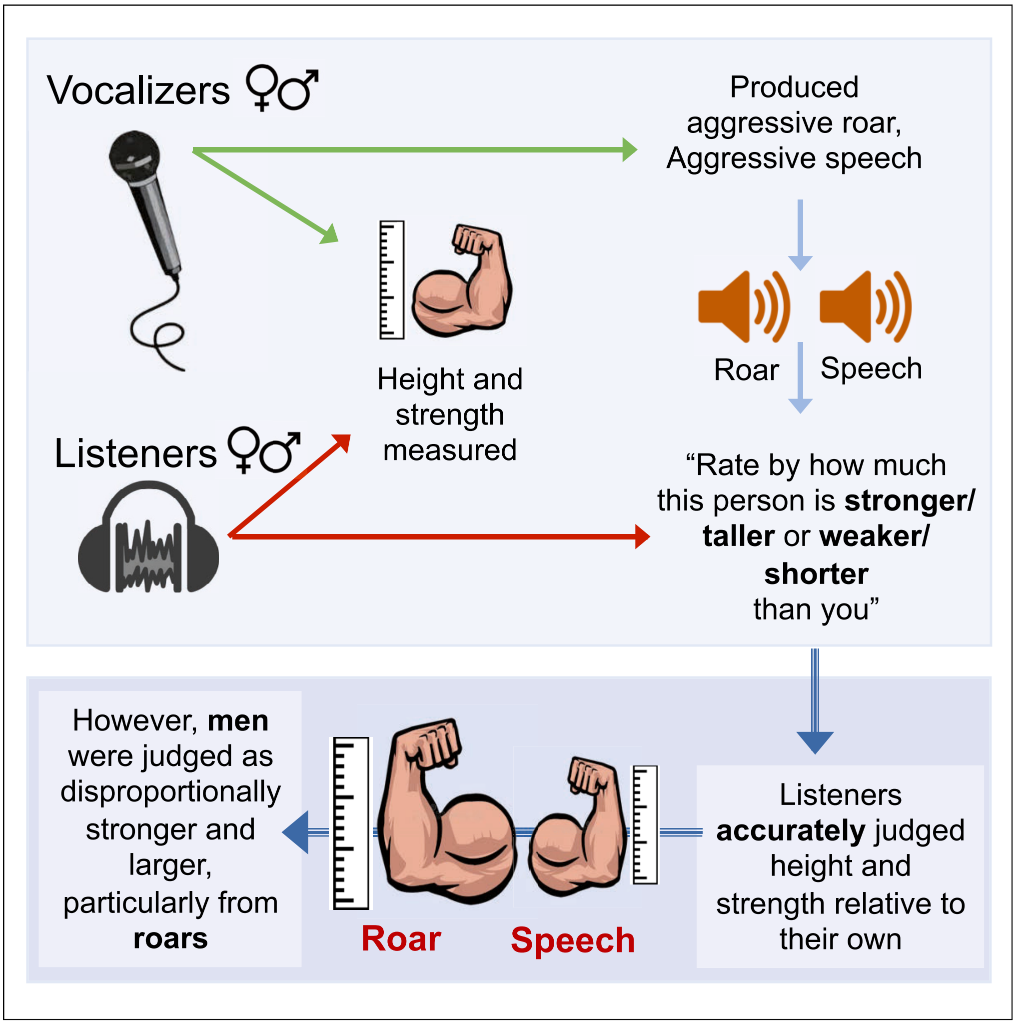
iScience

CellPress

Article

Human Listeners Can Accurately Judge Strength and Height Relative to Self from Aggressive Roars and Speech



Jordan Raine, Katarzyna Pisanski, Anna Oleszkiewicz, Julia Simner, David Reby

jordan@raineonline.co.uk

HIGHLIGHTS

We measured the strength and height of men and women (speakers and listeners)

Listeners rated the strength/height of speakers relative to their own, from roars and speech

Despite sex biases, listeners accurately judged relative strength/ height from voice

In males only, roars maximized the expression of threat compared to aggressive speech

Raine et al., iScience 4, 273–280
June 29, 2018
https://doi.org/10.1016/j.isci.2018.05.002

iScience

CellPress

Article

Human Listeners Can Accurately Judge Strength and Height Relative to Self from Aggressive Roars and Speech

Jordan Raine,[1,*] Katarzyna Pisanski,[1,2] Anna Oleszkiewicz,[2,3] Julia Simner,[1] and David Reby[1,4,5]

SUMMARY

Although animal vocalizations and human speech are known to communicate physical formidability, no previous study has examined whether human listeners can assess the strength or body size of vocalizers *relative to their own*, either from speech or from nonverbal vocalizations. Here, although men tended to underestimate women's formidability, and women to overestimate men's, listeners judged relative strength and height from aggressive roars and aggressive speech accurately. For example, when judging roars, male listeners accurately identified vocalizers who were substantially stronger than themselves in 88% of trials, and never as weaker. For male vocalizers only, roars functioned to exaggerate the expression of threat compared to aggressive speech, as men were rated as relatively stronger when producing roars. These results indicate that, like other mammals, the acoustic structure of human aggressive vocal signals (and in particular roars) may have been selected to communicate functional information relevant to listeners' survival.

INTRODUCTION

In nonhuman mammals, vocal cues to body size (a proxy of physical formidability and threat potential) mediate behavior in agonistic male-male interactions (koalas: Charlton et al., 2013b; sea lions: Charrier et al., 2011; fallow deer: Pitcher et al., 2015; red deer: Reby et al., 2005; domestic dogs: Taylor et al., 2010). While the nonverbal components of human speech also signal physical formidability, actual height and strength typically explain only a small proportion of variance in listeners' voice-based judgments of absolute height (Charlton et al., 2013a; Ives et al., 2005; Pisanski et al., 2014a; Rendall et al., 2007; Smith and Patterson, 2005), absolute strength (Sell et al., 2010), or relative height of two same-sex vocalizers (e.g., Charlton et al., 2013a, 2013b; Pisanski et al., 2014a; Rendall et al., 2007). To our knowledge, the capacity of listeners to assess the formidability of a vocalizer relative to their own, which should be particularly ecologically relevant in competitive or threatening contexts (to decide whether to flee or fight), as well as in mate choice contexts (e.g., assortative mating preferences for body size, Fink et al., 2007; Pawlowski, 2003), has not yet been investigated.

Here, to address this crucial shortcoming, we investigate whether listeners can estimate the strength and height of vocalizers relative to their own from two ecologically relevant vocal signals (aggressive roars and aggressive speech), recorded from 31 men and 30 women (see Supplemental Information for audio examples). We quantified the strength of vocalizers and listeners using a standardized amalgamated measure of flexed bicep circumference and handgrip strength and measured height via metric tape or self-report. In two playback experiments, we asked separate samples of listeners to estimate the strength (26 men, 19 women) or height (25 men, 31 women) of all vocalizers relative to their own for both speech types. Stimuli were rated on a sliding 101-point scale from −50 (much weaker/shorter) to 50 (much stronger/taller) and presented in a randomized order.

RESULTS AND DISCUSSION

Strength did not correlate with height among either male ($r = -.04$, p = .833) or female ($r = .083$, p = .655) vocalizers. Therefore, at least in our sample, these two physical measurements appear to characterize distinct aspects of physical formidability.

Judgments of Relative Strength

We ran a linear mixed multinomial logistic regression with the actual strength difference between vocalizer and listener, vocalizer sex, listener sex, and stimulus type (roar versus speech) as predictors, and included

[1]School of Psychology, University of Sussex, Brighton, UK

[2]Institute of Psychology, University of Wrocław, Wrocław, Poland

[3]Taste and Smell Center, Department of Otorhinolaryngology, Technische Universität Dresden, Dresden, Germany

[4]Senior author

[5]Lead Contact

*Correspondence: jordan@raineonline.co.uk

https://doi.org/10.1016/j.isci.2018.05.002

Source	df_1, df_2	F	p
i. Intercept	33, 5135	23.37	**<0.001**
ii. Actual strength difference	4, 5135	19.03	**<0.001**
iii. Vocalizer sex	1, 5135	78.59	**<0.001**
iv. Listener sex	1, 5135	3.73	0.054
v. Stimulus type	1, 5135	4.91	**0.027**
vi. Actual strength difference × vocalizer sex	4, 5135	3.25	**0.011**
vii. Actual strength difference × listener sex	4, 5135	2.97	**0.018**
viii. Actual strength difference × stimulus type	4, 5135	0.52	0.720
ix. Vocalizer sex × listener sex	1, 5135	4.21	**0.040**
x. Vocalizer sex × stimulus type	1, 5135	14.91	**<0.001**
xi. Listener sex × stimulus type	1, 5135	0.56	0.453
xii. Strength difference × vocalizer sex × listener sex	1, 5135	0.67	0.412
xiii. Strength difference× vocalizer sex× stimulus type	4, 5135	3.60	**0.006**
xiv. Strength difference × listener sex × stimulus type	4, 5135	0.37	0.832
xv. Vocalizer sex × listener sex × stimulus type	1, 5135	0.01	0.932
xvi. Strength difference × vocalizer sex × listener sex × stimulus type	1, 5135	1.30	0.255

Table 1. Mixed Multinomial Logistic Regression Examining Listeners' Strength Ratings as a Function of the Categorized Actual Difference in Strength Between Listener and Vocalizer, Vocalizer Sex, Listener Sex, and Stimulus Type

p value in bold are statistically significant at an alpha level of 0.05.

the relative strength difference as a categorical outcome variable. The model showed that, overall, the actual strength difference was a significant predictor of the perceived strength difference (Table 1, entry ii). Relatively stronger vocalizers were rated as relatively stronger, and vice versa (Figure 1). This demonstrates that listeners of both sexes are capable of making accurate functional judgments of the strength of other men and women, relative to their own, from both verbal and nonverbal vocal stimuli.

The model showed a significant main effect of vocalizer sex (Table 1, entry iii), with male vocalizers overall more likely to be judged as relatively stronger than females, and vice versa, independent of the actual strength difference between the vocalizer and listener (Figure 1). The main effects of vocalizer sex and actual strength difference interacted significantly (Table 1, entry vi), with listeners more likely to judge relatively weaker males, but relatively stronger females, as of similar strength to themselves than relatively stronger males or weaker females (Figure 1). We also observed a significant interaction between listener sex and actual strength difference (Table 2, entry vii). Female listeners were more likely to judge vocalizers as stronger or of similar strength to themselves than male listeners, except when the vocalizer was much weaker or much stronger.

The combined effects of vocalizer sex and listener sex resulted in a tendency for male listeners to underestimate the relative strength of female vocalizers (Figures 1A and 1C), and for female listeners to overestimate the relative strength of male vocalizers (Figures 1B and 1D). The significant interaction between listener sex and vocalizer sex (Table 1, entry ix) indicated that female listeners overestimated male vocalizers more than expected from the combined main effects (Figure 1). Together, these results suggest that listeners, particularly females, may overgeneralize population-level sex differences in strength (Bishop et al., 1987; see Lassek and Gaulin, 2009 for a review). Such sex-based overgeneralizations are common in human perception of nonverbal vocal cues (Reby et al., 2016; Rendall et al., 2007), and are likely to reflect stereotypical biases. The stronger bias among female than male listeners is consistent with previous indications that women perceive gender differences to be larger than do men, across a wide range of psychological traits (Zell et al., 2016).

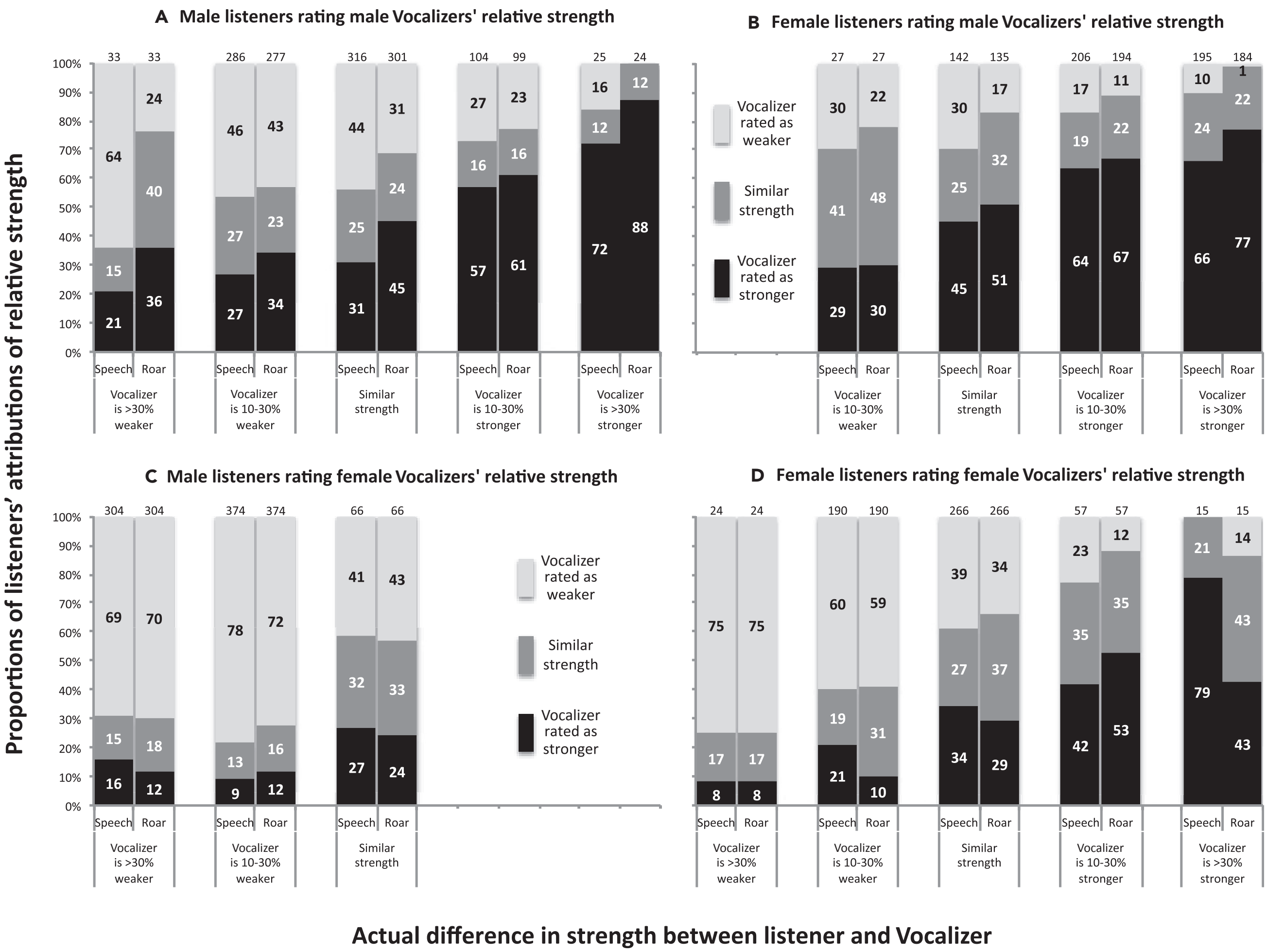


Figure 1. Relative Strength Ratings as a Function of Actual Strength Differences

Percentage of listeners judging vocalizers as relatively weaker (black), of similar strength (dark gray), or as relatively stronger (light gray) than themselves, as a function of the actual difference in strength between the listener and vocalizer. Separate graphs are reported for (A) male listeners rating male vocalizers, (B) female listeners rating male vocalizers, (C) male listeners rating female vocalizers, and (D) female listeners rating female vocalizers. Within each panel, for each actual strength difference category, separate bars are reported for listeners rating aggressive speech (left) and aggressive roars (right).

Finally, the model revealed a significant main effect of stimulus type, showing that overall, listeners were more likely to rate vocalizers as stronger or of similar strength to themselves when judging roars compared to speech. A significant interaction with vocalizer sex (Table 1, entry x) indicated that this was only the case when listeners rated male vocalizers (Figure 1). Furthermore, a three-way interaction between stimulus type, vocalizer sex, and actual strength difference indicated that this effect was strongest when male vocalizers were much weaker than male listeners, and was reversed in substantially stronger female vocalizers (Table 1, entry xiii, Figure 1). This suggests that although male roars increase the perceived difference in strength between listeners and vocalizers, compared to aggressive speech, this difference is particularly functional in the weakest male vocalizers. However, roaring in females does not appear to function to exaggerate perceived strength, and for particularly strong females, it may in fact minimize perceived strength.

Judgments of Relative Height

We ran a second linear mixed multinomial logistic regression with the actual height difference between vocalizer and listener, vocalizer sex, listener sex, and stimulus type as predictors, and included height difference as a categorical outcome variable. The model showed that overall, the actual height difference was a significant predictor of the perceived height difference (Table 2, entry ii). Relatively taller vocalizers were

Source	df_1, df_2	F	p
i. Intercept	33, 6738	31.51	<0.001
ii. Actual height difference	4, 6738	5.26	<0.001
iii. Vocalizer sex	1, 6738	193.37	<0.001
iv. Listener sex	1, 6738	25.43	<0.001
v. Stimulus type	1, 6738	3.62	0.057
vi. Actual height difference * vocalizer sex	3, 6738	0.60	0.616
vii. Actual height difference * listener sex	4, 6738	3.47	0.008
viii. Actual height difference * stimulus type	4, 6738	0.50	0.735
ix. Vocalizer sex * listener sex	1, 6738	0.60	0.438
x. Vocalizer sex * stimulus type	1, 6738	6.01	0.014
xi. Listener sex * stimulus type	1, 6738	0.01	0.951
xii. Height difference * vocalizer sex * listener sex	2, 6738	4.24	0.014
xiii. Height difference * vocalizer sex * stimulus type	3, 6738	0.34	0.794
xiv. Height difference * listener sex * stimulus type	4, 6738	0.32	0.865
xv. Vocalizer sex * listener sex * stimulus type	1, 6738	1.21	0.272
xvi. Height difference * vocalizer sex * listener sex * stimulus type	2, 6738	0.33	0.722

Table 2. Mixed Multinomial Logistic Regression Examining Listeners' Height Ratings as a Function of the Categorized Actual Difference in Height between Listener and Vocalizer, Vocalizer Sex, Listener Sex, and Stimulus Type

rated as relatively taller, and vice versa (Figure 2). This demonstrates that listeners of both sexes can judge the body size of other men and women, relative to their own, from both verbal and nonverbal stimuli.

This effect was qualified by an interaction with listener sex, whereby male listeners were more sensitive to relative size variation than were female listeners: as actual size differences increased, male listeners were increasingly more likely to rate the vocalizer as relatively taller than were female listeners. These findings support the hypothesis that size assessment abilities may have arisen primarily through male-male competition (see Puts, 2010 for additional discussion), and are consistent with previous observations that men are better than women at estimating body size from synthesized vocal stimuli (Charlton et al., 2013a). A significant three-way interaction between actual height difference, listener sex, and vocalizer sex indicated that the effect of actual height difference was minimal when female listeners rated female vocalizers (Figure 2D). This is consistent with evidence that male body size plays a role in female mate choice (Bruckert et al., 2006; Sell et al., 2017).

The model showed a significant main effect of vocalizer sex (Table 2, entry iii), with male vocalizers more likely to be judged as taller relative to the listener than female vocalizers, and vice versa, independent of the actual height difference between the vocalizer and listener (Figure 2). The main effect of listener sex was also significant (Table 2, entry iv), showing that female listeners were generally more likely to judge vocalizers as relatively taller than or of similar height to themselves than were male listeners (Figure 2). Thus, as with strength, male listeners tended to underestimate the relative height of female vocalizers (Figures 2A and 2C), and vice versa (Figures 1B and 1D). This suggests that sexual dimorphism in actual height in adult humans (i.e., men are approximately 7%–10% taller than women, Pisanski et al., 2014b) may induce disproportionate sex-dependent biases in listeners' relative height judgments.

Lastly, the interaction between stimulus type and vocalizer sex was significant (Table 2, entry x), with listeners more likely to rate male vocalizers (but not female vocalizers) as taller than or of similar height to

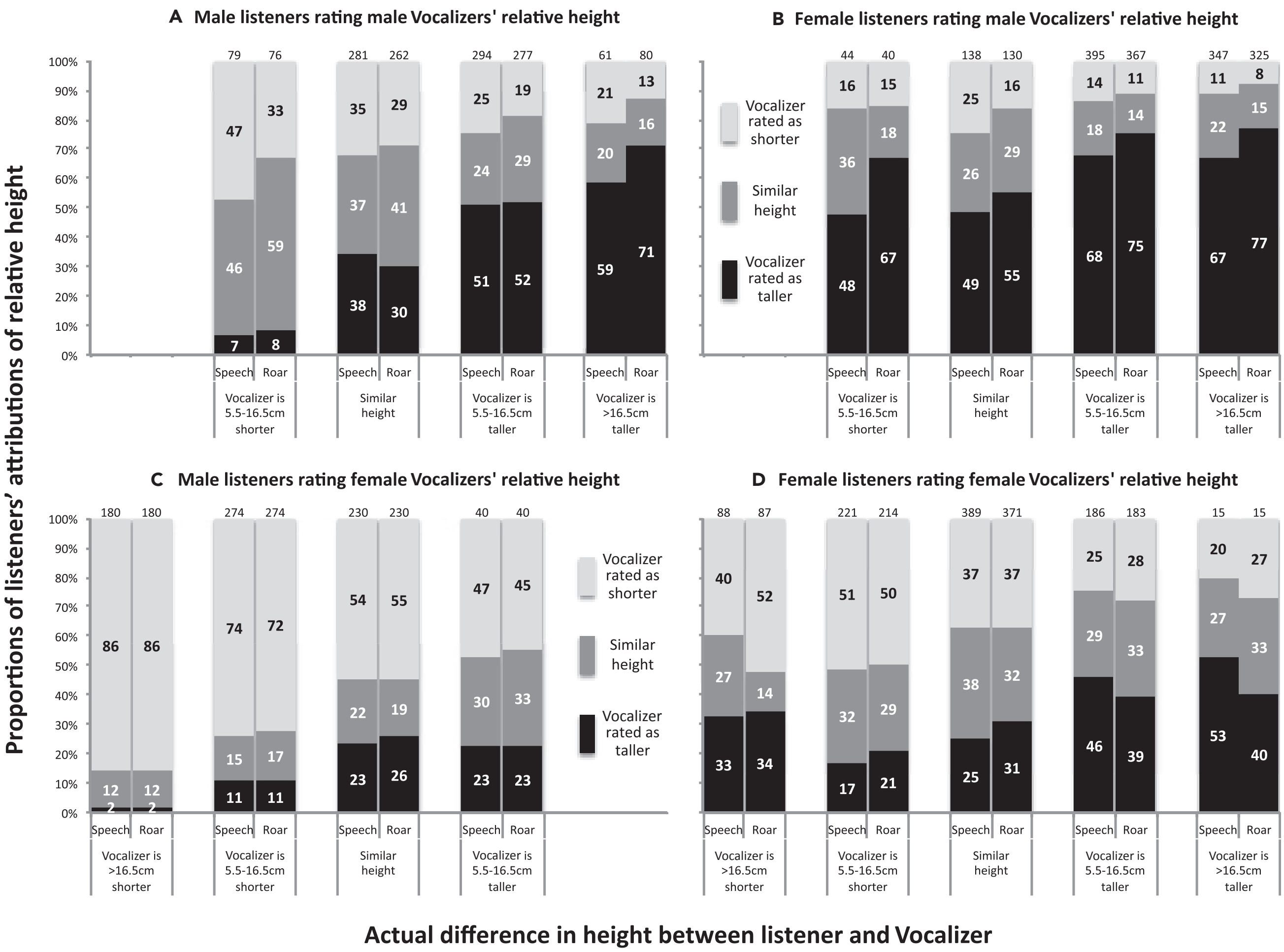


Figure 2. Relative Height Ratings as a Function of Actual Height Differences

Percentage of listeners judging vocalizers as relatively shorter (black), of similar height (dark gray), or as relatively taller (light gray) than themselves, as a function of the actual difference in height between the listener and vocalizer. Separate graphs are reported for (A) male listeners rating male vocalizers, (B) female listeners rating male vocalizers, (C) male listeners rating female vocalizers, and (D) female listeners rating female vocalizers. Within each panel, for each actual height difference category, separate bars are reported for listeners rating aggressive speech (left) and aggressive roars (right).

themselves when judging roars than speech (Figure 2). This is consistent with the hypothesis that roars serve to exaggerate physical formidability, as observed in nonhuman mammals (Charlton et al., 2011; Harris et al., 2006; Reby and McComb, 2003).

DISCUSSION

Earlier investigations of humans' capacity to estimate physical formidability from the voice have exclusively focused on absolute judgments of height or strength (e.g., Bruckert et al., 2006; Sell et al., 2010; Smith and Patterson, 2005) or comparisons between pairs of vocalizers (e.g., Charlton et al., 2013a; Pisanski et al., 2014a; Rendall et al., 2007). Our results provide the first evidence that listeners are able to estimate the formidability of vocalizers relative to their own, a judgment perhaps more closely aligned with the hypothesized central role of mate competition in selecting for the communication of formidability (Hill et al., 2013; Hill et al., 2017).

Indeed, whereas previous studies typically report that strength and height explain relatively modest proportions of variance in listeners' formidability judgments, we show that both male and female listeners can use available formidability cues conveyed in aggressive speech and roars to make ecologically relevant judgments about speakers with a high degree of accuracy. For example, listeners erroneously judged relatively stronger vocalizers as weaker on only 18% of trials, and substantially stronger vocalizers as weaker on only 6% of trials. Moreover, the finding that female listeners estimated strength (but not height) with high accuracy adds to a small but growing body of evidence suggesting that the capacity to assess strength may not only derive from sexual selection for mate competition, but also from female mate choice, with some researchers arguing that body size is less important than strength to females' judgments of males' attractiveness (Sell et al., 2017).

Male vocalizers were more likely to be perceived as stronger relative to listeners when producing roars than when producing aggressive speech. This effect was more pronounced when strength differences were extreme, with listeners almost never (less than 1% of cases) rating substantially stronger male vocalizers as weaker than themselves when judging roars. In turn, male listeners correctly identified substantially weaker vocalizers as weaker on only 24% of trials when judging roars. Our results thus support the hypothesis that men's roars, like many of their nonhuman analogs, are sexually selected to exaggerate formidability in male-male competitive interactions (Charlton et al., 2011; Harris et al., 2006; Morton, 1977; Reby and McComb, 2003), but may also afford advantages to males in mate choice contexts (Charlton et al., 2012; Charlton et al., 2007a, 2007b), likely as a result of resource holding potential benefits conferred by greater formidability (Brues, 1959; Frederick and Haselton, 2007; Gallup et al., 2007; Judge and Cable, 2004; Monden and Smits, 2009; Pisanski and Feinberg, 2013; Sell et al., 2017).

The observation that women were more likely to rate vocalizers as relatively stronger than were men at the same actual difference in strength is consistent with a general tendency for women to underestimate, and for men to overestimate, their skills and abilities (Bleidorn et al., 2016; Ehrlinger and Dunning, 2003; Erkut, 1983; Freund and Kasten, 2012; Gold et al., 1980; Kosakowska-Berezecka et al., 2017; Syzmanowicz and Furnham, 2011). Of particular interest is that women correctly identified relatively weaker male vocalizers on only 25% of trials, and tended to judge similar strength male vocalizers as stronger than themselves. Awareness of this negative bias may inform confidence-based interventions (already shown to ameliorate the "confidence gap" in cognitive tasks, Bench et al., 2015; Ehrlinger and Dunning, 2003; Estes and Felker, 2012) in sexual assault resistance programs (Jordan and Mossman, 2017; Senn et al., 2015, 2017; Wong and Balemba, 2016).

Future work could make use of outlier populations (e.g., bodybuilders) to examine the accuracy of and biases in strength estimation at extremes of strength and to ascertain how male listeners assess the relative strength of females who are stronger than them. In addition, given that in many nonhuman mammals acoustic cues to formidability mediate dyadic agonistic interactions between competing males (for example, large but not small dogs respond differentially to playback conditions simulating relatively smaller or larger conspecifics [Taylor et al., 2010]), it is assumed that nonhuman mammals are also able to assess the formidability of opponents relative to their own. To empirically verify this prediction, future research should now further examine how between-individual variation in the formidability of nonhuman receivers mediates vocal behavior (e.g., call response latency, calling rate, Charlton et al., 2013b; Reby et al., 2005).

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Transparent Methods and three data files and can be found with this article online at https://doi.org/10.1016/j.isci.2018.05.002.

ACKNOWLEDGMENTS

The authors thank James Alvarez for programming the experiments in the SynToolkit online platform. This work was supported by the University of Sussex through a scholarship to J.R., the European Commission through a Marie Skłodowska-Curie individual fellowship to K.P. (H2020-MSCA-IF-2014-655859), the Foundation for Polish Science through a START scholarship and the Ministry of Science and Higher Education through a scholarship (#626/STYP/12/2017) to A.O, and by the Iuventus Plus grant from the Polish Ministry of Science and Higher Education to K.P. (0619/IP3/2016/74).

AUTHORS CONTRIBUTIONS

J.R., D.R., and K.P designed the investigation. J.R., K.P., and A.O. collected the data. J.R performed acoustic and statistical analysis and created the figures. J.R and D.R. wrote the manuscript. J.S. provided customized access to a proprietary online platform to run playback experiments. The manuscript was reviewed, edited, and approved by all authors, who agree to be accountable for the work.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 26, 2018
Revised: April 2, 2018
Accepted: April 6, 2018
Published: June 28, 2018

REFERENCES

Bench, S.W., Lench, H.C., Liew, J., Miner, K., and Flores, S.A. (2015). Gender gaps in overestimation of math performance. Sex Roles *72*, 536–546.

Bishop, P., Cureton, K., and Collins, M. (1987). Sex difference in muscular strength in equally-trained men and women. Ergonomics *30*, 675–687.

Bleidorn, W., Arslan, R.C., Denissen, J.J.A., Rentfrow, P.J., Gebauer, J.E., Potter, J., and Gosling, S.D. (2016). Age and gender differences in self-esteem - a cross-cultural window. J. Pers. Soc. Psychol. *111*, 396–410.

Bruckert, L., Liénard, J.-S., Lacroix, A., Kreutzer, M., and Leboucher, G. (2006). Women use voice parameters to assess men's characteristics. Proc. R. Soc. Lond. B Biol. Sci. *273*, 83–89.

Brues, A. (1959). The spearman and the archer - an essay on selection in body build. Am. Anthropol. *61*, 457–469.

Charlton, B.D., Reby, D., and McComb, K. (2007a). Female perception of size-related formant shifts in red deer, *Cervus elaphus*. Anim. Behav. *74*, 707–714.

Charlton, B.D., Reby, D., and McComb, K. (2007b). Female red deer prefer the roars of larger males. Biol. Lett. *3*, 382–385.

Charlton, B.D., Ellis, W.A.H., McKinnon, A.J., Cowin, G.J., Brumm, J., Nilsson, K., and Fitch, W.T. (2011). Cues to body size in the formant spacing of male koala (*Phascolarctos cinereus*) bellows: honesty in an exaggerated trait. J. Exp. Biol. *214*, 3414–3422.

Charlton, B.D., Ellis, W.A.H., Brumm, J., Nilsson, K., and Fitch, W.T. (2012). Female koalas prefer bellows in which lower formants indicate larger males. Anim. Behav. *84*, 1565–1571.

Charlton, B.D., Taylor, A.M., and Reby, D. (2013a). Are men better than women at acoustic size judgements? Biol. Lett. *9*, 20130270.

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

Charrier, I., Ahonen, H., and Harcourt, R.G. (2011). What makes an Australian sea lion (*Neophoca cinerea*) male's bark threatening? J. Comp. Psychol. *125*, 385–392.

Ehrlinger, J., and Dunning, D. (2003). How chronic self-views influence (and potentially mislead) estimates of performance. J. Pers. Soc. Psychol. *84*, 5–17.

Erkut, S. (1983). Exploring sex differences in expectancy, attribution, and academic achievement. Sex Roles *9*, 217–231.

Estes, Z., and Felker, S. (2012). Confidence mediates the sex difference in mental rotation performance. Arch. Sex. Behav. *41*, 557–570.

Fink, B., Neave, N., Brewer, G., and Pawlowski, B. (2007). Variable preferences for sexual dimorphism in stature (SDS): further evidence for an adjustment in relation to own height. Personal. Individ. Differ. *8*, 2249–2257.

Frederick, D.A., and Haselton, M.G. (2007). Why is muscularity sexy? Tests of the fitness indicator hypothesis. Personal. Soc. Psychol. Bull. *33*, 1167–1183.

Freund, P.A., and Kasten, N. (2012). How smart do you think you are? A meta-analysis on the validity of self-estimates of cognitive ability. Psychol. Bull. *138*, 296–321.

Gallup, A.C., White, D.D., and Gallup, G.G. (2007). Handgrip strength predicts sexual behavior, body morphology, and aggression in male college students. Evol. Hum. Behav. *28*, 423–429.

Gold, A.R., Brush, L.R., and Sprotzer, E.R. (1980). Developmental changes in self-perceptions of intelligence and self-confidence. Psychol. Women Q. *5*, 231–239.

Harris, T.R., Fitch, W.T., Goldstein, L.M., and Fashing, P.J. (2006). Black and white colobus monkey (*Colobus guereza*) roars as a source of both honest and exaggerated information about body mass. Ethology *112*, 911–920.

Hill, A.K., Bailey, D.H., and Puts, D.A. (2017). Gorillas in our midst? Human sexual dimorphism and contest competition in men. In On Human Nature, M. Tibayrenc and F.J. Ayala, eds. (Academic Press), pp. 235–249.

Hill, A.K., Hunt, J., Welling, L.L.M., Cárdenas, R.A., Rotella, M.A., Wheatley, J.R., and Puts, D.A. (2013). Quantifying the strength and form of sexual selection on men's traits. Evol. Hum. Behav. *34*, 334–341.

Ives, D.T., Smith, D.R., and Patterson, R.D. (2005). Discrimination of speaker size from syllable phrases. J. Acoust. Soc. Am. *118*, 3816–3822.

Jordan, J., and Mossman, E. (2017). "Back off buddy, this is my body, not yours": empowering girls through self-defense. Violence Against Women, 1077801217741217. https://doi.org/10.1177/1077801217741217.

Judge, T.A., and Cable, D.M. (2004). The effect of physical height on workplace success and income: preliminary test of a theoretical model. J. Appl. Psychol. *89*, 428–441.

Kosakowska-Berezecka, N., Jurek, P., Besta, T., and Badowska, S. (2017). Self-presentation strategies, fear of success and anticipation of future success among university and high school students. Front. Psychol. *8*, 1884.

Lassek, W.D., and Gaulin, S.J.C. (2009). Costs and benefits of fat-free muscle mass in men: relationship to mating success, dietary requirements, and native immunity. Evol. Hum. Behav. *30*, 322–328.

Monden, C.W.S., and Smits, J. (2009). Maternal height and child mortality in 42 developing countries. Am. J. Hum. Biol. *21*, 305–311.

Morton, E.S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. Am. Nat. *111*, 855–869.

Pawlowski, B. (2003). Variable preferences for sexual dimorphism in height as a strategy for increasing the pool of potential partners in humans. Proc. R. Soc. Lond. B Biol. Sci. *270*, 709–712.

Pisanski, K., and Feinberg, D.R. (2013). Cross-cultural variation in mate preferences for averageness, symmetry, body size, and masculinity. Cross-Cult. Res. *47*, 162–197.

Pisanski, K., Fraccaro, P.J., Tigue, C.C., O'Connor, J.J.M., and Feinberg, D.R. (2014a). Return to Oz: voice pitch facilitates assessments of men's body size. J. Exp. Psychol. Hum. Percept. Perform. *40*, 1316–1331.

Pisanski, K., Fraccaro, P.J., Tigue, C.C., O'Connor, J.J.M., Röder, S., Andrews, P.W., and Feinberg, D.R. (2014b). Vocal indicators of body size in men and women: a meta-analysis. Anim. Behav. *95*, 89–99.

Pitcher, B.J., Briefer, E.F., and McElligott, A.G. (2015). Intrasexual selection drives sensitivity to pitch, formants and duration in the competitive calls of fallow bucks. BMC Evol. Biol. *15*, 149.

Puts, D.A. (2010). Beauty and the beast: mechanisms of sexual selection in humans. Evol. Hum. Behav. *31*, 157–175.

Reby, D., Levréro, F., Gustafsson, E., and Mathevon, N. (2016). Sex stereotypes influence adults' perception of babies' cries. BMC Psychol. *4*, 19.

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

Reby, D., McComb, K., Cargnelutti, B., Darwin, C., Fitch, W.T., and Clutton-Brock, T. (2005). Red deer stags use formants as assessment cues during intrasexual agonistic interactions. Proc. R. Soc. Lond. B Biol. Sci. *272*, 941–947.

Rendall, D., Vokey, J.R., and Nemeth, C. (2007). Lifting the curtain on the Wizard of Oz: biased voice-based impressions of speaker size. J. Exp. Psychol. Hum. Percept. Perform. *33*, 1208–1219.

Sell, A., Bryant, G.A., Cosmides, L., Tooby, J., Sznycer, D., von Rueden, C., and Gurven, M. (2010). Adaptations in humans for assessing physical strength from the voice. Proc. R. Soc. B Biol. Sci. *277*, 3509–3518.

Sell, A., Lukazsweski, A.W., and Townsley, M. (2017). Cues of upper body strength account for most of the variance in men's bodily attractiveness. Proc. Biol. Sci. *284*, 20171819.

Senn, C.Y., Eliasziw, M., Barata, P.C., Thurston, W.E., Newby-Clark, I.R., Radtke, H.L., and Hobden, K.L. (2015). Efficacy of a sexual assault resistance program for university women. N. Engl. J. Med. *372*, 2326–2335.

Senn, C.Y., Eliasziw, M., Hobden, K.L., Newby-Clark, I.R., Barata, P.C., Radtke, H.L., and Thurston, W.E. (2017). Secondary and 2-year outcomes of a sexual assault resistance program for university women. Psychol. Women Q. *41*, 147–162.

Smith, D.R., and Patterson, R.D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. J. Acoust. Soc. Am. *118*, 3177–3186.

Syzmanowicz, A., and Furnham, A. (2011). Gender differences in self-estimates of general, mathematical, spatial and verbal intelligence: four meta analyses. Learn. Individ. Differ. *21*, 493–504.

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

Wong, J.S., and Balemba, S. (2016). The effect of victim resistance on rape completion: a meta-analysis. Trauma Violence Abuse. https://doi.org/10.1177/1524838016663934.

Zell, E., Strickhouser, J.E., Lane, T.N., and Teeter, S.R. (2016). Mars, venus, or earth? Sexism and the exaggeration of psychological gender differences. Sex Roles *75*, 287–300.

ISCI, Volume *4*

Supplemental Information

Human Listeners Can Accurately Judge Strength and Height Relative to Self from Aggressive Roars and Speech

Jordan Raine, Katarzyna Pisanski, Anna Oleszkiewicz, Julia Simner, and David Reby

Supplemental Information

Transparent Methods

All experiments were approved by the University of Sussex's Life Sciences & Psychology Cluster-based Research Ethics Committee (C-REC) (Certificates of approval: ER/JR307/8, ER/JR307/9) and comply with the American Psychological Association's Ethical Principles of Psychologists and Code of Conduct.

Participants

Vocal stimuli were recorded from 30 male and 31 female (*M* age = 22.79 ± 1.12) drama or acting students from the Royal Central School of Speech and Drama and the University of Sussex, United Kingdom, who received monetary compensation in exchange for their participation.

We recruited separate samples of participants to provide voice-based assessments of the relative strength and height of vocalizers. The sample that rated strength (hereafter Experiment 1) consisted of 19 females and 26 males (age = 31.44 ± 8.33) recruited from Tromso and surrounding rural towns in Norway (*N* = 11, all fluent English speakers), and from the University of Sussex, UK (*N* = 34), in return for prize draw monetary compensations (5 x £20). The sample that rated height (hereafter Experiment 2) consisted of 31 females and 25 males (age = 34.27 ± 10.39), recruited from the USA using Amazon Mechanical Turk, and compensated with $1.75 USD. Participants from both experiments provided informed consent and completed the experiment online using a custom computer interface. Data from one female and male participant in Experiment 1, and from two female and two male participants in Experiment 2, who did not complete the experiment but rated more than half of the stimuli, were included in our analysis.

Materials

Vocal stimuli.

Vocalizers were audio recorded producing an aggressive roar and aggressive speech in a quiet, anechoic room, standing 150 cm from a Zoom H4n microphone. A chair was placed at this distance to restrict participants from moving closer to the microphone. Vocalizers were instructed to produce the speech sentence, ‘That’s enough, I’m coming for you!’, followed by a nonverbal vocalisation expressing the same motivation, while imagining themselves in a battle or war scenario, about to charge and attack. This resulted in a total of 122 vocal stimuli (see Electronic Supplementary Materials for examples of aggressive roars and aggressive speech).

To obtain realistic vocal stimuli, participants were encouraged to take as much time as they needed to immerse themselves in each imagined context, and to ‘let go of their inhibitions’. Participants were also given the option not to vocalise if they felt that they could not naturally produce the sentence or nonverbal vocalisation, and to repeat each sentence or vocalisation until they were satisfied with their portrayal. Recordings were saved as WAV files at 44.1 kHz sampling frequency and 16 bits amplitude resolution.

Physical formidability measures.

We measured the height of vocalizers using metric tape. The average height of our sample of vocalizers (male *M* = 182.03 ± 0.97 cm; female *M* = 167.10 ± 1.19 cm) compares well with that of the general UK population (male *M* = 175.3cm, female *M* = 161.9 cm, Moody, 2013). Flexed bicep circumference and handgrip strength were also measured, and these measurements were aggregated to produce a single, equally weighted, z-scored strength value for each subject (following Sell et al. 2009; Puts et al. 2011, and others). These measures explain approximately 55% and 24% of the variance in strength as measured by weight-lifting machines in male college students, respectively (Sell et al., 2009).

To measure flexed bicep circumference (male *M* = 32.09 ± 0.60 cm; female *M* = 28.96 ± 0.70 cm), participants were instructed to rest the elbow of their dominant arm on a table while seated, to clench their fist, and to curl their forearm perpendicular to the table. The experimenter measured the circumference of the bicep at its highest point. A Baseline hydraulic hand dynamometer in its standard use was used to measure the handgrip strength of participants' dominant arm (male *M* = 41.57 ± 1.36 kg; female *M* = 26.98 ± 1.06 kg). Each strength measure was recorded twice per subject and the highest achievable score, representing greatest strength, was used in analyses.

Procedure

All playback experiments were completed online on Syntoolkit, a dedicated online testing platform for psychology studies (e.g., Hughes, Gruffydd, Simner & Ward, in press; see Simner & Alvarez, in prep) that is particularly suited to running studies with sensory or multisensory stimuli. Listeners were instructed to use headphones and complete the experiment in a quiet place. To allow listeners to complete the experiment at a comfortable but audible volume, they were instructed to first set their volume to its lowest level. Listeners then heard a demo sound file (amalgamating a loud and quiet stimulus), and were instructed to raise their volume until they could clearly hear the quiet stimulus, while the louder stimulus did not cause discomfort. Following this, listeners were asked not to adjust the volume settings during the experiment unless it became too uncomfortable, and were asked at the end of the experiment if they had done so. Due to the agonistic nature of the stimuli, listeners were made aware that if they felt uncomfortable or distressed listening to the sounds, they could stop the experiment.

In playback experiments, vocal stimuli (n = 122) were blocked by sex and stimulus type (speech/roar). The order of blocks and stimuli within blocks was randomised. Before each block, participants were reminded to listen to each stimulus in full, and informed that they could take a break at any time. Listeners rated the physical strength (Experiment 1) or height (Experiment 2) of each voice stimulus ("Rate by how much this person is stronger/taller or weaker/shorter than you") on a 101-point scale from -50 (much weaker/shorter) to 50 (much stronger/taller). We set the slider's default position to 0 (described as 'same as you') and did not compel listeners to move the slider so as not to artificially force directional judgments.

Listeners were debriefed upon completion that the roars and screams were acted, and that the vocalizers were not really experiencing aggression or distress. We examined reaction times against stimulus durations to ensure that participants listened to the stimuli before entering their ratings. No participants responded before half of the stimulus had elapsed on more than five trials, thus no listeners were excluded.

To assess whether listeners could accurately judge the physical characteristics of vocalizers relative to their own, we measured listeners' own physical characteristics. In

Experiment 1, we used a tailor’s tape measure to measure bicep circumference (male M = 33.89 ± 0.46 cm; female M = 28.12 ± 0.57 cm), and a Takei hand dynamometer to measure handgrip strength (male M = 46.11 ± 1.67 kg; female M = 33.03 ± 1.10 kg), in identical fashion to measurements taken from vocalizers. These measures were taken in person, prior to the listener completing the playback experiment online at a time of their choosing. Both vocalizer and listener strength z-scores were calculated based on a pooled sample of the listeners’ and the vocalizers’ measurements. Experiment 2 relied on a self-report measure of height given at the start of the playback experiment (male M = 176.38 ± 1.30 cm; female M = 169.36 ± 1.48 cm). The validity of self-report measures of height has been extensively studied, and despite slight overestimations, self-reported height closely reflects measured height within the age range of our sample of listeners (Krul, Daanen, & Choi, 2011; Lim, Seubsman, & Sleigh, 2009; Parker, Dillard, & Phillips, 1994; Wada et al., 2005).

Coding and Statistical Analysis

To examine strength/height estimation in functionally relevant terms, we divided the actual difference in strength/height into five categories. In Experiment 1, percentage differences between -10% and 10% were coded as ‘similar strength’, differences between ± 10% and ± 30% were coded as ‘vocalizer is stronger (weaker) than listener’, and differences greater than ± 30% were coded as ‘vocalizer is much stronger (weaker) than listener’. In Experiment 2, we calculated by how many centimetres the vocalizer was taller than the listener. Values were coded into identical categories of 11 cm intervals. This interval was chosen as it produced a similar distribution to that observed for our actual strength difference categories.

In both experiments, we coded the rated difference in strength/height between listener and vocalizer into three categories. Ratings between 45 and 55 were categorised as ‘rated as similar strength’, and ratings above (below) this range were coded as ‘vocalizer rated as stronger (weaker)’. We computed a linear mixed multinomial logistic regression, testing the effects of the actual strength/height difference between listener and vocalizer, vocalizer sex, listener sex, and stimulus type on the rated difference between listener and vocalizer, excluding actual difference categories with sample sizes less than 15. In all models, we included listener identity as a subject variable, and vocalizer identity as a random factor, thus allowing the intercepts and slopes of the relationships between predictors and outcomes to vary between both vocalizers and listeners and testing null hypotheses based on the average of these intercepts and slopes.

List of Supplemental Audio Files

(F = female vocalizer; M = male vocalizer)

F01 Roar.wav

F01 Speech.wav

F02 Roar.wav

F02 Speech.wav

F03 Roar.wav

F03 Speech.wav

M01 Roar.wav

M01 Speech.wav

M02 Roar.wav

M02 Speech.wav

M03 Roar.wav

M03 Speech.wav

References

Hughes, J.E.A., Gruffydd, E., Simner, J., and Ward, J. (in press). Synaesthesia aids in savant-skill acquisition: training calendar calculation in sequence-space synaesthetes. Cortex.

Krul A.J., Daanen H.A.M. and Choi H., Self-reported and measured weight, height and body mass index (BMI) in Italy, The Netherlands and North America, *Eur. J. Public Health* **21**, 2011, 414–419.

Lim L.L., Seubsman S. and Sleigh A., Validity of self-reported weight, height, and body mass index among university students in Thailand: implications for population studies of obesity in developing countries, *Popul. Health Metr.* **7**, 2009, 15.

Moody, A. (2013). Adult anthropometric measures, overweight and obesity (Health Survey for England - 2012). NHS. Retrieved from http://digital.nhs.uk/catalogue/PUB13218.

Parker J.M., Dillard T.A. and Phillips Y.Y., Impact of using stated instead of measured height upon screening spirometry, *Am. J. Respir. Crit. Care Med.* **150** (6 Pt 1), 1994, 1705–1708.

Sell A., Cosmides L., Tooby J., Sznycer D., von Rueden C. and Gurven M., Human adaptations for the visual assessment of strength and fighting ability from the body and face, *Proc. R. Soc. Lond. B Biol. Sci.* **276**, 2009, 575–584.

Simner, J., & Alvarez, J. (in prep). The Syntoolkit: an online testing platform for behavioural research experiments.

Puts, D.A., Apicella, C.L., & Cárdenas, R.A., Masculine voices signal men's threat potential in forager and industrial societies, *Proc. R. Soc. Lond. B Biol. Sci.* rspb20110829.

Wada K., Tamakoshi K., Tsunekawa T., Otsuka R., Zhang H., Murata C. and Toyoshima H., Validity of self-reported height and weight in a Japanese workplace population, *Int. J. Obes. (Lond.)* **29**, 2005, 1093–1099.