Evolution and Human Behavior 44 (2023) 19–29

Contents lists available at ScienceDirect

Evolution and Human Behavior

journal homepage: www.elsevier.com/locate/ens

Cross-modal associations of human body odour attractiveness with facial and vocal attractiveness provide little support for the backup signals hypothesis: A systematic review and meta-analysis

Vít Třebický [a,b,c,*], Sylvain Delplanque [d], Camille Ferdenzi [e], Bernhard Fink [f,k], Lucie Jelínková [c], Žaneta Pátková [c], S. Craig Roberts [g], Susanne Röder [h], Tamsin K. Saxton [i], Dagmar Schwambergová [c], Zuzana Štěrbová [c,j], Jitka Třebická Fialová [c], Jan Havlíček [c]

[a] *Faculty of Physical Education and Sport, Charles University, Prague, Czech Republic*
[b] *National Institute of Mental Health, Klecany, Czech Republic*
[c] *Faculty of Science, Charles University, Prague, Czech Republic*
[d] *Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland*
[e] *Lyon Neuroscience Research Center, Bron, France*
[f] *Department of Evolutionary Anthropology, University of Vienna, Biosocial Science Information, Biedermannsdorf, Vienna, Austria*
[g] *Division of Psychology, University of Stirling, Stirling, United Kingdom*
[h] *Department of Psychiatry, Sozialstiftung Bamberg, Bamberg, Germany*
[i] *Department of Psychology, Northumbria University, Newcastle, United Kingdom*
[j] *Faculty of Arts, Charles University, Prague, Czech Republic*
[k] *Human Evolution and Archaeological Sciences (HEAS), University of Vienna, Vienna, Austria*

ARTICLE INFO

Keywords:
Mate choice
Face
Voice
Perception
Multiple messages

ABSTRACT

Assessing the attractiveness of potential mating partners typically involves multiple sensory modalities, including the integration of olfactory, visual, and auditory cues. However, predictions diverge on how the individual modalities should relate to each other. According to the *backup signals* hypothesis, multimodal cues provide redundant information, whereas the *multiple messages* hypothesis suggests that different modalities provide independent and distinct information about an individual's mating-related quality. The *backup signals* hypothesis predicts a positive association between assessments based on different modalities, whereas no substantial correlation across modalities is expected under the *multiple messages*hypothesis. Previous studies testing the two hypotheses have provided mixed results, and a systematic evaluation is currently missing.

We performed a systematic review and a meta-analysis of published and unpublished studies to examine the congruence in assessments between human body odour and facial attractiveness, and between body odour and vocal attractiveness. We found positive but weak associations between ratings of body odours and faces ($r = 0.1$, $k = 25$), and between body odours and voices ($r = 0.1$, $k = 9$). No sex differences were observed in the magnitude of effects.

Compared to judgments of facial and vocal attractiveness, our results suggest that assessment of body odour provides independent and non-redundant information about human mating-related quality. Our findings thus provide little support for the *backup signals* hypothesis and may be better explained by the *multiple messages* hypothesis.

1. Introduction

Across many different taxa, individuals assess potential mating partners via telereceptive senses such as vision, olfaction, and hearing (Aglioti & Pazzaglia, 2011). Although some vertebrates appear to rely predominantly on a single sense (Arakawa, Blanchard, Arakawa, Dunlap, & Blanchard, 2008; Candolin, 2003; Gosling & Roberts, 2001), most species, including humans, employ multiple senses (Candolin, 2003;

* Corresponding author at: Faculty of Physical Education and Sport, Charles University, José Martího 31, 162 52 Prague 6, Czech Republic.
E-mail address: vit.trebicky@ftvs.cuni.cz (V. Třebický).

https://doi.org/10.1016/j.evolhumbehav.2022.11.001
Received 8 February 2022; Received in revised form 28 October 2022; Accepted 7 November 2022
Available online 30 November 2022

Higham & Hebets, 2013) in their assessment. Frog calls, for example, are often accompanied by conspicuous vocal sac movements and/or water surface vibrations, while many bird species show complex, rhythmic and vigorous visual displays during courtship singing (for a review, see Halfwerk et al., 2019).

Perceived variation in these physical traits may provide information about an individual's mating-related quality, such as health and fertility (e.g., Grammer, Fink, Møller, & Thornhill, 2003; Rhodes, 2006; Thornhill & Gangestad, 1999b). As the judgment of an individual's attractiveness based on any single modality entails a certain level of error, using multiple sensory channels could enable a more reliable assessment (Møller & Pomiankowski, 1993). Two competing hypotheses have been proposed to explain the use of multiple modalities in the assessment of potential mates (Groyecka et al., 2017; Higham & Hebets, 2013). According to the '*backup signals*' hypothesis (Grammer, Fink, Juette, Ronzal, & Thornhill, 2001; also coined redundant signalling, Møller & Pomiankowski, 1993; Thornhill & Grammer, 1999), certain cues may provide similar (redundant) information; assessing this same information in several different modalities will then tend to reduce error and facilitate a more accurate overall assessment of underlying quality. In contrast, the *multiple messages* hypothesis (Cunningham, Barbee, & Pike, 1990; Møller & Pomiankowski, 1993) suggests that each trait provides distinct and independent (non-redundant) information about an individual's mating-related quality, but in combination, these can facilitate more accurate assessment of overall individual quality than any single cue in isolation. With all this in mind, we can make predictions to test these two ideas. One can expect that if attractiveness assessments based on different sensory channels are closely and positively associated, such congruence would suggest redundancy in information across traits and provide support for the backup hypothesis. Weak or absent cross-modal congruence (i.e. cues convey non-redundant information), however, would support the multiple messages hypothesis. The mating-related animal research provided some support for both of these hypotheses. The use of backup signals of quality was demonstrated, for instance, in *Drosophila saltans* where removing one courtship component (either visual, auditory, chemical or tactile) did not eliminate the female's decision to mate (Colyott, Odu, & Gleason, 2016). On the other hand, the study on peacock spiders (*Maratus volans*) showed that both visual and vibratory signalling is important for mating success supporting the multiple messages hypothesis (Girard, Elias, & Kasumovic, 2015). Overall, the majority of available animal research seems to provide more evidence in favour of the multiple messages hypothesis (Candolin, 2003).

Most research on human mate preferences has focused on visual cues, typically by investigating people's assessments of facial and/or body attractiveness. Although physical appearance certainly plays a prominent role (Groyecka et al., 2017; Herz & Inzlicht, 2002; Walter et al., 2020), the assessment of attractiveness in potential mating partners is undeniably multimodal. Research suggests that body odour (Havlíček et al., 2008; Roberts et al., 2011) and vocal cues (Hill & Puts, 2016; Pisanski, Feinberg, Oleszkiewicz, & Sorokowska, 2017; Zäske, Skuk, & Schweinberger, 2020) also contribute substantially to human mate preferences (Groyecka et al., 2017). However, studies that examine potential cross-modal congruency and redundancy of attractiveness judgments are scarce. In one of the first such studies, Rikowski and Grammer (1999) reported a positive relationship between judgments of women's faces and their body odour. They also found a similar association in men's faces and odour, when rated by women in the fertile phase of their menstrual cycle Note that authors assessed cycle phase based on counting methods which appears to be highly unreliable, see Gangestad et al., 2016; Havlíček & Roberts, 2022). Rikowski and Grammer concluded that human faces and body odours provide similar information about mate quality. Several other studies have subsequently reported positive associations between perceived attractiveness of faces and body odours (Mahmut & Stevenson, 2019; Roth, Samara, & Kret, 2021; Thornhill et al., 2003; Thornhill & Gangestad, 1999a), although the strengths of some associations were weak and two other studies (Roberts et al., 2011; Röder, Fink, & Jones, 2013) found no support for this association (see Table S0–6 and Fig. 2). Collectively, the available studies provide some support for both the *backup signals* and *multiple messages* hypotheses.

In view of this, we set out to conduct a systematic review and meta-analysis of the relationship between human body odour and facial attractiveness, to test between the two hypotheses. We collated the published studies and complemented these with unpublished datasets. During this process, we noticed that several of the unpublished datasets that we obtained from researchers also contained ratings of perceived vocal attractiveness. Therefore, we also performed meta-analyses of congruence between body odour and vocal attractiveness. As body odour perception and its relation to other modalities are still somewhat overlooked research topics, we focus our study primarily on the relationships between body odour attractiveness and other sensory modalities. Although of interest, the investigation of the association between facial and vocal attractiveness to a comparable extent (e.g. collecting both published and unpublished evidence) is beyond the scope of the current study.

2. Material and methods

2.1. Systematic review and Meta-analysis

2.1.1. Literature search and study selection

Following the PRISMA 2020 protocol (Page et al., 2021) and PRISMA 2020 checklists (see Supplementary material), we conducted a systematic literature search in July 2020 to identify empirical studies reporting data on the associations between perceived body odour and facial and/or vocal attractiveness. We searched the PubMed and Web of Science (WoS) databases. Topics (WoS) and all fields (PubMed) were searched using the keyword combinations *'odour AND face AND attractiveness', 'odour AND facial AND attractiveness', 'odour AND voice AND attractiveness'* and *'odour AND vocal AND attractiveness'* (WoS search query example TS = (odour) AND TS = (face) AND TS = (attractive); PubMed search query example ((odour[Title/Abstract]) AND (face[Title/Abstract])) AND (attractiveness[Title/Abstract]); results for each query and database are provided in the Supplementary material). Studies were also searched through cross-referencing and by direct correspondence with researchers who had published previously on body odour attractiveness. We contacted 13 authors, 7 of whom responded that they had no suitable data, and 6 of whom provided data.[1] Only articles and research papers written in English were reviewed. Both published and unpublished studies were considered. The complete list of search results is reported in Table S0–5 - Systematic literature search and Prisma Flow diagram (Supplementary material).

2.1.2. Inclusion criteria

A two-step selection process was adopted. First, titles and abstracts of studies identified by the search were screened for inclusion by one team member (VT). Studies were included if they met each of the following criteria: focused on humans (not other species); included ratings of body odour samples and either facial photographs or voice recordings (or both); provided data about perceived body odour attractiveness, and perceived facial and/or vocal attractiveness of the target participants. Second, all entries reporting the relevant data or unclear about reporting the relevant data were screened against the same criteria, where their full texts were examined for suitability. Studies were excluded from the meta-analysis if the key data (perceived body odour and facial or voice attractiveness) were collected but the relevant analyses were not

[1] All authors who provided unpublished data were offered co-authorship of the resulting manuscript. Their involvement in the study is described in the Author Contributions list.

conducted or not reported, unless the authors provided respective effect sizes or raw data for effect size calculations after we contacted them.

We used Pearson's *r* (correlation coefficient) as a measure of the effect size of the association between body odour and facial and/or vocal attractiveness. We excluded studies reporting effect size measures that could not be converted to Pearson's *r* and/or were not available from the authors.

For further details, see the PRISMA 2020 Flow Diagram and Table S0–5 (in the Supplementary material) that contains all selection steps.

2.1.3. Data extraction

Data extracted from the selected studies are reported in Table S0–6 - Summary of published and unpublished data. Two research team members (VT and JTF) individually extracted the data, summarised them, and verified their validity.

2.1.4. Analysis

All statistical tests within this article were performed in jamovi (The jamovi project, 2021). We used the MAJOR (Hamilton, 2021) jamovi module to perform a correlation coefficients meta-analysis, following recommendations by Harrer, Cuijpers, Furukawa, and Ebert (2021). The correlation coefficients of the associations between perceived body odour and facial attractiveness and body odour and vocal attractiveness were converted with Fisher's r-to-z transformation and accompanied by their 95% CI. Fisher's r-to-z transform is the recommended procedure for correcting for bias in studies with small sample sizes (Harrer et al., 2021).[2] Separate meta-analyses were performed for correlations between each pair of stimuli (body odour – facial attractiveness and body odour – vocal attractiveness). We performed each meta-analysis first for both target sexes combined and then separately for each target sex; the results for both sexes combined are reported in the main text, and the results for each sex are provided in the Supplementary material (Table S0–7 - Supplementary Meta-analyses results). We assumed that variation in effect sizes between studies was due to sampling error of true effect sizes or because of other (e.g., methodological) differences between studies. Therefore, we used the random-effects model with a restricted maximum-likelihood estimator (Harrer, Cuijpers, Furukawa, & Ebert, 2021) for heterogeneity statistics (Tau^2). Heterogeneity examines whether variation in the observed correlations results from sampling error. Cochran's Q (which tests whether effect size variability across samples is larger than would be expected by sampling error) and I^2 (which indicates the percentage of variability due to true heterogeneity; I^2 values of 25% are considered low, 50% moderate, and 75% high variability (Higgins, Thompson, Deeks, & Altman, 2003) were computed to quantify the proportion of variance in the observed effects attributable to sampling error (i.e., the extent to which true effect sizes vary within a meta-analysis) (Harrer et al., 2021). In the case of heterogeneity, the meta-analytic results are reported with their 95% prediction intervals (PI). We inspected small-study effects and between-study heterogeneity using contour-enhanced funnel plots and Egger's regression test for funnel plot asymmetry (Harrer et al., 2021); this test was carried out only for the association between perceived body odour and facial attractiveness as its usage is recommended when the number of studies (k) is ≥10 (Harrer et al., 2021; Sterne et al., 2011). To explore potential biases in published vs unpublished effects, we tested the moderator effect and performed separate meta-analyses for published and unpublished effects. Lastly, we also explored the potential moderating effect of the rating design (between- and within-subject design) on observed meta-analytic estimates. These comparisons were carried out only for the association between perceived body odour and facial attractiveness, as both published and unpublished effects were available for this association, and the number of available studies was $k \geq 10$.

2.1.5. Power analysis

We performed analyses of statistical power for the meta-analytic effects in both meta-analyses following Quintana (2015) and Quintana and Tiebel (2019). We conducted a sensitivity analysis to estimate what meta-analytical average effects we have the power to observe with the resulting number of effects per meta-analysis, the average number of stimuli per study (within a given meta-analysis), 5% α and β error rates ($p \leq 0.05$ in two-tailed tests, 1-β error probability ≤0.95 Power), and for potentially low, moderate, and high heterogeneity of the effects (Higgins et al., 2003) (Fig. 1).

2.1.6. Effect size distributions

We calculated effect size distributions (ESD) (e.g., Brydges, 2019; Gignac & Szodorai, 2016; Lovakov & Agadullina, 2021; Nordahl-Hansen, Cogo-Moreira, Panjeh, & Quintana, 2022; Quintana, 2017) for both investigated associations (body odour – facial attractiveness and body odour – vocal attractiveness). Alongside meta-analytic averages, ESD can facilitate more accurate power analyses to determine sample and effect sizes when planning future research in a particular area. The ESD primarily allows for the determination of empirically-based normative guidelines. Thus, instead of Cohen's (1988) traditional 'rule of thumb' conventions for correlations ($r \approx 0.10$: small effect; $r \approx 0.30$: moderate effect; $r \approx 0.50$: large effect), ESD serves as an evidence synthesis derived, field-specific benchmark against which effects from individual studies are compared (e.g., whether the observed effect size in a particular study is smaller, average/medium, or larger than in similar studies). We emphasise that the ESD provides effect size comparison with similar studies but is not designed to quantify the practical significance of observed effects.

To examine the distribution of correlation coefficient effect sizes, we calculated the 50th percentile, representing the average effect size, and the 25th and 75th percentiles, as these are equidistant from the average effect size representing small and larger effects size boundaries, respectively (Cohen, 1992; Quintana, 2017).

2.2. Analysis of the unpublished studies

Ten unpublished datasets (further referred to as Studies 1–10) were secured through personal communication. Data on the association between perceived body odour and facial attractiveness were available in all studies; five studies (Study 2, 5, 6, 7, 10) also included data on voice attractiveness. The Supplementary material contains a detailed description of the methods and results of each study, means per target (Table S0–1 - Means per target), and means per modality (Table S0–2 - Means per modality).

2.2.1. The stability and precision of mean rating estimates

To assess whether the number of ratings for each stimulus type within Studies 1–7, 9, 10 and part of Study 8 provided stable estimates, we calculated the point of stability (POS, a point at which means do not substantially change with additional observations) within a corridor of stability of a mean (COS) (Hehman, Xie, Ofosu, & Nespoli, 2018; Schönbrodt & Perugini, 2013) in R x64 (R Core Team, and Team, 2019) via RStudio (R Core Team, 2021). We used the settings following Hehman et al. (2018): for the 1–7 scale (Studies 1–4, 7, 9), the POS was specified as 95% CI of observed values falling within ±0.5 points (approximately 14%) (Fialová et al., 2020), for the 9-point scale (Study 5, −4 to +4 scale used for odour ratings) within ±0.6 points (∼ 14%), for the 0–1000 scale (Study 6) we set POS at 95% CI within ±70 points (∼ 14%), for the 1–10 scale (Study 8, the replication sample) we set POS

[2] Another approach is to use bias-corrected correlations. In the main paper, we report results using the Fisher's r-to-z transforms. We further ran the two presented meta-analyses with bias-corrected correlations for transparency and comparison between other meta-analyses and their effect size treatments; the analyses are reported in the Supplementary material. Both analyses produced essentially the same results with marginally smaller AIC values for Fisher's r-to-z transformed data.

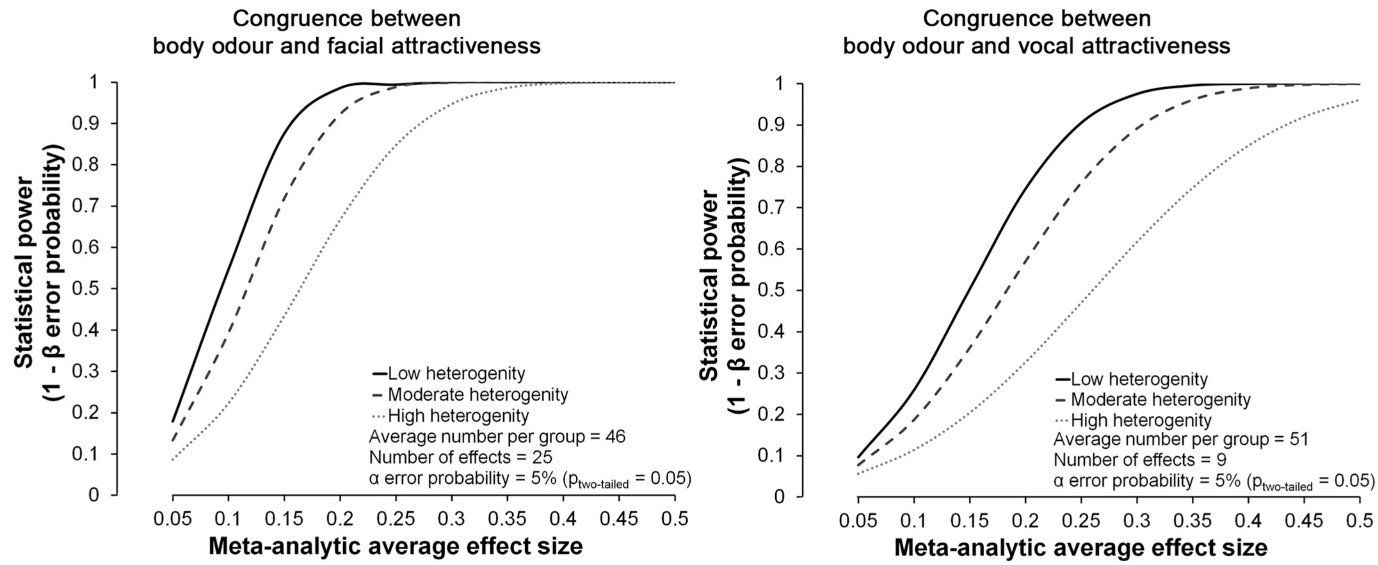

Fig. 1. Power curves for the sensitivity to detect meta-analytic effects as a function of heterogeneity. The plots display the sensitivity analysis for the meta-analysis of congruence between body odour and facial attractiveness (left panel) and between body odour and vocal attractiveness (right panel). Solid, dashed, and dotted curves represent low, moderate, and high heterogeneity. Power curve plots were generated in MS Excel 365 following Quintana *(2015) and* Quintana and Tiebel (2019) and edited in Adobe Photoshop CC2022.

at 95% CI within ±0.7 points (~ 14%) and for the 1–5 scale (Study 10) we set POS at 95% CI within ±0.35 points (~ 14%).

This analysis provided an estimate of the number of raters required to reach predefined POS (and allowed a comparison with the number of raters recruited and an estimation of the size of the raters' pool needed). We further calculated the mean rating precision each study reached with a COS of 95% CI, see Table S0–3 - Point of stability and Intra-class Correlation Coefficients (ICC) in the Supplementary material.

2.2.2. Assessment of inter-rater reliability

To assess inter-rater reliability for each stimulus type in Studies 1–7, 9, 10 and part of Study 8, we calculated the ICC (Koo & Li, 2016) using Reliability analysis in the SimplyAgree (version 0.0.2) jamovi module. We used a two-way random model for average agreement (type ICC2k) and followed recommended thresholds for values <0.5 as indicative of poor reliability, values between 0.5 and 0.75 as being of moderate reliability, values between 0.75 and 0.9 indicating good reliability and values >0.9 indicating excellent reliability (Koo & Li, 2016). See Table S0–3 – Point of stability and ICC in the Supplementary material for individual ICC values.

Further, using a linear mixed-effect model, we explored differences in ICCs for different stimulus types. Results are reported in the Supplementary material (ICC comparison).

2.2.3. Perceptual differences between rating sessions, side-related armpit differences, and an association between short- and long-term attractiveness ratings

In Studies 1, 2, 5, and 7, ratings were recorded in multiple sessions. To test for potential differences between sessions, we specified linear mixed-effect models. Attractiveness rating (for a specific modality) was set as the dependent variable, the number of sessions as a fixed effect factor, and both the rater and target participants' ID as random effects (example model syntax: Odour attractiveness rating ~ session + (1 | rater ID) + (1 | target ID)).

The raters in Study 5 were presented with the target's body odour samples from both armpits (separately, as two stimuli). Therefore, we used a bivariate correlation analysis (on aggregated ratings per armpit and target participant) to assess the association between the ratings of the two odour samples.

In several studies, body odour (Study 6–1, 6–2), facial (Study 4, 5, 6–1, 6–2, 9), and vocal stimuli (Study 6–1, 6–2) were rated for short- and long-term attractiveness. We used a bivariate correlation analysis (on aggregated ratings per scale type and target participant) to assess the association between these two scales. We initially set $r \geq 0.8$ (Brown, 2006) as the level at which we considered the two attractiveness scales as highly correlated and thus difficult to discriminate. In fact, ratings of short-term and long-term attractiveness were highly positively correlated with all r's ≥ 0.856, thus fulfilling our criteria to consider the two ratings numerically interchangeable. We therefore used the long-term attractiveness ratings for subsequent analyses and labelled these simply as 'attractiveness'.

All linear mixed effect models were run using GAMLj jamovi module (Gallucci, 2021) with REML fit; fixed effect factors were set as 'Simple' contrasts and covariate scaling was set to 'Centred'.

For the individual results, see the Methods and Results of each study in the Supplementary material.

2.2.4. Association between attractiveness of different modalities

Previous research reported positive associations between the attractiveness of body odour and facial images (Rikowski & Grammer, 1999; Thornhill et al., 2003; Thornhill & Gangestad, 1999a). Therefore, we ran one-tailed Pearson's r bivariate correlations ($r \geq \rho$) (on aggregated attractiveness ratings per stimulus type and per participant, i.e., the mean rating of a participant was the unit of analysis) between odour and face, and between odour and voice pairs, within each dataset. The resulting correlation coefficients are reported with 95% CI [lower limit, 1].

2.2.5. Power analysis

The current study used data from previous studies; therefore, we calculated the sensitivity to detect effects and their critical values for Exact Correlation (Bivariate normal model) using G*Power (Erdfelder, Faul, Buchner, & Lang, 2009; Faul, Erdfelder, Lang, & Buchner, 2007). The parameters were set to a one-tailed test ($r \geq \rho$), sample size (number of targets per individual dataset), 5% α error probability ($p = 0.05$) and

5% β error probability (1-β error probability = 0.95 Power).[3] For the sensitivity of individual studies, including observed effects and the power curves plot, see Table S0–4 - Power analysis, and Fig. S0–1 ibid. in the Supplementary material.

2.3. Data availability and supplementary materials

Datasets, tables of descriptive statistics, detailed descriptions of methods and statistical analyses of individual studies, literature review and meta-analysis methods, and jamovi outputs are all available in the Supplementary material.

3. Results

We extracted 25 effects for the relationship between body odour attractiveness and facial attractiveness, and 9 effects for body odour attractiveness and vocal attractiveness (Table S0–6). These were based on ten unpublished datasets and four published studies describing the association between body odour attractiveness and facial attractiveness, and between body odour attractiveness and vocal attractiveness (from 92 search results, see Table S0–5). The results reported below are based on 1001 target stimuli and 1350 raters.

3.1. Sensitivity to observe meta-analytical effects

With the 25 effects and an average sample size of 46 targets per group in the meta-analysis on the relationship between body odour and facial attractiveness, we reached a sensitivity to observe effects (with 5% α and β error rates) of 0.174, 0.214 and 0.303 for low, moderate, and high heterogeneity, respectively (Fig. 1 – left).

In the case of the meta-analysis on the relationship between body odour and vocal attractiveness, with 9 effects and an average sample size of 51 targets per group, we reached a sensitivity to observe effects (with 5% α and β error rates) of 0.276, 0.339 and 0.484 for low, moderate, and high heterogeneity, respectively (Fig. 1 – right).

Hence, effects smaller than those estimated by our sensitivity analysis would be observed with statistical power below 95%, following the associated curves in Fig. 1. For example, if the meta-analysis on the relationship between body odour and facial attractiveness would have small heterogeneity and observed effects of 0.2, 0.1, or 0.05, it would have ~99%, ~55%, or ~ 17% power to observe them, respectively.

3.2. Association between body odour and facial attractiveness

All 25 effects were included in the meta-analysis on the association between body odour and facial attractiveness. The observed correlation coefficients ranged from −0.436 to 0.867, with the majority of estimates (68%) above zero. The meta-analytical mean showed a statistically significant, weak positive correlation coefficient of 0.104 [0.034, 0.174], Z = 2.93, $p = 0.003$ (Table 1, Figs. 2 and 3). Although Cochran's Q test was not statistically significant, the effect tends to vary across the studies ($Q_{24} = 35.945$, $p = 0.056$), with small heterogeneity (Quintana & Tiebel, 2019) of about 22% attributable to sampling error. Based on the 95% PI, the *true* outcome is expected to be between −0.069 and 0.277. Results of the Egger's regression suggest no asymmetry in the funnel plot ($\beta_0 = 0.803$, $p = 0.422$, Fig. 3). For female (k = 8) and male (k = 17) targets, the meta-analytical means were 0.163 [0.011, 0.314] and 0.086 [0.005, 0.168], respectively (Table S0–7 - Supplementary meta-analyses results).

3.2.1. Comparison of published and unpublished effects

Considering only the published effects (k = 10), the meta-analytical mean showed a positive correlation coefficient of 0.185 [0.041, 0.328] with a moderate level of heterogeneity (50%). Based on a 95% PI, the *true* outcome thus can be expected between −0.156 and 0.526 (Table 2, Fig. 4). When only the unpublished effects (k = 15) are considered, the meta-analytic mean is 0.052 with 95% CI [−0.024, 0.128] overlapping 0, and 0% heterogeneity (Table 2, Fig. 4). When the publication status (published/unpublished) is used as a moderator, its effect is statistically non-significant (estimate = −0.128 [−0.259, 0.004], $p = 0.057$, heterogeneity $I^2 = 10.25\%$).

3.2.2. The effect of rating design

For studies (k = 16) using a between-subject rating design (different groups of participants provide attractiveness ratings for different stimulus types), the meta-analytical mean estimate for body odour and facial attractiveness was 0.089 with 95% CI [−0.05, 0.183] overlapping zero ($I^2 = 38.29\%$). Studies (k = 9) using a within-subject rating design (each participant judged both stimulus types) also showed a weak positive association between the modalities, 0.146 [0.036, 0.256] ($I^2 = 0\%$), (Table 3). When the rating design was used as moderator, its effect is statistically non-significant (estimate = −0.034 [−0.201, 0.134], $p = 0.692$, $I^2 = 0\%$), Table 3.

3.3. Association between body odour and vocal attractiveness

The association between body odour and vocal attractiveness (k = 9) was weakly positive and statistically significant. The observed correlation coefficients ranged from −0.189 to 0.297, with the majority of estimates (89%) above zero. The meta-analytical mean estimate was 0.098 [0.004, 0.192] with Z = 2.038, $p = 0.041$ (Table 1, Figs. 2 and 3). Cochran's Q ($Q_8 = 4.8$, $p = 0.779$) indicated that the effect did not vary between studies, with 0% of the observed effect attributable to sampling error. Considering females and males separately, the meta-analytical means were 0.143 [0.024, 0.263] for female targets (k = 5) and 0.024 [−0.128, 0.177] for male targets (k = 4) (Table S0–7 - Supplementary Meta-analyses results).

3.4. Effect size distributions

We constructed effect size distributions from all available effect sizes for the association between body odour and facial attractiveness ($n = 25$) and the association between body odour and vocal attractiveness ($n = 9$). In both cases, the 50th percentile values (average/medium effect size) are ~0.1 and equal to the meta-analytic averages (~ 0.1), the 25th percentile (small/below average effect size boundary) values are ~0, and the 75th percentile (above average/large effect size boundary) values are ~0.2. The distributions and percentiles for small (25th), medium (50th, median), and large (75th) effect sizes are presented in Fig. 5 and Table 4.

4. Discussion

Our results indicate that, although the association between body odour attractiveness and facial attractiveness is positive, the summary effect is relatively small ($r \sim 0.1$). We observed similar patterns and magnitudes of effects for female and male targets and also for the odour-voice attractiveness association. We suggest that body odour may provide distinct and non-redundant information about an individual's mating-related qualities compared to that available within either facial or vocal cues. Thus, concerning perceived attractiveness, body odour may provide different and non-redundant cues to an individual's mating-related qualities compared to cues communicated through the face and voice.

These findings contrast with those of Rikowski and Grammer (1999), who observed a strong positive correlation ($r_{19} = 0.7$) between facial

[3] We decided to choose a 1:1 ratio of the Type I and II error rates for all performed analyses, as we see committing both errors as of equal significance in this instance.

Table 1
Meta-analysis and heterogeneity results.

Congruence in	k	Estimate (Fisher's z)	95% CI		p	95% PI	
			LL	UL		LL	UL
Body odour and Facial attractiveness	25	0.104	0.034	0.174	0.003	−0.069	0.277
Body odour and Vocal attractiveness	9	0.098	0.004	0.192	0.042		

Heterogeneity Statistics	Tau	Tau^2	I^2 (%)	H^2	Q	df	p
Body odour and Facial attractiveness	0.079	0.0062	20.84	1.263	35.696	24	0.059
Body odour and Vocal attractiveness	0	0	0	1	4.8	8	0.779

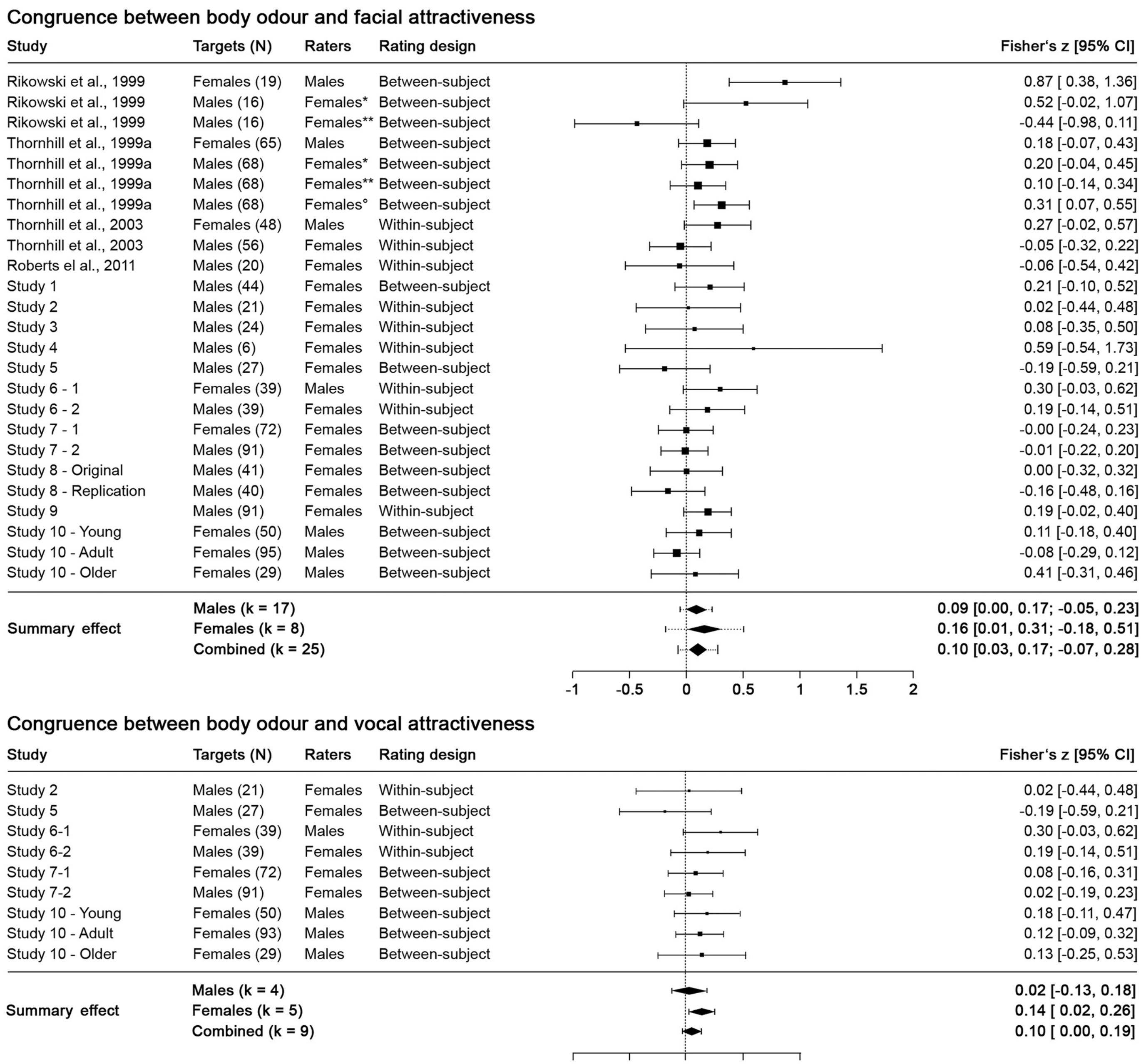

Fig. 2. Forest plots for congruence meta-analyses. Squares represent weighted mean effects of individual studies, and error bars their 95% confidence intervals. Diamonds represent summary effects, their width the 95% CIs, and dashed error bars their 95% PIs. *Female raters in fertile, **non-fertile phase of their menstrual cycle, and °hormonal contraception users. Summary effects are reported in Fisher's z-transformed correlation coefficients with 95% confidence intervals and in heterogeneous effects also followed with 95% prediction intervals. Forest plots were generated in jamovi, and edited in Adobe Photoshop CC2022.

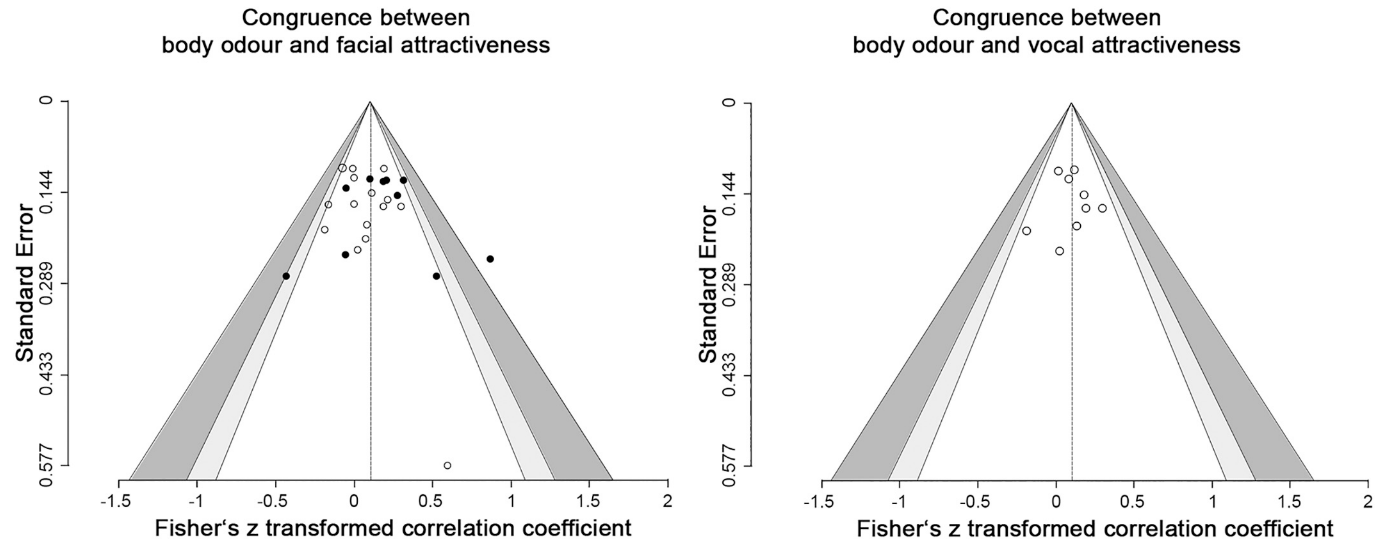

Fig. 3. Funnel plots for congruence meta-analyses. Area outside the contour-enhanced funnels represent *p* values <0.01, dark grey areas p values between 0.01 and 0.05, light grey p values between 0.05 and 0.1, and areas inside the funnel p values >0.1. Full circles illustrate published and empty circles unpublished studies. Dashed line show summary effect sizes; Y-axis is the standard error of Fisher's z. Funnel plots were generated in jamovi, and edited in Adobe Photoshop CC2022.

Table 2
Meta-analysis and heterogeneity results for published and unpublished effects.

Origin	k	Estimate (Fisher's z)	95% CI		p	95% PI	
			LL	UL		LL	UL
Published effects	10	0.185	0.041	0.328	0.012	−0.156	0.526
Unpublished effects	15	0.052	−0.024	0.128	0.182		
Moderator		−0.128	−0.259	0.004	0.057		

Heterogeneity Statistics	Tau	Tau^2	I^2 (%)	H^2	Q	df	p
Published effects	0.158	0.0249	49.91	1.996	19.813	9	0.019
Unpublished effects	0	0	0	1	11.92	14	0.613
Moderator	0.052	0.0027	10.25	1.114	31.733	24	0.106

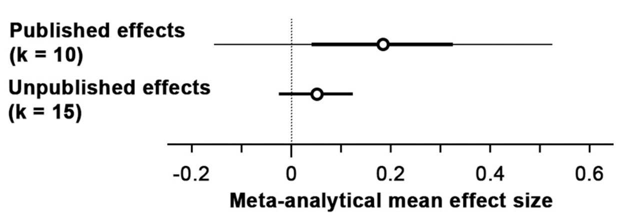

Fig. 4. Comparison of meta-analytic averages between published and unpublished effects. Circles represent mean effects. Thick error bars their 95% CI and thin error bars 95% PI. Due to observed heterogeneity only in the published effects, the mean effect is accompanied by 95% PI. The plot was generated in Adobe Photoshop CC2022.

and body odour attractiveness, but concur with more recent studies (Roth et al., 2021) that report a weak association between body odour, facial, and vocal attractiveness. Similarly, our findings are in line with those of two studies (Mahmut & Stevenson, 2019; Roth et al., 2021) that did not meet our formal inclusion criteria due to their non-parametric and non-frequentist data analysis (Table S0–5). In a sample of 82 female raters and 91 male donors, Mahmut and Stevenson (2019) reported Spearman's $\rho = 0.3$ for the association between body odour and facial sexiness. Using Bayesian analysis with a sample of 70 participants who served as both donors and raters, Roth et al. (2021), reported that body odour, facial, and vocal attractiveness were positively correlated but with small effect sizes. It is worth noting, however, that the authors discuss their findings of small and positive effects in favour of the *backup signals* hypothesis; we would disagree with this interpretation. The shared variability of attractiveness ratings resulting from the summary effects across the two pairs of modalities in the present meta-analyses was <1%, suggesting minimal (if any) redundancy in information transferred through these modalities.

In studies concerning an association between facial and vocal attractiveness, the current evidence shows inconsistent results, ranging from strong positive correlations in women only (Abend, Pflüger, Koppensteiner, Coquerelle, & Grammer, 2015; Collins & Missing, 2003; Wheatley et al., 2014) to weak (Zuckerman, Miyake, & Elkin, 1995) or no significant associations (Zäske et al., 2020). This range suggests that the overall pattern of relationships might be similar to that found in the present study between odour and these other modalities. However, there is currently no systematic investigation or meta-analysis available for the association between facial and vocal attractiveness to our best knowledge.

4.1. Notes on the meta-analyses and renumber other heading

Notes on the meta-analyses Although Fig. 4 shows a stronger (over 3×) positive mean effect for published effects than unpublished ones, but the meta-analytical mean of unpublished effects provides a more precise estimate: the mean effect (and over half of its 95% CI) falls within the 95% CI (and entirely within 95% PI) of the published effects. If the present study were based only on published evidence, it would

Table 3
Meta-analysis and heterogeneity results for between- and within-subject rating design.

Rating Design	k	Estimate (Fisher's z)	95% CI		p	95% PI	
			LL	UL		LL	UL
Between-subject	16	0.089	−0.05	0.183	0.062	−0.155	0.334
Within-subject	9	0.146	0.036	0.256	0.009		
Moderator		−0.034	−0.201	0.134	0.692		

Heterogeneity Statistics	Tau	Tau^2	I^2 (%)	H^2	Q	df	p
Between-subject	0.115	0.0133	38.29	1.62	29.439	15	0.014
Within-subject	0	0	0	1	5.605	8	0.691
Moderator	0.087	0.0076	24.52	1.325	35.708	24	0.044

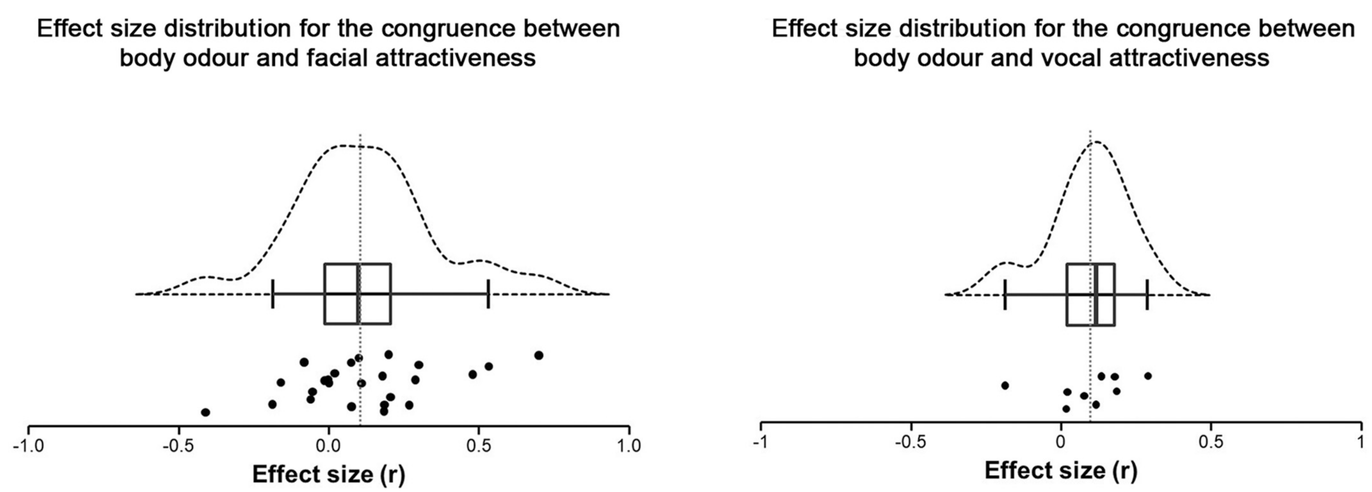

Fig. 5. Raincloud plots for effect size distribution. Density plots show effect sizes distribution, boxplots show median (thick line), 25th and 75th percentile (interquartile range, box), and minimum and maximum (error bars); jittered dots represent individual effect sizes; dotted vertical line shows effect size average for each meta-analysis (left 0.104, right 0.098). Raincloud plots were generated in JASP (0.16.2) and edited in Adobe Photoshop CC2022.

Table 4
Effect size distributions.

Congruence in	Number of effects	Percentiles		
		25th	50th	75th
Body odour and Facial attractiveness	25	−0.013	0.1	0.206
Body odour and Vocal attractiveness	9	0.02	0.116	0.178

thus report a stronger and less precise estimate of the meta-analytic effect for associations between assessments of body odour and facial attractiveness. Moreover, a meta-analysis of body odour and vocal attractiveness would not be possible as the literature search identified only a single study fulfilling the inclusion criteria (Roth et al., 2021 discussed above). This highlights the importance of considering unpublished data in quantifying effects through systematic reviews and evidence synthesis.

Although we generally observed low levels of heterogeneity in our meta-analyses, they rely on a relatively small number of effects and the sensitivity of our analyses is correspondingly low. In addition, the statistical power in many of the available studies is low, due to a relatively small number of stimuli (Table S0–4). The average number of raters per stimuli (mostly body odour stimuli) often resulted in wider corridors of rating stability (Hehman, Xie, Ofosu, & Nespoli, 2018) and thus less precise estimates of mean ratings (Table S0–3). This mainly arises from logistical limitations related to procedures employed in body odour sampling and rating. In contrast to facial images and vocal recordings, body odour stimuli can be used only a limited number of times due to microbial transformation and signal degradation (Lenochová, Roberts, & Havlíček, 2009). Furthermore, the number of odour stimuli that one rater can assess within a reasonable time is limited by olfactory adaptation (Köster & de Wijk, 1991). These issues hinder the accuracy of the present findings and represent challenges for further research.

In addition to the meta-analytical results, the current article presents a systematic overview of studies conducted over the last two decades, including data collection methods, sample sizes, populations, and observed ratings (Tables S0–6). We also included observed effect size distributions showing that commonly used correlation thresholds overestimate effect sizes observed in studies, where average and larger-than-average effects (50th and 75th percentile, respectively) are 'only' ~0.1 and ~ 0.2. Based on the unpublished datasets, where more detailed insight can be provided, the average number of stimuli used in this type of research is ~46 giving us sensitivity to observe correlations $\geq$0.49 (with 0.05 $p_{\text{two-tailed}}$ and 95% power, $\geq$ 0.39 with 80% power). On average, in these studies, body odour, and facial and vocal stimuli are rated for attractiveness by ~25, 31, and 32 raters, respectively, though based on our corridor of stability analysis samples $\geq$35 seem to be needed for more precise estimates. Overall, all three stimulus types seem to be rated with good reliability (mean ICC2k ~0.8), and we found no differences in reliability between stimulus types. See Tables S0–1, 2, 3 and 4, and ICC comparison in the Supplemental materials for further details. Future research investigating the association in attractiveness rating between modalities could benefit from this systematic overview, including effect size distributions, to plan and convey magnitudes of observed effects in comparison to the body of up-to-date literature.

4.2. Alternative reasons for the observed effects

It is conceivable that the associations between individual modalities are underestimated because (a) studies use 'snapshots' of an individual which might provide only a rough estimate of his or her mating-related qualities, and (b) these snapshots vary in duration across modalities. Odour stimuli are typically collected over a longer period (12-24 h) and may, therefore, provide a more reliable quality estimate. In contrast, vocal stimuli often last <1 min. and visual images capture less than a second. Previous studies testing the association between body odour attractiveness and physical attractiveness assessed from videos found a stronger correlation ($r = 0.32$) compared to the association between body odour attractiveness and facial attractiveness ($r = -0.08$) (Roberts et al., 2011). Thus, sampling time might influence the reliability of mating-related quality estimates. A reviewer also argued that the reason for the weak correlation between odour attractiveness and the two other modalities could be higher variability in ratings of body odour, perhaps because it is considered that olfactory judgments are either more difficult or more subjective. However, our ICC analysis shows that the level of agreement is comparable across the three modalities.

Similarly, the weak correlations that we observe between attractiveness assessments of different stimulus types might result from experimental (laboratory-based) settings and some variations in protocols. These include, for example, control over facial expressions during image acquisition, the volume of voice recordings, and dietary restrictions in body odour sampling. Although methodologically challenging, the use of more naturalistic stimuli with facial expressiveness, the prosody of speech and natural variation in body odour (Roberts et al., 2022) may provide additional insight into the patterns of associations and congruence across sensory modalities investigated here.

Further, earlier studies reporting positive associations between attractiveness and putative markers of mating-related quality had failed to replicate, especially when they were based on small samples. Many studies that were included in the current analysis had different groups of participants providing attractiveness ratings of the stimulus types (between-subject rating design). A high inter-individual variation in attractiveness ratings in some modalities would lead to a weak correlation between the modalities because the target is rated by some people in one modality and by other people in the other. Studies using a design where each participant judged all stimulus types (within-subject rating design) also tend to show a weak correlation between the modalities, meaning that weak correlations in individual studies cannot be solely due to study design.

An individual's mating-related quality may be perceived more accurately by combining cues from different modalities that independently correlate with mate preferences. However, most studies on physical attractiveness examine the influence of individual modalities separately, a design that lacks ecological validity because, in everyday life, we perceive others through multiple senses simultaneously (Groyecka et al., 2017). Similarly, the present meta-analysis is based on studies investigating several modalities separately, not on multimodal perception, which is a result of simultaneous perception across different sensory modalities. The resulting perception can differ qualitatively from the sum of the properties of its components and convey a unique message, or one modality can affect information transmitted by the other modalities, being different from the *backup*and *multiple messages*concepts (Halfwerk et al., 2019; Mitoyen, Cliodhna, & Leonida, 2019). How information based on different modalities contributes to overall attractiveness judgments is poorly understood (e.g., Ferdenzi, Delplanque, Atanassova, & Sander, 2016). Current research into the integration of human mate preferences indicates that they are best described by the Euclidean model (Conroy-Beam et al., 2019). Whether a similar pattern of integration can be expected in the case of physical attractiveness or whether it would follow another form, as explained by additive or threshold models, remains to be investigated (Csajbók, Bérkics, & Havlíček, 2022; Havlíček, Štěrbová, & Csajbók, 2022).

4.3. Theoretical implications

It has been proposed that attractiveness reflects an individual's mating-related qualities (e.g., in terms of health and fertility). Perceived facial attractiveness is influenced by several features, including symmetry, prototypicality, sexual dimorphism, adiposity, and skin condition. For instance, prototypicality is thought to be a marker of heterozygosity, symmetry a marker of developmental stability, while sexual dimorphism is a marker of sex hormone levels and skin quality is a marker of health status (for review, see Stephen & Luoto, 2022). Similarly, it has been suggested that body odour may also provide information about heterozygosity, developmental stability, sex hormones and health (for review, see Havlíček, Fialová, & Roberts, 2017). Hence, one might expect at least moderate associations between the attractiveness of these modalities, but we found only weak associations. Several associations between attractiveness and the proposed underlying qualities were recently revisited (Stephen & Luoto, 2022) and others are still debated. These include links between hormonal profiles and facial attractiveness (Jones, Jones, Shiramizu, & Anderson, 2021) or between body odour attractiveness and MHC heterozygosity (Havlíček, Winternitz, & Roberts, 2020).

Visual, olfactory, and acoustic modalities may provide unique (and non-redundant) information about an individual's mating-related quality. Our results are in line with the *multiple messages* hypothesis but seem to provide little support for the *backup signals* hypothesis. Moreover, they correspond with the majority of animal studies that have reported multiple traits to be unrelated, suggesting that backup signals are less common than multiple messages (Badyaev, Etges, Faust, & Martin, 1998; Candolin, 2003; Kraak, 1999). We speculate that facial appearance primarily provides cues to more stable characteristics such as the development of hormone-related secondary sexual characteristics and maturation (Marečková et al., 2011; Whitehouse et al., 2015). In contrast, body odour may provide cues to more variable characteristics, such as current health (Olsson et al., 2014; Sarolidou et al., 2020) and fertility status (Gildersleeve, Haselton, Larson, & Pillsworth, 2012; Havlíček, Dvořákova, Bartoš, & Flegr, 2006). These are provocative and open questions that require in-depth investigations.

In conclusion, the present study found weak congruence between attractiveness assessments of human body odours and those of faces or voices. These results provide little support for the *backup signals* hypothesis in explaining the use of multiple modalities in attractiveness assessments, but favour the *multiple messages* hypothesis, suggesting that body odour provides information about mating-related quality different from that of faces or voices.

Ethics

All procedures within the individual studies were carried out following the Declaration of Helsinki, and Institutional Review Boards approved each study. Individual approvals can be found in the Supplemental Materials.

Funding

This study was supported by the Czech Science Foundation GAČR P407/19/11822S, the Charles University Research Centres UNCE/HUM/025 (UNCE 204056), and the Charles university grant agency GA UK No 1374319.

CRediT authorship contribution statement

Vít Třebický: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Validation, Visualization, Writing – original draft, Writing – review & editing. **Sylvain Delplanque:** Investigation, Methodology, Resources, Validation, Writing – review & editing. **Camille Ferdenzi:** Investigation, Methodology, Resources, Validation, Writing – review & editing. **Bernhard Fink:** Investigation, Methodology, Resources, Validation, Writing – review & editing. **Lucie Jelínková:** Investigation, Methodology, Resources, Validation, Writing – review & editing. **Žaneta Pátková:** Investigation, Methodology, Resources, Validation, Writing – review & editing. **S. Craig Roberts:** Investigation, Methodology, Resources, Supervision, Validation, Writing – review & editing. **Susanne Röder:** Investigation, Methodology, Resources, Validation, Writing – review & editing. **Tamsin K. Saxton:** Investigation, Methodology, Resources, Validation, Writing – review & editing. **Dagmar Schwambergová:** Investigation, Methodology, Resources, Validation, Writing – review & editing. **Zuzana Štěrbová:** Investigation, Methodology, Resources, Validation, Writing – review & editing. **Jitka Třebická Fialová:** Conceptualization, Funding acquisition, Investigation, Methodology, Resources, Validation, Writing – original draft, Writing – review & editing. **Jan Havlíček:** Conceptualization, Funding acquisition, Investigation, Methodology, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declared no potential conflicts of interest concerning the research, authorship, and/or publication of this article.

Acknowledgements

We thank Radim Kuba, Tereza Nevolová, David Stella, Pavel Šebesta, and other members of the Human Ethology group (www.etologiecloveka.cz) for their help with data collection. We also express our gratitude to all our volunteers for their participation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.evolhumbehav.2022.11.001.

References

Abend, P., Pflüger, L. S., Koppensteiner, M., Coquerelle, M., & Grammer, K. (2015). The sound of female shape: A redundant signal of vocal and facial attractiveness. *Evolution and Human Behavior, 36*(3), 174–181. https://doi.org/10.1016/j.evolhumbehav.2014.10.004

Aglioti, S. M., & Pazzaglia, M. (2011). Sounds and scents in (social) action. *Trends in Cognitive Sciences, 15*(2), 47–55. https://doi.org/10.1016/j.tics.2010.12.003

Arakawa, H., Blanchard, D. C., Arakawa, K., Dunlap, C., & Blanchard, R. J. (2008). Scent marking behavior as an odorant communication in mice. *Neuroscience & Biobehavioral Reviews, 32*(7), 1236–1248. https://doi.org/10.1016/j.neubiorev.2008.05.012

Badyaev, A. V., Etges, W. J., Faust, J. D., & Martin, T. E. (1998). Fitness correlates of spur length and spur asymmetry in male wild turkeys. *Journal of Animal Ecology, 67*(6), 845–852. https://doi.org/10.1046/j.1365-2656.1998.6760845.x

Brown, T. A. (2006). In *Confirmatory factor analysis for applied research*. The Guilford Press.

Brydges, C. R. (2019). Effect size guidelines, sample size calculations, and statistical power in gerontology. *Innovation in Aging, 3*(4). https://doi.org/10.1093/geroni/igz036

Candolin, U. (2003). The use of multiple cues in mate choice. *Biological Reviews, 78*(4), 575–595. https://doi.org/10.1017/S1464793103006158

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Collins, S. A., & Missing, C. (2003). Vocal and visual attractiveness are related in women. *Animal Behaviour, 65*(5), 997–1004. https://doi.org/10.1006/anbe.2003.2123

Colyott, K., Odu, C., & Gleason, J. M. (2016). Dissection of signalling modalities and courtship timing reveals a novel signal in Drosophila saltans courtship. *Animal Behaviour, 120*, 93–101. https://doi.org/10.1016/j.anbehav.2016.07.015

Conroy-Beam, D., Buss, D. M., Asao, K., Sorokowska, A., Sorokowski, P., Aavik, T., … Zupančič, M. (2019). Contrasting computational models of mate preference integration across 45 countries. *Scientific Reports, 9*(1), 16885. https://doi.org/10.1038/s41598-019-52748-8

Csajbók, Z., Bérkics, M., & Havlíček, J. (2022). Meeting minimum mate preference thresholds can be more important than the overall score. *Personality and Individual Differences, 159*, 111675. https://doi.org/10.1016/j.paid.2022.111675

Cunningham, M. R., Barbee, A. P., & Pike, C. L. (1990). What do women want? Facialmetric assessment of multiple motives in the perception of male facial physical attractiveness. *Journal of Personality and Social Psychology, 59*(1), 61–72. https://doi.org/10.1037/0022-3514.59.1.61

Erdfelder, E., Faul, F., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146

Ferdenzi, C., Delplanque, S., Atanassova, R., & Sander, D. (2016). Androstadienone's influence on the perception of facial and vocal attractiveness is not sex specific. *Psychoneuroendocrinology, 66*, 166–175. https://doi.org/10.1016/j.psyneuen.2016.01.016

Fialová, J., Třebický, V., Kuba, R., Stella, D., Binter, J., & Havlíček, J. (2020). Losing stinks! The effect of competition outcome on body odour quality. *Philosophical Transactions of the Royal Society B: Biological Sciences, 375*(1800), 20190267. https://doi.org/10.1098/rstb.2019.0267

Gallucci, M. (2021). *GAMLj: General analyses for linear models (2.3.0)*.

Gangestad, S. W., Haselton, M. G., Welling, L. L. M., Gildersleeve, K., Pillsworth, E. G., Burriss, R. P., … Puts, D. A. (2016). How valid are assessments of conception probability in ovulatory cycle research? Evaluations, recommendations, and theoretical implications. *Evolution and Human Behavior, 37*(2), 85–96. https://doi.org/10.1016/j.evolhumbehav.2015.09.001

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences, 102*, 74–78. https://doi.org/10.1016/j.paid.2016.06.069

Gildersleeve, K. A., Haselton, M. G., Larson, C. M., & Pillsworth, E. G. (2012). Body odor attractiveness as a cue of impending ovulation in women: Evidence from a study using hormone-confirmed ovulation. *Hormones and Behavior, 61*(2), 157–166. https://doi.org/10.1016/j.yhbeh.2011.11.005

Girard, M. B., Elias, D. O., & Kasumovic, M. M. (2015). Female preference for multi-modal courtship: Multiple signals are important for male mating success in peacock spiders. *Proceedings of the Royal Society B: Biological Sciences, 282*(1820), 20152222. https://doi.org/10.1098/rspb.2015.2222

Gosling, L. M., & Roberts, S. C. (2001). Scent-marking by male mammals: Cheat-proof signals to competitors and mates. *Advances in the Study of Behavior, 30*, 169–217. https://doi.org/10.1016/s0065-3454(01)80007-3

Grammer, K., Fink, B., Juette, A., Ronzal, G., & Thornhill, R. (2001). Female faces and bodies: N-dimensional feature space and attractiveness. In G. Rhodes, & L. Zebrowitz (Eds.), *Advances in visual cognition* (pp. 91–125). Ablex Publishing.

Grammer, K., Fink, B., Møller, A. P., & Thornhill, R. (2003). Darwinian aesthetics: Sexual selection and the biology of beauty. *Biological Reviews, 78*(3), 385–407. https://doi.org/10.1017/S1464793102006085

Groyecka, A., Pisanski, K., Sorokowska, A., Havlíček, J., Karwowski, M., Puts, D. A., … Roberts, S. C. (2017). Attractiveness is multimodal: Beauty is also in the nose and ear of the beholder. *Frontiers in Psychology, 8*(778), 1–5. https://doi.org/10.3389/fpsyg.2017.00778

Halfwerk, Wouter, Michiel, Blaas, Lars, Kramer, Nadia, Hijner, Paula, A. Trillo, Ximena, E. Bernal, … Jacintha, Ellers (2019). Adaptive Changes in Sexual Signalling in Response to Urbanization. *Nature Ecology & Evolution, 3*(3), 374–380. https://doi.org/10.1038/s41559-018-0751-8. https://www.nature.com/articles/s41559-018-0751-8

Hamilton, W. K. (2021). *MAJOR - Meta-analysis for JAMOVI (1.1.1)*.

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2021). *Doing Meta-analysis with R, June, 2020–2023*. https://doi.org/10.1201/9781003107347

Havlíček, J., Dvořákova, R., Bartoš, L., & Flegr, J. (2006). Non-advertized does not mean concealed: Body odour changes across the human menstrual cycle. *Ethology, 112*(1), 81–90. https://doi.org/10.1111/j.1439-0310.2006.01125.x

Havlíček, J., Fialová, J., & Roberts, S. C. (2017). Individual variation in body odor. *In: Springer handbook of odor* (A. Buettner, pp. 943–955). New York: Springer.

Havlíček, J., & Roberts, S. C. (2022). Menstrual cycle variation in women's mating psychology: Empirical evidence and theoretical considerations. In J. K. Mogilski, & T. K. Shackelford (Eds.), *The Oxford handbook of evolutionary psychology and romantic relationships* (pp. 643–674). Oxford University Press.

Havlíček, J., Saxton, T. K., Roberts, S. C., Jozífková, E., Lhota, S., Valentová, J., & Flegr, J. (2008). He sees, she smells? Male and female reports of sensory reliance in mate choice and non-mate choice contexts. *Personality and Individual Differences, 45*(6), 565–570. https://doi.org/10.1016/j.paid.2008.06.019

Havlíček, J., Štěrbová, Z., Csajbók, Z., d'Ettorre, P., & Freeberg, T. M. (2022). Human mate choice. In T. M. T.M. Freeberg, A. Ridley, & P. d'Ettorre (Eds.), *Routledge handbook of comparative psychology* (pp. 338–354). Routledge.

Havlíček, J., Winternitz, J., & Roberts, S. C. (2020). Major histocompatibility complex-associated odour preferences and human mate choice: Near and far horizons. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 375*(1800), 20190260. https://doi.org/10.1098/rstb.2019.0260

Hehman, E., Xie, S., Ofosu, E., & Nespoli, G. (2018). Assessing the point at which averages are stable: A tool illustrated in the context of person perception. *PsyArXiv Preprints.*. https://doi.org/10.31234/osf.io/2n6jq

Herz, R. S., & Inzlicht, M. (2002). Sex differences in response to physical and social factors involved in human mate selection. *Evolution and Human Behavior, 23*(5), 359–364. https://doi.org/10.1016/S1090-5138(02)00095-8

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ, 327*(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557

Higham, J. P., & Hebets, E. A. (2013). An introduction to multimodal communication. *Behavioral Ecology and Sociobiology, 67*(9), 1381–1388. https://doi.org/10.1007/s00265-013-1590-x

Hill, A. K., & Puts, D. A. (2016). Vocal attractiveness. In T. K. Shackelford, & V. A. Weekes-Shackelford (Eds.), *Encyclopedia of evolutionary psychological science* (pp. 1–3). Springer International Publishing.

Jones, B. C., Jones, A. L., Shiramizu, V., & Anderson, C. (2021). What does women ' s facial attractiveness signal? Implications for an evolutionary perspective on appearance enhancement. *Archives of Sexual Behavior*. https://doi.org/10.1007/s10508-021-01955-4, 0123456789.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Köster, E. P., & de Wijk, R. A. (1991). Olfactory adaptation. In *The human sense of smell* (pp. 199–215). Berlin Heidelberg: Springer. https://doi.org/10.1007/978-3-642-76223-9_10.

Kraak, S. B. M. (1999). Sexual selection in sticklebacks in the field: Correlates of reproductive, mating, and paternal success. *Behavioral Ecology, 10*(6), 696–706. https://doi.org/10.1093/beheco/10.6.696

Lenochová, P., Roberts, S. C., & Havlíček, J. (2009). Methods of human body odor sampling: The effect of freezing. *Chemical Senses, 34*(2), 127–138. https://doi.org/10.1093/chemse/bjn067

Lovakov, A., & Agadullina, E. R. (2021). Empirically derived guidelines for effect size interpretation in social psychology. *European Journal of Social Psychology, 51*(3), 485–504. https://doi.org/10.1002/ejsp.2752

Mahmut, M. K., & Stevenson, R. J. (2019). Do single men smell and look different to partnered men? *Frontiers in Psychology, 10*(261), 1–9. https://doi.org/10.3389/fpsyg.2019.00261

Marečková, K., Weinbrand, Z., Chakravarty, M. M., Lawrence, C., Aleong, R., Leonard, G., … Paus, T. (2011). Testosterone-mediated sex differences in the face shape during adolescence: Subjective impressions and objective features. *Hormones and Behavior, 60*(5), 681–690. https://doi.org/10.1016/j.yhbeh.2011.09.004

Mitoyen, Clémentine, Cliodhna, Quigley, & Leonida, Fusani (2019). Evolution and Function of Multimodal Courtship Displays. *Ethology, 125*(8), 503–515. https://doi.org/10.1111/eth.12882. https://onlinelibrary.wiley.com/doi/full/10.1111/eth.12882

Møller, A. P., & Pomiankowski, A. (1993). Why have birds got multiple sexual ornaments? *Behavioral Ecology and Sociobiology, 32*(3), 167–176. https://doi.org/10.1007/BF00173774

Nordahl-Hansen, A., Cogo-Moreira, H., Panjeh, S., & Quintana, D. S. (2022). Redefining effect size interpretations for psychotherapy RCTs in depression. *OSF Preprints.*. https://doi.org/10.31219/osf.io/erhmw

Olsson, M. J., Lundström, J. N., Kimball, B. A., Gordon, A. R., Karshikoff, B., Hosseini, N., … Lekander, M. (2014). The scent of disease. *Psychological Science, 25*(3), 817–823. https://doi.org/10.1177/0956797613515681

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *The BMJ, 372*. https://doi.org/10.1136/bmj.n71

Pisanski, K., Feinberg, D., Oleszkiewicz, A., & Sorokowska, A. (2017). Voice cues are used in a similar way by blind and sighted adults when assessing women's body size. *Scientific Reports, 7*(10329). https://doi.org/10.1038/s41598-017-10470-3

Quintana, D. S. (2015). From pre-registration to publication: A non-technical primer for conducting a meta-analysis to synthesize correlational data. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.01549

Quintana, D. S. (2017). Statistical considerations for reporting and planning heart rate variability case-control studies. *Psychophysiology, 54*(3), 344–349. https://doi.org/10.1111/psyp.12798

Quintana, D. S., & Tiebel, J. (2019). *How to calculate statistical power for your meta-analysis*. https://doi.org/10.17605/OSF.IO/5C7UZ

R Core Team. (2021). *RStudio: Integrated development for R*.

R Core Team, & Team, R. C. (2019). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing.

Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology, 57*, 199–226. https://doi.org/10.1146/annurev.psych.57.102904.190208

Rikowski, A., & Grammer, K. (1999). Human body odour, symmetry and attractiveness. *Proceedings of the Royal Society B: Biological Sciences, 266*(1422), 869–874. https://doi.org/10.1098/rspb.1999.0717

Roberts, S. C., Třebická Fialová, J., Sorokowska, A., Langford, B., Sorokowski, P., Třebický, V., & Havlíček, J. (2022). Emotional expression in human odour. *Evolutionary Human Sciences, 4*, E44. https://doi.org/10.1017/ehs.2022.44

Roberts, S. C., Kralevich, A., Ferdenzi, C., Saxton, T. K., Jones, B. C., DeBruine, L. M., … Havlíček, J. (2011). Body odor quality predicts behavioral attractiveness in humans. *Archives of Sexual Behavior, 40*(6), 1111–1117. https://doi.org/10.1007/s10508-011-9803-8

Röder, S., Fink, B., & Jones, B. C. (2013). Facial, olfactory, and vocal cues to female reproductive value. *Evolutionary Psychology: An International Journal of Evolutionary Approaches to Psychology and Behavior, 11*(2), 392–404. https://doi.org/10.1177/147470491301100209

Roth, T. S., Samara, I., & Kret, M. E. (2021). Multimodal mate choice: Exploring the effects of sight, sound, and scent on partner choice in a speed-date paradigm. *Evolution and Human Behavior, 42*(5), 461–468. https://doi.org/10.1016/j.evolhumbehav.2021.04.004

Sarolidou, G., Axelsson, J., Kimball, B. A., Sundelin, T., Regenbogen, C., Lundström, J. N., … Olsson, M. J. (2020). People expressing olfactory and visual cues of disease are less liked. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 375*(1800), 20190272. https://doi.org/10.1098/rstb.2019.0272

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality, 47*(5), 609–612. https://doi.org/10.1016/j.jrp.2013.05.009

Stephen, I. D., & Luoto, S. (2022). Physical cues of partner quality. In J. K. Mogilski, & T. K. Shackelford (Eds.), *The Oxford handbook of evolutionary psychology and romantic relationships* (pp. 42–73). Oxford University Press.

Sterne, J. A. C., Sutton, A. J., Ioannidis, J. P. A., Terrin, N., Jones, D. R., Lau, J., … Higgins, J. P. T. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ, 343*. https://doi.org/10.1136/bmj.d4002

The jamovi project. (2021). Jamovi (Version 2.0.0). https://www.jamovi.org.

Thornhill, R., & Gangestad, S. W. (1999a). The scent of symmetry: A human sex pheromone that signals fitness? *Evolution and Human Behavior, 20*(3), 175–201.

Thornhill, R., & Gangestad, S. W. (1999b). Facial attractiveness. *Trends in Cognitive Sciences, 3*(12), 452–460. https://doi.org/10.1016/S1364-6613(99)01403-5

Thornhill, R., Gangestad, S. W., Miller, R., Scheyd, G., McCollough, J. K., & Franklin, M. (2003). Major histocompatibility complex genes, symmetry, and body scent attractiveness in men and women. *Behavioral Ecology, 14*(5), 668–678. https://doi.org/10.1093/beheco/arg043

Thornhill, R., & Grammer, K. (1999). The body and face of woman: One ornament that signals quality? *Evolution and Human Behavior, 20*(2), 105–120. https://doi.org/10.1016/S1090-5138(98)00044-0

Walter, K. V., Conroy-Beam, D., Buss, D. M., Asao, K., Sorokowska, A., Sorokowski, P., … Zupančič, M. (2020). Sex differences in mate preferences across 45 countries: A large-scale replication. *Psychological Science, 31*(4), 408–423. https://doi.org/10.1177/0956797620904154

Wheatley, J. R., Apicella, C. A., Burriss, R. P., Cárdenas, R. A., Bailey, D. H., Welling, L. L. M., & Puts, D. A. (2014). Women's faces and voices are cues to reproductive potential in industrial and forager societies. *Evolution and Human Behavior, 35*(4), 264–271. https://doi.org/10.1016/j.evolhumbehav.2014.02.006

Whitehouse, A. J. O., Gilani, S. Z., Shafait, F., Mian, A., Tan, D. W., Maybery, M. T., … 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

Zäske, R., Skuk, V. G., & Schweinberger, S. R. (2020). Attractiveness and distinctiveness between speakers' voices in naturalistic speech and their faces are uncorrelated. *Royal Society Open Science, 7*(12), Article 201244. https://doi.org/10.1098/rsos.201244

Zuckerman, M., Miyake, K., & Elkin, C. S. (1995). Effects of attractiveness and maturity of face and voice on interpersonal impressions. *Journal of Research in Personality, 29*(2), 253–272. https://doi.org/10.1006/jrpe.1995.1015