This is a preprint of a manuscript that has been accepted for publication in

Behavioral Ecology.

A copy editor may make further changes to improve the readability of this manuscript.

Linking men's voice pitch to actual and perceived trustworthiness across domains

Christoph Schild[1*], Julia Stern[2*], & Ingo Zettler[1]

[1] Department of Psychology
University of Copenhagen
Øster Farimagsgade 2a, 1353 Copenhagen, Denmark

[2] Department of Psychology & Leibniz ScienceCampus Primate Cognition
University of Goettingen
Gosslerstrasse 14, 37073 Goettingen, Germany

*Christoph Schild and Julia Stern share the first authorship.

Corresponding authors: Christoph Schild (cs@psy.ku.dk) and Julia Stern (julia.stern@psych.uni-goettingen.de)

This research was funded by a grant from the Carlsberg Foundation (CF16-0444) to the last author.

Abstract

Previous research suggests that judgments about a male speaker's trustworthiness vary due to the speaker's voice pitch (mean *F0*) and differ across domains. Mixed results in terms of the direction and extent of such effects have been reported, however. Moreover, no study so far has investigated whether men's mean *F0* is, indeed, a valid cue to their self-reported and behavioral trustworthiness, and whether trustworthiness judgments are accurate. We tested the relation between mean *F0* and actual general, economic and mating-related trustworthiness in 181 men, as well as trustworthiness judgments of 95 perceivers across all three domains. Analyses show that men's mean *F0* is not related to Honesty-Humility (as a trait indicator of general trustworthiness), trustworthy intentions, or trust game behavior, suggesting no relation of mean *F0* to general or economic trustworthiness. In contrast, results suggest that mean *F0* might be related to mating-related trustworthiness (as indicated by self-reported relationship infidelity). However, lower mean *F0* was judged as more trustworthy in economic, but less trustworthy in mating-related domains and rather weakly related to judgments of general trustworthiness. Trustworthiness judgments were not accurate for general or economic trustworthiness, but exploratory analyses suggest that women might be able to accurately judge men's relationship infidelity based on their voice pitch. Next to these analyses, we report exploratory analyses involving and controlling for additional voice parameters.

Keywords: trustworthiness, voice pitch, social perception, fundamental frequency, Honesty-Humility, F0, trust game, infidelity, cooperation

Linking men‘s voice pitch to actual and perceived trustworthiness across domains

Theoretical Background

Humans are a highly reciprocal species with individuals being dependent on functioning interpersonal relationships (West, El Mouden, & Gardner, 2011). In order to initiate and maintain interpersonal relationships in different domains such as mating or trading, it is important for a person to be judged as trustworthy (e.g. Ewing, Caulfield, Read, & Rhodes, 2015; O'Connor, Re, & Feinberg, 2011). Indeed, current research has linked judged trustworthiness to social outcomes such as corporate hierarchy (Linke, Saribay, & Kleisner, 2016), leadership roles (Chen, Jing, & Lee, 2014; Little, Roberts, Jones, & DeBruine, 2012), and managerial pay awards (Fruhen, Watkins, & Jones, 2015). In a similar vein, research has indicated that individuals who are judged as trustworthy are, in fact, more trusted with money in economic games (Ewing et al., 2015; Van't Wout & Sanfey, 2008).

While being judged as trustworthy can be influenced by various aspects such as discreet and receptive behaviors (e.g., Levin, Whitener, & Cross, 2006) or facial impressions (e.g., Oosterhof & Todorov, 2009), one crucial characteristic that seems to influence whom people tend to trust is the voice. An individual's voice is generally one of the most important providers of information considered when forming social impressions (Belin, Bestelmeyer, Latinus, & Watson, 2011; Tsantani, Belin, Paterson, & McAleer, 2016). Although the human voice is characterized by several aspects, the mean fundamental frequency (mean *F0*), defined as the rate of vocal fold vibrations and judged as voice pitch (Titze, 1994), seems to play a central role in social judgments of human male voices. Mean *F0* is highly sexually dimorphic (Puts, Jones, & DeBruine, 2012; Puts et al., 2016) and has repeatedly been found to influence judgments of men across different domains, such as attractiveness (e.g., Hodges-Simeon, Gaulin, & Puts, 2010; Jünger et al., 2018; Puts et al., 2016), competence (e.g., Klofstad, Anderson, & Nowicki, 2015; Oleszkiewicz, Pisanski, Lachowicz-Tabaczek, & Sorokowska, 2017), and dominance (e.g., Puts et al., 2016; Saxton, Mackey, McCarty, & Neave, 2016;

Vukovic et al., 2011). That is, lower mean *F0* has been linked to higher levels of attractiveness, competence, and dominance. Further, mean *F0* has been linked to important social outcomes, e.g., courtship outcomes (Leongómez et al., 2014; Pisanski, Oleszkiewicz, Plachetka, Gmiterek, & Reby, 2018), mating success (Puts, 2005), or reproductive success (Apicella, Feinberg, & Marlowe, 2007; Rosenfield, Sorokowska, Sorokowski, & Puts, 2019). In line with such findings, it has been argued that mean *F0* was shaped by sexual selection to communicate men's masculinity, dominance, and genetic quality (Puts et al., 2016). Interestingly, beyond mating-related outcomes, mean *F0* does even seem to influence voters' behavior for politicians, in a way that people prefer to vote for candidates with a lower mean *F0* (Klofstad et al., 2015; Tigue, Borak, O'Connor, Schandl, & Feinberg, 2012).

While a relation between lower mean *F0* and higher attractiveness and dominance in men has been found across studies quite consistently (e.g., Hodges-Simeon, Gaulin, & Puts, 2010; Puts, Jones, & DeBruine, 2012; Saxton, Mackey, McCarty, & Neave, 2016), evidence on the relation between mean *F0* and being judged as trustworthy is rather mixed and seems to be affected by the domain and study design. Specifically, the most commonly used study design in investigating the relation between mean *F0* and being judged as trustworthy uses a forced-choice task in which raters listen to two versions of the same recording, one with experimentally raised (feminized) and one with experimentally lowered (masculinized) levels of *F0*. After that, raters usually have to choose between both versions (e.g., "*Which voice did you perceive as more trustworthy*?"). Using this kind of task for judgments of general trustworthiness, two studies found that especially male voices with lowered levels are judged as more trustworthy (Tigue et al., 2012; Tsantani et al., 2016), one study found that male voices with increased levels are judged as more trustworthy (Ponsot, Burred, Belin, & Aucouturiera, 2018), whereas two other studies did not observe any effect (O'Connor & Barclay, 2017; Vukovic et al., 2011). Using manipulated voice recordings but a rating design—i.e., rating each voice file separately for trustworthiness using a Likert scale—,

another study also found that male voices with lowered levels were judged as more trustworthy (Oleszkiewicz et al., 2017). In contrast, a study that implemented a correlational design—i.e., rating non-manipulated voices separately—found that higher pitched voices were judged as more trustworthy (McAleer, Todorov, & Belin, 2014). The same voice recordings were, however, also rated by another group of (non-native) raters (Baus, McAleer, Marcoux, Belin, & Costa, 2019) failing to replicate the finding. Two other correlational studies (Baus et al., 2019; Mahrholz, Belin, & McAleer, 2018) also found no significant relation between mean *F0* and trustworthiness judgments. However, all these correlational studies (Baus et al., 2019, Studies 1 & 2; Mahrholz et al., 2018; McAleer et al., 2014) are based on relatively small sample sizes (*N*s < 33).[1]

Given this inconclusive pattern within and across study designs, recent studies have argued that being judged as trustworthy might be affected by the domain (e.g. mating or economic domain; O'Connor & Barclay, 2017). Specifically, two studies, implementing a forced-choice task, found that in a (hypothetical) trust game participants trusted higher pitched voices over lower pitched voices (Montano, Tigue, Isenstein, Barclay, & Feinberg, 2017; O'Connor & Barclay, 2017), suggesting that higher voice pitch is more trustworthy in economic domains. Further, three studies consistently found that women judge male voices with lower mean *F0* as less trustworthy in relationship domains, meaning that higher infidelity risk was attributed to men with lower levels of mean *F0* (O'Connor & Barclay, 2017; O'Connor, Pisanski, Tigue, Fraccaro, & Feinberg, 2014; O'Connor, Re, & Feinberg, 2011).

[1] McAleer et al. (2014) and Baus et al. (2019) do not report the direct relations between *F0* and trustworthiness ratings. Thus, we contacted the authors and reanalyzed their datasets. Mahrholz et al. (2018) do also not report the correlations but their data is publicly available. Indeed, *F0* was found to be positively correlated with trustworthiness ratings ($r = .51$, $p < .01$) in McAleer et al. (2014). The same voice recordings were, however, also rated by another group of (non-native) raters (Baus et al., 2019). Here, no significant relation was found ($r = .08$, $p = .65$). A second study (Baus et al., 2019) on a different set of voices also showed no significant relation between *F0* and trustworthiness ratings ($r = .30$, $p = .10$). In Mahrholz et al. (2018), relations ranged from $r = -.18$ to $r = .34$ across four different recordings provided by the same 30 speakers. However, none of these relations was significant (*p*s > .069).

Summarizing previous results, the evidence for a link between *F0* and being judged as trustworthy in general is mixed, whereas the evidence for being judged as trustworthy in specific domains (i.e., the economic and the mating domain) rather suggests that higher levels of *F0* are linked to being judged as more trustworthy. However, the latter line of research is strongly dominated by experimental designs in which mean *F0* levels are manipulated in isolation. This is, experimental studies typically focus on manipulating mean F0 levels only while keeping other vocal characteristics stable. Because other vocal characteristics, such as formant position (Pf), do also influence social perceptions (e.g., Puts et al., 2012; Sorokowski et al., 2019) and are related to relevant speaker characteristics (e.g., Pisanski et al., 2016), experimental studies alone are not sufficient to describe and explain a specific vocal parameters' role in vocal perception in natural settings (i.e., outside experiments when vocal parameters can very freely). Thus, especially in order to investigate the relative influence of one vocal parameter, correlational approaches complementing experimental approaches are needed. Next to the inconclusive pattern about the relation between mean *F0* and judged trustworthiness (using different designs) of previous findings, no study so far has investigated whether (inter-)individual differences in men's self-reported or behavioral trustworthiness do, indeed, mirror in differences in mean *F0*. That it, it is currently unknown whether a speakers' mean *F0* can be interpreted as a valid indicator of (un)trustworthy behavior, such as trait trustworthiness, acting reliable in a financial exchange situation, or committing sexual infidelity in a committed relationship. There are two reasons why men's mean *F0* and actual trustworthiness might be related: First, as stated above, some previous research reported that mean *F0* is used as a cue to judge one's trustworthiness, which is only adaptive if the cue is valid for actual behavior. Second, mean *F0* and trustworthiness development in men have been discussed to have the same biological basis in that both are reported to be influenced by testosterone levels (O'Connor & Barclay, 2017).

The present investigation

The present study has three aims:[2] First, we investigate whether mean *F0* is a valid cue of men's trustworthiness. We hypothesize *F0* to be a valid cue of men's trustworthiness in a way that there is a positive relation between *F0* and general trustworthiness. Specifically, we expect mean *F0* to be positively related to the basic personality dimension Honesty-Humility of the HEXACO model (Lee & Ashton, 2004), as this trait has been linked to trustworthy behaviors (Thielmann & Hilbig, 2015; Hypothesis 1). Further, from an exploratory point of view, we link mean *F0* to the recently introduced trustworthy intentions scale (Levine, Bitterly, Cohen, & Schweitzer, 2018) which has also been linked to trustworthy behaviors and aims to measure trait trustworthiness.[3] On the behavioral level, we expect a positive relation between *F0* and economic trustworthiness, as operationalized by trust-game behavior (Thielmann & Hilbig, 2015; Hypothesis 2). We also analyze whether mean *F0* is positively related to mating-related trustworthiness in an exploratory manner.[4]

Second, we investigate the link between mean *F0* and judged trustworthiness. Based on previous research on trustworthiness perception, we expect to replicate a positive relation between mean *F0* and other-rated general trustworthiness (Hypothesis 3; McAleer et al., 2014), mating-related trustworthiness (Hypothesis 4; O'Connor & Barclay, 2017), and economic trustworthiness (Hypothesis 5; Montano et al., 2017).

Third, in an exploratory manner, we investigate whether four other vocal parameters, namely, *F0 CV*, Formant position (*Pf*), *F0* min, and *F0* max, as well as duration do influence

[2] Please note that this study focusses solely on men's trustworthiness, as voice pitch is highly sexually dimorphic and hypotheses for relationships between men's voices and trustworthiness cannot easily be translated to women. Further, stating hypotheses for this relationship in women is not trivial, as previous studies rather focused on men, and because the relationship between voice parameters and social outcomes (e.g. attractiveness) seems to be less consistent in women than in men. However, we think that the relationship between voice and trustworthiness is equally relevant across sexes and encourage future studies to investigate this relationship in women.

[3] We became aware of this scale shortly after the data collection was started. Deviating from our preregistration, we decided to include it for exploratory analyses. The scale was filled in by 166 participants (out of 181).

[4] We did not preregister a hypothesis on the relation between *F0* and mating-related trustworthiness, because we expected that a minority of our target participants would report ever having cheated in a committed relationship. Thus, analyses for this hypothesis might be underpowered.

trustworthiness perceptions and/or are valid indicators of self-reported or behavioral trustworthiness. Further, as these vocal parameters have been found to influence social perceptions of various traits (e.g., Pisanski et al., 2018; Puts et al., 2012; Sorokowski et al., 2019), including these vocal parameters in extensive analyses allows to strengthen conclusions regarding the role of mean *F0* in trustworthiness domains.

Finally, as we measured both Honesty-Humility and trust game behavior, we expect to replicate the finding from Thielmann and Hilbig (2015; Study 2) that Honesty-Humility and the average return in an economic trust game are positively related (Hypothesis 6).

Methods

This study has been pre-registered online at the Open Science Framework (OSF; https://osf.io/khua4/), before any data were collected. The raw data, analyses script, administered questionnaires, and the instruction material are also provided in the OSF folder alongside this manuscript.

Participants and procedure

Two groups of subjects participated: Targets and raters of targets' voice recordings. All participants were recruited at the University of Göttingen and signed a written consent form. Being part of a research project on *F0* and personality in general, this study was approved by the Institutional Review Board of the University of Copenhagen.

Targets

The sample size was based on a power analysis using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007). For a statistical power of 80% to detect a small-to-medium effect size (r = .20) for the relation between mean *F0* and trustworthiness (with alpha = .05), the required sample size was N = 153, which we, following the preregistration, slightly oversampled. Out of 186 people recruited via a local participant pool, a total of 181 men (aged 18-56 years, M = 25.45, SD = 6.17) participated as targets. One participant dropped out because of not wanting

his voice recording being presented and rated, two participants dropped out because of not filling out the self-report personality questionnaires, and two participants dropped out because of a programming error on our end. Target participants first provided demographic information, then filled out different personality questionnaires administered in paper-pencil format, including the German version of the HEXACO-60 (Moshagen, Hilbig, & Zettler, 2014), a self-translated German version of the trustworthy intentions scale (Levine et al., 2018), two self-designed items assessing mating-related trustworthiness, and the revised Sociosexual Orientation Inventory (SOI-R; Penke & Asendorpf, 2008) that is not part of the current study. Next, target participants completed an economic trust game, being in the role of the trustee (as in Berg, Dickhaut, & McCabe, 1995; Thielmann & Hilbig, 2015). Finally, target participants' voices were recorded via a studio microphone (Rode® NT 1-A; the distance from the microphone was approx. 5-10cm) and the software Audacity® at a sampling rate of 44.1 kHz and 32-bit amplitude quantization. Files were saved as uncompressed WAV files. Participants spoke a German translation of the rainbow passage (Fairbanks, 1960). The translation aimed to reflect the original version content-wise while maintaining a phonetically balanced structure that reflects the German language. The full translation is available in the study material on the OSF. Target participants were thanked and rewarded with 3-12€, depending on their behavior in the economic trust game.

Raters

The sample size was based on previous studies showing very high levels of interrater agreements (alphas > .80, *ps* <.001) for voice ratings with 15 raters (Kordsmeyer, Hunt, Puts, Ostner, & Penke, 2018). Further, simulations sampling from a population of 2,513 raters, each of whom had rated the attractiveness of 102 faces, indicate that > 99% of 1,000 random samples of 15 raters produce Cronbach's alphas > .8 (DeBruine, & Jones, 2018). In a similar vein, Hehman et al. (2018) report that increasing the number of raters providing trustworthiness ratings for faces has a negligible effect on the ratings once ratings have been

collected from 26 raters. Thus, we aimed to recruit 15 male and 15 female raters per rating condition[5] (see below). Finally, ninety-five people (46 male, 49 female, aged 18 to 60 years, *M* = 24.91, *SD* = 6.80) participated in the following rating procedure:

First, all rater participants read the instructions and provided demographic information about themselves. Then, the voice recordings (of target participants) were presented via headphones (JVC® HA-RX300) in a randomized order using the open source framework Alfred (Treffenstaedt & Wiemann, 2018), which is based on the programming language Python (version 2.7, www.python.org). Because of time constraints, only one sentence of the recorded passage was presented (‘*Der Regenbogen ist eine Teilung von weißem Licht in viele schöne Farben’ ~ ‘The rainbow is a division of white light into many beautiful colors’*; Fairbanks, 1960). After hearing one voice (only once, in full length), raters were asked to rate the speaker’s trustworthiness in one out of three conditions: 1) general trustworthiness, 2) mating-related trustworthiness, or 3) economic trustworthiness. Raters were randomly assigned to one of the three conditions, resulting in *n* = 31 (15 male, 16 female) for general trustworthiness; *n* = 33 (16 male, 17 female) for mating-related trustworthiness, and *n* = 31 (15 male, 16 female) for economic trustworthiness. Detailed information about each rating condition are given in the “Ratings on Targets Trustworthiness” section. After completing the rating task, rater participants were thanked and rewarded with 5€ or course credit.

Measures

Target participants’ vocal parameters

Voice recordings were analyzed using Praat software (version 6.0.49; Boersma & Weenink, 2019). Four parameters of fundamental frequency, namely, mean (mean *F0*; *M* = 121.22, range = 87.05–173.84; *SD* = 16.46), minimum (*F0* min; *M* = 81.17, range = 67.07–117.94; *SD*

[5] Importantly, an adequate number of raters is not influenced by the number of voices rated (except for cases in which there are several different sets of stimuli, in which only parts of the stimuli is rated by randomly assigned raters, or in which rater fatigue can be expected). Power analyses have to be done for both raters and stimuli separately. In the best case, researchers recruit a large number of raters to ensure accurate assessments of stimuli as well as a large number of stimuli to ensure variance in the stimuli material and thus generalizability of findings.

= 9.25), maximum (*F0* max; *M* = 190.87, range = 120.99 – 302.35; *SD* = 44.74) and standard deviation (*F0* s.d., *M* = 19.94, range = 7.50 – 38.90; *SD* = 6.50), as well as the duration of the recording (in seconds; *M* = 4.12, range = 2.89–5.89; *SD* = 0.49) were derived via the "voice report" function in Praat, using standard settings for male voices (pitch floor and ceiling were 75 Hz and 300 Hz, in accordance with programmers' recommendations; otherwise default settings were used). Using *F0* s.d. and mean *F0*, the coefficient of variation (*F0 CV*; given by F0 s.d./F0 mean; *M* = 0.16, range = 0.07 – 0.29; *SD* = 0.04) was calculated. Lastly, formants (f1, f2, f3, and f4) were measured at each glottal pulse and averaged across measurements. f1, f2, f3, and f4 were used to calculate standardized formant position (*Pf*; following Puts, Apicella, & Cárdenas, 2012, we calculated this by standardizing each formant measure and dividing the sum of these four measures by four; *M* = 0, range = -1.42–1.87; *SD* = 0.58).

Perceptually, mean F0 represents the average voice pitch, while *F0* min and *F0* max represent its lower and upper ranges. *F0* s.d. and *F0 CV* represent measures of pitch variability with *F0 CV* being the more reliable indicator of perceptual salience of F0 variability (see Pisanski et al., 2018). Thus, *F0 CV* is used in the further analyses. *Pf* is a commonly used measure of formant structure which reflects vocal tract resonances. All vocal parameters mentioned above as well as the duration of the recording have been shown to influence different social perceptions (e.g., Pisanski et al., 2018; Puts et al., 2012; Sorokowski et al., 2019).

Importantly, mean *F0* measurements of the whole recording and the rated segment correlated very highly ($r = .98$, $p < .001$). Consequently, the mean *F0* measurements of the rated segment were used for the main analyses. Importantly, using the mean *F0* measurement of the whole recording did not yield different statistical conclusions. While Pf measurements ($r = .92$, $p < .001$) and *F0 CV* measurements ($r = .73$, $p < .001$) were also highly correlated, *F0* min measurements ($r = .19$, $p = .009$) and *F0* max measurements ($r = .34$, $p < .001$) were only moderately related. Thus, exploratory analyses including these parameters as predictors

of speaker characteristics were run twice, once with the measurements of the rated segment and once with the measurements of the whole recording.

Target participants' actual trustworthiness

Several measures were used to assess target participants' actual trustworthiness, including both personality questionnaires and actual behavior (King, 2010) in a behavioral task. To start with, target participants filled out the German (Moshagen et al., 2014) Honesty-Humility items from the 60-item version of the HEXACO Personality Inventory-Revised (HEXACO-60; Ashton & Lee, 2009). Sample items for Honesty–Humility are "I wouldn't pretend to like someone just to get that person to do favors for me." or "I would never accept a bribe, even if it were very large". Further, a self-translated German version of the trustworthy intentions scale (Levine et al., 2018) was administered. Sample items are "If I borrowed something of value and returned it broken, I would offer to pay for the repairs." or "If I decided to meet someone for lunch, I would definitely be there." Responses on both questionnaires were given on a 5-point Likert Scale ranging from '*strongly disagree*' to '*strongly agree*', and mean factor scores for each participant were computed (by averaging the Likert Scale responses for the ten Honesty-Humility and eight trustworthy intentions items, respectively). Mating-related trustworthiness was assessed with a self-designed item asking if the participant has ever cheated sexually in a committed relationship ('*Have you ever been sexually unfaithful in an exclusive, committed relationship*?' – '*yes*' or '*no*'). Another self-designed item asked how long the participants have been in committed relationships (in months, total sum score). This item was needed to exclude participants who have not been in a relationship before and consequently did not have the chance to cheat in a committed relationship. Further, it was needed as a control variable, since the longer participants have been in committed relationships, the more time they had to cheat.

Economic trustworthiness was assessed via a trust game (Thielmann & Hilbig, 2015). Each participant (in the role of the trustee) was randomly matched to an unknown trustor from

a previous study, so that no deception was involved. More precisely, the trustees (participants) were informed that the trustor (a randomly assigned participant of a previous study) had decided how much of an endowment of 3 € (in 0.50 € increments) they want to transfer to the participant. The transfer was then tripled accordingly (i.e., max. 9 €). The task of the participants was then to indicate how much of the (tripled) transfer they wanted to return to the trustor (e.g., "Assuming Player 1 transferred 3.00€ of his/her 3.00€ to you. You thus receive 9.00€. How much (between 0 and 9.00€) would you like to return to Player 1?"). As in Thielmann and Hilbig (2015), we used the strategy method, in which the participants were not aware of the trustor's actual transfer and had to specify their return for each of the six potential (tripled) amounts the trustor could transfer (i.e., between 1.5€ and 9€, in 0.50€ increments). Thus, the outcome variable for economic trustworthiness was the average return (in %). Importantly, this game was not hypothetical, such that the participants were actually incentivized with the amount of money they decided to keep for themselves in one randomly chosen scenario (i.e., depending on what the matched participant decided) in addition to a flat fee of 3€.

Ratings on targets trustworthiness

Rating questions differed between the three rating conditions: 1) general trustworthiness ('*How trustworthy does the speaker sound?*'), 2) mating-related trustworthiness ('*How likely is it that the speaker would cheat on his spouse/girlfriend in a committed, romantic relationship?*'), and 3) economic trustworthiness ('*Imagine that you could trust this person with some money. By doing this, you could either earn some additional money (i.e. you would earn some money by „investing" or the person could keep parts of the money and cause a monetary loss for you. How likely is it that you trust this person with money?*'). Ratings were provided on seven-point Likert scales ranging from -3 ('*very untrustworthy*') to +3 ('*very trustworthy*') in condition 1), and from -3 ('*very unlikely*') to +3 ('*very likely*') in conditions 2) and 3). Further, to be consistent across conditions, ratings in

condition 2) were reverse coded. Interrater agreement was high for general (α = .88) and for economic (α = .87) trustworthiness, and ratings between male and female raters correlated highly (r = .76, p < .001, for general trustworthiness, r = .77, p = < .001 for economic trustworthiness). Ratings of all raters were aggregated for these conditions. Interrater agreement (α = .75) was moderate for mating-related trustworthiness and ratings for mating-related trustworthiness did correlate substantially between the two sexes (r = .69, p <.001). However, as we preregistered to average ratings by men and women only if the correlation between their ratings is higher than .70, mating-related trustworthiness ratings were averaged for both sexes separately (with α = .52 for male raters; α = .62 for female raters) and statistical analyses for this condition were thus conducted and reported separately.

Statistical analyses

Whenever directed hypothesis were tested, one-tailed tests were used.[6] Zero-order correlations of the continuous variables are presented in Table 1. All analyses in the current manuscript were calculated with the statistic software R 3.5.1 (R Core Team, 2018).

Results

Mean F0 and target trustworthiness

First, we linked mean *F0* to our four measures of target trustworthiness. Here, mean *F0* was not found to be related to Honesty-Humility (r = .07, p = .184) or trustworthy intentions (r = .08, p = .144). Further, mean *F0* was not found to be related to the average return in the trust game (r = –.04, p = .681). Therefore, Hypotheses 1 and 2 were not supported, suggesting that mean *F0* is not related to general or economic trustworthiness. In contrast, in an exploratory logistic regression investigating the relation between mean *F0* and mating-related

[6] We preregistered to use one-tailed tests in the first preregistration (on the relationship between *F0* and target participants trustworthiness) and did our power analyses based on one-tailed tests. However, we preregistered to use two-tailed tests in the second preregistration (on the relationship between *F0* and rater participants' ratings). To stay concise, we decided to report one-tailed tests for all directed hypotheses here. Importantly, all analyses were repeated using two-tailed tests and conclusions did not change substantially. Details can be found in the supplement.

trustworthiness, mean *F0* did negatively (OR[7] = 0.50, p = .013) predict self-reported infidelity in a committed relationship, whereas relationship length did not predict self-reported infidelity (OR = 1.41, p = .078). For this and upcoming analyses including self-reported infidelity, we excluded participants who reported that they have never been in a committed relationship (N = 21) and consequently did not have the chance to actually cheat. Further, we controlled for relationship length because there was strong variance within the measure (M = 47.11 months, SD = 39.91) and participants that have been in relationships longer arguably have had more opportunities to cheat. The relation between mean *F0* and self-reported infidelity in a committed relationship is illustrated in Figure 1.

----Figure 1----

Figure 1. Relations between F0 and self-reported infidelity. Vertical dashed lines represent mean F0 means of cheaters (red) and non-cheaters (blue).

Mean F0 and judged trustworthiness

Second, we linked mean *F0* to judgments of trustworthiness. Here, mean *F0* was not found to be related to judgments of general trustworthiness (r = –.09, p = .889). In contrast, mean *F0* was positively related to judgments of mating-related trustworthiness for both male (r = .37, p < .001) and female (r = .24, p < .001) raters. Specifically, men and women rated higher pitched voices as less likely to cheat in a committed romantic relationship. Therefore, no support was found for Hypothesis 3, whereas Hypothesis 4 was supported. Lastly, mean *F0* was negatively related to economic trustworthiness (r = –.32, p < .001). Specifically, for Hypothesis 5, a strong effect in the opposite direction was found, in that men with lower mean *F0* were rated as being more likely to be trusted with money. Relations between judgments and mean *F0* are illustrated in Figure 2. Replicating findings by O'Connor and Barclay (2017), judgments of general trustworthiness were strongly related to judgments of economic

[7] Because scaled predictors were used, ORs refer to log odds ratios for one SD change of the specific predictor.

trustworthiness (r =.87, p < .001). In contrast, judgments of general trustworthiness and economic trustworthiness were not related to judgments of mating-related trustworthiness by male (r = .08, p = .287 for general trustworthiness, r = -.02, p = .762 for economic trustworthiness) and female raters (r = .07, p = .367 for general trustworthiness, r = .05, p = .545 for economic trustworthiness).

----Figure 2----

Figure 2. Correlations between *F0 (in Hz)* and judgments of General Trustworthiness (A), Mating-related Trustworthiness by female (B) and male raters (C), and Economic Trustworthiness (D)

After inspection of the scatterplot of the relation between mean *F0* and mating-related trustworthiness, we decided to run exploratory multiple regressions including the quadratic term of mean *F0* to predict both ratings by male and female raters. Curvilinear models described the relation between mean *F0* and judged mating-related trustworthiness better than linear models for both male ($F_{2,178}$ = 6.59, p = .011) and female ratings ($F_{2,178}$= 28.62, p < .001). No curvilinear relations between mean *F0* and any other ratings or any self-reported or behavioral trustworthiness measures were found. Details are reported in the supplement.

Judged trustworthiness and target trustworthiness

We further investigated how trustworthiness judgments were related to actual target trustworthiness in an exploratory manner: Judgments of general trustworthiness were not related to Honesty-Humility or trustworthy intentions (rs ≤ .12, ps ≥ .120). Further, judgments of economic trustworthiness were not related to trust game behavior (r = –.03, p = .657). In contrast, a logistic regression showed that judgments of mating-related trustworthiness by female raters (OR = 0.63, p =.049) and relationship length (OR = 1.52, p =.035) predicted self-reported infidelity in a committed relationship. Another logistic regression model, including judgments by male raters (OR = 0.72, p =.155) and relationship length (OR = 1.49,

p = .042), did only show a significant effect for the latter, however. These results indicate that women, rather than men, may be able to validly judge relationship infidelity based on men's mean *F0*.

Honesty-Humility and target trustworthiness

Finally, replicating a finding by Thielmann and Hilbig (2015), a significant positive relation between Honesty-Humility and the average return in the trust game was found (r = .23, p = .001). Additionally, when Honesty-Humility and trustworthy intentions were entered into a multiple regression to predict the average return in the trust game ($F_{2,160}$= 4.82, p = .009, R^2 = .05), Honesty-Humility positively predicted the average return in the trust game (β = .22, p = .007), but trustworthy intentions did not (β = .03, p = .683). Hypothesis 6 was thus supported. An additional logistic regression showed that relationship length (*OR* = 1.63, p =.024), Honesty-Humility (*OR* = 0.55, p =.032) and trustworthy intentions (*OR* = 0.58, p =.025) were significant predictors of self-reported infidelity in a committed relationship.

----Table 1----

Exploratory analyses including further vocal parameters

We ran seven multiple linear regressions and one multiple logistic regression which, next to mean *F0*, included *F0* min, *F0 max*, *Pf*, as well as duration as predictors of both judged and target trustworthiness. While *F0* and formant measures are theoretically and practically independent (Titze, 1994), other vocal parameters can be expected to covary. We checked the variance inflation factors (VIF's) in each model to detect multicollinearity. However, in none of our models VIF's higher than the commonly used threshold of 5 were observed. Thus, no predictors were dropped.

----Figure 3----

Figure 3. Regression coefficient plots of multiple regression with vocal parameters as predictors and trustworthiness ratings as outcomes. Each color reflects a different model. Thick lines reflect .90 CIs, thin lines reflect .95 CIs.

Overviews of the results of the multiple linear regressions predicting judged and target trustworthiness are illustrated in Figures 3 and 4. Corresponding tables, including exact *p* values, can be found in the supplementary material. In summary, multiple other vocal parameters were significantly related to judgements of trustworthiness. *F0* min was positively linked to judged mating-related trustworthiness (both by male and female raters), while *F0* CV was positively related to judged general and economic trustworthiness. This means that speakers with lower minimum pitch were judged as less trustworthy in mating contexts, while more monotonic voices were judged as less trustworthy in general and economic contexts. Recording duration, an indirect indicator of speech rate, was negatively related to general and economic trustworthiness but positively related to mating-related trustworthiness. This means that speakers with faster speech rate were judged as more trustworthy in mating-related contexts, but less trustworthy in general and economic trustworthiness. *F0* max and *Pf,* however, were not found to be related to any judgements. Mean *F0,* in line with previous results, was related to judged economic and mating-related trustworthiness. In contrast to our main analyses, however, mean *F0* was also negatively, albeit weakly, linked to judged general trustworthiness. Interestingly, mean *F0* was no significant predictor of mating-related trustworthiness when other vocal parameters were entered as predictors into the linear regression. However, in line with previous results, entering the quadratic term of mean *F0* into the model led to a highly significant R2 increase ($F_{2,173}$= 22.63, $p < .001$).

No vocal parameters were significantly linked to Honesty-Humility, Trustworthy Intentions, or Trust game behavior as illustrated in Figure 3B. The same pattern was found in a multiple logistic regression predicting self-reported infidelity (all *p*s >.107). Thus, in

contrast to previous analyses, mean *F0* was no significant predictor of self-reported relationship infidelity (*OR* = 0.50, *p* = .107) when other vocal parameters were entered into the logistic regression simultaneously. Please note that our study was underpowered for the multiple linear and logistic regressions reported here and thus results should be interpreted with caution.

----Figure 4----

Figure 4. Regression coefficient plots of multiple regression with vocal parameters as predictors and target trustworthiness as outcomes. Each color reflects a different model. Thick lines reflect .90 CIs, thin lines reflect .95 CIs.

Discussion

This study investigated whether mean *F0* is a valid cue to men's trustworthiness and whether men's mean *F0* influences judged trustworthiness. Results suggest that, other than hypothesized, mean *F0* is not a valid cue to general or economic trustworthiness based on both self-report questionnaires as well as actual behavior, whereas exploratory analyses suggest that mean *F0* may be related to mating-related trustworthiness (measured via one self-report item). Moreover, analyses of trustworthiness judgments showed that mean *F0* is not substantially related to judgments of general trustworthiness, but that men with lower pitched voices are judged as being less trustworthy in committed relationships and more trustworthy in economic domains. Further, judged general trustworthiness was unrelated to actual general trustworthiness as measured via questionnaires (i.e., self-reported Honesty-Humility and trustworthy intentions), and judged economic trustworthiness was unrelated to actual trust game behavior. Interestingly, judged mating-related trustworthiness was related to self-reported infidelity in committed relationships, but only when rated by women, not by men. Other vocal parameters, F0 min and *F0 CV*, were found to influence trustworthiness perceptions above mean *F0* but did not provide valid information about a speaker's

trustworthiness. Finally, Honesty-Humility predicted actual behavior in the economic trust game (Thielmann & Hilbig, 2015) and infidelity in committed romantic relationships (Hilbig, Moshagen, & Zettler, 2015). In the following, we interpret these findings and highlight implications for future research.

Is F0 a valid cue to men's trustworthiness?

Previous research stated that mean *F0* might be a valid cue to men's trustworthiness, because of both variables having the same biological foundations: The development of men's vocal parameters (e.g., mean *F0*) and trustworthy behaviors have been discussed to be associated with testosterone levels in men (O'Connor & Barclay, 2017). Evidence for the latter is mostly based on studies linking mean *F0* or testosterone to mating behavior in that men with higher levels of testosterone and men with lower voice pitch report a higher number of (extra-pair) sex partners (Booth & Dabbs, 1993; Fisher et al., 2012, 2009; Puts, 2005). Two other studies suggest that testosterone levels are also related to offer acceptance and return rate in an economic trust game (Burnham, 2007; Takagishi, Takahashi, & Yamagishi, 2011).

However, this is the first study investigating more straightforwardly whether mean *F0* is, indeed, a valid cue to men's trustworthiness. Contrary to the hypotheses, we did not observe a significant relation between mean *F0* and general trustworthiness as measured by (questionnaire-based) Honesty-Humility or trustworthy intentions, or between mean *F0* and economic trustworthiness as measured by actual behavior in a trust game. These results indicate that mean *F0* is not a valid cue to men's general or economic trustworthiness. Interestingly, exploratory analyses suggest that mean *F0* might be related to men's mating-related trustworthiness, and, thus, a valid cue to infidelity in committed relationships. This finding fits to the line of research relating testosterone and mean *F0* to relationship infidelity and to the number of sex partners. Further, as low mean *F0* is reported to be perceived as more dominant and attractive (e.g., Hill et al., 2013; Jünger et al., 2018; Puts et al., 2016) men

with low mean *F0* are likely to have more mating opportunities and thus also have more opportunities to cheat. Nevertheless, we want to emphasize that this finding is exploratory. Given that we did not have a high number of participants in our sample who actually reported having cheated in a previous relationship (n = 25 out of N = 181), this finding should be interpreted with caution. Further, our measure of relationship infidelity did only ask whether cheating had occurred. A more precise attempt would have been to ask how often and in how many relationships infidelity was committed. These limitations imply that the relationship between mean *F0* and relationship infidelity should be replicated in a well powered study with a higher number of target participants who committed relationship infidelity before.

F0 and trustworthiness judgments

The second reason why mean *F0* was proposed to be a valid cue to men's trustworthiness is that previous research reported that mean *F0* is used to form trustworthiness judgments (McAleer et al., 2014; Oleszkiewicz et al., 2017; Tigue et al., 2012; Tsantani et al., 2016), which would only be adaptive if the cue leads to valid judgments. In contrast to this idea, we did only find weak evidence that suggests that listeners use mean *F0* to judge speakers' general trustworthiness. The mixed pattern of the results from other studies and our research might be explained by different methods. In most previous studies, participants had to choose which out of two experimentally manipulated versions of the same voice sounds more trustworthy (O'Connor & Barclay, 2017; Tigue et al., 2012; Tsantani et al., 2016; Vukovic et al., 2011) or rate the trustworthiness of manipulated stimuli (Oleszkiewicz et al., 2017). In these studies, mean *F0* was found to be either negatively related or unrelated to trustworthiness judgments. However, when experimentally manipulating the pitch of voices, the strength of pitch manipulations differed (voices were raised or lowered by different amounts of Hz), which might explain differences in results (e.g. increased likelihood to detect a significant effect with stronger manipulation). However, three previous studies implemented a correlational design (a rating design using non-manipulated stimuli; Baus et al., 2019,

Mahrholz et al., 2018; McAleer et al., 2014). Although it was not the authors' aim to investigating trustworthiness judgments based on mean *F0*, their data suggests a rather positive relation between both variables. Yet, their sample sizes of target participants were comparably small ($N < 33$ males vs. $N = 181$ in our sample).

Mixed results in previous studies have also been explained by domain-specific effects of trustworthiness judgments, in that mean *F0* is not used to form judgments on general trustworthiness, but on economic or mating-related trustworthiness (O'Connor & Barclay, 2017). Our results support this claim. Concerning economic-related trustworthiness, we observed, however, a negative relation between mean *F0* and economic trustworthiness, in contrast to previous findings (Montano et al., 2017; O'Connor & Barclay, 2017). Participants in our study were actually more likely to trust a speaker with lower *F0* with money, although it does not seem to be a valid cue.

Our results might, again, differ from those of previous studies because of differences in design (rating vs. choosing) and stimuli manipulation (natural stimuli vs. experimentally manipulated stimuli). Moreover, the scenario that we introduced differed importantly from those used previously: Participants decided whether they would invest in a person, resulting in losses, no changes, or benefits. In contrast, previous studies rather introduced a trust game where equal or unequal distributions of money were possible (Montano et al., 2017; O'Connor & Barclay, 2017), but not really gains and losses. As we framed the economic-trustworthiness judgment as an investment decision, it seems plausible that participants were more likely to invest in men with lower mean *F0*, as previous research suggests that these men are judged as being more trustworthy leaders (Klofstad, Anderson, & Peters, 2012), more dominant (e.g., Borkowska & Pawlowski, 2011; Collins, 2000; Hodges-Simeon, Gaulin, & Puts, 2010; Puts, Gaulin, & Verdolini, 2006; Puts, Hodges, Cárdenas, & Gaulin, 2007), and having a higher ability to acquire resources (Smith, Olkhov, Puts, & Apicella, 2017). Thus, different study designs and introduced economic scenarios might lead to inconsistent results.

As opposed to the finding that lower mean *F0* was related to higher judged economic trustworthiness, women judged men with lower mean *F0* as less trustworthy in committed relationships, replicating previous findings on infidelity judgments based on mean *F0* (O'Connor & Barclay, 2017; O'Connor et al., 2014, 2011). However, additional analyses revealed that the relation between mean *F0* and mating-related trustworthiness might also be curvilinear, with very high and very low mean *F0* being judged as less trustworthy. While we cannot rule out that this finding could be a false positive, it leads to implications for future research to further investigate possible curvilinear relations.

For the moment, our exploratory analyses suggest that men's mean *F0* might, indeed, be a valid cue to relationship infidelity, and it seems adaptive that participants use it to make judgments about relationship infidelity. Sexual infidelity can be costly to the in-pair woman (because of the man investing in other women and the potential loss of a partner), as well as to the extra-pair man (as a potential intrasexual competitor). However, lower mean *F0* has also been reported to be judged as more attractive (e.g., Hodges-Simeon et al.,2010; Jünger et al., 2018; Puts et al., 2016), and associated with higher mating success (Puts, 2005). Thus, the relationship between mean *F0* and infidelity judgments might be mediated by attractiveness judgments (or judgments of accessibility to infidelity), which should be investigated in future research.

Judgments and their accuracy

Is a man's general, economic, or mating-related trustworthiness accurately perceivable from his voice? Our results suggest that judgments of general and economic trustworthiness are not accurate, indicating that this vocal cue was not selected to evolve for signaling these characteristics. Nevertheless, in line with previous research (Hughes & Harrison, 2017), there is some evidence that men's relationship infidelity is accurately perceivable, but probably only by women, not other men. It might be especially adaptive for women to be able to validly judge a men's mating-related trustworthiness, because infidelity may cause high

fitness costs, such as the loss of protection and provisioning (Geary, Vigil, & Byrd-Craven, 2004) as well as parental and relationship investment (O'Connor et al., 2011). However, men did descriptively also use mean *F0* to judge relationship infidelity (*OR* = 0.72). Accordingly, in another study, male and female listeners both judged more masculine voices as less trustworthy in mating-related domains (O'Connor & Barclay, 2017), arguing that relationship infidelity might also cause fitness costs for other men. Since our sample had a rather low percentage of men who indicated actually having committed relationship infidelity before, future research should conduct well-powered replication studies, ideally with a higher number of target men who have cheated sexually in a committed relationship.

Other vocal parameters and actual and perceived trustworthiness

Given our correlational design, other vocal parameters did also vary freely. Since other vocal parameters, namely, *F0* min, *F0* max, *F0* CV, *Pf*, as well as recording duration have been shown to influence a broad set of social evaluations (e.g., Pisanski et al., 2018; Puts et al., 2012; Sorokowski et al., 2019), we included their measurements in further exploratory analyses. This served two main purposes. First, controlling for other vocal parameters allows to further define the exact role of mean *F0* in trustworthiness contexts. Second, because other vocal parameters are arguably understudied in trustworthiness contexts, this allowed us to explore whether they explain variance above mean *F0* and are valid indicators of a speaker's trustworthiness. Here, *F0* CV, F0 min, and duration—but not *Pf* and F0 max— were indeed found to be related to different trustworthiness perceptions. However, none of these parameters did explain variance in regard to any measure of target trustworthiness. Future studies might build upon these exploratory findings to broaden our understanding of voice based trustworthiness perceptions.

Limitations

There are five limitations of our study that we would like to mention. First, we did only investigate actual trustworthiness and its perception from voices in men. It remains an open question whether results would differ for women. Second, it might be that perceived and actual economic trustworthiness were not measuring the same quality. This is, perceived economic trustworthiness might not only have been based on trustworthiness but also other characteristics such as dominance or competence. Third, we used a context-free, standardized sentence to investigate the relation between voice and trustworthiness, whereas a form of speech in that participants somehow address trustworthiness (e.g. an offer to invest money) or nonverbal vocalizations (Raine et al., 2019) related to trustworthiness might have led to different results (but see Mahrholz et al., 2018, who found that judgments on different speech contents are highly correlated). Thus, we encourage future studies to replicate our findings using different speech content (both standardized and free speech) that is more directly related to trustworthiness. Fourth, as there is no measure to objectively assess sexual infidelity, we used self-reported infidelity to assess mating-related trustworthiness. Clearly, this self-report measure bears the risk of being biased (e.g., some participants might have not responded to this truthfully because of social desirability). However, we tried to address this concern by administering completely anonymous surveys as well as by presenting the corresponding item in between the items of the SOI-R to make this question sound more natural. Fifth, related to this, our item asked for sexual unfaithfulness in a committed relationship which might depend on personal boundaries and definitions of sexual unfaithfulness in a committed relationship. That is, some might consider a kiss as sexual unfaithfulness whereas others would refer to actual sexual intercourse only. Thus, future studies should be more clear about what is meant by sexual infidelity.

Conclusion

In summary, our findings indicate that mean *F0* is not a valid cue to men's general or economic trustworthiness, but possibly to men's mating-related trustworthiness. Mean *F0* is

not substantially used as a cue to judge men's general trustworthiness, but for forming judgments on men's mating-related and economic trustworthiness. In addition, other vocal characteristics, namely, *F0* CV and F0 min, do explain variance in trustworthiness judgements above mean *F0*. Raters were not able to accurately judge trustworthiness from voice recordings, in spite of women perceiving sexual infidelity. From a functional perspective, women who can accurately judge sexual infidelity from a man's voice at zero acquaintance can benefit from this accuracy by rather selecting a trustworthy partner for a committed relationship. Future studies should test the robustness of this finding and also investigate women as target participants.

Data Accessibility

Analyses reported in this article can be reproduced using the data provided by Schild, Stern, and Zettler (2019). Additionally, the raw data, analyses script, administered questionnaires, and the instruction material are also provided in the OSF folder alongside this manuscript (https://osf.io/khua4/).

References

Apicella, C. L., Feinberg, D. R., & Marlowe, F. W. (2007). Voice pitch predicts reproductive success in male hunter-gatherers. *Biology Letters*, *3*(6), 682–684. https://doi.org/10.1098/rsbl.2007.0410

Ashton, M. C., & Lee, K. (2009). The HEXACO–60: A Short Measure of the Major Dimensions of Personality. *Journal of Personality Assessment*, *91*(4), 340–345. https://doi.org/10.1080/00223890902935878

Baus, C., McAleer, P., Marcoux, K., Belin, P., & Costa, A. (2019). Forming social impressions from voices in native and foreign languages. *Scientific Reports*, *9*(1), 414. https://doi.org/10.1038/s41598-018-36518-6

Belin, P., Bestelmeyer, P. E. G., Latinus, M., & Watson, R. (2011). Understanding Voice Perception. *British Journal of Psychology*, *102*(4), 711–725. https://doi.org/10.1111/j.2044-8295.2011.02041.x

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, Reciprocity, and Social History. *Games and Economic Behavior*, *10*(1), 122–142. https://doi.org/10.1006/game.1995.1027

Booth, A., & Dabbs, J. M. (1993). Testosterone and Men's Marriages. *Social Forces*, *72*(2), 463–477. https://doi.org/10.1093/sf/72.2.463

Borkowska, B., & Pawlowski, B. (2011). Female voice frequency in the context of dominance and attractiveness perception. *Animal Behaviour*, *82*(1), 55–59. https://doi.org/10.1016/j.anbehav.2011.03.024

Burnham, T. C. (2007). High-testosterone men reject low ultimatum game offers. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1623), 2327–2330. https://doi.org/10.1098/rspb.2007.0546

Chen, F. F., Jing, Y., & Lee, J. M. (2014). The looks of a leader: Competent and trustworthy, but not dominant. *Journal of Experimental Social Psychology*, *51*, 27–33. https://doi.org/10.1016/j.jesp.2013.10.008

Collins, S. A. (2000). Men's voices and women's choices. *Animal Behaviour*, *60*(6), 773–780. https://doi.org/10.1006/anbe.2000.1523

Ewing, L., Caulfield, F., Read, A., & Rhodes, G. (2015). Appearance-based trust behaviour is reduced in children with autism spectrum disorder. *Autism*, *19*(8), 1002–1009. https://doi.org/10.1177/1362361314559431

Fisher, A. D., Bandini, E., Rastrelli, G., Corona, G., Monami, M., Mannucci, E., & Maggi, M. (2012). Sexual and Cardiovascular Correlates of Male Unfaithfulness. *The Journal of Sexual Medicine*, *9*(6), 1508–1518. https://doi.org/10.1111/j.1743-6109.2012.02722.x

Fisher, A. D., Corona, G., Bandini, E., Mannucci, E., Lotti, F., Boddi, V., … Maggi, M. (2009). Psychobiological Correlates of Extramarital Affairs and Differences between Stable and Occasional Infidelity among Men with Sexual Dysfunctions. *The Journal of Sexual Medicine*, *6*(3), 866–875. https://doi.org/10.1111/j.1743-6109.2008.01140.x

Fruhen, L. S., Watkins, C. D., & Jones, B. C. (2015). Perceptions of facial dominance, trustworthiness and attractiveness predict managerial pay awards in experimental tasks. *The Leadership Quarterly*, *26*(6), 1005–1016. https://doi.org/10.1016/j.leaqua.2015.07.001

Geary, D. C., Vigil, J., & Byrd-Craven, J. (2004). Evolution of human mate choice. *The Journal of Sex Research*, *41*(1), 27–42. https://doi.org/10.1080/00224490409552211

Hilbig, B. E., Moshagen, M., & Zettler, I. (2015). Truth Will Out: Linking Personality, Morality, and Honesty Through Indirect Questioning. *Social Psychological and Personality Science*, *6*(2), 140–147. https://doi.org/10.1177/1948550614553640

Hill, A. K., Hunt, J., Welling, L. L. M., Cárdenas, R. A., Rotella, M. A., Wheatley, J. R., … Puts, D. A. (2013). Quantifying the strength and form of sexual selection on men's traits. *Evolution and Human Behavior*, *34*(5), 334–341. https://doi.org/10.1016/j.evolhumbehav.2013.05.004

Hodges-Simeon, C. R., Gaulin, S. J. C., & Puts, D. A. (2010). Different Vocal Parameters Predict Perceptions of Dominance and Attractiveness. *Human Nature*, *21*(4), 406–427. https://doi.org/10.1007/s12110-010-9101-5

Hughes, S. M., & Harrison, M. A. (2017). Your Cheatin' Voice Will Tell on You: Detection of Past Infidelity from Voice. *Evolutionary Psychology*, *15*(2), 1474704917711513. https://doi.org/10.1177/1474704917711513

Jünger, J., Motta-Mena, N. V., Cardenas, R., Bailey, D., Rosenfield, K. A., Schild, C., … Puts, D. A. (2018). Do women's preferences for masculine voices shift across the ovulatory cycle? *Hormones and Behavior*, *106*, 122–134. https://doi.org/10.1016/j.yhbeh.2018.10.008

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

Klofstad C. A., Anderson R. C., & Peters S. (2012). Sounds like a winner: voice pitch influences perception of leadership capacity in both men and women. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1738), 2698–2704. https://doi.org/10.1098/rspb.2012.0311

Kordsmeyer, T. L., Hunt, J., Puts, D. A., Ostner, J., & Penke, L. (2018). The relative importance of intra- and intersexual selection on human male sexually dimorphic traits. *Evolution and Human Behavior*, *39*(4), 424–436. https://doi.org/10.1016/j.evolhumbehav.2018.03.008

Lee, K., & Ashton, M. C. (2004). Psychometric Properties of the HEXACO Personality Inventory. *Multivariate Behavioral Research*, *39*(2), 329–358. https://doi.org/10.1207/s15327906mbr3902_8

Leongómez, J. D., Binter, J., Kubicová, L., Stolařová, P., Klapilová, K., Havlíček, J., & Roberts, S. C. (2014). Vocal modulation during courtship increases proceptivity even

in naive listeners. *Evolution and Human Behavior*, *35*(6), 489–496. https://doi.org/10.1016/j.evolhumbehav.2014.06.008

Levin, D. Z., Whitener, E. M., & Cross, R. (2006). Perceived trustworthiness of knowledge sources: The moderating impact of relationship length. *Journal of Applied Psychology*, *91*(5), 1163–1171. https://doi.org/10.1037/0021-9010.91.5.1163

Levine, E. E., Bitterly, T. B., Cohen, T. R., & Schweitzer, M. E. (2018). Who is trustworthy? Predicting trustworthy intentions and behavior. *Journal of Personality and Social Psychology*, *115*(3), 468–494. https://doi.org/10.1037/pspi0000136

Linke, L., Saribay, S. A., & Kleisner, K. (2016). Perceived trustworthiness is associated with position in a corporate hierarchy. *Personality and Individual Differences*, *99*, 22–27. https://doi.org/10.1016/j.paid.2016.04.076

Little, A. C., Roberts, S. C., Jones, B. C., & DeBruine, L. M. (2012). The perception of attractiveness and trustworthiness in male faces affects hypothetical voting decisions differently in wartime and peacetime scenarios. *The Quarterly Journal of Experimental Psychology*, *65*(10), 2018–2032. https://doi.org/10.1080/17470218.2012.677048

Mahrholz, G., Belin, P., & McAleer, P. (2018). Judgements of a speaker's personality are correlated across differing content and stimulus type. *PLOS ONE*, *13*(10), e0204991. https://doi.org/10.1371/journal.pone.0204991

McAleer, P., Todorov, A., & Belin, P. (2014). How Do You Say 'Hello'? Personality Impressions from Brief Novel Voices. *PLOS ONE*, *9*(3), e90779. https://doi.org/10.1371/journal.pone.0090779

Montano, K. J., Tigue, C. C., Isenstein, S. G. E., Barclay, P., & Feinberg, D. R. (2017). Men's voice pitch influences women's trusting behavior. *Evolution and Human Behavior*, *38*(3), 293–297. https://doi.org/10.1016/j.evolhumbehav.2016.10.010

Moshagen, M., Hilbig, B. E., & Zettler, I. (2014). Faktorenstruktur, psychometrische Eigenschaften und Messinvarianz der deutschsprachigen Version des 60-Item HEXACO Persönlichkeitsinventars. *Diagnostica*, *60*(2), 86–97. https://doi.org/10.1026/0012-1924/a000112

O'Connor, J. J. M., & Barclay, P. (2017). The influence of voice pitch on perceptions of trustworthiness across social contexts. *Evolution and Human Behavior*, *38*(4), 506–512. https://doi.org/10.1016/j.evolhumbehav.2017.03.001

O'Connor, J. J. M., Pisanski, K., Tigue, C. C., Fraccaro, P. J., & Feinberg, D. R. (2014). Perceptions of infidelity risk predict women's preferences for low male voice pitch in short-term over long-term relationship contexts. *Personality and Individual Differences*, *56*, 73–77. https://doi.org/10.1016/j.paid.2013.08.029

O'Connor, J. J. M., Re, D. E., & Feinberg, D. R. (2011). Voice Pitch Influences Perceptions of Sexual Infidelity. *Evolutionary Psychology*, *9*(1), 147470491100900100. https://doi.org/10.1177/147470491100900109

Oleszkiewicz, A., Pisanski, K., Lachowicz-Tabaczek, K., & Sorokowska, A. (2017). Voice-based assessments of trustworthiness, competence, and warmth in blind and sighted adults. *Psychonomic Bulletin & Review*, *24*(3), 856–862. https://doi.org/10.3758/s13423-016-1146-y

Oosterhof, N. N., & Todorov, A. (2009). Shared perceptual basis of emotional expressions and trustworthiness impressions from faces. *Emotion (Washington, D.C.)*, *9*(1), 128–133. https://doi.org/10.1037/a0014520

Penke, L., & Asendorpf, J. B. (2008). Beyond global sociosexual orientations: A more differentiated look at sociosexuality and its effects on courtship and romantic relationships. *Journal of Personality and Social Psychology*, *95*(5), 1113–1135. https://doi.org/10.1037/0022-3514.95.5.1113

Pisanski, K., Jones, B. C., Fink, B., O'Connor, J. J. M., DeBruine, L. M., Röder, S., & Feinberg, D. R. (2016). Voice parameters predict sex-specific body morphology in men and women. *Animal Behaviour*, *112*, 13–22. https://doi.org/10.1016/j.anbehav.2015.11.008

Pisanski, K., Oleszkiewicz, A., Plachetka, J., Gmiterek, M., & Reby, D. (2018). Voice pitch modulation in human mate choice. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1893), 20181634. https://doi.org/10.1098/rspb.2018.1634

Ponsot, E., Burred, J. J., Belin, P., & Aucouturier, J.-J. (2018). Cracking the social code of speech prosody using reverse correlation. *Proceedings of the National Academy of Sciences*, *115*(15), 3972–3977. https://doi.org/10.1073/pnas.1716090115

Puts, D. A., Hill, A. K., Bailey, D. H., Walker, R. S., Rendall, D., Wheatley, J. R., … Ramos-Fernandez, G. (2016). Sexual selection on male vocal fundamental frequency in humans and other anthropoids. *Proc. R. Soc. B*, *283*(1829), 20152830. https://doi.org/10.1098/rspb.2015.2830

Puts, D. A., Apicella, C. L., & Cárdenas, R. A. (2012). Masculine voices signal men's threat potential in forager and industrial societies. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1728), 601–609. https://doi.org/10.1098/rspb.2011.0829

Puts, D. A., Jones, B. C., & DeBruine, L. M. (2012). Sexual Selection on Human Faces and Voices. *The Journal of Sex Research*, *49*(2–3), 227–243. https://doi.org/10.1080/00224499.2012.658924

Puts, D. A. (2005). Mating context and menstrual phase affect women's preferences for male voice pitch. *Evolution and Human Behavior*, *26*(5), 388–397. https://doi.org/10.1016/j.evolhumbehav.2005.03.001

Puts, D. A., Gaulin, S. J. C., & Verdolini, K. (2006). Dominance and the evolution of sexual dimorphism in human voice pitch. *Evolution and Human Behavior*, *27*(4), 283–296. https://doi.org/10.1016/j.evolhumbehav.2005.11.003

Puts, D. A., Hodges, C. R., Cárdenas, R. A., & Gaulin, S. J. C. (2007). Men's voices as dominance signals: vocal fundamental and formant frequencies influence dominance attributions among men. *Evolution and Human Behavior*, *28*(5), 340–344. https://doi.org/10.1016/j.evolhumbehav.2007.05.002

Raine, J., Pisanski, K., Bond, R., Simner, J., & Reby, D. (2019). Human roars communicate upper-body strength more effectively than do screams or aggressive and distressed speech. *PLOS ONE*, *14*(3), e0213034. https://doi.org/10.1371/journal.pone.0213034

Rosenfield, K. A., Sorokowska, A., Sorokowski, P., & Puts, D. A. (2019). Sexual selection for low male voice pitch among Amazonian forager-horticulturists. *Evolution and Human Behavior*. https://doi.org/10.1016/j.evolhumbehav.2019.07.002

Saxton, T. K., Mackey, L. L., McCarty, K., & Neave, N. (2016). A lover or a fighter? Opposing sexual selection pressures on men's vocal pitch and facial hair. *Behavioral Ecology*, *27*(2), 512–519. https://doi.org/10.1093/beheco/arv178

Schild, C., Stern, J., & Zettler, I., (2019). Data from: Linking men's voice pitch to actual and perceived trustworthiness across domains. *Behavioral Ecology*. http://doi.org/10.5061/dryad.q2db26m

Smith, K. M., Olkhov, Y. M., Puts, D. A., & Apicella, C. L. (2017). Hadza Men With Lower Voice Pitch Have a Better Hunting Reputation. *Evolutionary Psychology*, *15*(4), 1474704917740466. https://doi.org/10.1177/1474704917740466

Sorokowski, P., Puts, D., Johnson, J., Żółkiewicz, O., Oleszkiewicz, A., Sorokowska, A., … Pisanski, K. (2019). Voice of Authority: Professionals Lower Their Vocal Frequencies When Giving Expert Advice. *Journal of Nonverbal Behavior*. https://doi.org/10.1007/s10919-019-00307-0

Takagishi, H., Takahashi, T., & Yamagishi, T. (2011). Testosterone diminishes reciprocity in a trust game. *Neuroscience Research*, *71*, e282. https://doi.org/10.1016/j.neures.2011.07.1230

Thielmann, I., & Hilbig, B. E. (2015). The Traits One Can Trust: Dissecting Reciprocity and Kindness as Determinants of Trustworthy Behavior. *Personality and Social Psychology Bulletin*, *41*(11), 1523–1536. https://doi.org/10.1177/0146167215600530

Tigue, C. C., Borak, D. J., O'Connor, J. J. M., Schandl, C., & Feinberg, D. R. (2012). Voice pitch influences voting behavior. *Evolution and Human Behavior*, *33*(3), 210–216. https://doi.org/10.1016/j.evolhumbehav.2011.09.004

Titze, I. R. (1994). *Principles of voice production*. Englewood Cliffs, N.J: Prentice Hall.

Tsantani, M. S., Belin, P., Paterson, H. M., & McAleer, P. (2016). Low Vocal Pitch Preference Drives First Impressions Irrespective of Context in Male Voices but Not in Female Voices. *Perception*, *45*(8), 946–963. https://doi.org/10.1177/0301006616643675

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

Vukovic, J., Jones, B. C., Feinberg, D. R., DeBruine, L. M., Smith, F. G., Welling, L. L. M., & Little, A. C. (2011). Variation in perceptions of physical dominance and trustworthiness predicts individual differences in the effect of relationship context on women's preferences for masculine pitch in men's voices. *British Journal of Psychology*, *102*(1), 37–48. https://doi.org/10.1348/000712610X498750

West, S. A., El Mouden, C., & Gardner, A. (2011). Sixteen common misconceptions about the evolution of cooperation in humans. *Evolution and Human Behavior*, *32*(4), 231–262. https://doi.org/10.1016/j.evolhumbehav.2010.08.001

Table 1

Means, standard deviations, and correlations with confidence intervals

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1.Age	25.45	6.17														
2.Mean F0	121.22	16.46	-.03													
3.F0 CV	0.16	0.04	.21**	.30**												
4.F0 min	81.17	9.25	-.09	.48**	-.22**											
5.F0 max	190.87	44.74	.05	.59**	.55**	.19**										
6.Pf	0.00	0.58	-.06	.28**	.08	.02	.05									
7.Duration	4.12	0.49	.14	.02	.22**	.03	.17*	-.12								
8.HH	3.24	0.69	.01	.07	.04	.13	.11	.00	-.07							
9.Trustw. Int	4.31	0.41	.11	.08	.20**	-.04	.17*	.03	.11	.36**						
10.TG giving	0.45	0.22	-.05	-.04	-.10	.02	-.10	.10	.01	.23**	.11					
11.Gen. Trust	4.28	0.64	.08	-.09	.21**	-.15	.04	.01	-.17*	.11	.12	-.01				
12.Mate Trust (m)	3.04	0.56	.15*	.37**	.20**	.26**	.25**	.15*	.24**	.09	.18*	-.00	.08			
13.Mate Trust (f)	3.47	0.55	.25**	.24**	.20**	.23**	.26**	.05	.24**	.09	.12	-.02	.07	.69**		
14.Eco. Trust	3.97	0.62	.11	-.32**	.13	-.22**	-.09	-.11	-.12	.08	.08	-.03	.87**	-.02	.05	
15.Rel. length	41.39	40.45	.52**	-.03	.01	-.02	-.04	.02	-.01	.01	.10	.01	.03	.03	.03	.10

Note. HH = Honesty-Humility, Trustw. Int = Trustworthy Intentions, TG giving = Trust game giving, Gen. Trust = General trustworthiness, Mate Trust = mating related trustworthiness, Eco. Trust = Economic trustworthiness, Rel. length = Relationship length. * $p < .05$. ** $p < .01$.