Perception of Emergent Leaders’ Faces and Evolution of Social Cheating: Cross-Cultural Experiments

Victoria V. Rostovtseva¹, Anna A. Mezentseva¹, and Marina L. Butovskaya¹,²

Abstract
The aim of the present study was to investigate whether neutral faces of individuals with different propensities for leadership may convey information about their personal qualities, and are there impacts of sex, population and social environment on the facial perception. This study is based on a previous experiment (Rostovtseva et al., 2022), where emergent leadership in the context of male group cooperation was investigated in Buryats (Mongolian population of Siberia). In the previous study three behavioural types of participants were revealed: non-leaders, prosocial leaders and leaders-cheaters, each having a set of distinguishing personality, communicative, and cooperative features. In the current study, three composite portraits representing different leadership qualities of Buryat men from the prior experiment were created. The composites were then scored on a number of traits by male and female Russian and Buryat independent raters (N = 435). The results revealed that ratings on masculinity, physical strength, dominance, competitiveness, and perceived leadership were positively correlated, while perceived trustworthiness was negatively associated with these traits. However, the composite portraits of actual leaders generally were scored as more trustworthy, masculine, and physically strong, with the prosocial leaders’ portrait being perceived as healthier than others. Surprisingly, the composite of leaders-cheaters was scored as the most trustworthy and generous, and the least competitive than others. No significant effects of raters’ sex, origin, or degree of familiarity with Mongolian appearance were revealed. We conclude that static facial morphology contributes to appearing trustworthy, which may allow exploitation of others.

Keywords
facial perception, leadership, trustworthiness, perceived masculinity, perceived physical strength, perceived health, cheating, deception, cheater detection, social environment

Introduction
The aim of the present study was to investigate the role of facial appearance as a source of information about individual qualities, and its’ relation to the phenomenon of cheating in the context of human leadership. The problem of free-riding has been the most comprehensively studied within the framework of cooperative behaviour (Ensminger & Henrich, 2014; Fehr & Gächter, 2000; Heckathorn, 1989; Ledyard, 1994; Marlowe, 2009; Yang et al., 2018). From an evolutionary perspective, both cheating and free-riding constitute similar behavioural constructs, which are ultimately aimed at achieving individual rather than social benefits. However, the phenomenon of cheating implies not just “consumer” behaviour, but rather braking social norms and own promises. Although it is usually assumed that leaders should act in the interests of a group and facilitate coordination between group members for achieving common good (Van Vugt & von Rueden, 2020), a number of empirical studies suggest that cheating may also occur in the context of human leadership, where it is closely related to specific personality features of emergent leaders (Bereczkei, 2018; O’Reilly & Doerr, 2020; Rostovtseva et al.,

¹Institute of Ethnology and Anthropology, Russian Academy of Sciences, Moscow 119334, Russia
²Russian State University for the Humanities, Moscow, 125047, Russia

Corresponding Author: Victoria V. Rostovtseva, Leninskiypr-t, 32a, (IEA RAS), Moscow 119334, Russia. Email: victoria.v.rostovtseva@gmail.com

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (https://us.sagepub.com/en-us/nam/open-access-at-sage).
2022). For instance, individuals who score high on Macheviallanism are perceived by others as good and charismatic leaders (Deluga, 2001; Jones & Paulhus, 2009), they can occupy positions of top managers, especially in young companies, whereas in a long-term perspective such leaders have a detrimental effect on the careers and well-being of their wards (Bereczkei, 2018; Volmer et al., 2016).

The present study builds on the previous experiment, in which emergent leadership in the context of male group cooperation was investigated (Rostovtseva et al., 2022). In the previous study, based on the face-to-face iterated Public Goods Game (iPGG) (conducted in initially leaderless groups of four male strangers), three types of behaviour related to emergent leadership were revealed: 60% of the male participants were classified as non-leaders, 30% expressed potential for prosocial leadership, and 10% appeared to be leaders-cheaters (for details see Methods). Independent analysis revealed that each of these behavioural types had a set of distinguishing personality, communicative, and cooperative features. Both prosocial leaders and leaders-cheaters were equally followed by other group members (on average in 88% of cases), indicating successful communicative strategies of leaders-cheaters, which allowed them to evoke trust in others despite actual deceptive intentions (Rostovtseva et al., 2022). This experiment was conducted in a specific population – among young Buryats of Southern Siberia (traditionally nomadic pastoralists of Mongolian origin). Although the obtained results demonstrated that some communicative features of leaders-cheaters were perceived by others as indicators of trustworthiness (see also Okubo et al., 2012, 2013, 2017, 2018), it was not clear whether such individual features represent only behavioural components, or they may also be imbedded into static individual morphology. To make this point clear, we have conducted this follow-up study.

The present study was designed to extend our previous findings. Here we address four new questions: (1) do neutral faces of individuals defined as non-leaders, prosocial leaders, and leaders-cheaters have common morphometric features (per type), which can be perceived by others in a consistent way?; (2) if so, which physical and behavioral traits are attributed to the faces of individuals belonging to each of these behavioural types?; (3) are there sex and population differences in facial perception of non-leaders, prosocial leaders, and leaders-cheaters?; (4) is there an impact of social environment on perception of faces of different population (racial) origin?

To answer the new research questions, we have created three composite portraits of Buryat men with different propensities for leadership and leadership styles based on individual photographs obtained in the prior experiment. The study was restricted only to male portraits since leadership qualities in the previous study were measured only in men, and generally leadership is of greater concern to men compared to women, which is deeply rooted in human phylogeny (Smith et al., 2020; Van Vuurt & von Rueden, 2020; von Rueden et al., 2018). Moreover, according to the findings by other authors, behavioural traits are better recognized in male, but not female faces (Tognetti et al., 2013). The three composite portraits were then presented to independent raters (Buryats and Russians of both sexes) using an online form. Russian raters were originating from areas with different levels of urbanization and familiarity with Mongolian appearance in everyday life. The composite portraits were scored on a number of physical and behavioural characteristics (masculine, physically strong, healthy, attractive, trustworthy, generous, leader, competitive, and dominant), without providing participants with information about behaviour of individuals constituting each of the stimulus portraits. Intuitively, it may be expected that those Russian raters who were more exposed to perception of Mongolian appearance in everyday life, living side by side with Buryats (Russians from Buryatia), will be more accurate in judging personal traits based on facial appearance (for instance, such tendencies have been previously described for accuracy of emotion recognition; Elfenbein & Ambady, 2002).

The physical and behavioural traits, used for facial judgments, were selected relying on the large body of previous findings. It has been recently shown that body masculinity and facial masculinity are positively related to male cooperativeness in the context of inter-group competition (Muñoz-Reyes et al., 2020). Physical strength and perceived masculinity are correlated with each other (Windhager et al., 2011). Previous findings by other authors suggest that leaders with masculine traits (e.g. low-pitched voices) are preferred over others since they are perceived as physically stronger (Klofstad et al., 2015). Preference for masculine, physically strong, and dominant leaders has been reported to be especially pronounced under condition of inter-group conflicts (Laustsen & Petersen, 2016, 2017; Little et al., 2007). It has been proposed earlier that humans choose leaders based on particular situation, favouring physical parameters under threatening and competitive conditions, and prosocial traits in cooperative tasks (Little, 2014). However, studies investigating perception of leader traits, by and large are based on stated preferences for leadership. In the current study, we address the question of whether individuals who actually emerged as leaders in male groups (under condition of cooperative interactions with no pressure of inter-group competition) were perceived by others as more masculine, physically strong, and dominant based on their static facial morphology.

The relation between masculinity and health is another question of interest. Facial and body masculinity represent exposure to male sex hormones (in particular, testosterone), which is costly to produce and metabolize (Alonso-Alvarez et al., 2007). A number of studies report positive association between male-specific traits and actual men’s health (Boothroyd et al., 2013; Thornhill & Gangestad, 2006), however findings on perceived health and masculinity remain somewhat contradictory (Boothroyd et al., 2005; Henderson et al., 2016). Health was also studied with regard to prosociality and altruism, and it is well established that good health and happiness are linked to prosocial and altruistic behaviour (Post, 2007). However, it is still difficult to conclude what might be the cause and what the consequence.
Such traits as trustworthy, generous and competitive were chosen as measures of perceived prosociality. A number of studies have previously demonstrated that humans can deduce prosocial intentions of others based on silent videotapes (Fetchenhauer et al., 2010; Oda, Naganawa, et al., 2009; Oda, Yamagata, et al., 2009), and even static facial images (Little et al., 2013; Tognetti et al., 2013; Verplaetse et al., 2007), and that such abilities are cross-population. At the same time, cheating behaviour is very problematic for non-verbal recognition, which has been demonstrated in numerous works (Brown et al., 2003; Okubo et al., 2012, 2013, 2017, 2018; Verplaetse et al., 2007). Moreover, studies show that generally people tend to trust attractive-looking individuals even if they don’t actually behave trustworthy (Rezlescu et al., 2012; Shinada & Yamagishi, 2014; Takahashi et al., 2006; Wilson & Eckel, 2006).

Here we hypothesize that: (1) observers will be able to differentiate leadership potential and leader styles by certain physical and behavioral characteristics based only on static facial information; (2) leaders’ faces are expected to be perceived as more masculine, physically strong, dominant, and healthy, compared to non-leaders; given a high rate of following leaders’ suggestions in the game (88%), we also hypothesize that faces of both prosocial leaders and leaders-cheaters should be scored high on attractiveness, generosity, trustworthiness, and low on competitiveness; (3) we expect Buryat raters to be more accurate in judging individual traits (based on Buryat photographs), compared to Russian raters; and (4) we expect Russian raters from Buryatia to be more accurate in their judgements of Buryat portraits than Russians from areas where contacts with individuals of Mongolian origin are scarce. There was no clear hypothesis concerning the possible sex effects in perceiving the appearance of male individuals distinguished by the propensities for leadership and leadership styles. On the one hand, leadership is generally more closely associated with male activities (Smith et al., 2020; Van Vugt & von Rueden, 2020; von Rueden et al., 2018), which may cause better recognition of individual qualities in the male-male context, but on the other hand, women are generally known to be better at processing non-verbal signals and mind-reading (Stevens & Hamann, 2012; Thompson & Voyer, 2014), which could lead to the opposite effect.

The present study is the first to our knowledge to test whether neutral facial images may convey information about actual (not perceived) personal trustworthiness in the context of emergent leadership among males. Determining, which traits are attributed to leaders with cheating propensities will help to get an insight into complex phenomenon of successful social cheating.

**Methods**

**Measuring Leadership Potential**

The leadership potential of young male Buryats was assessed under experimental conditions – in the context of the face-to-face iPGG (Chaudhuri, 2011; Ledyard, 1994) conducted in groups of four males who were strangers to each other. Participants of the experiment were young male Buryats (N = 104; aged 17–28 years, mean age 20 ± 2 years), representatives of Mongolian population of nomadic pastoralists living at the Baikal Lake area. Only those individuals who stated that both of their parents were Buryats were included into the study.

The iPGG was held in two parts: (1) under silent condition and (2) with option of verbal communication between members of a group. Each part had three repeated interactions conducted in a same way. In each interaction, each participant was given 20 initial tokens (with real monetary equivalent), and had to decide privately how many tokens (from 0 to 20) to invest in the “common pool”. After all investments were made, the common pool was doubled by the experimenter, and equally distributed between all four members of a group. Tokens that were not invested were kept by participants.

The iPGG represents a social dilemma, where individual interests meet group-level benefits. From the individual perspective it is more beneficial not to invest anything into a common pool and gain profit by free-riding on the costly efforts of others. At the same time, common cooperation brings the maximum profit to a group, and in this case all participants end up with high payoffs. In the second part of the experiment participants of our study were allowed to negotiate prior to each decision and try to achieve some consensus. However, after negotiations, individual decisions were made privately, so that any agreements could be privily violated.

Each experimental group was placed in a separate room equipped with two cameras (Web-camera Logitech Pro C920, HD 1080p). Experimental interactions were videotaped for post-hoc behavioral analysis.

Classification and definition of leaders (and non-leaders) used in the current study was implemented in several steps and was based on the full results of the former experiment. Literature suggests that in leaderless groups or dyads, leader-follower structure emerges automatically (Bass, 1949; Guastello, 2010; Harcourt et al., 2009; Van Vugt, 2006), based on common needs and individual qualities. On the first step, individual leadership potential was assessed using participants’ behaviour in the second part of the iPGG (part with negotiations). Participants, who initiated negotiations by proposing exact solutions, reasoned, actively participated in discussions, and summarized agreements before each decision-making were classified as individuals with leadership potential. Those who did not suggest anything, although could participate in the negotiations in a passive manner (agreed with solution proposed by a “leader” or expressed distrust without making any own propositions), or kept silent, were classified as individuals without leadership potential. During the observational analysis it became clear that individuals with leadership potential could be classified into two different types, based on the content of their verbal behaviour. Those who “did not insist in their suggestions,
were ready to adjust the proposed solution if other group members expressed distrust or criticized it, reasoned without excessive eagerness” were defined as “creative leaders”, whereas those who “actively agitated to follow their strategy (always to invest maximum), did not compromise with other group members, insisted on proposed solutions by resorting to a variety of arguments” were defined as “leaders-stimulators” (Rostovtseva et al., 2022). To assess the validity of the method, one more rater was invited for independent classification of 40 randomly selected participants based on the videotapes’ observations. Both raters were professionals in anthropology and human ethology. Agreement between the two raters was estimated using Cohen’s Kappa (κ) test, and was substantial enough to consider the method reliable (κ = 0.751, p < 0.001) (for reference: the “almost perfect” agreement starts from κ = 0.81 [Landis & Koch, 1977; Sim & Wright, 2005]). The inter-rater agreement on the classification of “leaders-stimulators” was 100%.

Leadership is a phenomenon, which implies followership, and both of these components are inseparable. Therefore, on the next step we estimated willingness of the group members to follow propositions of each leader. This was possible since solutions proposed by leaders in all cases contained exact amounts of tokens suggested to invest into a common pool. Likelihood of being followed (for each leader) was calculated as a percentage of investment decisions of other group members, which supported the amount of tokens initially proposed by a leader (over all interaction decisions of other group members, which supported the number of leaders-cheaters in the general sample (seven). Hence, all leaders-cheaters were included in the composite portrait of this type, and for prosocial leaders and non-leaders 7 photographs per each were selected randomly. Equal number of individual images constituting the composites of each type was necessary to ensure equal quality of the stimulus portraits.

Seventy-one anthropometric facial landmarks and semilandmarks were placed on each of the selected photographs in tpsDig 2.17 (Rohlf, 2015) following configuration development by Windhager et al. (2011), and used previously in our earlier studies (Butovskaya et al., 2018; Rostovtseva et al., 2020a, 2021a, 2021b). The facial configurations were then standardized for the position, orientation, and scale using Generalized Procrustes superimposition with sliding of the semi-landmarks in tpsRelw 1.67 (Rohlf, 2015). The composite portraits were then created by unwarping and averaging individual images upon mean facial configurations of each type (non-leaders, prosocial leaders, and leaders-cheaters) in tpsSuper 2.04 (Rohlf, 2015). The composites used in the online form are presented in Figure 1.

The approach using averaged, but not individual images, as a stimulus material targets the existence of common features of facial morphology characteristic of individuals with certain types of behaviour, which can be perceived by others in a consistent manner.

**Composite Portraits of Men with Different Leadership Potential and Styles**

All participants of the experiment were photographed in frontal perspective, with neutral facial expression, sitting on a fixed chair, with a head positioned in frontal horizontal plane. The camera was set at the eyes’ height at the fixed distance from an object. Each photograph included a scale bar (in centimetres).

Composite portraits of men with different propensities for leadership and leadership styles were created based on 7 photographs per each of three types (non-leaders, prosocial leaders, and leaders-cheaters). The number of photographs was dictated by the number of leaders-cheaters in the general sample (seven). Hence, all leaders-cheaters were included in the composite portrait of this type, and for prosocial leaders and non-leaders 7 photographs per each were selected randomly. Equal number of individual images constituting the composites of each type was necessary to ensure equal quality of the stimulus portraits.

Seventy-one anthropometric facial landmarks and semilandmarks were placed on each of the selected photographs in tpsDig 2.17 (Rohlf, 2015) following configuration development by Windhager et al. (2011), and used previously in our earlier studies (Butovskaya et al., 2018; Rostovtseva et al., 2020a, 2021a, 2021b). The facial configurations were then standardized for the position, orientation, and scale using Generalized Procrustes superimposition with sliding of the semi-landmarks in tpsRelw 1.67 (Rohlf, 2015). The composite portraits were then created by unwarping and averaging individual images upon mean facial configurations of each type (non-leaders, prosocial leaders, and leaders-cheaters) in tpsSuper 2.04 (Rohlf, 2015). The composites used in the online form are presented in Figure 1.

The approach using averaged, but not individual images, as a stimulus material targets the existence of common features of facial morphology characteristic of individuals with certain types of behaviour, which can be perceived by others in a consistent manner.

**Rating of the Composite Portraits**

The three composite portraits of young Buryats representing three behavioral types (Figure 1) were scored by 435 independent raters (251 women) of Buryat (N = 184; 110 women) and Russian (N = 251; 141 women) population origin. Age of the raters was in the range of 17 to 40 years (mean age 23 ± 3 y.). The Buryat raters were from Ulan-Ude in Southern Siberia (a residential Buryat location with a mixed Russian-Buryat population of around 400 thousand), whereas Russian raters were represented by subjects from Buryatia and neighbouring areas, as well as from Moscow (the capital of Russian Federation, with population of around 12 million), and from a provincial Russian town
populated predominantly by Russians (Tula, with population of around 500 thousand); a small number of subjects were from other Russian towns with a predominantly Russian population. These data allowed for the analysis of the effect of social environment on judgements of the composite portraits. Raters were then divided into three groups: (1) very familiar with Mongolian appearance based on everyday life experience (Russians from Buryatia; \(N = 116\)); (2) marginally familiar with large variety of populations, including those with Mongolian appearance (Russians from Moscow; \(N = 73\)); not very familiar with Mongolian appearance based on everyday life experience (Russians from provinces with predominantly Russian population; \(N = 62\)).

Ratings of the composite portraits were obtained through an online form created with SoSci Survey (Leiner, 2019) and made available to the participants on www.soscisurvey.com. Each composite was presented on a separate page of the online form, and was rated using a slider from 1 to 100 on nine traits (trustworthy, healthy, generous, leader, masculine, competitive, dominant, physically strong, attractive), with all traits presented on the same page with a judged portrait. The order of the portraits (pages) was randomly shuffled, while the order of the judged traits was given as presented above. The latter was done for two reasons: (1) trustworthiness was always rated first (as the main trait of interest), which prevented judgements on the other traits to “color” the judgements on trustworthiness; (2) attractiveness was always rated last, which prevented judgements on the other traits to be “colored” by a rather general judgement on attractiveness. To test for a general order effect (systematic increase or decrease in given scores with respect to the order of the nine judged traits), all individual judgements were pooled and a linear regression analysis was performed. The scores for the judged traits were set as a response variable, and their order (from 1 to 9) as a continuous independent variable. No significant effect of the order of the judged traits was revealed, the regression line was flat with less than 1% of the variance explained. The raters were not provided with information about behaviour of individuals constituting each of the stimulus portraits. Prior to making judgements, demographic information about participants (age, sex, population origin, and location of origin) was collected.

Participation in the study was voluntary and participants gave consent to the use of their answers by submitting their online forms. All information was anonymous, with no personal details. The study was approved by the scientific board of the Institute of Ethnology and Anthropology of the Russian Academy of Sciences.

**Statistical Analysis**

Since we used continuously-scaled sliders (from 0 to 100) for scoring the composites on the nine traits, a very large variation in general individual levels of ratings occurred (with some raters giving low scores to all portraits on all traits, and some giving generally high scores). To level this inter-individual variation, we have standardized the raw scores within each rater. Within-individual distributions of judgements were symmetrical enough to apply z-score transformation. All scores given by an individual to all three composites on all judged traits were z-transformed within that individual (mean value of all judgements of an individual was set to zero, and all other values of his/her judgements were transformed into new values in terms of (+/-) standard deviations from that mean; the procedure was held separately per each individual). Within-individual
standardization allowed eliminating individual differences in the general levels of ratings, saving initial variation patterns in the data. This procedure was necessary to cope with the fact that some participants used the bottom end of the scale and some generally judged at the top of the scale, which resulted in automatic statistical artifacts (correlation between the raw scores on the judged traits resulted in highly significant positive associations between all traits). After applying within-individual z-transformation, associations between ratings on different traits were analyzed using partial Pearson’s correlation. For this purpose, ratings (z-transformed) of all three composite portraits were pooled within each trait, and correlated pairwise with control for sex, population origin of the raters, and the type of the portrait judged.

Associations between judgements on each single trait with a number of independent factors and their interactions were analyzed using general linear models (GLM) and the Tukey-Kramer test. Using raw scores in this case would also introduce additional noise to the data, since individual contributions to the trait variable values would remain unweighted. Therefore, this analysis was also performed on the within-individual z-transformed data.

The only procedure, which involved raw ratings was the analysis of intraclass correlation coefficients (ICC) assessing the inter-rater agreement of judgements of the three portraits on each trait. Since this procedure was based on the consistency estimates it was unaffected by the general inter-individual variation in judgements’ levels. The ICC analysis was implemented based on the two-way random effects model (Koo & Li, 2016). Obtained ICC values were interpreted in accordance with Cicchetti (1994).

The significance level was set to 0.05. The analysis was performed in SPSS (IBM Corp. Released 2019. IBM SPSS Statistics for Windows, Version 26.0. Armonk, NY: IBM Corp). The data that support the findings of this study are available from the corresponding author upon reasonable request.

**Results**

**Correlation Between Judged Traits**

First of all, we assessed correlations between ratings on different traits, controlling for sex and population origin of the raters, as well as for the type of the composite portrait judged. Since there was very large variation in general individual levels of ratings, within-individual z-scores were used in the analysis (see Methods). The results are presented in Figure 2 (for exact values see Supplementary Table 1).

The results demonstrated that as a general tendency (independent of sex, population origin of the raters, or type of the judged composite) there were sets of traits, which were perceived in association with each other. These associations generally were rather weak, but significant. Ratings on masculinity, physical strength, dominance, competitiveness, and leadership were positively correlated with each other, at the same time, perceived trustworthiness was negatively associated with these traits. Healthy- and attractive-looking portraits were scored higher on trustworthiness.

**Inter-Rater Consistency of Judgements**

An essential part of the analysis was the assessment of the inter-rater consistency of judgements of the three portraits on each trait per different sets of raters. For this analysis we used raw scores without z-transformation. The results of the intraclass correlation analysis are presented in Figure 3. According to the results raters agreed the most at judging physical characteristics (such as masculinity, physical strength, and health) attributed to certain composites. Good consistency was also observed for such behavioural traits as trustworthiness, generosity, and competitiveness. At the same time, leadership, dominance and attractiveness were judged very inconsistently, in some cases even negative values of mean covariance coefficients occurred. The latter indicates very large individual differences in perceiving such traits as leadership, dominance and attractiveness in our sample. A closer look at the pattern of distribution of ICC values reveals that Buryats (especially women) agreed the most on judging masculinity and physical strength, whereas Russians generally better agreed in judging behavioural traits. We did not find substantial evidence for Russians with different degrees of familiarity with Mongolian appearance to demonstrate distinctive patterns in consistency of their judgements of Buryat portraits.

Further analysis was performed using within-individual z-scores, only on those traits, which revealed excellent and good between-rater consistency, according to Cicchetti (1994) (average ICC > 0.6).

**An Impact of Social Environment**

To test the effect of social environment on the ratings of the composite portraits representing individuals with different leadership potential we applied GLM, where each judged trait (masculine, physically strong, healthy, trustworthy, generous, competitive) were one by one set as single dependent variables, whereas familiarity with Mongolian appearance (“very
familiar” for Buryats, and Russians from Buryatia; “marginally familiar” for Russians from Moscow; and “not very familiar” for Russians from provinces), population origin (Buryats, Russians), sex of the raters, and type of the judged portrait (non-leader, prosocial leader, and leader-cheater) were set as independent factors. Interactions between factors were also included into the models. In all six models the only significant factor affecting judgments of the composites was the “Type of Leadership Potential”: (1) Masculinity ($F = 9.018$, $p < 0.001$, $R^2_{model} = 0.04$); (2) Physical Strength ($F = 21.400$, $p < 0.001$, $R^2_{model} = 0.10$); (3) Health ($F = 30.758$, $p < 0.001$, $R^2_{model} = 0.11$); (4) Trustworthiness ($F = 9.777$, $p < 0.001$, $R^2_{model} = 0.05$); (5) Generosity ($F = 15.075$, $p < 0.001$, $R^2_{model} = 0.10$); (5) Competitiveness ($F = 10.343$, $p < 0.001$, $R^2_{model} = 0.11$) (for detailed statistics on all insignificant effects see Supplementary Table 2, 3). We conclude that originating from environments with different degrees of familiarity with Mongolian appearance in everyday life did not affect perception of the studied traits by Russian raters. Thereafter, we pooled judgements of Russian raters for further analysis.

**Association of the Facial Ratings with the Type of Leadership Potential, Raters’ Sex, Age and Population Origin**

Next, we looked at the associations between judgements of the composite portraits representing individuals with different leadership potential and leader styles with regard to age, sex, and population origin of the raters.

Figure 4 represents the results of the GLM, where ratings on the physical characteristics (masculine, physically strong, healthy) were set as single dependent variables, and behavioral type represented by the composite portrait, raters’ sex, age, and population origin (and interactions between these factors) as multiple independent variables. The composite portrait of leaders-cheaters was scored as the most masculine (especially by younger raters) (Figure 4, a, b). According to the Tukey-Kramer test, ratings on Masculinity significantly differed between all three stimulus portraits: the non-leaders’ composite was scored as less masculine than the prosocial leaders’ composite portrait ($p = 0.03$), and less masculine than the leaders-cheaters’ portrait ($p < 0.001$); at the same time, the composite portrait of leaders-cheaters was scored as more masculine than the prosocial leaders’ composite ($p = 0.01$). The effect was significant and independent of sex or population origin of the raters (Table 1), however, it was quite weak as it explained only 4% of the overall ratings’ variance. The composite portraits of leaders in general were scored as more physically strong than the composite of non-leaders (Tukey-Kramer test: $p < 0.001$ for both cases of leaders), however, ratings on this trait did not differ significantly between the prosocial leaders’ and the leaders-cheaters’ composite portraits (Figure 4, c). The effect explained 10% of the overall ratings’ variance (Table 1). The prosocial leaders’ composite was scored as significantly healthier than both non-leaders’ ($p < 0.001$) and leaders-cheaters’ ($p <$

![Figure 3. Inter-rater consistency of judgements on each trait. ICC – intraclass correlation coefficients (mean covariance coefficients), based on two-way random effects model, consistency, average measures.](image-url)
0.001) composite portraits (Figure 4d), which was especially the case for female raters (Table 1). Older raters generally scored the portraits as healthier than younger raters (Figure 4e). The GLM model for the trait Healthy explained 11% of the overall judgements’ variance (Table 1).

When testing associations between ratings on prosocial qualities and a set of independent variables, age and population origin of the raters and their interactions with any of other predictors were not significant in any of the models. Therefore, raters’ age and origin were further excluded from the list of independent variables.

The composite portraits of leaders in general were scored higher on trustworthiness, than the non-leaders’ composite. However, the ratings on prosocial qualities converged at the point that the leaders-cheaters’ portrait was perceived as more trustworthy, more generous, and also less competitive than others (Figure 5), and there were no highly significant effects or interactions for sex (Table 2). According to the

---

**Figure 4.** Association of ratings on physical traits (z-scores) with the leadership potential, raters’ sex, age and origin. Significant differences according to Tukey-Kramer test are shown between x-axis conditions: *p* < 0.05; **p** < 0.01.
Table 1. Association of Ratings on Physical Traits (z-Scores) with the Type of Leadership Potential, Raters’ sex, age and Origin.

| Dependent var.: Masculinity (perceived) | F     | Sig.  | R²   |
|----------------------------------------|-------|-------|------|
| Independent var.                        |       |       |      |
| Type of Leadership Potential           | 6.683 | 0.001*| 0.04 |
| Age of the raters                      | 0.157 | 0.692 |      |
| Sex of the raters                      | 0.701 | 0.403 |      |
| Origin of the raters                   | 3.487 | 0.062*|      |
| Type of Leadership Potential * Age of the raters | 4.289 | 0.014*|      |
| Type of Leadership Potential * Sex of the raters | 0.554 | 0.575 |      |
| Type of Leadership Potential * Origin of the raters | 0.829 | 0.437 |      |

| Dependent var.: Physical Strength (perceived) | F     | Sig.  | R²   |
|----------------------------------------------|-------|-------|------|
| Independent var.                            |       |       |      |
| Type of Leadership Potential                | 7.940 | <0.001*| 0.10 |
| Age of the raters                           | 2.120 | 0.146 |      |
| Sex of the raters                           | 0.209 | 0.647 |      |
| Origin of the raters                        | 0.108 | 0.742 |      |
| Type of Leadership Potential * Age of the raters | 1.041 | 0.353 |      |
| Type of Leadership Potential * Sex of the raters | 1.192 | 0.304 |      |
| Type of Leadership Potential * Origin of the raters | 0.024 | 0.977 |      |

| Dependent var.: Health (perceived)          | F     | Sig.  | R²   |
|---------------------------------------------|-------|-------|------|
| Independent var.                            |       |       |      |
| Type of Leadership Potential                | 2.403 | 0.091*| 0.11 |
| Age of the raters                           | 7.832 | 0.005*|      |
| Sex of the raters                           | 2.785 | 0.095*|      |
| Origin of the raters                        | 0.273 | 0.601 |      |
| Type of Leadership Potential * Age of the raters | 1.561 | 0.210 |      |
| Type of Leadership Potential * Sex of the raters | 3.540 | 0.029*|      |
| Type of Leadership Potential * Origin of the raters | 1.661 | 0.191 |      |

Tukey-Kramer test, the leaders-cheaters’ composite was rated as more trustworthy than the non-leaders’ composite (p < 0.001), and also than the prosocial leaders’ composite portrait (p = 0.05); the leaders-cheaters’ composite was also distinguished from others as more generous (p < 0.001). The composite portrait of leaders-cheaters was perceived as less competitive than both prosocial leaders’ (p = 0.009), and non-leaders’ (p < 0.001) composites. However, in all these cases the effect of leaders’ appearance, revealed through GLM analysis, explained only 4% of the overall ratings’ variance (Table 2).

Discussion

The goal of the present study was to assess how actual prosocial/cheating behaviour in the context of emergent leadership among men (revealed earlier in experimental setting) is related to perception of their neutral faces by representatives of own and contrasting populations, as well as by raters of different sexes. Although many studies have been previously conducted investigating different aspects of human faces’ perception, very few were focusing on matching facial perception with actual prosocial behaviour of judged subjects (Tognetti et al., 2013; Verplaetse et al., 2007), and even those few were targeting the ability to distinguish between limited types of behaviour (prosocial/selfish) based on facial images. Here we applied the different approach, estimating which physical and behavioural traits are attributed to averaged faces of individuals with different types of leader/follower and prosocial/cheating behaviour. The present study involved raters of different population origin: Buryats (Mongolian people from Eastern Siberia, representatives of the population from the stimulus portraits) and Russians (people of Caucasian origin, representing contrasting population).

Our results revealed that raters of the both sexes and from both populations highly agreed at judging composite portraits of individuals with different behavioural backgrounds. In particular, it was manifested in judgements on physical traits (masculine, physically strong, healthy), and a number of behavioural traits (trustworthy, generous and competitive), meaning that the same kinds of composites received equally high or low judgements on these traits independently of sex or population origin of the raters (Figure 2). However, ratings on such traits as leadership, dominance, and attractiveness were very inconsistent, especially among Buryat raters. This suggests that judgements on these traits are highly dependent on personal characteristics of the raters (Little et al., 2011), which should be taken into account in future research. Russian raters in our study differed in the degree of familiarity with Mongolian appearance in everyday life. Contrary to what was expected, our results demonstrated that living in the mixed-racial social environment also did not affect judgements of the composite portraits: Russians from Buryatia (living side by side with Buryats), and Russians from multi-ethnic megalopolis (Moscow), and Russian provinces (with mostly Russian population) did not differ either in ratings of the Buryat portraits or in the consistency of judgements. These results together support the notion that humans perceive facial appearance of others in a similar way (Van’t Wout & Sanfey, 2008), and that perception of at least male faces is not strongly sex- or population-specific (Tognetti et al., 2013). The similarity of the mechanisms of facial perception in humans was previously proposed by many authors in relation to the recognition of facial expressions (Ekman et al., 1969; Ekman & Friesen, 1971, 1975; Sorenson, 1975). However, in the more recent literature addressing this issue, the Universality Thesis (as the claim that basic emotions are universally recognized by humans regardless of cultural and population backgrounds) has been widely debated and criticized (Matsumoto & Assar, 1992; Nelson & Russell, 2013). Recent studies revealed considerable cultural specificity, when remote (isolated) societies were tested (Crivelli et al., 2016; Gendron et al., 2014, 2020). These results questioned
whether and to what extent specific aspects of facial perception differ across populations. One of possible arguments for high similarity of mechanisms of human perception of static faces across different populations may be based on the fact that certain behavioural features (e.g. cooperativeness or aggressiveness) (Carré & McCormick, 2008; Geniole et al., 2015; Rostovtseva et al., 2021b; Stirrat & Perrett, 2010, 2012; Wen & Zheng, 2020), as well as physical characteristics (e.g. physical strength) (Butovskaya et al., 2018) are known to be incorporated into individual facial morphology, and display similar patterns at cross-population level (although see: Butovskaya et al., 2021). This may partly explain why people react similarly to certain types of facial appearance (such reactions may be driven by their actual life experiences). However, this may also be partly a result of high exposure to the global international mass media (television, Internet, cinematograph), which introduces a large diversity of the humankind to the audience from different parts of the world. To clarify to what extent unified perception of neutral facial traits may be the consequence of the global information environment, studies in societies with limited access to international mass media, should be conducted.

The composite portraits of leaders generally were scored as more physically strong and more masculine than the non-leaders' portrait, which was significant with no respect to age, sex, and population origin of the raters. This result demonstrates that preference for physically developed leaders may occur not only under condition of explicit intergroup competition, which is in line with findings by other authors (Little, 2014). Possibly, one of the reasons that individuals with high perceived physical strength and masculinity emerged as leaders during group male cooperative tasks lies in a specific cultural background of Buryats. In the recent past Buryats were traditionally nomadic pastoralists with highly developed male warfare practices, and male collective hunting (more details on their culture can be found elsewhere: Rostovtseva et al., 2020b). Such cultural traditions may still have echoes in their modern social behaviour. At the same time, the leaders-cheaters' composite portrait was perceived as the most masculine by younger raters (< 28 years), whereas older ones tended to score the prosocial

![Figure 5. Association of ratings on behavioral traits (z-scores) with the leadership potential and raters' sex. Significant differences according to Tukey-Kramer test are shown between x-axis conditions: * p < 0.05; ** p < 0.01.](image-url)
Table 2. Association of Ratings on Behavioral Traits (z-Scores) with the Type of Leadership Potential and Raters’ sex.

| Dependent var. | F       | Sig.   | R²   |
|---------------|---------|--------|------|
| **Trustworthiness (perceived)** |         |        |      |
| Type of Leadership Potential | 16.779  | <0.001 | 0.04 |
| Sex of the raters | 0.196   | 0.658  |      |
| Type of Leadership Potential * Sex of the raters | 2.960   | 0.052* |      |
| **Generosity (perceived)** |         |        |      |
| Type of Leadership Potential | 17.705  | <0.001 | 0.04 |
| Sex of the raters | 1.576   | 0.210  |      |
| Type of Leadership Potential * Sex of the raters | 0.442   | 0.643  |      |
| **Competitiveness (perceived)** |         |        |      |
| Type of Leadership Potential | 21.957  | <0.001 | 0.04 |
| Sex of the raters | 0.883   | 0.348  |      |
| Type of Leadership Potential * Sex of the raters | 0.053   | 0.949  |      |

Three independent GLMs are presented. Type of Leadership Potential (non-leader, prosocial leader, leader-cheater). Statistical trends (p < 0.1) are marked with *; highly significant associations (p < 0.001) are bold and marked with ***.

leaders’ composite higher on masculinity (Figure 4b). Visual inspection of the stimulus portraits (Figure 1) suggests that the main features distinguishing leaders-cheaters and prosocial leaders are the shape of the eyes, lips, jaw, and brows (with leaders-cheaters having larger and more rounded eyes, fuller lips, more rounded jaw, and more prominent brows). In previous research on facial sexual dimorphism of Buryats it was revealed that more rounded shape of the eyes, and more rounded jaw are among the main female sex-specific features characteristic of this population (Rostovtseva et al., 2021b). It has been previously demonstrated that feminized male faces are perceived as more attractive and honest (Perrett et al., 1998), which may partly explain perception of the leaders-cheaters’ composite as the most trustworthy. But still the question arises why younger raters tended to score their portrait as the most masculine, whereas it was characterized by more feminine facial appearance. Since stimulus portraits were based on the faces of young Buryats (17 –25 y.), we can conclude that the effect of perceiving the leaders-cheaters’ facial appearance as more masculine occurred only within peer environment. This result suggests some implicit mechanisms of peer facial perception (Rhodes & Anastasi, 2012), exploited by individuals with cheating intentions. However, yet we can only speculate on this issue, and more research on morphology and perception is needed to clarify it.

The composite portrait of prosocial leaders was judged as the most healthy. This is in line with evidence proposing that prosociality and altruism are positively related to health (Post, 2007). Moreover, as a general trend, perceived health was also positively associated with scores on trustworthiness (Figure 2). However, associations of perceived health with other traits were very weak (this is the trait with the lowest correlations with other traits among all tested in this study) (Figure 2). Perceived health was not related to perceived masculinity, generosity, competitiveness, and leadership. This result corresponds to the previous findings that health judgements and masculinity are not directly associated (Henderson et al., 2016). However, perceived health was also not related to attractiveness, which contradicts earlier findings by other authors (Jones et al., 2001).

Considering human prosociality, which is known to be a cornerstone of the Homo sapiens species, large work has been conducted to investigate those mechanisms, which theoretically could lead to high density of prosocial individuals in human societies (see reviews: Apicella & Silk, 2019; Nowak, 2006). However, there always a certain proportion of cheaters and free-riders in human populations. One of the reasons behind that are sophisticated mechanisms underlying choice of interaction partners, including recognition of social traits by perceived cues, but at the same time, a kind of prosocial mimicry or camouflage, which allow selfish agents to lure the victims for exploitation (Dawkins, 1976; Gambetta, 2005; Mokkonen & Lindstedt, 2016). In line with this theoretical background, the composite portrait of leaders-cheaters in our study was scored as the most trustworthy, most generous and least competitive compared to all other composites. These results indicate that leaders-cheaters had certain features of static facial morphology, which made them to be perceived by others as more prosocial. Although the respondents did not directly assess leaders-cheaters as better leaders, something in their appearance made people trust them more. This may partially explain why they could successfully free-ride on the goodwill of others. This effect was weak, but significant independently of sex, and population origin of the raters.

However, this study has certain limitations. The stimulus composite portraits were created based only on 7 individual photographs per type, which can be considered as relatively few. The number of individual photographs for averaging is limited by the minimal number of individuals per behavioural type, but due to frequency dependent selection (Avilés, 2002; Pruitt & Riechert, 2009) leaders-cheaters are expected to remain in great minority. Unfortunately, this limitation is difficult to overcome within a single experimental study, using design with face-to-face interactions and ethnological analysis of behaviour. One of the possible solutions could be prescreening individuals to gain larger sample sizes. Another limitation is that, as a certain compromise, the order of judged traits was not randomized in our study. Finally, one should remember that the aim of our study was to investigate emergent leadership in groups of individuals who were strangers to each other. This implies that reputation (which is known to be one of the most important aspects of leadership (Anderson et al., 2015; Henrich & Gil-White, 2001; Van Vugt & von Rueden, 2020; von Rueden & Van Vugt, 2015), was not involved in the experimental interactions. We assume that if reputation was involved, it would override the influence of all other factors (both at the level of leadership and cooperation). Our study targets the
processes of social self-organization in initially leaderless groups, and the obtained results cannot be generalized to communities with already established social hierarchy. However, the results may apply to social processes in modern urban environments, where interactions among strangers are very common. More independent studies are required to test reproducibility of the obtained results.

To sum up, our results reveal that individual features of static facial morphology also contribute to the phenomenon of successful cheating in the context of male emergent leadership, along with personality and communicative traits (Okubo et al., 2012, 2013, 2017, 2018; Rostovtseva et al., 2022). Our study opens up new perspectives for further research, invoking investigation of particular facial structures, which distinguish individuals with different leadership qualities, and their specific contribution to facial perception.

Acknowledgments
We want to thank all our subjects for their interest and willingness to participate in the project. Special thanks to the East-Siberian State Institute of Culture (VSGIK) and the Faculty of Natural Sciences of the Tula State Lev Tolstoy Pedagogical University for collaboration. We also express our gratitude to the anonymous reviewer and associate editor Lisa Welling for proofreading the English writing.

Ethics
Participation in the study was voluntary and participants gave consent to the use of their answers by submitting their online-forms. All information was anonymous, with no personal details. The study was approved by the scientific board of the Institute of Ethnology and Anthropology of the Russian Academy of Sciences.

Data, Code and Materials
The data that support the findings of this study are available from the corresponding author upon reasonable request.

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

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Russian Science Foundation, (grant number 18-18-00075).

ORCID iD
Victoria V. Rostovtseva https://orcid.org/0000-0002-1846-9865

Supplemental material
Supplemental material for this article is available online.

Note
1. Although the raters’ age and its interactions with other independent variables were not significant in any of the models, adding age to the models affected the significance of the other predictors. In order to ensure that the raters’ age does not make a significant contribution, additional statistical procedures were performed. The general sample was split into two groups (under 25 years old [majority], and over 25 years old) and the GLM models were performed on these two age cohorts separately. In this case, the age variation in each cohort was rather low and age was not added to the list of predictors. The results revealed only one significant predictor, “Leadership Potential (three types)”. It was highly significant in all three models (trustworthiness, generosity, and competitiveness) and had the same direction in both age cohorts, while other predictors remained insignificant. To capture possible non-linear effects, associations between age (in the general sample) and judgements on behavioral traits (per each type of the composite portraits) were tested with various non-linear models (quadratic, cubic, logarithmic). Results revealed no (even nearly) significant associations (regression lines were flat in all cases). These results indicate that the age of the raters did not interfere with the main findings.

References
Alonso-Alvarez C., Bertrand S., Faiivre B., Chastel O., & Sorci G. (2007). Testosterone and oxidative stress: The oxidation handicap hypothesis. *Proceedings of the Royal Society B: Biological Sciences*, 274(1611), 819–825. https://doi.org/10.1098/rspb.2006.3764
Anderson C., Hildreth J. A. D., & Howland L. (2015). Is the desire for status a fundamental human motive? A review of the empirical literature. *Psychological Bulletin*, 141(3), 574. https://doi.org/10.1037/a0038781
Apicella C. L., & Silk J. B. (2019). The evolution of human cooperation. *Current Biology*, 29(11), R447–R450. https://doi.org/10.1016/j.cub.2019.03.036
Avišés L. (2002). Solving the freeloaders paradox: Genetic associations and frequency-dependent selection in the evolution of cooperation among nonrelatives. *Proceedings of the National Academy of Sciences*, 99(22), 14268–14273. https://doi.org/10.1073/pnas.212408299
Bass B. M. (1949). An analysis of the leaderless group discussion. *Journal of Applied Psychology*, 33(6), 527. https://doi.org/10.1037/h0058164
Bereczkei T. (2018). *Machiavellianism: The psychology of manipulation*. Routledge. https://doi.org/10.4324/9781315106922
Boothroyd L. G., Jones B. C., Burt D. M., Cornwell R. E., Little A. C., Tiddeman B. P., & Perrett D. I. (2005). Facial masculinity is related to perceived age but not perceived health. *Evolution and Human Behavior*, 26(5), 417–431. https://doi.org/10.1016/j.evolhumbehav.2005.01.001
Boothroyd L. G., Scott I., Gray A. W., Coombes C. I., & Pound N. (2013). Male facial masculinity as a cue to health outcomes. *Evolutionary Psychology*, 11(5), 147470491301100508. https://doi.org/10.1177/147470491301100508
Brown W. M., Palameta B., & Moore C. (2003). Are there nonverbal cues to commitment? An exploratory study using the zero-acquaintance video presentation paradigm. *Evolutionary Psychology*, 1(1), 147470490300100104. https://doi.org/10.1177/147470490300100104
Butovskaya M. L., Mezentseva A., Mabulla A., Shackelford T. K., Schaefer K., Fink B., & Windhager S. (2021). Facial cues to physical strength increase attractiveness but decrease aggressiveness assessments in male Maasai of Northern Tanzania. *Evolution and
Human Behavior. https://doi.org/10.1016/j.evolveumbehav.2021.11.006

Butovskaya M. L., Windhager S., Karelin D., Mezentseva A., Schaefer K., & Fink B. (2018). Associations of physical strength with facial shape in an African pastoralist society, the Maasai of Northern Tanzania. PloS One, 13(5), e0197738. https://doi.org/10.1371/journal.pone.0197738

Carrié J. M., & McCormick C. M. (2008). In your face: Facial metrics predict aggressive behavior in the laboratory and in varsity and professional hockey players. Proceedings of the Royal Society B: Biological Sciences, 275(1651), 2651–2656. https://doi.org/10.1098/rspb.2008.0873

Chaudhuri A. (2011). Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. Experimental Economics, 14(1), 47–83. https://doi.org/10.1007/s10683-010-9257-1

Cicchetti D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological Assessment, 6(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284

Costa P. T. Jr., & McCrae R. R. (1992). The NEO-PI/NEO-FFI manual supplement. Psychological Assessment Resources.

Crivelli C., Jarillo S., Russell J. A., & Fernández-Dols J. M. (2016). Reading emotions from faces in two indigenous societies. Journal of Experimental Psychology: General, 145(7), 830. https://doi.org/10.1037/xge0000172

Dawkins R. (1976). The selfish gene. Oxford University Press.

Deluga R. J. (2001). American Presidential Machiavellianism: Implications for charismatic leadership and rated performance. The Leadership Quarterly, 12(3), 339–363. https://doi.org/10.1016/S1048-9843(01)00082-0

Ekman P., & Friesen W. (1975). Unmasking the face. Prentice-Hall.

Ekman P., & Friesen W. V. (1971). Constants across cultures in the face and emotion. Journal of Personality and Social Psychology, 17(2), 124–129. https://doi.org/10.1037/h0030377

Ekman P., Sorenson E. R., & Friesen W. V. (1969). Pan-cultural elements in facial displays of emotion. Science (New York, N.Y.), 164(3875), 86–88. https://doi.org/10.1126/science.164.3875.86

Elfenbein H. A., & Ambady N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. Psychological Bulletin, 128(2), 203. https://doi.org/10.1037/0033-2909.128.2.203

Ensminger J., & Henrich J. (Eds.). (2014). Experimenting with social norms: Fairness and punishment in cross-cultural perspective. Russell Sage Foundation.

Fehr E., & Gächter S. (2000). Cooperation and punishment in public goods experiments. American Economic Review, 90(4), 980–994. https://doi.org/10.1257/aer.90.4.980

Fetchenhauer D., Groothuis T., & Pradel J. (2010). Not only states but traits—humans can identify permanent altruistic dispositions in 20 s. Evolution and Human Behavior, 31(2), 80–86. https://doi.org/10.1016/j.evolveumbehav.2009.06.009

Gambetta D. (2005). Deceptive mimicry in humans. In Hurley S., & Chater N. (Eds.), Perspectives on imitation: Imitation, human development, and culture (pp. 221–241). The MIT Press.

Gendron M., Hoemann K., Crittenden A. N., Mangola S. M., Ruark G. A., & Barrett L. F. (2020). Emotion perception in Hadza Hunter-Gatherers. Scientific Reports, 10(1), 1–17. https://doi.org/10.1038/s41598-020-62057-2

Gendron M., Roberson D., van der Vyver J. M., & Barrett L. F. (2014). Perceptions of emotion from facial expressions are not culturally universal: Evidence from a remote culture. Emotion (Washington, D.C.), 14(2), 251. https://doi.org/10.1037/a0036052

Geniole S. N., Denson T. F., Dixon B. J., Carré J. M., & McCormick C. M. (2015). Evidence from meta-analyses of the facial width-to-height ratio as an evolved cue of threat. PloS One, 10(7), e0132726. https://doi.org/10.1371/journal.pone.0132726

Guastello S. J. (2010). Self-organization and leadership emergence in emergency response teams. Nonlinear Dynamics, Psychology, and Life Sciences, 14(2), 179–204.

Harcourt J. L., Ang T. Z., Sweetman G., Johnstone R. A., & Manica A. (2009). Social feedback and the emergence of leaders and followers. Current Biology, 19(3), 248–252. https://doi.org/10.1016/j.cub.2008.12.051

Heckathorn D. D. (1989). Collective action and the second-order free-rider problem. Rationality and Society, 1(1), 78–100. https://doi.org/10.1177/1043463189001001006

Henderson A. J., Holzleitner I. J., Talamas S. N., & Perrett D. I. (2016). Perception of health from facial cues. Philosophical Transactions of the Royal Society B: Biological Sciences, 371(1693), 20150380. https://doi.org/10.1098/rspb.2015.0380

Henrich J., & Gil-White F. J. (2001). The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. Evolution and Human Behavior, 22(3), 165–196. https://doi.org/10.1016/S1090-5138(00)00071-4

Jones B. C., Little A. C., Penton-Voak I. S., Tiddeman B. P., Burt D. M., & Perrett D. I. (2001). Measured facial asymmetry and perceptual judgments of attractiveness and health. Evolution and Human Behavior, 22, 417–429. https://doi.org/10.1016/S1090-5138(01)00083-6

Jones D. N., & Paulhus D. L. (2009). Machiavellianism. In Leary M. R., & Hoyle R. H. (Eds.), Individual differences in social behavior (pp. 93–108). Guilford.

Klofstad C. A., Anderson R. C., & Nowicki S. (2015). Perceptions of competence, strength, and age influence voters to select leaders with lower-pitched voices. PloS One, 10(8), e0133779. https://doi.org/10.1371/journal.pone.0133779

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

Landis J. R., & Koch G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33(1), 159–174. https://doi.org/10.2307/2529310

Laustsen L., & Petersen M. B. (2016). Winning faces vary by ideology: How nonverbal source cues influence election and communication success in politics. Political Communication, 33(2), 188–211. https://doi.org/10.1080/10584609.2015.1050565

Laustsen L., & Petersen M. B. (2017). Perceived conflict and leader dominance: Individual and contextual factors behind preferences for dominant leaders. Political Psychology, 38(6), 1083–1101. https://doi.org/10.1111/pop.12403

Ledyard J. O. (1994). Public goods: A survey of experimental research (pp. 111–194). Princeton University Press.
Sim J., & Wright C. C. (2005). The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical therapy*, 85(3), 257–268. https://doi.org/10.1093/ptj/85.3.257

Smith J. E., Ortiz C. A., Buhbe M. T., & van Vugt M. (2020). Obstacles and opportunities for female leadership in mammalian societies: A comparative perspective. *The Leadership Quarterly*, 31(2), 101267. https://doi.org/10.1016/j.leaqua.2018.09.005

Sorenson E. R. (1975). Culture and the expression of emotion. In Williams T. R. (Ed.), *Psychological anthropology* (pp. 351–372). Aldine. https://doi.org/10.1515/9783110802818.361

Stirrat M., & Perrett D. I. (2012). Face structure predicts cooperation: Attractiveness readable in static facial features? An inter-cultural approach. *Evolution and Human Behavior*, 34(6), 427–432. https://doi.org/10.1016/j.evolhumbehav.2013.08.002

Van 't Wout M., & Sanfey A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition*, 108(3), 796–803. https://doi.org/10.1016/j.cognition.2008.07.002

Van Vugt M. (2006). Evolutionary origins of leadership and followership. *Personality and Social Psychology Review*, 10(4), 354–371. https://doi.org/10.1207/s15327957pspr1004_5

Van Vugt M., & van Rueden C. R. (2020). From genes to minds to cultures: Evolutionary approaches to leadership. *The Leadership Quarterly*, 31(2), 101404. https://doi.org/10.1016/j.leaqua.2020.101404

Verplaetse J., Vanneste S., & Braeckman J. (2007). You can judge a book by its cover: The sequel: A kernel of truth in predictive cheating detection. *Evolution and Human Behavior*, 28(4), 260–271. https://doi.org/10.1016/j.evolhumbehav.2007.04.006

Volmer J., Koch I. K., & Görtz A. S. (2016). The bright and dark sides of leaders’ dark triad traits: Effects on subordinates’ career success and well-being. *Personality and Individual Differences*, 101, 413–418. https://doi.org/10.1016/j.paid.2016.06.046

von Rueden C., Alami S., Kaplan H., & Gurven M. (2018). Sex differences in political leadership in an egalitarian society. *Evolution and Human Behavior*, 39(4), 402–411. https://doi.org/10.1016/j.evolhumbehav.2018.03.005

Von Rueden C., & Van Vugt M. (2015). Leadership in small-scale societies: Some implications for theory, research, and practice. *The Leadership Quarterly*, 26(6), 978–990. https://doi.org/10.1016/j.leaqua.2015.10.004

Wen G., & Zheng L. (2020). Facial width to height ratio predicts physical aggression in committed relationships in men and dominance in women in China. *Personality and Individual Differences*, 157, 109832. https://doi.org/10.1016/j.paid.2020.109832

Wilson R. K., & Eckel C. C. (2006). Judging a book by its cover: Beauty and expectations in the trust game. *Political Research Quarterly*, 59(2), 189–202. https://doi.org/10.1177/106591290605900202

Windhager S., Schaefer K., & Fink B. (2011). Geometric morphometrics of male facial shape in relation to physical strength and perceived attractiveness, dominance, and masculinity. *American Journal of Human Biology*, 23(6), 805–814. https://doi.org/10.1002/ajhb.21219

Yang F., Choi Y. J., Mischa A., Yang X., & Dunham Y. (2018). In defense of the commons: Young children negatively evaluate and sanction free riders. *Psychological Science*, 29(10), 1598–1611. https://doi.org/10.1177/0956797618779061