ARTICLE IN PRESS

Evolution and Human Behavior xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Evolution and Human Behavior

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

# Male voice pitch mediates the relationship between objective and perceived formidability

Toe Aung, Kevin A. Rosenfield, David Puts*

*Department of Anthropology, Pennsylvania State University, University Park, PA, USA*

ARTICLE INFO

*Keywords:*
Dominance
Formidability
Fundamental frequency
Honest signaling
Sexual selection
Voice pitch

ABSTRACT

Acoustic signals function in intrasexual mating competition in a wide variety of species, including humans. The low voice pitch of human males has been proposed to represent an honest signal of formidability. Although voice pitch in men affects perceptions of size and dominance, it is relatively weakly associated with objective measures of formidability such as body size and strength. As a result, some authors have argued that low male voice pitch is not a valid signal of formidability but is deceptive and salient only because it hijacks a tendency to perceive lower frequency sounds as emanating from larger sources. In this paper, we consider theoretical and empirical issues associated with this perceptual exploitation hypothesis and ask whether male voice pitch transmits information about formidability. We utilize mediation models to investigate whether male voice pitch is an honest signal of formidability in data collected from university students in the U.S. ($n$ = 231 male speakers, 565 male raters) and Canada ($n$ = 74 male speakers, 108 female raters, 65 male raters). In both data sets, male voice pitch mediated the relationship between objective (measured by height) and perceived formidability. Collectively, these results indicate that men's voice pitch transmits information about formidability from signaler to receiver.

## 1. Introduction

The acoustic properties of vertebrate vocalizations are often sexually dimorphic and show evidence of having been shaped by sexual selection (Charlton & Reby, 2016; Ryan, Fox, Wilczynski, & Rand, 1990). Sex differences in fundamental frequency ($f_o$, the rate of vocal fold vibration during phonation) derive from differences in the length and thickness of the vocal folds (Titze, 2000). Larger vocal folds vibrate at a lower rate, and this is perceived as lower pitch (Titze, 2000). As vocal folds vibrate when air is pushed through the larynx, the propagated air is filtered by the supralaryngeal vocal tract, producing a distinct set of resonant frequencies known as formants (Titze, 2000). Longer vocal tract lengths (VTL) generate lower formant frequencies (Titze, 2000), and both fundamental and formant frequencies have been shown to be important vocal parameters that mediate agonistic male-male competition among nonhuman mammals, including koalas (Charlton, Whisson, & Reby, 2013), sea lions (Charrier, Ahonen, & Harcourt, 2011), fallow deer (Vannoni & McElligott, 2008), red deer (Reby et al., 2005), black-tailed gazelles (Frey, Volodin, Volodina, Soldatova, & Juldaschev, 2011), domestic dogs (Taylor, Reby, & McComb, 2010), and giant pandas (Charlton, Zhihe, & Snyder, 2010). In many anthropoid primates, low $f_o$ appears to be selected via male contest competition (Puts et al., 2016), perhaps because it exaggerates perceived body size (Delgado, 2006; Puts et al., 2016; Rendall, Vokey, & Nemeth, 2007) and hence conveys the appearance of dominance and formidability (Darwin, 1871; Puts et al., 2016).

However, although larger species produce lower $f_o$ across carnivores and primates (Bowling et al., 2017), there is little evidence within nonhuman primate species regarding whether low-frequency vocalizations honestly signal formidability or dominance. Before considering this question further, it is useful to define some of the above terms. Formidability refers to the ability to inflict physical damage on others in combat (Sell, Tooby, & Cosmides, 2009), whereas dominance refers to rank attainment based on threat of force (Cheng, Tracy, Foulsham, Kingstone, & Henrich, 2013). Dominance and formidability however are often used synonymously (e.g., Cheng et al., 2013; Šebesta, Třebický, Fialová, & Havlíček, 2019). A signal is a phenotype produced by an individual (the signaler) that functions to influence the behavior of other individuals (receivers) by transmitting information (Lachmann, Számado, & Bergstrom, 2001; Marler, 1961; Maynard Smith & Harper, 1995; Seyfarth & Cheney, 2017). A signal must convey information that, on average, increases the fitness of both (1) the sender and (2) the receiver, and (3) it must have evolved to do so. However, because some deceptive signaling is likely to evolve, and because there is noise in any signaling system, the fidelity of information transmission will vary across signals. Thus, although an "honest" signal may be defined as one

* Corresponding author at: 409 Carpenter Building, Department of Anthropology, The Pennsylvania State University, University Park, PA 16802, USA.
*E-mail address:* dap27@psu.edu (D. Puts).

https://doi.org/10.1016/j.evolhumbehav.2020.08.007
Received 3 December 2019; Received in revised form 11 August 2020; Accepted 13 August 2020

that conveys information at a rate that is better than chance (Dawkins & Guilford, 1991), honest signals will differ considerably in their degree of honesty. In comparison, cues are defined as traits or behaviors that convey information but have not evolved for that function.

Human beings comprise an invaluable model organism for investigating whether low-frequency male vocalizations honestly signal formidability. Not only are humans of special interest, but there is also strong evidence implicating sexual selection in the evolution of human $f_o$, as well as a unique richness to the data available for addressing questions regarding honest signaling. Multiple lines of evidence suggest that low male $f_o$ may have evolved for intimidating competitors. This evidence includes the hypertrophic growth of males' vocal folds at puberty (Titze, 2000), the high degree of $f_o$ sexual dimorphism in humans compared to our closest living relatives (Puts et al., 2016), and the strong effects of $f_o$ on social perceptions (Klofstad, Anderson, & Nowicki, 2015; Puts, Gaulin, & Verdolini, 2006; Tsantani, Belin, Paterson, & McAleer, 2016), particularly perceptions of body size, dominance, and fighting ability (Feinberg, Jones, Little, Burt, & Perrett, 2005; Puts et al., 2006; Puts, Hodges, Cárdenas, & Gaulin, 2007; Wolff & Puts, 2010).

These studies provide evidence that low male $f_o$ functions to intimidate rivals, but there is debate over whether male $f_o$ is an honest signal. Because correlations between men's $f_o$ and objective measures of formidability such as size and strength are relatively weak (Armstrong, Lee, & Feinberg, 2019; Pisanski et al., 2014), some authors (Armstrong et al., 2019; Feinberg, Jones, & Armstrong, 2018, 2019) have concluded that men's $f_o$ does not signal formidability. According to these authors, the tendency to infer formidability from low $f_o$ is a byproduct of a bias in the human perceptual system that associates lower frequencies with larger sound sources. In this view, low male $f_o$ may have evolved because it exaggerated the apparent size of its bearer to competitors and/or potential mates (Puts et al., 2016; Rendall et al., 2007), but the appearance of formidability is purely deceptive. According to these authors, "voice pitch is not an honest indicator of physical dominance" (Armstrong et al., 2019, p.49), and there is "no evidence that male voice pitch signals formidability" (Feinberg et al., 2019, p. 190).

In response, Aung and Puts (2020; Puts & Aung, 2019) noted some theoretical and empirical hurdles that this hypothesis must overcome. First, game-theory models suggest that such deceptive signaling must be infrequent or costly for the signaling system to be evolutionarily stable (Grafen, 1990; Johnstone & Grafen, 1993; Polnaszek & Stephens, 2013; Sell et al., 2010; Titze, 2000; Webster, Ligon, & Leighton, 2018; cf.Szamado, 2000). Game theory models are not without limitations; for example, they often focus on fitness outcomes without considering mechanisms underlying the trait, treat stabilizing strategies within isolated contexts which ignore counter-selection forces or other biological constraints, and overlook individual differences (van den Berg & Weissing, 2015).

Nevertheless, because male traits such as dominance displays that arise through contest competition are frequently tested by competitors, they should evolve to be partly honest (Laidre, 2005; Parker & Ligon, 2002; Rohwer & Rohwer, 1978; Webster et al., 2018). To the extent that male traits are honest indicators of condition, females may secondarily evolve preferences for them (Berglund & Pilastro, 1996). In other cases, males may evolve sexual ornaments and displays that take advantage of a female sensory bias. In either case, continued sensitivity to a male trait is likely to be proportional to the degree to which the trait reveals relevant information about the male (Aung & Puts, 2020). In *Physalaemus* frogs, females prefer low frequencies in the chuck portion of the male's call, and phylogenetic reconstruction indicates that the female preference preceded the evolution of the trait itself; hence, the male chuck possibly originated to exploit a female sensory bias (Ryan et al., 1990). Nevertheless, low chuck frequencies in the Túngara frog (*Physalaemus pustulosus*) indicate larger male body size (Ryan, 1980), and larger males leave fewer eggs unfertilized (Ryan et al., 1990); thus, the advantages of mating with larger males likely play a role in maintaining female sensitivity to chuck frequency. By contrast, in several Goodeinae fishes (Garcia & Ramirez, 2005), a terminal yellow band on the male's tail mimics characteristics of their larvae prey and elicits both feeding and sexual responses from females. Although terminal yellow bands serve as stronger stimuli than larvae, females of species with more elaborate male terminal yellow bands show reduced feeding responsiveness to conspicuous terminal yellow bands than females of species with poorly developed or absent terminal yellow bands. Hence, males likely evolved conspicuous terminal yellow bands to elicit sexual responses from females, and females appear to have evolved the ability to partially discriminate the deceptive signal from similar environmental stimuli that provide accurate fitness-relevant information (Garcia & Ramirez, 2005).

Among men, a lower $f_o$ predicts greater resource access (Mayew, Parsons, & Venkatachalam, 2013), social status (Klofstad, 2016), mating success (Puts, 2005), self-reported infidelity (Schild, Stern, & Zettler, 2020), and reproductive success (Apicella, Feinberg, & Marlowe, 2007; Rosenfield, Sorokowska, Sorokowski, & Puts, 2020). It is plausible that these benefits were obtained in part through the relatively strong effects of $f_o$ on relevant social perceptions, such as formidability (Aung & Puts, 2020; Zhang & Reid, 2017). If so, then the tendency to perceive low male $f_o$ as formidable and dominant is costly, and this tendency should be eliminated by selection unless there are compensatory benefits. That is, the loss of status, mating, and reproduction incurred by deference to individuals with low $f_o$ (Apicella et al., 2007; Klofstad, 2016; Mayew et al., 2013; Puts, 2005) should favor inattention to men's $f_o$ unless such attention is adaptive because $f_o$ is actually related to social and/or physical power (Aung & Puts, 2020).

In fact, meta-analyses indicate that low $f_o$ modestly predicts some correlates of formidability and underlying condition such as size (Pisanski et al., 2014), strength, and testosterone (Aung & Puts, 2020) in men. Limited evidence also suggests that lower male $f_o$ may indicate greater immunocompetence. According to the immunocompetence handicap hypothesis (ICHH; Folstad & Karter, 1992), testosterone mediates a relationship between underlying immune function and the expression of sexually dimorphic traits. Growing evidence indicates that glucocorticoids such as the stress hormone cortisol may also negatively interact with testosterone in mediating this relationship (Bortolotti, Mougeot, Martinez-Padilla, Webster, & Piertney, 2009; Moore et al., 2011; Roberts, Buchanan, Hasselquist, & Evans, 2007). Consistent with the hypothesis that lower $f_o$ reflects underlying immunocompetence, lower $f_o$ has been associated with testosterone levels during development (Hodges-Simeon, Gurven, & Gaulin, 2015) and in adulthood (see Aung & Puts, 2020 for meta-analysis), and was more strongly related to testosterone levels in men with lower cortisol levels in two samples (Puts et al., 2016). In a recent study, male $f_o$ also negatively predicted salivary immunoglobulin-A (sIgA), a marker of mucosal immunity (Arnocky, Hodges-Simeon, Ouellette, & Albert, 2018). However, although dominance ratings of men's voices predicted their self-reported health (Albert, Arnocky, Puts, & Hodges-Simeon, 2020), $f_o$ did not (Albert et al., 2020; Arnocky et al., 2018).

During social interactions, men have also been found to lower their speaking $f_o$ in relation to perceived relative formidability (Leongómez, Mileva, Little, & Roberts, 2017; Puts, Gaulin, & Verdolini, 2006) and authoritativeness (Sorokowski et al., 2019), and such modulation appears to influence perceptions of dominance (Fraccaro et al., 2013) and authoritativeness (Sorokowski et al., 2019), as well as eventual rank (Cheng, Tracy, Ho, & Henrich, 2016). If $f_o$ is an honest signal, then its honestly may in part be maintained socially. Under a sexual priming condition, low-$f_o$ male voices elicited aggressive cognitions and intent in men who perceived themselves to be more dominant and stronger (Zhang & Reid, 2017). Males who falsely signal dominance through low $f_o$ could incur dangerous retribution. Recently, pitch modulation has been shown to predict men's perceived aggressive intent, independent of perceived fighting ability (Zhang, Hodges-Simeon, Gaulin, & Reid, 2020). Thus, some evidence indicates that both habitual $f_o$ and $f_o$

modulation are correlated with proxies of formidability and/or dominance, and hence that attention to men's $f_o$ is functional.

Despite these theoretical and empirical considerations, more data are necessary to clarify whether $f_o$ is a partly honest signal, or whether it is purely deceptive as Feinberg et al., 2019 suggest. To this end, Armstrong et al. (2019) tested whether perceived or measured height mediated the relationship between $f_o$ and perceptions of dominance. In other words, within each sex, are people with lower $f_o$ perceived as more dominant because they are taller, or merely because they are *perceived* to be taller? Armstrong et al. (2019) used height as a proxy for dominance because body size is "the primary indicator of physical dominance" (p.43) across species (Darwin, 1871; Ellis, 1994; French & Smith, 2005). In humans, height reflects good nutrition and low stress during development (Deaton, 2007; Perkins, Subramanian, Smith, & Özaltin, 2016), as well as genetic and endocrine correlates of immune function (Leongómez et al., 2020; Zaidi et al., 2019). Taller men have a fighting advantage (Beaver, Connolly, & Schwartz, 2015; Carrier, 2011) and are more likely to engage in direct physical aggression (Archer & Thanzami, 2007). In addition, taller men are more likely to be authoritative; height predicted authoritative behavior in professional refereeing (Stulp, Buunk, Verhulst, & Pollet, 2012). Taller men are also more likely to win dyadic non-physical confrontations in confined spaces (Stulp, Buunk, Verhulst, & Pollet, 2015) and be less sensitive to dominance signals in other men (Watkins et al., 2010; cf. Wolff & Puts, 2010). Hence, although height is not equivalent to physical dominance, it can serve as a reasonable proxy.

Armstrong et al. (2019) found that both measured and perceived height mediated the relationship between men's $f_o$ and dominance ratings of their voices, but perceived height played a substantially stronger role. Armstrong et al. (2019) also found that residual height perception (residuals after regressing perceived height on measured height) more strongly predicted perceived dominance than measured height did. They concluded that dominance ratings of voices are based on the misperception that people with low-$f_o$ voices are tall, and that $f_o$ is not an honest indicator of physical dominance.

The study by Armstrong et al. (2019) is important because it provides some of the strongest available evidence that low male $f_o$ increases the appearance of dominance by making the vocalizer sound large, a critical prediction of a proposed mechanism through which low male relative to female $f_o$ evolved in humans and many other catarrhine primates (Puts et al., 2016). However, despite the strength of these data, the analyses may be suboptimal for testing whether $f_o$ is an honest signal. First, perceptions of size and dominance are highly intercorrelated (Batres, Re, & Perrett, 2015; van Quaquebeke & Giessner, 2010) and strongly influenced by $f_o$ (Feinberg et al., 2005; Puts et al., 2006). It is therefore unsurprising (and hence not particularly informative) that perceived height would strongly mediate the relationship between $f_o$ and perceived dominance—that is, that one perception would mediate the relationship between $f_o$ and a closely related perception. Indeed, it is also likely that the reverse would apply—that perceived dominance would mediate the relationship between height and perceived height—although this was not tested. Likewise, given the modest correlation between measured height and $f_o$ (Pisanski et al., 2014, 2016), it is unsurprising that measured height only weakly mediated the relationship between $f_o$ and perceived dominance. Second, mediation in simplest form represents an independent variable (X) that causes the mediator (M), which causes the dependent variable (Y) (MacKinnon, Fairchild, & Fritz, 2007). The analysis of Armstrong et al. treated $f_o$ as the independent variable (X) and height as the mediator (M), yet it is logically incorrect to assume that $f_o$ causes height. Third, it is unsurprising that residual height perception more strongly mediated the relationship between $f_o$ and perceived dominance than measured height did. As Armstrong et al. (2019) state, "there is little selection pressure to accurately assess the size of other humans from voice alone, simply because we can see height better than we can hear it" (p.48). Hence, even if humans extract size information from voice, they are demonstrably poor at doing so (Collins, 2000; Gonzalez, 2003; Rendall et al., 2007), and thus residual height perception should be highly correlated with raw height perception (i.e., before regressing on measured height). Once again, one perception mediated the relationship between $f_o$ and a closely related perception, and the implications for the honesty of $f_o$ as a dominance signal are unclear.

## 2. The present research

A signal must alter or manipulate receivers' behavior by transmitting information (Lachmann et al., 2001; Maynard Smith & Harper, 1995). The transfer of information is receiver-dependent and, under an adaptationist framework, is best viewed as an emergent and important feature of communication, as it influences the receiver's behavior (Hauser, 1996; Scott-Phillips, 2008). The most common approach to providing evidence of signaling is to demonstrate the information content of the putative signal, for example by showing that the frequency of loud calls in crested macaques (*Macaca nigra*) reflects dominance rank (Neumann, Assahad, Hammerschmidt, Perwitasari-Farajallah, & Engelhardt, 2010), or that train length in peacocks (*Pavo cristatus*) reflects immunocompetence (Møller & Petrie, 2002). The conspicuousness of a putative signal is often taken as prima facie evidence that the information carried by it is received, and hence that the putative signal influences the behavior of others. More thorough studies demonstrate both (1) the information content of a putative signal and (2) that the putative signal is perceived by others. For example, lower $f_o$ both indicates higher testosterone level and is perceived as more masculine in pre-pubertal boys (Cartei et al., 2020).

Showing that a putative signal both carries relevant information and influences the behavior of receivers comprises strong evidence that it transmits the information to the receiver. As noted above, abundant evidence indicates that $f_o$ influences the behavior of human receivers, and some evidence suggests that it carries relevant information. However, it is also important to demonstrate that these associations are not trivial, i.e., that (3) receivers can in fact detect the relevant information above chance levels, for example because normal variation in other components of the signaling modality does not wash out the effects of the putative signal, and (4) that the ability to detect the relevant information is in part attributable to the putative signal rather than to other signals. The latter evidence can be provided through statistical mediation analysis, by demonstrating that the putative signal statistically mediates the relationship between the information and the receiver's ability to detect the information. For example, Cartei et al. (2020) reported that $f_o$ mediated the relationship between testosterone levels and dominance ratings among pre-pubertal males, but not females.

Thus, a test of whether $f_o$ signals formidability in men constitutes asking the following: (1) Can listeners extract information about formidability from the voice? If so, then (2) Is this ability partly mediated by $f_o$? If the answers to both questions are yes, then this suggests that $f_o$ conveys information about formidability and supports the hypothesis that it functions as a signal of formidability. We therefore tested whether $f_o$ mediates the relationship between a proxy of formidability (adult stature) and dominance perceptions in two samples.

In Study 1, we tested whether $f_o$ mediates the relationship between measured height and physical dominance ratings in a large sample (Puts et al., 2016; Puts, Apicella, & Cárdenas, 2012a; Wolff & Puts, 2010). The production of $f_o$ is determined by the vocal folds which lie low in the vocal tract and outside the bony structures of the skull (Fitch, 2000); hence, $f_o$ is relatively unconstrained by overall skeletal growth (Rendall et al., 2007). Formant frequencies are filtered by the VTL, which is largely constrained by the length of the neck and the size of the skull. In our analyses, we controlled for apparent VTL, a correlate of both height and perceptions of size and dominance, and hence a potential confounding variable. In Study 2, we used data from Armstrong et al. (2019) to replicate the results of Study 1, as well as to test whether

$f_o$ mediates the relationship between measured height and perceived height.

## 3. Study 1

### 3.1. Methods

Data were obtained from a study in which some data have been previously published (e.g., Puts et al., 2016). Here, we summarize relevant stimulus collection and production, experimental procedures, and voice measures.

#### 3.1.1. Stimuli

Voice recordings of 231 men ages 18-26 from Michigan State University were used in our analyses. Recordings came from 176 male participants reported in Puts et al. (2012a) and an additional 55 brothers of those participants whose data were previously unused in Puts et al. (2012a). The recordings were collected as part of a study approved by the Michigan State University Institutional Review Board. Speakers were recorded reading an excerpt from the Rainbow Passage (Fairbanks, 1960) using a Shure SM58 vocal cardioid microphone in a quiet, anechoic chamber. Each speaker provided two voice recordings and height measurements during two different sessions. To minimize rater fatigue, we extracted the first sentence of each recording "When the sunlight strikes raindrops in the air, they act as a prism and form a rainbow" and adjusted mean amplitude of each to 71.5 ± 2.4 dB to equilibrate audibility for ratings.

#### 3.1.2. Participants

A total of 565 male students at The Pennsylvania State University rated the voice recordings in a study approved by the Pennsylvania State University Institutional Review Board. Each rater assessed one of 30 stimulus sets comprising of approximately 25 voice recordings. Recordings were allocated randomly to a set, with the condition that only one recording per speaker was included in each set. Each stimulus set was rated on dominance using seven-point Likert scales (7 = very dominant) by at least 15 raters [mean ± SD ratings per stimulus: 35.24 ± 8.30, range: 15-52]. All ratings from every rater were included in our analyses.

#### 3.1.3. Voice measures

Each recording (mean duration = 30.6 ± 3.8 s) was analyzed using the 'voice report' function and a pitch range set to 75-300 Hz in Praat version 5.3. We measured $f_o$ and first four formant frequencies. Formants were measured at each glottal pulse and averaged across measurements (Puts et al., 2012a) and then used to estimate VTL using the method of Reby and McComb (2003). Acoustic measures of estimated VTL and $f_o$ are available in our Supplementary Materials.

#### 3.1.4. Statistical analyses

We conducted statistical analyses in R and ran linear mixed effect models using the 'lme4' (Bates, Maechler, Bolker, & Walker, 2015) and 'lmerTest' (Kuznetsova, Brockhoff, & Christensen, 2017) packages. Following Armstrong et al. (2019), we z-scored height, $f_o$, VTL, and dominance ratings of voice recordings. To account for non-independence among recordings of sibling pairs ($n$ = 55), we controlled for sibling group as a covariate in our models. In each model, random intercepts were specified separately for speakers and raters. For random slope specifications, the recording session was allowed to vary across speakers, and the predictive effect of each independent variable in the model was allowed to vary across raters. If the addition of $f_o$ as a potential mediator reduced the predictive power of independent variable, we ran a formal mediation analysis via the 'mediation' package (Tingley, Yamamoto, Hirose, Keele, & Imai, 2014). In addition, we used the 'mediation' package to test the mediating effect of VTL, independent of $f_o$. The 'mediation' package has limitations, and it was not possible to include both random effects groups specified in our multi-level models. As in Armstrong et al. (2019), we dropped random effects of speakers and used only random effects of raters in our mediation analyses. We used 1,000 bootstrap samples and 95% confidence intervals for our mediation analyses. The output, specifications, and scripts for all of our models can be found in the Supplementary Materials.

### 3.2. Results

#### 3.2.1. $f_o$ and VTL mediate the relationship between height and physical dominance ratings

To test the mediating effect of $f_o$, independent of VTL, we first regressed physical dominance ratings of voice on height and VTL. Both height (estimate = 0.06, SE = 0.03, $p$ = .032) and VTL (estimate = 0.11, SE = 0.03, $p < .001$) significantly predicted physical dominance ratings. Adding $f_o$ to the model (estimate = -0.13, SE = 0.02, $p < .001$) reduced the predictive power of height on physical dominance ratings (estimate = 0.04, SE = 0.03, $p$ = .216); hence, we ran a formal mediation analysis, with $f_o$ as a mediating variable between height and physical dominance ratings. The average mediation effect of $f_o$ on height and physical dominance ratings was 0.025 ($p < .001$), mediating 38.75% of the total proportion with 95% CI [0.26, 0.60]. We also tested the mediating effect of VTL, independent of $f_o$. The average mediation effect of VTL on height and physical dominance ratings was 0.01

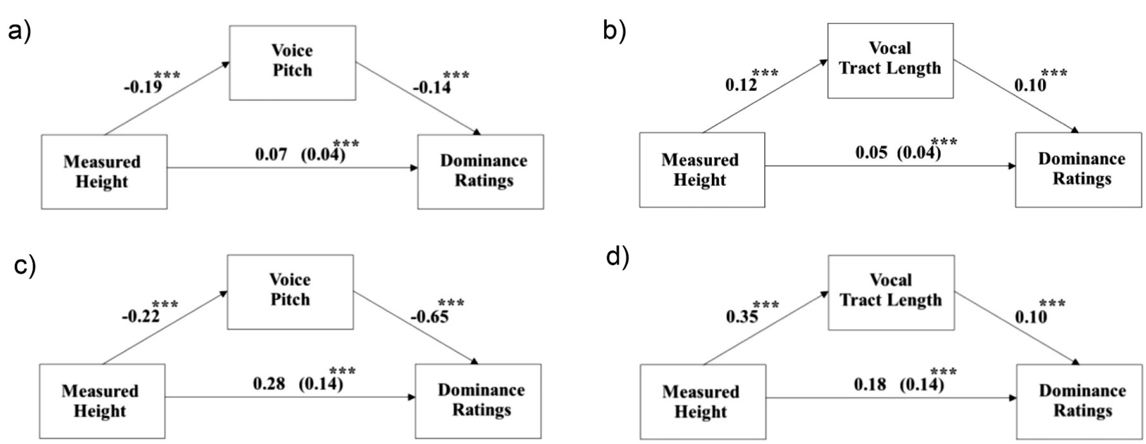


**Fig. 1.** Standardized regression coefficients for the relationship between height and dominance ratings as mediated by voice pitch and vocal tract length. In Study 1, (a) voice pitch mediated about 40%, and (b) vocal tract length mediated about 24%, of the total proportion. In Study 2, (c) voice pitch mediated about 50%, and (d) vocal tract length mediated about 21%, of the total proportion. The average direct effect, controlling for voice pitch and vocal tract length, is in parentheses. ***$p < .001$

($p < .001$), mediating 24.34% of the total proportion with 95% CI [0.15, 0.45]. Both analyses suggest independent mediating effects of $f_o$ and VTL on the relationship between actual and perceived formidability (Fig. 1).

## 4. Study 2

### 4.1. Methods

Data were obtained from a previously-published study (Armstrong et al., 2019). Here, we summarize relevant stimulus collection and production, experimental procedures, and voice measures.

#### 4.1.1. Stimuli

Voice recordings were collected from 74 men aged 17-30, each producing the English vowels /a/, /e/, /i/, /o/, and /u/ in a consistent order. All voices were recorded via an MKH 800 studio condenser microphone in an anechoic chamber. The height of each speaker was also reported in Armstrong et al. (2019) and is referred to in our analyses as "measured height."

#### 4.1.2. Participants

A total of 173 raters (n = 108 females) from McMaster University in Canada listened to the voice stimuli and rated these voices in separate blocks on three different attributes: "perceived height" (1 = very short; 7 = very tall), "perceived physical dominance" (1 = very submissive; 7 = very dominant), and "perceived social dominance (1 = very submissive; 7 = very dominant)." The order of the rating blocks, as well as the order of voices within each block, was randomized. Participants chose to complete one, two, or three blocks of ratings. Physical dominance was defined as "a physically dominant person is someone who if they were in a fistfight with an average undergraduate male, they would probably win". Here, we are interested in ratings of "perceived height" (rated by 54 women [mean ± SD age: 18.7 ± 1.33] and 31 men [mean ± SD age: 19.1 ± 1.13]) and "perceived physical dominance" (rated by 55 women [mean ± SD age: 19.1 ± 2.39] and 35 men [mean ± SD age: 18.8 ± 1.30]). A total of 125 unique raters (n = 78 females) who contributed data for both perceived height and physical dominance ratings were included in our analyses.

#### 4.1.3. Voice measures

We used measures of $f_o$ and estimated VTL for 74 male recordings from Armstrong et al. (2019) in our analyses. In Armstrong et al. (2019), $f_o$ was measured via the auto-correlation algorithm in Praat software (Boersma & Weenink, 2013), with a setting range of 65-300 Hz. Estimated VTL was calculated using the same method in Reby and McComb (2003) via mean values for formant frequencies $F_1$-$F_4$, using the Burg linear predictive coding (LPC) algorithm in Praat (Boersma & Weenink, 2013), with a maximum formant setting of 5000 Hz. Formant frequencies are resonant frequencies that influence the perception of vocal timbre and are primarily determined by the size and shape of vocal tract (Titze, 2000). Previous meta-analyses (Pisanski et al., 2014) suggest that VTL explains 10-15% of the variance in body size among same-sex adults.

#### 4.1.4. Statistical analyses

We conducted statistical analyses in R (R Core Team, 2017) and ran linear mixed effect models using the 'lme4' (Bates et al., 2015) and 'lmerTest' (Kuznetsova et al., 2017) packages. As in Armstrong et al. (2019), we z-scored each variable of perceived height, measured height, $f_o$, and VTL. If the addition of $f_o$ as a potential mediator reduced the predictive power of the independent variable, then we ran a formal mediation analysis via the 'mediation' package (Tingley et al., 2014), using 1,000 bootstrap samples and 95% confidence intervals. As in Armstrong et al. (2019), we included the sex of rater and VTL as covariates. For random slope specifications, the predictive effect of each independent variable (except sex of rater) in the model could vary across raters. We also used the 'mediation' package to test the mediating effect of VTL, independent of $f_o$. The 'mediation' package has limitations, and it was not possible to include both random effects groups specified in our multi-level models. As in Armstrong et al. (2019), we dropped random effects of speakers and used only random effects of raters in our mediation analyses. The output, specifications, and scripts for all our models can be found in the Supplementary Materials.

### 4.2. Results

#### 4.2.1. $f_o$ and VTL mediate the relationship between height and perceived dominance

We tested whether $f_o$ and VTL independently mediated the relationship between measured height and physical dominance ratings of voice stimuli. Measured height (estimate = 0.28, SE = 0.02, $p < .001$) and VTL predicted male physical dominance ratings (estimate = 0.34, SE = 0.03, $p < .001$). When $f_o$ (estimate = -0.65, SE = 0.03, $p < .001$) was added to the model, the predictive power of height on dominance rating was reduced (estimate = 0.14, SE = 0.02, $p < .001$). Thus, we ran a mediation analysis, with $f_o$ as a potential mediating variable and VTL as a covariate (Fig. 1). The average mediation effect of $f_o$ on height and dominance rating was 0.14 ($p < .001$), mediating 50.90% of the total proportion with 95% CI [0.44, 0.59]. We also tested the mediating effect of VTL, independent of $f_o$ (Fig. 1). In this model, the average mediation effect of VTL on height and dominance rating was 0.04 ($p < .001$), mediating 20.95% of the total proportion with 95% CI [0.12, 0.32].

#### 4.2.2. $f_o$ and VTL mediate the relationship between measured height and perceived height

We also tested whether $f_o$ mediates the relationship between measured height and perceived height. Both measured height (estimate = 0.24, SE = 0.01, $p < .001$) and VTL predicted perceived height (estimate = 0.48, SE = 0.01, $p < .001$). When $f_o$ (estimate = -0.69, SE = 0.01, $p < .001$) was added to the model, the predictive power of measured height on perceived height was reduced (estimate = 0.09, SE =

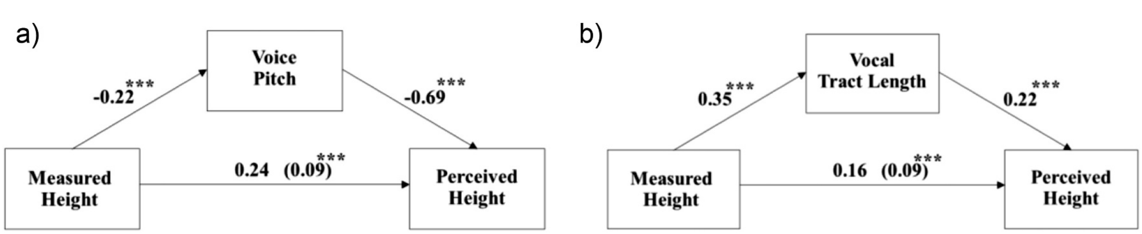


**Fig. 2.** Standardized regression coefficients for the relationship between measured height and perceived height as mediated by vocal pitch (a) and vocal tract length (b) in Study 2. Vocal pitch mediates about 64%, and vocal tract length about 48%, of the total proportion. The average direct effect, controlling for voice pitch and vocal tract length, is in parentheses. ***$p < .001$

0.01, $p < .001$). Hence, we ran a mediation analysis, with $f_o$ as a potential mediating variable and VTL as a covariate (Fig. 2). The average mediation effect of $f_o$ on the relationship between measured height and perceived height was 0.15 ($p < .001$), mediating 63.97% of the total proportion with 95% CI [0.60, 0.68]. We also tested the mediating effect of VTL, independent of $f_o$ (Fig. 2). In this model, the average mediation effect of VTL on height and perceived height was 0.08 ($p < .001$), mediating 48.21% of the total proportion with 95% CI [0.44, 0.53].

## 5. Discussion

In the two present studies, $f_o$ mediated the relationship between objective and perceived dominance, suggesting that listeners perceive formidable men as being more formidable in part because of their voice pitch. Although mediation analysis suggests causality, it is possible that listeners utilized another, correlated acoustic feature to assess formidability. However, in our analyses, the mediating effect of $f_o$ on perceived height and dominance remained after controlling for the effects of VTL, a correlate of size (Pisanski et al., 2014) and the most likely such confounding variable. In fact, the mediating effect of $f_o$ was stronger than that of VTL; $f_o$ mediated 39-40% of the total effect observed between height and perceived dominance, whereas VTL mediated about 21-24% of the total effect. Although other acoustic correlates of $f_o$ and VTL could mediate the relationship between size and perceptions, the perceptual salience of $f_o$ and formants, determined by VTL, make this possibility less compelling. In addition, $f_o$ mediated about 64% of the total effect observed between measured height and perceived height, whereas VTL mediated about 48% of the total proportion. Previously, Armstrong et al. (2019) reported a much stronger effect of unmanipulated male pitch on perceived body size ($R^2 = 0.71$) than actual body size ($R^2 = 0.16$). Here, we provided additional insights by showing that $f_o$ mediated the relationship between measured height and perceived height, and that the exaggeration effect of $f_o$ is size-dependent; taller men are perceived to be taller because of their lower voice pitch, and shorter people are perceived to be shorter because of their higher voice pitch. Even though the two studies employed different stimuli (continuous speech vs. vowels) and groups of raters (male only vs. male and female), their results are highly similar. Overall, this evidence supports the view that $f_o$ is a VTL-independent signal that provides information about the vocalizer's formidability.

A widely favored hypothesis is that, in some species, low male relative to female $f_o$ has evolved to exploit a pre-existing perceptual association of low frequencies with large sound sources. Large objects occupy more of the visual field and produce lower frequency sounds, on average. When there is a fitness advantage to exaggerating size, animals tend to evolve visual and acoustic exaggerations along these dimensions. Yet, such exaggerations could work only if receivers' visual and acoustic processing systems were already adapted to perceive such stimuli as emanating from larger objects (Grassi, 2005; Perrott, Musicant, & Bettina, 1980; Rendall et al., 2007). Hence, multiple researchers have suggested that low-frequency male vocalizations in humans (e.g., Rendall et al., 2007; Puts et al., 2016; Armstrong et al. (2019) and other mammals (Charlton, Reby, & McComb, 2008; Puts et al., 2016) evolved to exploit such pre-existing perceptual associations. To test this hypothesis, future studies should measure responses to low-$f_o$ vocalizations across primate species and use phylogenetic reconstruction to infer whether deference to low $f_o$ is likely to have preceded the evolution of relatively low male $f_o$. However, in our view, this hypothesis is uncontroversial, and the current debate instead concerns the veracity of the claim that "voice pitch is not an honest indicator of physical dominance" (Armstrong et al., 2019, p.49).

The present work is the first to our knowledge to show that men's $f_o$ mediates the relationship between an objective proxy measure of formidability and relevant perceptions. Thus, rather than suggesting that voice pitch does not indicate formidability, data from Armstrong et al. (2019) and similar data from a larger study appear to indicate the opposite. Many studies have shown that lower $f_o$ increases perceptions of formidability and physical dominance (Feinberg et al., 2005; Puts et al., 2006; Puts et al., 2007; Saxton, Mackey, McCarty, & Neave, 2016; Wolff & Puts, 2010), and the present results provide evidence that there is some "truth in advertising" (Kodric-Brown & Brown, 1984).

Indeed, given the apparent costs of deferring to males with low $f_o$ in terms of resources (Mayew et al., 2013), status (Klofstad, 2016), mates (Puts, 2005; Rosenfield et al., 2020), and reproduction (Apicella et al., 2007; Rosenfield et al., 2020; cf. Atkinson et al., 2012), it is unlikely that deference to low $f_o$ would be maintained by selection if low $f_o$ were a purely deceptive exaggeration of size (Puts & Aung, 2019). Previous results suggest that low $f_o$ signals formidability and/or underlying condition: for example, meta-analytical results indicate that low $f_o$ predicts greater height (Pisanski et al., 2014), testosterone levels (Aung & Puts, 2020), and upper-body strength (Aung & Puts, 2020). However, these associations, including those in the present studies, are modest in magnitude; the observed proportional mediating effect of $f_o$ between height and perceived dominance was substantial (39-40%), but the overall mediating effect itself was relatively weak. If a low $f_o$ signals formidability, then why are these associations not stronger?

There are several possible and non-mutually exclusive answers to this question. First, it is important to emphasize that across ages and sexes, $f_o$ is in fact a strong predictor of formidability. This is a result of the precipitous decline in male $f_o$ at puberty; a person with a deep voice is highly likely to be an adult male, which is strongly indicative of formidability. For example, $f_o$ explained over 60% of the variance in upper body strength among a mixed-sex sample of university students and over 70% of the variance among peripubertal male forager-horticulturists (Aung & Puts, 2020; Hodges-Simeon, Gurven, Puts, & Gaulin, 2014; Puts et al., 2016). Attention to $f_o$ is thus likely to be favored in part because of the strong associations between $f_o$ and formidability among humans generally.

Second, the correlation between $f_o$ and formidability may be weakened by the presence of "cheating", falsely advertising formidability with a low $f_o$. Following Dawkins and Guilford (1991), we have considered a signal honest if it indicates a quality at a significantly better-than-chance correlation, but signaling systems are open to cheating, especially when the assessment of the signal is itself costly to the receiver (Dawkins & Guilford, 1991). When receivers cannot avoid paying the full costs of signal assessment, honest signals are often "corrupted" into conventional signals whereby cheating becomes more common and has a negative frequency-dependent advantage. As long as the costs of cheating rise as cheating becomes more common, and enough honest signalers exist to inflict costs to probing by receivers (Dawkins & Guilford, 1991), an honest signaling system with cheating is likely to evolve and persist. If cheating spreads through the population completely, then the signal becomes useless, and receivers will no longer attend to the signal (Maynard Smith & Harper, 1988). Signalers may evolve to send signals at lower costs, and receivers in turn evolve to better discriminate between a high- and low-quality signal (for review see McCoy & Haig, 2020). This scenario is likely given that human vocal folds are relatively unconstrained by body size, and pitch can be modulated considerably within the constraints of a given vocal anatomy; hence, a lower voice pitch can perhaps be produced with little physiological cost. Across mammals, the capacity for vocal modulation may evolve via sexual selection for such "dishonest" signaling (Garcia & Ravignani, 2020) and/or to adjust dominance signaling with greater sensitivity across social contexts. Nevertheless, producing a lower pitch is likely to be costlier for lower-quality signalers in general because of the potential retaliation costs (Zhang & Reid, 2017).

Third, signals of formidability and/or underlying condition are likely to be multi-modal and involve multiple components within sensory modalities. Numerous other cues and signals, including aspects of physical appearance, behavior, and many other acoustic and linguistic variables influence perceptions related to formidability. Assuming that these cues and signals are partially non-redundant, each is expected to explain only a fraction of the variance in both perceived and objectively measured formidability (Aung & Puts, 2020). Further, because there is noise in any informational system, and because biological signals will have additional noise due to deception, the proportion of variance in objectively measured formidability that is uniquely explained by each characteristic should be somewhat less than the variance it explains in perceptions of formidability (Aung & Puts, 2020).

Finally, the expectation that $f_o$ should strongly predict objective measures of formidability may in part be an artifact of laboratory studies that artificially inflate the salience of acoustic variables by omitting cues that are normally available during social interactions (Aung & Puts, 2020). In more naturalistic contexts where male formidability is judged from videos or by familiar peers, the relationships between male $f_o$ and perceived formidability (correlations of approximately 0.2: Hill et al., 2013; Kordsmeyer, Hunt, Puts, Ostner, & Penke, 2018) are much closer in magnitude to relationships between proxies of dominance and male $f_o$, such as those in the present studies.

Low male voice pitch may signal formidability and underlying condition to both same-sex competitors and potential mates. However, a large body of evidence suggests that men's secondary sex traits (Hill et al., 2013; Hill, Bailey, & Puts, 2016; Kordsmeyer et al., 2018; Puts et al., 2016; Puts, Bailey, & Reno, 2015), and deep voices in particular (Feinberg et al., 2005; Puts et al., 2016; Puts, Doll, & Hill, 2014; Saxton et al., 2016), function more effectively at intimidating and/or winning contests with rivals than in mate attraction. For example, among Bolivian forager-horticulturists, lower male $f_o$ increased the appearance of physical formidability to other men but decreased attractiveness to women (Rosenfield et al., 2020). Low male relative to female vocalization pitch probably evolved in early catarrhine primates as a result of male-male competition (Puts et al., 2016). Nevertheless, female mate choice was likely important in shaping men's traits, as well, and several (Feinberg et al., 2005; Puts, 2005; Puts et al., 2016), but not all (Apicella & Feinberg, 2009; Rosenfield, Sorokowska, Sorokowski, & Puts, 2020; Shirazi, Puts, & Escasa-Dorne, 2018), studies have found that women prefer deeper voices.

A Fisherian mate choice model via runaway selection has also been suggested as a possible mechanism favoring low male $f_o$ (Puts, 2005; Puts et al., 2014) but is unlikely to be a primary driver of selection for low male $f_o$ in humans. Otherwise, we would expect females to prefer the lowest male pitches. Evidence instead suggests that females do not necessarily prefer lower male pitch (Apicella & Feinberg, 2009; Rosenfield et al., 2020; Shirazi et al., 2018), especially compared to consistently strong negative linear effects on perceptions of dominance across studies (Aung & Puts, 2020; Puts, Jones, & DeBruine, 2012b), and if anything, the relationship between pitch and vocal attractiveness appears to be negatively quadratic, such that women prefer voices around the male average (Puts et al., 2014; Puts et al., 2012b).

Additional work is needed before firm conclusions are possible. In a recent study of boys aged 3-10 (Cartei et al., 2020), $f_o$ mediated the relationship between perceived masculinity and testosterone levels but not height, likely because developmental changes in testosterone are more critical than size to male voice pitch (Hodges-Simeon, Gurven, Cárdenas, & Gaulin, 2013). Future studies should also sample across a variety of ages and cultural contexts, and multiple sensory components or modalities should be measured to investigate the extent to which $f_o$, among other traits, provides useful information about the signaler (Aung & Puts, 2020). We used height as a proxy for formidability and underlying condition, as height is related to physical prowess and reflects good nutrition and low stress during development, as well as genetic predictors of immune function. Nevertheless, height is only one measure of formidability and underlying condition, and future research should utilize additional measures, such as hormone levels, health history, biomarkers of infectious disease, and genetic predictors of immune function such as heterozygosity at major histocompatibility loci. Finally, the present research highlights the utility of a mediation-analytical approach to testing signal honesty across taxa.

## Declaration of Competing Interest

None

## Acknowledgements

We thank Greg Bryant, Ed Hagen, and two anonymous reviewers for helpful comments on our manuscript.

The authors report no funding source for this work.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.evolhumbehav.2020.08.007.

## References

Albert, G., Arnocky, S., Puts, D. A., & Hodges-Simeon, C. R. (2020). Can listeners assess men's self-reported health from their voice? *Evolution and Human Behavior* in press.

Apicella, C. L., & Feinberg, D. R. (2009). Voice pitch alters mate-choice-relevant perception in hunter-gatherers. *Proceedings of the Royal Society B: Biological Sciences, 276*(1659), 1077–1082.

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.

Archer, J., & Thanzami, V. (2007). The relation between physical aggression, size, and strength among a sample of young indian man. *Personality and Individual Differences, 43*(3), 627–633.

Armstrong, M. M., Lee, A. J., & Feinberg, D. R. (2019). A house of cards: Bias in perception of body size mediates the relationship between voice pitch and perceptions of dominance. *Animal Behaviour, 147*, 43–51. https://doi.org/10.1016/j.anbehav.2018.11.005.

Arnocky, S., Hodges-Simeon, C., Ouellette, D., & Albert, G. (2018). Do men with more masculine voices have better immunocompetence? *Evolution & Human Behavior, 39*(6), 602–610. https://doi.org/10.1016/j.evolhumbehav.2018.06.003.

Atkinson, J., Pipitone, R. N., Sorokowska, A., Sorokowski, P., Mberira, M., Bartels, A., & Jr, G. G. G. (2012). Voice and handgrip strength predict reproductive success in a group of indigenous african females. *PLoS One, 7*(8), 1–9. https://doi.org/10.1371/journal.pone.0041811.

Aung, T., & Puts, D. A. (2020). Voice pitch: A window into the communication of social power. *Current Opinion in Psychology, 33*, 154–161. https://doi.org/10.1016/j.copsyc.2019.07.028.

Bates, D., Maechler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed effects models using lme4. *Journal of Statistical Software, 67*(1), 1e48.

Batres, C., Re, D. E., & Perrett, D. I. (2015). Influence of perceived height, masculinity, and age on each other and on perceptions of dominance in male faces. *Perception, 44*(11), 1293–1309. https://doi.org/10.1177/0301006615596898.

Beaver, K. M., Connolly, E. J., & Schwartz, J. A. (2015). Male physical fighting ability during adolesence influlenced by height and bulk. *Journal of Developmental and Life-Course Criminology, 1*, 434–446.

Berglund, A., & Pilastro, A. (1996). Armaments and ornaments : An evolutionary explanations of traits of dual utility. *Biological Journal of the Linnean Society, 58*(4), 385–399.

Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer [Computer program]. Retrieved from: http://www.fon.hum.uva.nl/praat/.

Bortolotti, G. R., Mougeot, F., Martinez-Padilla, J., Webster, L. M., & Piertney, S. B. (2009). Physiological stress mediates the honesty of social signals. *PLoS ONE, 4*(3), Article e4983. https://doi.org/10.1371/journal.pone.0004983.

Bowling, D. L., Garcia, M., Dunn, J. C., Ruprecht, R., Stewart, A., Frommolt, K. H., & Fitch, W. T. (2017). Body size and vocalization in primates and carnivores. *Scientific Reports, 7*, 1–11. https://doi.org/10.1038/srep41070.

Carrier, D. R. (2011). The advantage of standing up to fight and the evolution of habitual bipedalism in hominins. *PLoS ONE, 6*(5), Article e19630.

Cartei, V., Banerjee, R., Garnham, A., Oakhill, J., Roberts, L., Anns, S., & Reby, D. (2020). Physiological and perceptual correlates of masculinity in children's voices. *Hormones and Behavior, 117*, 104616. https://doi.org/10.1016/j.yhbeh.2019.104616.

Charlton, B. D., & Reby, D. (2016). The evolution of acoustic size exaggeration in terrestrial mammals. *Nature Communications, 7*, 1–8. https://doi.org/10.1038/ncomms12739.

Charlton, B. D., Reby, D., & McComb, K. (2008). Effect of combined source (F0) and filter (formant) variation on red deer hind responses to male roars. *The Journal of the Acoustical Society of America, 123*, 2936–2943.

Charlton, B. D., Whisson, D. A., & Reby, D. (2013). Free-ranging male koalas use size-related variation in formant frequencies to assess rival males. *PLoS One, 8*, Article e70279.

Charlton, B. D., Zhihe, Z., & Snyder, R. (2010). Giant pandas perceive and attend to formant frequency variation in male bleats. *Animal Behaviour, 79*(6), 1221–1227.
Charrier, I., Ahonen, H., & Harcourt, R. G. (2011). What makes an Australian sea lion (*Neophoca cinerea*) male's bark threatening? *Journal of Comparative Psychology, 125*, 385–392.
Cheng, J. T., Tracy, J. L., Foulsham, T., Kingstone, A., & Henrich, J. (2013). Evidence that dominance and prestige are distinct yet viable avenues to social rank and influence. *Journal of Personality and Social Psychology, 104*(1), 103–125.
Cheng, J. T., Tracy, J. L., Ho, S., & Henrich, J. (2016). Listen, follow me: Dynamic vocal signals of dominance predict emergent social rank in humans. *Journal of Experimental Psychology: General, 145*(5), 1–12. https://doi.org/10.1037/xge0000166.
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.
Darwin, C. (1871). *The descent of man, and selection in relation to sex.* London, U.K.: Penguin Books.
Dawkins, M. S., & Guilford, T. (1991). The corruption of honest signaling. *Animal Behaviour, 41*(5), 865–873.
Deaton, A. (2007). Height, health, and development. *Proceedings of the National Academy of Sciences, 104*(33), 13232–13237. https://doi.org/10.1073/pnas.0611500104.
Delgado, R. A. (2006). Sexual selection in the loud calls of male primates: Signal content and function. *International Journal of Primatology, 27*(1), 5–25. https://doi.org/10.1007/s10764-005-9001-4.
Ellis, L. (1994). The high and the mighty among man and beast: How universal is the relationship between height (or body size) and social status? In L. Ellis (Vol. Ed.), *Social stratification and socioeconomic inequality. 2. Social stratification and socioeconomic inequality* (pp. 93–112). Westport, CT: Praeger.
Fairbanks, G. (1960). Recent experimental investigations of vocal pitch in speech. *Journal of the Acoustical Society of America, 11*, 457–466.
Feinberg, D. R., Jones, B. C., & Armstrong, M. M. (2018). Sensory exploitation, sexual dimorphism, and human voice pitch. *Trends in Ecology and Evolution, 33*(12), 901–903.
Feinberg, D. R., Jones, B. C., & Armstrong, M. M. (2019). No evidence that men's voice pitch signals formidability. *Trends in Ecology and Evolution, 34*(3), 190–192. https://doi.org/10.1016/j.tree.2018.12.014.
Feinberg, D. R., Jones, B. C., Little, A. C., Burt, D. M., & Perrett, D. I. (2005). Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices. *Animal Behaviour, 69*(3), 561–568. https://doi.org/10.1016/j.anbehav.2004.06.012.
Fitch, W. T. (2000). The evolution of speech: A comparative review. *Trends in Cognitive Science, 4*(7), 258–267.
Folstad, I., & Karter, A. J. (1992). Parasites, bright males, and the immunocompetence handicap. *American Naturalist, 139*(3), 603–622. https://doi.org/10.2307/2462500.
Fraccaro, P. J., O'Connor, J. J. M., Re, D. E., Jones, B. C., DeBruine, L. M., & Feinberg, D. R. (2013). Faking it: Deliberately altered voice pitch and vocal attractiveness. *Animal Behaviour, 85*(1), 127–136. https://doi.org/10.1016/j.anbehav.2012.10.016.
French, A. R., & Smith, T. B. (2005). Importance of body size in determining dominance hierarchies among diverse tropical frugivores. *Biotropica, 37*(1), 96–101.
Frey, R., Volodin, I., Volodina, E., Soldatova, N. V., & Juldaschev, E. T. (2011). Descended and mobile larynx, vocal tract elongation and rutting roars in male goitred gazelles (*Gazella subgutturosa*). *Journal of Anatomy, 218*, 566–585.
Garcia, M., & Ravignani, A. (2020). Acoustic allometry and vocal learning in mammals. *Biology Letters, 16*, 20200081.
Garcia, M. C., & Ramirez, E. (2005). Evidence that sensory traps can evolve into honest signals. *Nature, 434*, 501–505. https://doi.org/10.1038/nature03351.1.
Gonzalez, J. (2003). Estimation of speakers' weight and height from speech: A re-analysis of data from multiple studies by lass and colleagues. *Perceptual and Motor Skills, 96*(1), 297. https://doi.org/10.2466/pms.96.1.297-304.
Grafen, A. (1990). Biological signals as handicaps. *Journal of Theoretical Biology, 144*, 517–546. https://doi.org/10.1016/s0022-5193(05)80088-8.
Grassi, M. (2005). Do we hear size or sound ? Balls dropped on plates. *Perception & Psychophysics, 67*(2), 274–284. https://doi.org/10.3758/BF03206491.
Hauser, M. D. (1996). *The evolution of communication.* Cambridge, Massachusetts: MIT Press.
Hill, A. K., Bailey, D. H., & Puts, D. A. (2016). Gorillas in our midst? human sexual dimorphism and contest competition in men. *On human nature: Biology, psychology, ethics, politics, and religion* (pp. 235–249). . https://doi.org/10.1016/B978-0-12-420190-3.00015-6.
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., Gurven, M., Cárdenas, R. A., & Gaulin, S. J. C. (2013). Voice change as a new measure of male pubertal timing: A study among Bolivian adolescents. *Annals of Human Biology, 40*(3), 209–219. https://doi.org/10.3109/03014460.2012.759622.
Hodges-Simeon, C. R., Gurven, M., & Gaulin, S. J. C. (2015). The low male voice is a costly signal of phenotypic quality among Bolivian adolescents. *Evolution and Human Behavior, 36*(4), 294–302. https://doi.org/10.1016/j.evolhumbehav.2015.01.002.
Hodges-Simeon, C. R., Gurven, M., Puts, D. A., & Gaulin, S. J. (2014). Vocal fundamental and formant frequencies are honest signals of threat potential in peripubertal males. *Behavioral Ecology, 25*(4), 984–988.
Johnstone, R. A., & Grafen, A. (1993). Dishonesty and the handicap. *Animal Behaviour, 46*, 759–764.
Klofstad, C. A. (2016). Candidate voice pitch influences election outcomes. *Political Psychology, 37*(5), 725–738. https://doi.org/10.1111/pops.12280.
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*, Article e0133779.
Kodric-Brown, A., & Brown, J. H. (1984). Truth in advertising: The kinds of traits favored by sexual selection. *The American Naturalist, 124*(3), 309–323.
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.
Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest: Tests in linear mixed effect models. *Journal of Statistical Software, 82*(13).
Lachmann, M., Számado, S., & Bergstrom, C. T. (2001). Cost and conflict in animal signals and human language. *Proceedings of the National Academy of Sciences of the United States of America, 98*(23) 13489–13194.
Laidre, M. E. (2005). Honest signaling of intentions: The importance of performance risk, receiver response, and prior knowledge of opponents in the threat displays of mandrills. *Behaviour, 142*, 455–476.
Leongómez, J. D., Mileva, V. R., Little, A. C., & Roberts, S. C. (2017). Perceived differences in social status between speaker and listener affect the speaker's vocal characteristics. *PLoS ONE, 12*(6), Article e0179407.
Leongómez, J. D., Sánchez, O. R., Vásquez-Amézquita, M., Valderrama, E., Castellanos-Chacón, A., Morales-Sánchez, L., ... González-Santoyo, I. (2020). Self-reported health is related to body height and waist circumference in rural indigenous and urbanised latin-american populations. *Scientific Reports, 10*, 4391. https://doi.org/10.1038/s41598-020-61289-4.
MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology, 58*, 593–614. https://doi.org/10.1146/annurev.psych.58.110405.085542.
Marler, P. (1961). The logical analysis of animal communication. *Journal of Theoretical Biology, 1*, 295–317.
Mayew, W. J., Parsons, C. A., & Venkatachalam, M. (2013). Voice pitch predicts labor market success. *Evolution and Human Behavior, 34*(4), 243–248.
Maynard Smith, J., & Harper, D. G. (1988). The evolution of aggression: Can selection generate variability? *Philosophical Transactions of the Royal Society B, 319*, 557–570.
Maynard Smith, J., & Harper, D. G. (1995). Animal signals: Models and terminology. *Journal of Theoretical Biology, 177*(3), 305–311.
McCoy, D. E., & Haig, D. (2020). Embryo selection and mate choice: Can "honest signals" be trusted? *Trends in Ecology & Evolution, 35*(4), 308–318.
Møller, A. P., & Petrie, M. (2002). Condition dependence, multiple sexual signals, and immunocompetence in peacocks. *Behavioral Ecology, 13*(2), 248–253. https://doi.org/10.1093/beheco/13.2.248.
Moore, F. R., Cornwell, R. E., Smith, M. J., Al Dujaili, E. A., Sharp, M., & Perrett, D. I. (2011). Evidence for the stress-linked immunocompetence handicap hypothesis in human male faces. *Proceedings of the Royal Society B: Biological Sciences, 278*(1706), 774–780. https://doi.org/10.1098/rspb.2010.1678.
Neumann, C., Assahad, G., Hammerschmidt, K., Perwitasari-Farajallah, D., & Engelhardt, A. (2010). Loud calls in male crested macaques, *Macaca nigra*: A signal of dominance in a tolerant species. *Animal Behaviour, 79*(1), 187–193. https://doi.org/10.1016/j.anbehav.2009.10.026.
Parker, T. H., & Ligon, D. J. (2002). Dominant male red junglefowl (*Gallus gallus*) test the dominance status of other males. *Behavioral Ecology and Sociobiology, 53*(1), 20–24.
Perkins, J. M., Subramanian, S. V., Smith, G. D., & Özaltin, E. (2016). Adult height, nutrition, and population health. *Nutrition Reviews, 74*(3), 149–165. https://doi.org/10.1093/nutrit/nuv105.
Perrott, D. R., Musicant, A., & Bettina, S. (1980). The expanding-image effect: The concept of tonal volume revisited. *Journal of Auditory Research, 20*(1), 43–55.
Pisanski, K., Fraccaro, P. J., Tigue, C. C., O'Connor, J. J. M., Röder, S., Andrews, P. W., ... Feinberg, D. R. (2014). Vocal indicators of body size in men and women: A meta-analysis. *Animal Behaviour, 95*, 89–99. https://doi.org/10.1016/j.anbehav.2014.06.011.
Pisanski, K., Mora, E. C., Pisanski, A., Reby, D., Sorokowski, P., Frackowiak, T., & Feinberg, D. R. (2016). Volitional exaggeration of body size through fundamental and formant frequency modulation in humans. *Scientific Reports, 6*, 1–8. https://doi.org/10.1038/srep34389.
Polnaszek, T. J., & Stephens, D. W. (2013). Why not lie? Costs enforce honesty in an experimental signalling game. *Proceedings of the Royal Society B: Biological Sciences, 281*(1774), https://doi.org/10.1098/rspb.2013.2457.
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., Apicella, C. L., & Cárdenas, R. A. (2012a). Masculine voices signal men's threat potential in forager and industrial societies. *Proceedings of the Royal Society B: Biological Sciences, 278*(1710), 1414–1418. https://doi.org/10.1098/rspb.20.
Puts, D. A., & Aung, T. (2019). Does men's voice pitch signal formidability? A reply to Feinberg et al. *Trends in Ecology and Evolution, 34*(3), 189–190. https://doi.org/10.1016/j.tree.2018.12.004.
Puts, D. A., Bailey, D. H., & Reno, P. L. (2015). Contest competition in men. In D. M. Buss (Ed.). *The handbook of evolutionary psychology* (pp. 1–18). Inc: John Wiley & Sons.
Puts, D. A., Doll, L. M., & Hill, A. K. (2014). Sexual selection on human voices. In V. A. Weekes-Shackelford, & T. K. Shackelford (Eds.). *Evolutionary perspectives on human sexual psychology and behavior* (pp. 69–86). Springer.
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., 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. *Proceedings of the Royal Society B: Biological Sciences,*

*283*(1829), 0–7. https://doi.org/10.1098/rspb.2015.2830.

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.

Puts, D. A., Jones, B. C., & DeBruine, L. M. (2012b). Sexual selection on human faces and voices. *Journal of Sex Research, 49*, 227–243.

Reby, D., & McComb, K. (2003). Anatomical constraints generate honesty: Acoustic cues to age and weight in the roars of red deer stags. *Animal Behaviour, 65*, 519–530.

Reby, D., McComb, K., Cargnelutti, B., Darwin, C., Fitch, W. T., & Clutton-Brock, T. (2005). Red deer stags use formants as assessment cues during intrasexual agonistic interactions. *Proceedings of the Royal Society B: Biological Sciences, 272*, 941–947.

Rendall, D., Vokey, J. R., & Nemeth, C. (2007). Lifting the curtain on the Wizard of Oz: Biased voice-based impressions of speaker size. *Journal of Experimental Psychology: Human Perception and Performance, 33*(5), 1208–1219. https://doi.org/10.1037/0096-1523.33.5.1208.

Roberts, M. L., Buchanan, K. L., Hasselquist, D., & Evans, M. R. (2007). Effects of testosterone and corticosterone on immunocompetence in the zebra finch. *Hormones and Behavior, 51*(1), 126–134. https://doi.org/10.1016/j.yhbeh.2006.09.004.

Rohwer, S., & Rohwer, F. C. (1978). Status signalling in harris sparrows. *Animal Behaviour, 26*, 1012–1022.

Rosenfield, K. A., Sorokowska, A., Sorokowski, P., & Puts, D. A. (2020). Sexual selection for low male voice pitch among amazonian forager-horticulturists. *Evolution and Human Behavior, 41*(1), 3–11. https://doi.org/10.1016/j.evolhumbehav.2019.07.002.

Ryan, M. J. (1980). Female mate choice in a neotropical frog. *Nature, 209*, 523–525.

Ryan, M. J., Fox, J. H., Wilczynski, W., & Rand, A. S. (1990). Sexual selection for sensory exploitation in the frog. *Nature, 343*, 412–415.

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. (2020). Linking men's voice pitch to actual and perceived trustworthiness across domains. *Behavioral Ecology, 31*(1), 164–175.

Scott-Phillips, T. C. (2008). Defining biological communiation. *Journal of Evolutionary Biology, 21*, 387–395. https://doi.org/10.1111/j.1420-9101.2007.01497.x.

Šebesta, P., Třebický, V., Fialová, J., & Havlíček, J. (2019). Roar of a champion: Loudness and voice pitch predict perceived fighting ability but not success in MMA fighters. *Frontiers in psychology, 10*, 859. https://doi.org/10.3389/fpsyg.2019.00859.

Sell, A., Bryant, G. A., Cosmides, L., Tooby, J., Sznycer, D., Von Rueden, C., ... Gurven, M. (2010). Adaptations in humans for assessing physical strength from the voice. *Proceedings of the Royal Society B: Biological Sciences, 277*(1699), 3509–3518. https://doi.org/10.1098/rspb.2010.0769.

Sell, A., Tooby, J., & Cosmides, L. (2009). Formidability and the logic of human anger. *Proceedings of the National Academy of Sciences, 106*(35), 15073–15078.

Seyfarth, R. M., & Cheney, D. L. (2017). The origin of meaning in animal signals. *Animal Behaviour, 124*, 339–346.

Shirazi, T. N., Puts, D. A., Escasa-Dorne, M. J., et al. (2018). Filipino women' s preferences for male voice pitch : intra-individual, life history, and hormonal predictors. *Adaptive Human Behavior and Physiology, 4*(2), 188–206.

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, 43*(2), 257–269. https://doi.org/10.1007/s10919-019-00307-0.

Stulp, G., Buunk, A. P., Verhulst, S., & Pollet, T. V. (2012). High and mighty: Height increases authority in professional refereeing. *Evolutionary Psychology, 10*(3), 588–601.

Stulp, G., Buunk, A. P., Verhulst, S., & Pollet, T. V. (2015). Human height Is positively related to interpersonal dominance in dyadic interactions. *PLoS One, 10*(2), 1–18. https://doi.org/10.1371/journal.pone.0117860.

Szamado, S. (2000). Cheating as a mixed strategy in a simple model of aggressive communication. *Animal Behaviour, 59*, 221–230.

Taylor, A. M., Reby, D., & McComb, K. (2010). Size communication in domestic dog, *Canis familiaris*, growls. *Animal Behaviour, 79*, 205–210.

Team, R. C. (2017). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation : R package for causal mediation analysis. *Journal of Statistical Software, 59*(5), https://doi.org/10.18637/jss.v059.i05.

Titze, I. R. (2000). *Principles of voice production.* National Center for Voice and Speech.

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.

van den Berg, P., & Weissing, F. (2015). Evolutionary game theory and personality. In V. Zeigler-Hill, L. Welling, & T. Shackelford (Eds.). *Evolutionary perspectives on social psychology*. Springer.

van Quaquebeke, N., & Giessner, S. R. (2010). How embodied cognitions affect judgments: height-related attribution bias in football foul calls. *Journal of Sport and Exercise Psychology, 32*(1), 3–22. https://doi.org/10.1123/jsep.32.1.3.

Vannoni, E., & McElligott, A. G. (2008). Low frequency groans indicate larger and more dominant fallow deer (*Dama dama*) males. *PLoS One, 3*(9), Article e3113. https://doi.org/10.1371/journal.pone.0003113.

Watkins, C. D., Fraccaro, P. J., Smith, F. G., Vukovic, J., Feinberg, D. R., Debruine, L. M., & Jones, B. C. (2010). Taller men are less sensitive to cues of dominance in other men. *Behavioral Ecology, 21*(5), 943–947. https://doi.org/10.1093/beheco/arq091.

Webster, M. S., Ligon, R. A., & Leighton, G. M. (2018). Social costs are an under-appreciated force for honest signalling in animal aggregations. *Animal Behaviour, 143*, 167–176. https://doi.org/10.1016/j.anbehav.2017.12.006.

Wolff, S. E., & Puts, D. A. (2010). Vocal masculinity is a robust dominance signal in men. *Behavioral Ecology and Sociobiology, 64*(10), 1673–1683. https://doi.org/10.1007/s00265-010-0981-5.

Zaidi, A. A., White, J. D., Mattern, B. C., Liebowitz, C. R., Puts, D. A., Claes, P., & Shriver, M. D. (2019). Facial masculinity does not appear to be a condition-dependent male ornament and does not reflect MHC heterozygosity in humans. *Proceedings of the National Academy of Sciences, 116*(5), 1633–1638. https://doi.org/10.1073/pnas.1808659116.

Zhang, J., Hodges-Simeon, C., Gaulin, S. J. C., & Reid, S. A. (2020). Pitch lowering enhances men's perceived aggressive intent, not fighting ability. *Evolution and Human Behavior.*. https://doi.org/10.1016/j.evolhumbehav.2020.07.007.

Zhang, J., & Reid, S. A. (2017). Aggression in young men high in threat potential increases after hearing low-pitched male voices: Two tests of the retaliation-cost model. *Evolution and Human Behavior, 38*(4), 513–521. https://doi.org/10.1016/j.evolhumbehav.2017.02.005.