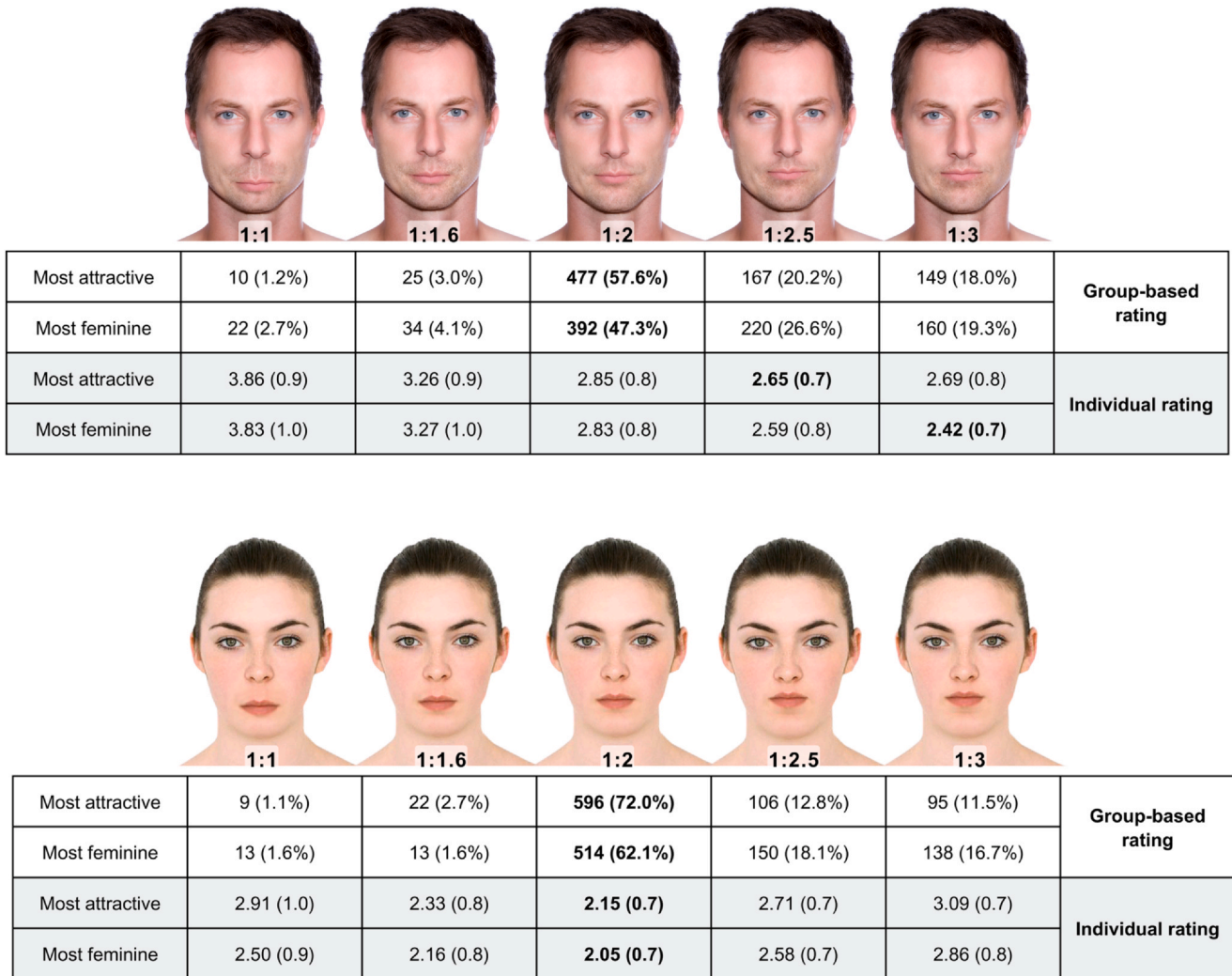
Figure 1 The influence of different proportions in the vertical lip position on the perception of attractiveness and masculinity/femininity in group-based versus individual ratings. The most attractive/masculine/feminine ratio is highlighted in bold text.

origin. Conducted at the University Hospital of Medical University of Łódź, Poland, the research spanned from September 2022 to April 2023.

Prior to their inclusion into the study, volunteers were provided with comprehensive information regarding the objectives and scope pertinent to this study. Each volunteer provided written informed consent explicitly for the use of their demographic and analytical data solely within the confines of this study. The research received ethical approval from the Bioethical Commission of the Medical University of Łódź, with the protocol number RNN/217/22/KE.

Experimental study setup

All 727 volunteers were asked to evaluate 40 frontal facial images (20 male, 20 female) based on their attractiveness, as well as for their perceived masculinity/femininity. The rating of the volunteers was based on a five-point Likert scale ranging from 1 (very attractive/very masculine/very feminine) to 5 (very unattractive/very unmasculine/very unfeminine). The 40 images were presented online via a questionnaire (accessible at: https://forms.gle/mUcPiGZNbsZwyLcW7), and results were collected likewise online following the completion of the rating (Table 1).

The evaluated images were frontal photographs of one male and one female, both of Caucasian ethnic background. The images were digitally modified using Adobe Photoshop Version 21 (Adobe Software, San José, CA, USA). The proportions modified were the length of the upper and lower ergotrid to the respective vermilion thickness, the width of the oral commissure to the respective mandibular angle and the position of the mouth between chin and nose. The following four proportions with five respective ratios were used:

- *Vertical lip position*: Philtrum length: Lower lip vermilion-gnathion (1:1; 1:1.6; 1:2; 1:2.5; 1:3)
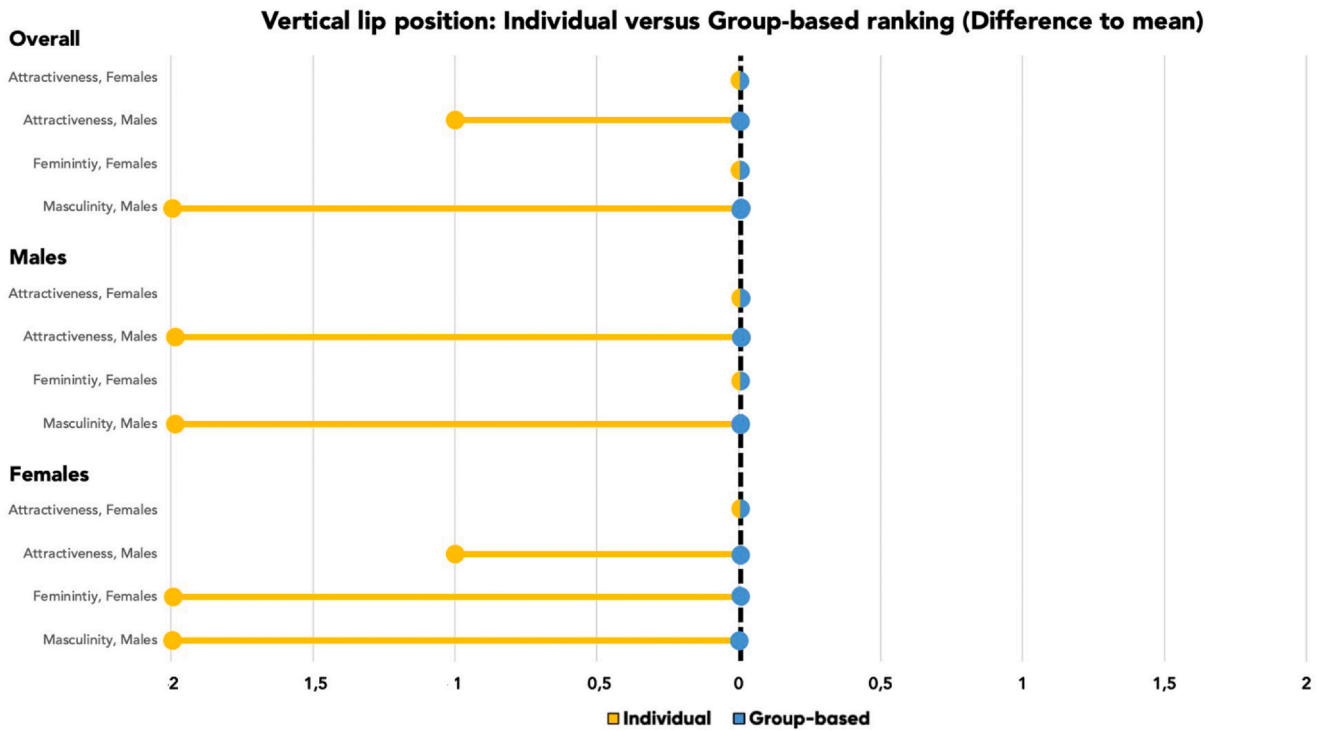
Figure 2 Dumbbell plot visualizing the average difference of the image rated as most attractive/masculine/feminine to the central image for different proportions in the vertical lip position, in group-based versus individual ratings for overall, male and female raters.

- *Lips width*: Cheilion-cheilion: Gonion-gonion (1:1.8; 1:2; 1:2.2; 1:2.5; 1:3)
- *Upper lip esthetics*: Upper lip: Philtrum length (1:1; 1:1.6; 1:2; 1:3; 1:4)
- *Lower lip esthetics*: Lower lip: Lower lip vermilion-gnathion (1:1; 1:1.6; 1:2; 1:3; 1:4)

The image display was based on the study hypothesis and varied between (1) individual image display (and rating) versus (2) group-based image display (and rating).

Ad (1): The raters observed and rated each image individually without the option to see the other ratios (see above). This required the raters to evaluate each image on an individual basis without the possibility to compare.

Ad (2): The raters observed and rated the respective series of modified images at the same time and had the possibility to compare between the differently modified ratios (see above). In other words, the raters were able to see the entire series of generated ratios at the same time and therefore had the possibility to compare.

Therefore, the ratings of both image presentation modalities (individual vs. group-based) resulted in the rating of each image twice.

Statistical analysis

The investigative approach of this study compared the ratings of attractiveness, masculinity, and femininity of 40 modified facial images between a group-based vs. individual image presentation. To detect such differences, the interclass correlation coefficient (ICC) was calculated, and a proprietary method was applied, termed difference to the mean calculation: the five displayed images were ranked according to their distance to the central image of the image series. The image in the center (of the five) received rank 0, the two images on both sides of the central image received rank 1, and the images furthest away from the center received rank 2; all numbers were two-sided. The distance of the most attractive/masculine/feminine ratio to the central image was calculated for the group-based and individual ratings. The difference between ranks was used to calculate the central tendency bias between the group-based vs. individual image presentation. The non-parametric Wilcoxon signed rank test was calculated to identify differences between sex and esthetic category. All computations were executed using SPSS Statistics 27 (IBM, Armonk, NY, USA), with statistical significance set at a probability level of ≤ 0.05 to guide conclusions.

Results

Demographic description of the study volunteers

This study analyzed the responses of n = 727 volunteers comprising 106 males and 621 females with an average age of 29.5 (9.9) years (range: 16–71).

Lip width						
	1:1.8	**1:2**	**1:2.2**	**1:2.5**	**1:3**	
Most attractive	22 (2.7%)	5 (0.6%)	88 (10.6%)	352 (42.5%)	**361 (43.6%)**	Group-based rating
Most masculine	32 (3.9%)	26 (3.1%)	153 (18.5%)	**336 (40.6%)**	281 (33.9%)	
Most attractive	4.27 (0.7)	3.72 (0.8)	2.94 (0.8)	**2.51 (0.8)**	2.97 (0.8)	Individual rating
Most masculine	3.63 (1.0)	3.24 (0.9)	2.68 (0.8)	**2.25 (0.7)**	2.89 (0.9)	

	1:1.8	**1:2**	**1:2.2**	**1:2.5**	**1:3**	
Most attractive	1 (0.1%)	67 (8.1%)	**424 (51.2%)**	314 (37.9%)	22 (2.7%)	Group-based rating
Most feminine	5 (0.6%)	63 (7.6%)	**356 (43.0%)**	335 (40.5%)	69 (8.3%)	
Most attractive	4.24 (0.7)	3.57 (0.8)	**2.75 (0.8)**	2.83 (0.7)	3.54 (0.8)	Individual rating
Most feminine	3.97 (0.9)	3.35 (1.0)	2.63 (0.9)	**2.55 (0.8)**	3.06 (1.0)	

Figure 3 The influence of different proportions in the lip width on the perception of attractiveness and masculinity/femininity in group-based versus individual ratings. The most attractive/masculine/feminine ratio is highlighted in bold text.

Lip esthetics

Vertical lip position
The most attractive vertical lip position was in a group-based evaluation, the 1:2 ratio (males and females), whereas it was the 1:2 ratio for females and the 1:2.5 ratio for males when assessed individually. The most masculine vertical lip position was the 1:2 ratio in a group-based rating and the 1:3 ratio in an individual rating. The most feminine vertical lip position was the 1:2 ratio for both group-based and individual ratings (Figures 1 and 2).

Lip width
The most attractive lip width was, in a group-based evaluation, the 1:3 ratio for males and was the 1:2.2 ratio for females. In an individual rating, it was the 1:2.5 ratio for males and was the 1:2.2 ratio for females. The most masculine lip width was the 1:2.5 ratio evaluated both in a group-based and in an individual rating. The most feminine lip width was the 1:2.2 ratio for a group-based rating, whereas it was the 1:2.5 ratio for an individual rating (Figures 3 and 4).

Upper lip esthetics
The most attractive upper lip ratio was, in a group-based evaluation, the 1:2 ratio for both males and females. In an individual rating, it was the 1:3 ratio for males and the 1:2 ratio for females. The most masculine upper lip ratio was the 1:2 ratio evaluated in a group-based and the 1:4 ratio in an individual rating. The most feminine upper lip ratio was the 1:2 ratio in both the group-based and individual ratings (Figures 5 and 6).

Lower lip esthetics
The most attractive lower lip ratio was in a group-based evaluation: the 1:3 ratio for both males and females. In an
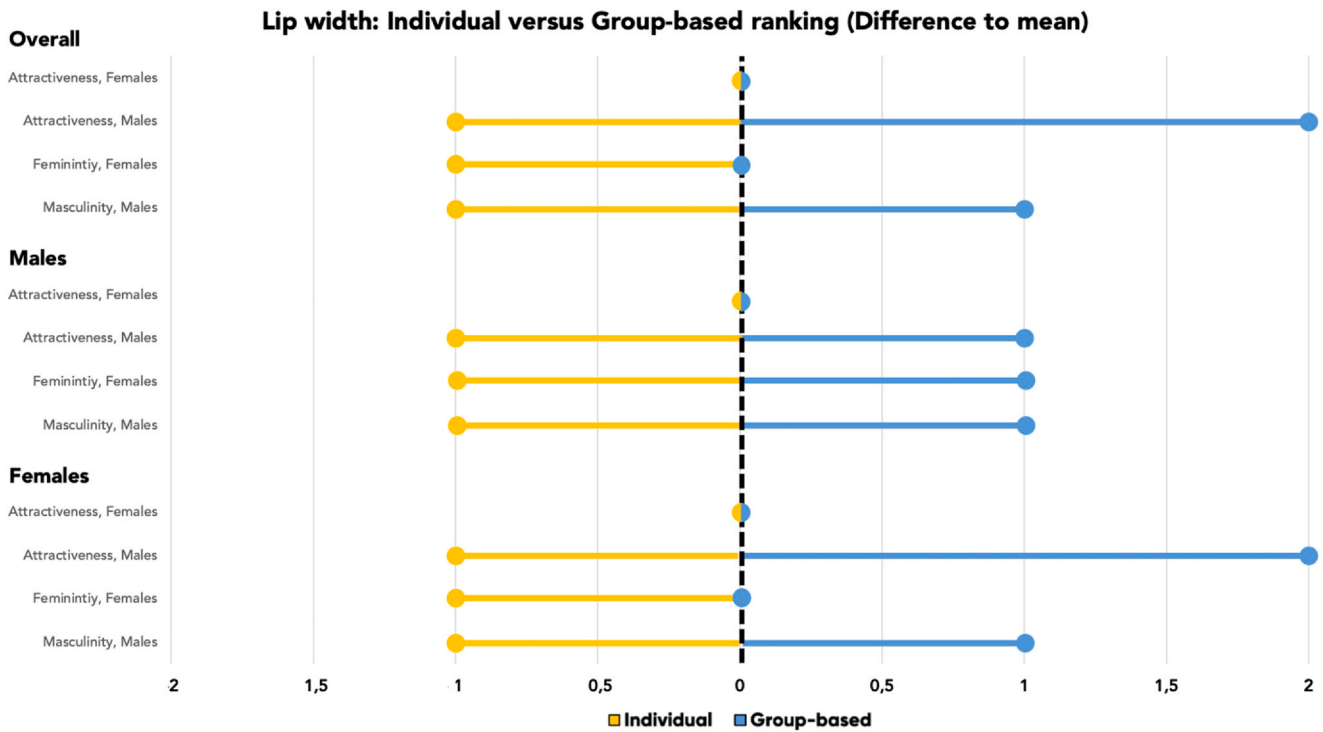
Figure 4 Dumbbell plot visualizing the average difference of the image rated as most attractive/masculine/feminine to the central image for different proportions in the lip width, in group-based versus individual ratings for overall, male and female raters.

individual rating, it was the 1:4 ratio for both males and females. The most masculine lower lip ratio was the 1:4 ratio evaluated in both the group-based and individual ratings. The most feminine lower lip ratio was the 1:3 ratio in both the group-based and individual ratings (Figures 7 and 8).

Differences between group-based vs. individual rating

Investigating the overall central tendency bias (= whether observers gravitate to the central image of the five-image panel in a group-based rating), it was identified that for the group-based rating, the average distance to the central image was 0.50 (0.71) (range: 0-2), whereas the distance for the individual rating was 1.00 (0.79) (range: 0-2) with $p = 0.033$.

Investigating the difference between the rating of male vs. female observers, it was identified that the central tendency bias was in males (group-based vs. individual rating) 0.63 (0.72) (range: 0-2) vs. 1.13 (0.86) (range: 0-2) with $p = 0.038$, whereas it was in females 0.63 (0.78) (range: 0-2) vs. 1.13 (0.78) (range: 0-2) with $p = 0.103$. No gender difference in both the group-based vs individual ratings was identified with $p = 1.000$.

The sex independent central tendency bias for attractiveness (group-based vs. individual rating) was 0.50 (0.73)

(range: 0-2) vs. 0.88 (0.76) (range: 0-2) with $p = 0.180$, whereas for masculinity it was 0.75 (0.83) (range: 0-2) vs. 1.75 (0.43) (range: 1-2) with $p = 0.157$, and for femineity it was 0.25 (0.43) (range: 0-1) vs. 0.50 (0.5) (range: 0-1) with $p = 0.317$.

Utilizing the ICC as a measure of agreement between the group-based and the individual rating, it was identified that the degree of agreement was 0.549, which represents only moderate agreement.

Discussion

The objective of this study was to identify whether a difference can be detected between the group-based and the individual rating of facial frontal images displaying different ratios in four lip-related proportions. The 727 study volunteers were asked to evaluate 40 modified facial images for their attractiveness, masculinity, and femininity twice: when displayed in a series of five images vs. when displayed individually. The null hypothesis of this study was defined as no difference between the two performed ratings (group-based vs. individual). When calculating the ICC as a degree of internal consistency, it was identified that the degree of agreement was 0.549. This finding indicates that only a moderate agreement is present between the two rating modalities despite the same rater evaluating the same image.[7] Therefore, the only difference between these two

Upper lip Aesthetics

	1:1	1:1.6	1:2	1:3	1:4	
Most attractive	34 (4.1%)	88 (10.6%)	**279 (33.7%)**	247 (29.8%)	180 (21.7%)	**Group-based rating**
Most feminine	19 (2.3%)	51 (6.2%)	**277 (33.5%)**	233 (28.1%)	248 (30.0%)	
Most attractive	3.86 (0.9)	3.26 (0.9)	2.85 (0.8)	**2.65 (0.7)**	2.69 (0.8)	**Individual rating**
Most feminine	3.83 (1.0)	3.27 (1.0)	2.83 (0.8)	2.59 (0.8)	**2.42 (0.7)**	

	1:1	1:1.6	1:2	1:3	1:4	
Most attractive	157 (19.0%)	283 (34.2%)	**296 (35.7%)**	56 (6.8%)	36 (4.3%)	**Group-based rating**
Most feminine	**254 (30.7%)**	240 (29.0%)	**254 (30.7%)**	39 (4.7%)	41 (5.0%)	
Most attractive	2.91 (1.0)	2.33 (0.8)	**2.15 (0.7)**	2.71 (0.7)	3.09 (0.7)	**Individual rating**
Most feminine	2.50 (0.9)	2.16 (0.8)	**2.05 (0.7)**	2.58 (0.7)	2.86 (0.8)	

Figure 5 The influence of different proportions in upper lip esthetics on the perception of attractiveness and masculinity/femininity in group-based versus individual ratings. The most attractive/masculine/feminine ratio is highlighted in bold text.

ratings was whether the displayed facial image was presented individually or as a part of an image series along with four other additional images for comparison (= group-based rating).

When applying the difference to the mean calculation method, it was identified that the most esthetically pleasing images were positioned closer to the central image when rated during a group-based image presentation compared with an individual image presentation: 0.50 vs. 1.00 with p = 0.033. A similar trend was observed for each of the investigated esthetic subcategories comprising attractiveness (0.50 vs. 0.88), masculinity (0.75 vs. 1.75), and femininity (0.25 vs. 0.50). Interestingly, no difference was detected when comparing the rating of male vs. female observers with p = 1.000, thereby indicating that the underlying mechanism is sex independent.

The propensity of raters in esthetic assessments to gravitate toward a mean (i.e., central image) when presented in a series might be drawn back to the influence of the "central tendency bias" influencing perceptual judgments. This fallacy has been documented across numerous research fields[8-13] and potentially extends to domains as subjective as the appraisal of facial beauty.

The findings from this study could significantly influence clinical practice and research methodologies. The observed bias toward a mean (i.e., central image) in group-based presentations may profoundly alter our understanding and interpretation of treatment outcomes within the body of
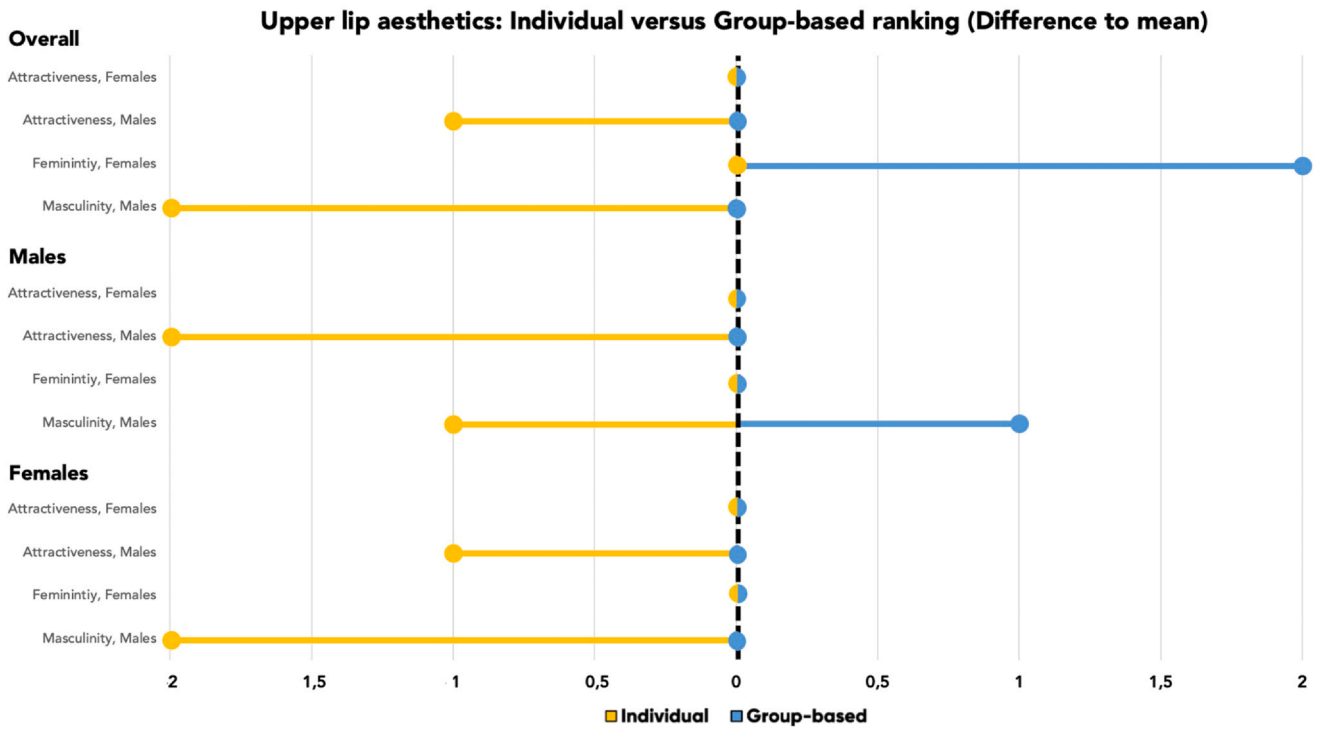
Figure 6 Dumbbell plot visualizing the average difference of the image rated as most attractive/masculine/feminine to the central image for different proportions in upper lip esthetics, in group-based versus individual ratings for overall, male and female raters.

existing literature and clinical practice. Past research that relied on group-based image evaluations for evaluating the success of esthetic procedures may have inadvertently misrepresented treatment effects, potentially leading to an under- or overestimation. Consequently, the selection of representative images plays a pivotal role in accurately portraying desired treatment outcomes. Future research designs might consider acknowledging this effect and adopting individual assessments to enhance accuracy.

In clinical settings, the tendency of comparative assessments to be influenced by surrounding images further raises concerns about patient decision-making processes. Patients exposed to a spectrum of outcomes may develop unrealistic expectations, often centered around the most average result depicted. Leveraging the insights from this study, practitioners could improve counseling by emphasizing individual rather than comparative assessments, thereby helping patients to set more realistic expectations. In general, a more personalized counseling strategy that focuses on the individual esthetic situation rather than comparative visual benchmarks is warranted.

Intriguingly, our findings also revealed a tendency for raters to be more "dareful" in individual assessments, often choosing more pronounced "non-average" appearances. This inclination underscores a discrepancy between an "intrinsic esthetic blueprint," which may be more diverse and genuine, and the homogenized standards often perpetuated by cultural norms and social media. This discrepancy potentially highlights the influence of the comparative environment surrounding us every day on such an "intrinsic esthetic blueprint." Pugach et al. were able to show that

esthetic preferences do not seem to be fixed but rather fluctuate throughout the lifespan, with heightened stability in young adulthood and decreased stability in later years, thereby challenging the common belief that adolescents are mercurial while children and older adults are more adamant.[14] Whether and to what extent such "intrinsic esthetic blueprint", particularly concerning the esthetic perception of facial characteristics, is being sculpted by external factors over a lifetime warrants further exploration.[15] Further research is required to understand how exposure to a broader range of esthetic standards might shape individual preferences and influence decisions regarding cosmetic procedures. In this regard, the homogenous study sample of $n = 727$ volunteers of Caucasian ethnic background can be considered a double-edged sword for this study. On the one hand, the findings obtained are robust due to the large sample size and uniform cultural background, thereby minimizing intercultural differences in the esthetic perception of faces. On the other hand, this homogeneity limits the study by precluding interethnic comparisons that might validate whether the identified effects are consistent across different ethnic groups. Therefore, the generalizability of our findings, including the results obtained and the central tendency bias observed in this study, to other ethnic groups remains an open question. The results presented herein were obtained in a specific demographic context, raising the possibility of varying applicability across different ethnic backgrounds. Future studies should aim to examine this effect in a broader ethnical context, not only in the patient images but also among the observers, to thoroughly investigate and potentially validate our findings across a

Lower lip Aesthetics

	1:1	1:1.6	1:2	1:3	1:4	
Most attractive	14 (1.7%)	9 (1.1%)	77 (9.3%)	**373 (45.0%)**	355 (42.9%)	**Group-based rating**
Most feminine	13 (1.6%)	6 (0.7%)	67 (8.1%)	323 (39.0%)	**419 (50.6%)**	
Most attractive	4.59 (0.6)	3.15 (0.9)	3.15 (0.9)	2.27 (0.7)	**2.14 (0.7)**	**Individual rating**
Most feminine	4.16 (0.9)	3.06 (1.0)	3.06 (1.0)	2.21 (0.7)	**2.12 (0.6)**	

	1:1	1:1.6	1:2	1:3	1:4	
Most attractive	12 (1.4%)	55 (6.6%)	164 (19.8%)	**503 (60.7%)**	94 (11.4%)	**Group-based rating**
Most feminine	53 (6.4%)	86 (10.4%)	234 (28.3%)	**375 (45.3%)**	80 (9.7%)	
Most attractive	4.44 (0.6)	3.25 (0.9)	2.95 (0.9)	2.70 (0.7)	**2.69 (0.7)**	**Individual rating**
Most feminine	3.68 (1.1)	2.85 (1.0)	2.65 (0.9)	**2.55 (0.7)**	2.56 (0.8)	

Figure 7 The influence of different proportions in lower lip esthetics on the perception of attractiveness and masculinity/femininity in group-based versus individual ratings. The most attractive/masculine/feminine ratio is highlighted in bold text.

more diverse population. Such studies are required to comprehend the full spectrum of implications of this effect and ensure that our findings are robust and universally relevant. Further, the images presented to the raters can be regarded as another limitation of this study. While the images were adjusted to showcase five different ratios across four distinct lip proportions, the study was constrained by its use of only a single male and a single female photograph, both individuals of Caucasian ethnic background. Incorporating a more diverse assortment of ratios and proportions of facial features alongside a broader range of facial examples and ethnicities would have significantly enhanced the study. This expansion would facilitate a more comprehensive exploration into whether the observed effects are consistent across a universal spectrum of human faces.

Based on the findings obtained in this study, the use of group-based comparisons for the esthetic evaluation of patients should be reconsidered. It was shown that individual ratings differed significantly from group-based ratings, thereby potentially highlighting an influence of the comparative context on the more accurate unbiased perception of beauty and attractiveness. The development and implementation of novel tools and technologies, such as 3D imaging and morphometric analysis, have the potential to augment current subjective assessment frameworks. Acknowledging the importance of individual esthetic assessment, rather than group-based esthetic assessment, in conjunction with the additional dimension of objective assessments might refine the accuracy of pre- and post-procedural evaluations, ultimately aiming to improve esthetic outcomes and patient satisfaction.
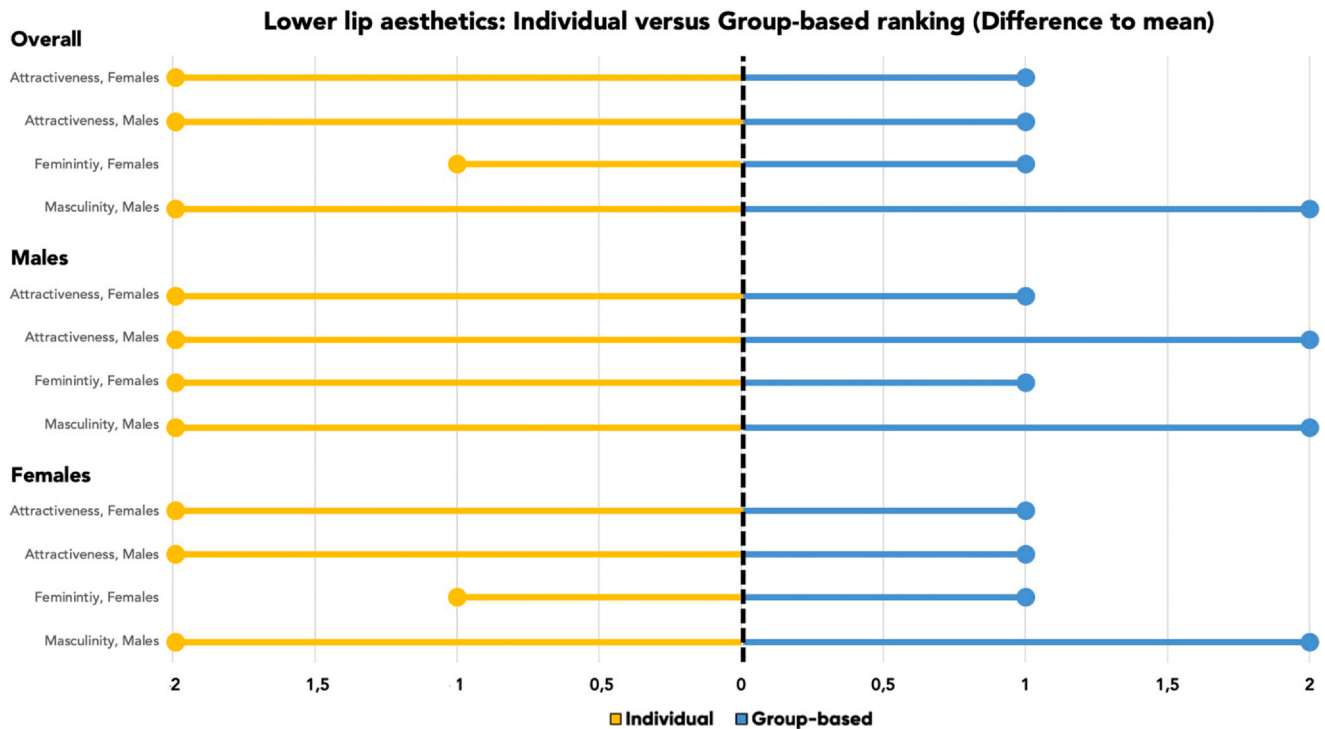
Figure 8 Dumbbell plot visualizing the average difference of the image rated as most attractive/masculine/feminine to the central image for different proportions in lower lip esthetics, in group-based versus individual ratings for overall, male and female raters.

Conclusion

This study highlighted the importance of different image presentation modalities in facial esthetic perception. A significant central tendency bias, as seen in raters gravitating toward more central images of an image series, was noted for group-based ratings. This tendency suggests an unconscious bias toward more normative or "average" features when evaluating faces in a group context. Contrastingly, the individual assessment revealed a distinct set of esthetic preferences with a more pronounced preference for "non-average" appearances, thereby diverging from group trends. These findings point to an underlying "intrinsic esthetic blueprint" that appears to be malleable, influenced by the comparative frameworks prevalent in our visual environment. The perception of beauty is not solely an innate, individual judgment but is significantly shaped by the collective visual narratives we encounter. The design of future studies will need to account for this bias when evaluating pre vs. post images, especially when presented in an image series of gradual differences, to not confound outcomes and to provide an accurate evaluation of a true change.

Ethical approval statement

This project received ethical approval from the Bioethical Commission of the Medical University of Łódź, with the protocol number RNN/217/22/KE.

Patient consent

The images used in this study are modified images generated via computer modification from a previously obtained stock image from Adobe Stock Exchange. With the purchased license, free use of the image for research publication purposes is warranted, and no patient consent is required.

Funding

Author disclosure

References

1. The American Society of Plastic Surgeons. 2022 ASPS Procedural Statistics Release. Published 2022. Accessed October 16, 2023. ⟨https://www.plasticsurgery.org/documents/News/Statistics/2022/plastic-surgery-statistics-report-2022.pdf⟩.
2. Lei Y, He X, Zhao T, Tian Z. Contrast effect of facial attractiveness in groups. *Front Psychol* 2020;**11**:2258. https://doi.org/10.3389/fpsyg.2020.02258.

3. Kenrick DT, Gutierres SE. Contrast effects and judgments of physical attractiveness: When beauty becomes a social problem. *J Pers Soc Psychol* 1980;**38**(1):131–40. https://doi.org/10.1037/0022-3514.38.1.131.

4. Zell E, Balcetis E. The influence of social comparison on visual representation of one's face. *PLoS One* 2012;**7**(5):e36742. https://doi.org/10.1371/journal.pone.0036742.

5. Kedia G, Mussweiler T, Mullins P, Linden DEJ. The neural correlates of beauty comparison. *Soc Cognit Affect Neurosci* 2014;**9**(5):681–8. https://doi.org/10.1093/scan/nst026.

6. Bzdok D, Langner R, Caspers S, et al. ALE meta-analysis on facial judgments of trustworthiness and attractiveness. *Brain Struct Funct* 2011;**215**(3-4):209–23. https://doi.org/10.1007/s00429-010-0287-4.

7. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;**15**(2):155–63. https://doi.org/10.1016/j.jcm.2016.02.012.

8. Attali Y, Bar-Hillel M. Guess where: The position of correct answers in multiple-choice test items as a psychometric variable. *J Educ Meas* 2003;**40**(2):109–28. https://doi.org/10.1111/j.1745-3984.2003.tb01099.x.

9. Douven I. A Bayesian perspective on likert scales and central tendency. *Psychon Bull Rev* 2018;**25**(3):1203–11. https://doi.org/10.3758/s13423-017-1344-2.

10. Klement J. Social Desirability and Central Tendency Bias in the Survey and Variation of Life Satisfaction. Published online 2022. ⟨doi:10.21203/rs.3.rs-1906292/v1⟩.

11. Riskey DR, Parducci A, Beauchamp GK. Effects of context in judgments of sweetness and pleasantness. *Percept Psychophys* 1979;**26**(3):171–6. https://doi.org/10.3758/BF03199865.

12. Roberson D, Damjanovic L, Pilling M. Categorical perception of facial expressions: Evidence for a "Category Adjustment" model. *Mem Cognit* 2007;**35**(7):1814–29. https://doi.org/10.3758/bf03193512.

13. Corbin JC, Crawford LE, Vavra DT. Misremembering emotion: Inductive category effects for complex emotional stimuli. *Mem Cognit* 2017;**45**(5):691–8. https://doi.org/10.3758/s13421-017-0690-7.

14. Pugach C, Leder H, Graham DJ. How stable are human aesthetic preferences across the lifespan? *Front Hum Neurosci* 2017;**11**:289. https://doi.org/10.3389/fnhum.2017.00289.

15. Sola CA, Fabi SG. Perception drift. *Dermatol Surg Publ Am Soc Dermatol Surg* 2019;**45**(12):1747–8. https://doi.org/10.1097/DSS.0000000000001823.