PROCEEDINGS OF THE ROYAL SOCIETY B

FirstCite e-publishing

Proc. R. Soc. B
doi:10.1098/rspb.2010.0769
Published online

Adaptations in humans for assessing physical strength from the voice

Aaron Sell[1,2,*], Gregory A. Bryant[4], Leda Cosmides[1,2], John Tooby[1,3], Daniel Sznycer[1,3], Christopher von Rueden[1,3], Andre Krauss[5] and Michael Gurven[1,3]

[1]*Center for Evolutionary Psychology,* [2]*Department of Psychology, and* [3]*Department of Anthropology, University of California, Santa Barbara, CA 93106, USA*
[4]*Department of Communication Studies, University of California, Los Angeles, CA 90095, USA*
[5]*Department of Psychology, 4DPPD, University of Timisoara, 300223 Timisoara, Romania*

Recent research has shown that humans, like many other animals, have a specialization for assessing fighting ability from visual cues. Because it is probable that the voice contains cues of strength and formidability that are not available visually, we predicted that selection has also equipped humans with the ability to estimate physical strength from the voice. We found that subjects accurately assessed upper-body strength in voices taken from eight samples across four distinct populations and language groups: the Tsimane of Bolivia, Andean herder-horticulturalists and United States and Romanian college students. Regardless of whether raters were told to assess height, weight, strength or fighting ability, they produced similar ratings that tracked upper-body strength independent of height and weight. Male voices were more accurately assessed than female voices, which is consistent with ethnographic data showing a greater tendency among males to engage in violent aggression. Raters extracted information about strength from the voice that was not supplied from visual cues, and were accurate with both familiar and unfamiliar languages. These results provide, to our knowledge, the first direct evidence that both men and women can accurately assess men's physical strength from the voice, and suggest that estimates of strength are used to assess fighting ability.

Keywords: resource holding power; physical strength; evolutionary psychology; voice; formidability

1. INTRODUCTION

Multiple converging lines of evidence indicate that the ability to fight has been a powerful selective force acting on human males over evolutionary time. Males have been shown to set thresholds for acceptable resource division (i.e. welfare trade-off ratios; see Tooby *et al.* 2008) based partly on physical strength: stronger men are more prone to anger (Archer & Thanzami 2007; Sell *et al.* 2009*b*). Stronger men also prevail more in conflicts of interest, and have a higher sense of entitlement to contested resources (Sell *et al.* 2009*b*). Stronger males are also rated as more physically attractive, have more sexual partners and lose their virginity at earlier ages than weaker men (Sell 2006; Frederick & Haselton 2007; Gallup *et al.* 2007). Most significantly, cross-cultural evidence indicates that a man's fighting ability is a powerful determinant of his access to resources in most, if not all, cultures (Daly & Wilson 1988). Documented cases include the Yanomamo of Venezuela (Chagnon 1983), the Achuar of Ecuador (Patton 2000), the Tsimane of Bolivia (von Rueden *et al.* 2008), the Dani of highland New Guinea (Sargent 1974), the Samoan islanders (Freeman 1983), Mexican gangs (Lewis 1961), the Montenegrins of Eastern Europe (Boehm 1984), the Inuit (Balikci 1970) and American gangs (Toch 1969). Finally, a large range of anatomical and physiological sex differences testify to an evolutionary past in which males frequently engaged in interpersonal aggression. For example, men mature later, have larger bodies (Plavcan & Van Schaik 1997), have higher basal metabolic rates (Garn & Clark 1953), larger hearts, better heat dissipation, more haemoglobin, more muscle, less fat and denser bones (Archer 2009; Lassek & Gaulin 2009).

Given the importance of men's strength and fighting ability in resolving conflicts and in female mate preferences, it would be surprising if both men and women did not have adaptive specializations for detecting relevant cues of these attributes. Such abilities have been routinely documented across a broad array of non-human species (for review, see Huntingford & Turner 1987; Krebs & Davies 1993). In fact, recent research has shown that both men and women can judge physical strength from photographs of conspecific males (Sell *et al.* 2009*a*); approximately 40 per cent of the variance in adult male strength can be detected from static photographs of the body, and 25 per cent from static photographs of the face alone.

(a) *Are there acoustic cues of fighting ability in the human voice?*

Ancestrally, there would have been many occasions when the assessment of fighting ability from visual cues would have been problematic: e.g. distance, darkness and intervening physical obstructions (as when chimpanzee male coalitions call from neighbouring territories). Such

* Author for correspondence (sell@psych.ucsb.edu).

Received 10 April 2010
Accepted 25 May 2010

conditions, if sufficiently recurrent, would have selected for supplementary assessment systems that did not rely on sight. Moreover, it seems probable that the voice conveys an array of different kinds of information relevant to conflict that is largely absent from a visual inspection of the body. For example, the vocal cords are regulated by the vagus nerve, which receives information from the viscera, and is also involved in the parasympathetic–sympathetic regulation of heart rate, blood pressure and other aspects of the so-called fight-or-flight system. The voice is likely to reveal short run changes in the maintenance or loss of fine-motor control compared with gross motor control (relevant to rage); parasympathetic/sympathetic system state; and the current tension of the individual's muscular system—itself an indicator of fear. As muscular tension goes up, pitch rises (Titze 1994). The voice may reveal longer term, more stable characteristics as well, such as anatomical relationships related to resonance, and the degree of exposure to testosterone during development. The vocal cords have androgen receptors, and sex differences in vocal fold size are surprisingly large (Titze 1994; Newman *et al.* 2000).

If the voice contains adaptively relevant information not available visually, then it seems likely that specializations evolved for extracting this information. For these reasons, it becomes important to ask whether there are non-visual cues of fighting ability to which humans may have been designed to detect and respond. A number of lines of evidence suggest that normal adult male speech may contain cues of formidability.

— Many non-human species have been shown to assess fighting ability via auditory cues; e.g. red deer (Clutton-Brock & Albon 1979), common loons (Mager *et al.* 2007), toads (Davies & Halliday 1978), baboons (Kitchen *et al.* 2003), croaking gourami (Ladich 1998) and owls (Hardouin *et al.* 2007).
— Testosterone is known to have pronounced effects on male voices (Dabbs & Mallinger 1999) and across development could plausibly affect physical strength and fighting ability (Bhasin *et al.* 1996), and aggressiveness (Archer 2006).
— Women's mate choices are affected by voice 'quality', suggesting that features of the male voice correlate with desirable attributes (Collins 2000; Hughes *et al.* 2004; Puts 2005).
— Thirty years of auditory perception research suggests a possible ability to detect height and weight from the voice, though with frequent failures to replicate (e.g. Lass *et al.* 1979*a*,*b*; Collins 2000; Gonzalez & Oliver 2004).

Measuring fighting ability in humans is ethically difficult and complicated by an evolutionary history of weaponry. Upper-body strength, however, has been shown to be a good proxy for fighting ability: it correlates with a history of aggression, self-reported success in conflicts and objective ratings of fighting ability by third parties viewing photographs of the targets (Sell *et al.* 2009*a*). An engineering analysis also shows that it is primarily upper-body strength that determines the hitting power of the weaponry available to human ancestors (Brues 1959).

(b) *Predictions*

In the eight studies presented below, we tested the following predictions derived from the theory that humans possess perceptual adaptations for extracting vocal cues of strength and fighting ability.

(i) People should be able to assess physical strength from the speaking voice of adult males.
(ii) If the ability to assess strength from the voice was the result of a general ability to detect correlations, then individuals would probably perform best on the targets they were trained on—members of their own culture. If the ability to assess strength from the voice is a specialization that was built by evolution operating ancestrally across a diverse range of cultures and languages, then the system should focus on cues that are predictive across the species range. This, in turn, predicts that people should be able to detect strength from voices drawn from other cultures and populations, speaking unfamiliar languages.
(iii) Because males deployed physical aggression more than females ancestrally, selection would have probably designed individuals to be more accurate at assessing strength from male voices than from female voices.
(iv) By hypothesis, we expect that there is information available from the voice that is not available visually. If this is true, then strength assessments are predicted to be substantially more accurate when both auditory and visual channels are available to raters.
(v) Physical strength is probably a better predictor of fighting ability than height or weight (Sell *et al.* 2009*a*,*b*). Therefore, if raters can assess height and weight from the voice at all, raters should be able to assess strength independent of height and weight.

In addition to these primary predictions, we analysed the voice samples acoustically to determine the relationship of fundamental frequency (F0) and formant dispersion (Df) to voice ratings and physical strength and body size (i.e. height and weight).

2. MATERIAL AND METHODS

(a) *Stimulus subjects*

Voice samples, body and strength measurements were taken from eight populations. For each sample, strength measures were *z*-scored and then averaged to create a single strength score for each subject that weighted each strength measure equally (Sell *et al.* 2009*a*).

(i) *Sample 1*

Sixty-three male psychology students from the University of California, Santa Barbara (UCSB), (mean age = 18.7, s.d. = 0.88, range: 18–22) were given course credit for participating.

Strength assessments: (Cronbach's $\alpha = 0.78$).
— Handgrip strength ($M = 53.2$ kg, s.d. = 8.4). A Rolyan hydraulic hand dynamometer was used to measure grip strength.
— Flexed bicep circumference ($M = 33.5$ cm, s.d. = 3.5). Approximately 50 per cent of the variance in weight-lifting strength among male college students is tracked by flexed bicep circumference (Sell *et al.* 2009*a*).
— Photo ratings. Subjects are highly accurate at judging physical strength from full body photographs (Sell *et al.* 2009*a*).

Ratings of physical strength from full body pictures explain about 50 per cent of the variance in weight-lifting strength. Full body photographs of the stimulus subjects (i.e. no shirt, standardized gym shorts) were shown to 50 UCSB undergraduates (18 males and 32 females) who rated each photograph on physical strength from 1 (very weak) to 7 (very strong). Ratings were averaged so that each stimulus subject had a single photo-strength rating.

Voice samples.

Subjects were instructed to say in their normal speaking voice, 'This is an experiment, over and out'. Voices were recorded using the built in microphone on a portable audio cassette recorder (Sony CFD-S350; frequency range = 80 Hz–10 kHz). Acoustic analyses were performed on recordings acquired in a variety of ways, many using relatively low sampling rates, and three using compression (samples 3, 5 and 7). Research has shown that F0 measurements are highly comparable across formats in the range of sampling rates used in the current studies, including recordings using lossy compression (mp3; e.g. Gonzalez *et al.* 2003; Deliyski *et al.* 2005).

Aggression measures.

In sample 1 only, subjects were asked to report how many physical fights they had engaged in during the last 4 years ($M = 4.0$, s.d. = 7.4) and were given the Physical Aggression subscale of the Buss–Perry Aggression Questionnaire ($M = 2.5$, s.d. = 0.75; Buss & Perry 1992). This scale contains items designed to measure the tendency to react with overt physical attacks such as, 'If somebody hits me, I hit back' and 'Once in a while I can't control the urge to strike another person'.

(ii) *Sample 2*

Forty-nine adult male Tsimane Indians provided voice samples in their native Tsimane language and had their strength and body size measurements taken as part of a larger project (mean age = 35.8, s.d. = 13.5, range: 19–68; von Rueden *et al.* 2008). The Tsimane language resembles Moseten (the language of a Bolivian group similar to the Tsimane), but otherwise these two are distantly related to other South American languages.

Strength assessments: (Cronbach's $\alpha = 0.79$).

— Chest strength ($M = 28.9$ kg, s.d. = 8.3). The subject pressed a Lafayette Manual Muscle Tester between the palms of his hands, with the elbows perpendicular to the torso at mid-chest height.

— Shoulder strength ($M = 10.2$ kg, s.d. = 2.1). The experimenter held the muscle tester on the subject's wrist while the subject's arm was outstretched at a right angle from his torso. The subject then raised his arm against the experimenter's resistance.

— Handgrip strength ($M = 39.5$ kg, s.d. = 7.3). A Smedley III handgrip strength dynamometer was used to measure grip strength.

— Flexed bicep circumference ($M = 31.0$ cm, s.d. = 1.9).

Voice samples.

Subjects were instructed to say in a normal speaking voice, 'Nobi cojiro tsun quin dyem' venchuban aca'yaty anic fer no'bacni tsun', which translates as, 'We will cross the river and then arrive home; it was a tough crossing for us'. Voices were recorded with a portable digital recorder (iPod with Belkin Voice Recorder; frequency range = 500 Hz–12 kHz).

(iii) *Sample 3*

Twenty male Andean herder-horticulturalists from the villages of Gobernador Solá and Ingeniero Maury in the province of Salta, Argentina, provided voices and were measured on strength, height and weight (mean age = 34.8; s.d. = 19.1, range 15–71). Their voice samples were given in Spanish.

Strength assessments: (Cronbach's $\alpha = 0.57$).

— Chest strength ($M = 38.8$ kg, s.d. = 10.4). A Rolyan hydraulic hand dynamometer was used as a measure of chest strength. The handle was inverted and the subjects held the dynamometer to their chest and pressed the bars together. Subjects were told to hold the device to their chest with elbows at 90° from the torso and press in as hard as possible (Sell *et al.* 2009*a*; electronic supplementary material).

— Flexed bicep circumference ($M = 32.3$ cm, s.d. = 3.6).

Voice Samples.

Subjects were instructed to speak in a normal speaking voice, 'Cuando llueve se inundan las chacras, y la gente junta el maíz y prende fuego', which translates to, 'When it rains the ranches get flooded, and the people gather the maize and light fires'. Voices were recorded using a built-in microphone on a portable digital recorder (Olympus WS-100; frequency response: 100 Hz–12 kHz).

(iv) *Samples 4 and 6*

Fifty male students (sample 4) and 50 female students (sample 6) were taken from the student population at UCSB (males: mean age = 20.2, s.d. = 2.24, range: 18–31; females: mean age = 18.8, s.d. = 0.95, range: 18–22) and were paid $10 for their participation.

Strength assessments: (Cronbach's $\alpha = 0.75$ males, 0.63 females).

— Chest strength dynamometer (men: $M = 55.2$ kg, s.d. = 14.1; women: $M = 26.3$ kg, s.d. = 6.1), as with sample 3.

— Flexed bicep circumference (men: $M = 33.5$ cm, s.d. = 3.2; women: $M = 28.6$ cm, s.d. = 2.6).

— Photo ratings. Twenty-four UCSB undergraduates (12 female) rated body photographs as in sample 1.

Voice samples.

Subjects were asked to read the first sentence of the Rainbow Passage (Fairbanks 1960), 'When the sunlight strikes raindrops in the air, they act like a prism and form a rainbow'. The recording was taken with the built-in microphone on a portable cassette recorder (Sony Pressman TCM-400DV; frequency range: 250–6300 Hz).

(v) *Samples 5 and 7*

Forty-four male students (sample 5) and 30 female students (sample 7) from the University of Timisoara in Romania (males: mean age = 21.7, s.d. = 3.48, range: 20–38; females: mean age = 21.1, s.d. = 1.89, range 20–29) participated as part of a course requirement.

Strength assessments: (Cronbach's $\alpha = 0.68$ males, 0.64 females).

— Chest strength dynamometer (men: $M = 46.3$ kg, s.d. = 10.7; women: $M = 26.6$ kg, s.d. = 4.9) as with sample 3.

— Handgrip strength (men: $M = 52.6$ kg, s.d. = 7.7; women: 32.6 kg, s.d. = 5.2).

— Flexed bicep circumference (men: $M = 32.9$ cm, s.d. $= 2.7$; women: 27.6 cm, s.d. $= 2.5$).

Height and weight measures were not available for these subjects.

Voice samples.

Subjects were instructed to say in a normal speaking voice, 'Iesi si taci', which translates to, 'Get out and be quiet'. Voices were recorded with the built-in microphone on a portable digital recorder (Olympus WS-100; frequency response: 100 Hz–12 kHz).

(vi) *Sample 8*

Fifty-four male students from the student population at UCSB (mean age: 19.9, s.d. $= 2.0$, range: 18–23) were brought to a sound studio on campus to give voice samples and have strength measurements taken. High-quality recordings are needed to get accurate Df measures, so this sample served primarily to determine the relationship between Df and voice ratings and body measurements.

Strength assessments: (Cronbach's $\alpha = 0.61$).

— Chest strength dynamometer ($M = 53.6$ kg, s.d. $= 12.5$) as with sample 3.

— Flexed bicep circumference ($M = 31.7$ cm, s.d. $= 2.9$).

Voice samples.

Subjects were instructed to produce five vowels ('a' as in bait, 'e' as in beet, 'i' as in bite, 'o' as in boat and 'oo' as in boot). Accurate measures of Df require averaging across a range of vocal tract configurations. Voices were recorded digitally (44.1 kHz, 16 bit) using AKG C 414B-XL II microphones with the polar pattern set to Cardioid (frequency range: 20 Hz–20 kHz).

(b) *Voice ratings and acoustic analysis*

Undergraduates from UCSB were instructed to rate the voices on physical strength, height and weight using a 7-point scale. Types of ratings were done separately and randomly so that a given rater would rate all voices in a given sample on one variable at random (e.g. strength), then the same voices on the next variable (e.g. height) and then on the final variable (e.g. weight). Voices were also randomized within each block. Because height and weight were not available for samples 5 and 7, these voices were rated on physical strength only. For sample 1 only, raters also assessed the voices on 'how tough he would be in a physical fight'. Each sample was rated by a different group.

Raters for sample 1. Fifty-three undergraduates (22 female, mean age: 20.7, s.d.: 2.8) were paid $5 to rate the 63 United States (US) male voices.

Raters for sample 2. Thirty-one undergraduates (10 female, mean age: 19.7, s.d.: 1.5) were paid $5 to rate the 49 male Tsimane voices.

Raters for sample 3. Thirty undergraduates (17 female, mean age: 18.8, s.d.: 1.1) were paid $5 or given course credit to rate the 20 Andean voices.

Raters for sample 4. Fifty-four undergraduates (42 female, mean age: 20.5, s.d.: 5.7) were paid $5 to rate the 50 US male voices.

Raters for sample 5. Twenty undergraduates (12 female, mean age: 19.1, s.d.: 1.0) were given course credit to rate the 44 Romanian male voices.

Raters for sample 6. Forty-seven undergraduates (30 female, mean age: 20.2, s.d.: 1.9) were paid $5 to rate the 50 US female voices.

Raters for sample 7. Twenty-one undergraduates (14 female, mean age: 19.0, s.d.: 0.9) were given course credit to rate the 30 Romanian female voices.

Raters for sample 8. Thirty-six undergraduates (25 female, mean age: 19.4, s.d.: 1.4) were paid $5 to rate the recordings of vowel sounds from the 54 US males.

(i) *Acoustic analysis*

All analogue (i.e. cassette recorded) speech samples were digitized (44.1 kHz, 16 bit) and resampled to 11.025 kHz with an anti-aliasing filter. Acoustic analysis was performed using PRAAT (v. 4.6.03). Average (F0) was calculated on entire utterances using PRAAT's autocorrelation algorithm with a search setting of 75–500 Hz, as recommended for adult males, or 100–600 Hz for female voices. In sample 8, Df was calculated as $(F4 - F3) + (F3 - F2) + (F2 - F1)/3$ with separate calculations of all formants for each vowel, and those values averaged for the final Df value. Overall F0 was calculated by averaging F0 values across the five vowels.

3. RESULTS AND DISCUSSION

All samples were analysed independently. In order to compare our results with previous studies that averaged ratings together and correlated them with body variables, we computed Pearson correlations between averaged voice ratings and their actual counterparts. This is the traditional technique for assessing the accuracy of voice ratings, but it provides ambiguous information about the accuracy of individual raters because a few raters with large correlations can cause sizeable aggregate correlations even if the average rater is not accurate at all. A more informative measure of the average individual accuracy can be done with hierarchical linear regression models (HLM) that we used to compute the average within-rater slopes of the relationship between voice ratings as a dependent variable and actual body characteristics as level 1 predictors (Raudenbush & Bryk 2002). For example, a regression slope relating ratings of strength to actual strength can be computed for each of the 53 raters in study 1. Hierarchical regression then computes a regression coefficient (γ) that represents the average of those 53 slopes and tests the null hypothesis that this average slope is zero. All variables were standardized (z-scored) before entry into the HLM, making the gamma statistic interpretable as a standardized regression coefficient. All HLM analyses controlled for sex of rater effects (all non-significant). All HLM calculations were performed with HLM 6.0 software from Scientific Software International, Inc.

(a) *Accuracy of strength, height and weight assessment*

So that our data could be compared with previous studies, we have reported the Pearson correlations between actual body characteristics and the average ratings of those characteristics from the voice on the left side of table 1. HLMs with ratings of a trait as the dependent variable and the actual measured trait as the level 1 predictor variable were used to estimate the average individual accuracy of strength, height and weight assessment from the voice and are reported in the three middle columns of table 1.

Table 1. Summary of strength, height and weight assessment across seven samples. (p-values have not been corrected for family-wise error. n.a., not applicable.)

	accuracy of averaged ratings (r)			individual accuracy of ratings (γ)			independent accuracy of strength, height and weight ratings γ[a]		
	strength	height	weight	strength	height	weight	strength c/ht and c/wt	height c/str and c/wt	weight c/str and c/ht
sample 1: US males	0.45, $p = 0.001$	0.28, $p = 0.028$	0.32, $p = 0.011$	0.23, $p = 10^{-14}$	0.12, $p = 10^{-5}$	0.16, $p = 10^{-11}$	0.21, $p = 10^{-11}$	0.06, $p = 0.014$	−0.003, $p = 0.89$
sample 2: Tsimane males	0.35, $p = 0.02$	0.13, $p = 0.41$	0.22, $p = 0.15$	0.25, $p = 10^{-8}$	0.08, $p = 0.012$	0.12, $p = 0.0003$	0.27, $p = 10^{-6}$	0.04, $p = 0.23$	0.06, $p = 0.16$
sample 3: Andean males	0.46, $p = 0.04$	0.29, $p = 0.22$	0.62, $p = 0.004$	0.24, $p = 10^{-5}$	0.15, $p = 0.002$	0.28, $p = 10^{-6}$	0.21, $p = 0.001$	0.08, $p = 0.13$	0.16, $p = 0.012$
sample 4: US males	0.51, $p = 0.001$	0.45, $p = 0.001$	0.37, $p = 0.008$	0.30, $p = 10^{-11}$	0.23, $p = 10^{-11}$	0.21, $p = 10^{-9}$	0.28, $p = 10^{-8}$	0.14, $p = 10^{-5}$	−0.01, $p = 0.79$
sample 5: Romanian males	0.48, $p = 0.001$	n.a.	n.a.	0.31, $p = 10^{-6}$	n.a.	n.a.	n.a.	n.a.	n.a.
sample 6: US females	0.26, $p = 0.07$	0.23, $p = 0.10$	0.38, $p = 0.006$	0.14, $p = 0.0001$	0.17, $p = 0.0001$	0.15, $p = 10^{-5}$	−0.14, $p = 0.001$	0.15, $p = 0.001$	0.12, $p = 0.005$
sample 7: Romanian females	0.32, $p = 0.08$	n.a.	n.a.	0.21, $p = 0.0001$	n.a.	n.a.	n.a.	n.a.	n.a.
sample 8: US males—vowels	0.30, $p = 0.03$	0.38, $p = 0.004$	0.21, $p = 0.13$	0.18, $p = 0.001$	0.23, $p = 0.0001$	0.13, $p = 0.005$,	0.19, $p = 0.014$	0.27, $p = 10^{-7}$	0.03, $p = 0.54$

[a]Also controls for subject age in samples 2 and 3 in which age varied substantially.

These gammas represent the estimated accuracy for a random individual rater.

The results confirmed predictions (i) through to (v).

- (i) Average individual estimates of strength from the voice were accurate and highly significant across all six male samples ranging from $\gamma = 0.18$ to 31. This accuracy is similar to the accuracy of strength assessment from static visual images of the face, but lower than estimation from images of the body (Sell *et al.* 2009*a*).
- (ii) Accuracy of strength estimation was similar across both familiar and unfamiliar languages.
- (iii) While strength was accurately estimated from women's voices in both the US and Romanian samples, the effect was about half as large as for their male counterparts. The same pattern is found when assessing strength from visual images of the face (Sell *et al.* 2009*a*).
- (iv) Estimates of strength were notably enhanced when both auditory and visual channels were available (see below).
- (v) Assessments of strength remained significant, controlling for both height and weight (see below).

(b) *Is perceived fighting ability related to the other voice ratings?*

Yes. In sample 1, raters assessed 'how tough he would be in a physical fight' in addition to strength, height and weight. Perceptions of the targets' fighting ability were virtually identical to the perceptions of physical strength ($r = 0.98$), showing that—regardless of their relationship in the real world—raters were treating physical strength and fighting ability as synonymous. This same relationship was found when assessing fighting ability from photographs (Sell *et al.* 2009*a*).

(c) *Are subjects' perceptions of men's fighting ability related to those men's actual history of fighting?*

Yes. In sample 1, the only sample for which the data were available, the average perceived fighting ability scores for target males were positively correlated with their fighting history (how many fights they reported having been in during the last 4 years), $r = 0.36$, $p = 0.002$, and their score on the Physical Aggression subscale of the Buss–Perry Aggression Questionnaire, $r = 0.49$, $p = 0.0001$. In other words, raters are detecting a cue in the voice that tracks actual measures of fighting behaviour. Although this finding suggests that they are tracking actual fighting ability, it ought to be treated cautiously because we do not know how many of these fights the targets won. Nonetheless, research has shown that more formidable individuals are those more likely to engage in fights (Archer & Thanzami 2007; Sell *et al.* 2009*b*).

(d) *Do listeners extract additional formidability information from the voice that is not supplied from visual cues?*

Yes. To test prediction (iv), our strength measures for samples 1 and 4 were recomputed without the photograph rating component. The new strength measure was then predicted in a simultaneous regression analysis by both the average voice rating of strength and the average photograph ratings of strength. The results of the model showed independent contributions from both the photo ratings (sample 1: $\beta = 0.52$, $p = 0.00001$; sample 4: $\beta = 0.50$, $p = 0.0002$) and voice ratings (sample 1: $\beta = 0.25$, $p = 0.022$; sample 4: $\beta = 0.27$, $p = 0.032$) on actual strength. These results demonstrate that there are cues in the voice which predict strength that are not available during visual inspection of static photographs, and suggest that the system evolved to take advantage of this additional information.

(e) *Is the ability to estimate strength a by-product of height and weight detection?*

Because height, weight and strength are positively related to each other, the ability to assess strength from the male voice could be a by-product of height and weight detection. Alternatively, the ability to accurately estimate height and weight from the voice could be a by-product of a mechanism for strength assessment. Our prediction is that a mechanism designed by natural selection to assess formidability would extract cues of physical strength rather than just height or weight. This prediction stems from previous research showing that physical strength, more than height or weight, is related to anger and aggression (Sell *et al.* 2009*a*,*b*). To distinguish between these hypotheses, we constructed three HLM models each with actual strength, height and weight predicting one of the three ratings. The independent accuracy of each of the three body measurements across six studies are shown on the right side of table 1. The results indicate:

- — ratings of strength from the voice, when controlling for height and weight, continue to predict actual strength across all male samples with roughly the same effect size. The ability to assess strength is *not* dependant on height or weight assessment;
- — ratings of height inconsistently predict actual height when controlling for strength and weight;
- — ratings of weight largely fail to predict actual weight when controlling for strength and height; and
- — in contrast to male voices, the ability to estimate female strength from the voice completely disappeared, in fact reversed, when controlling for height and weight.

The ability to estimate strength, in men at least, is not a by-product of height and weight assessment; the data are more consistent with the proposal that height and weight assessment are by-products of strength assessment.

Furthermore, raters seemed to be giving the same ratings regardless of whether they were asked to rate height, weight or strength. The inter-correlations between the average ratings of height, weight and strength are extremely high, especially for samples with more raters: Cronbach's α (sample $1 = 0.97$; sample $2 = 0.85$; sample $3 = 0.90$; sample $4 = 0.98$; sample $5 = 0.91$, sample $8 = 0.94$; all $n = 3$). In summary, regardless of whether you ask individuals to rate height, weight, strength or fighting ability, they will produce the same ratings, which track cues of physical strength independent of height and weight.

(f) *Is voice assessment accuracy the result of men who think they are strong modulating their voice?*

No. In four of the six male samples, the stimulus subjects were asked to report their own physical strength (samples 1, 3, 4 and 5). The speakers rated their physical strength relative to other males—e.g. 'I am stronger than ____% of others of my sex'. This self-report measure has been shown to correlate with actual lifting strength in previous studies (Sell *et al.* 2009*a*; electronic supplementary material). For each of the four studies in which self-report was available, a simultaneous regression analysis was performed, predicting the voice rating of strength from both self-reported strength and actual strength. In all four samples, self-reported strength did not account for any significant variance in voice ratings when actual strength was in the model: sample 1: actual strength $\beta = 0.47$, $p = 0.002$, self-reported strength $\beta = -0.083$, $p = 0.57$; sample 3: actual strength $\beta = 0.47$, $p = 0.047$, self-reported strength $\beta = -0.05$, $p = 0.82$; sample 4: actual strength $\beta = 0.51$, $p = 0.005$, self-reported strength $\beta = 0.004$, $p = 0.98$; sample 5: actual strength $\beta = 0.53$, $p = 0.0005$, self-reported strength $\beta = -0.17$, $p = 0.22$. In other words, actual strength determined how strong the subject's voice sounded to others, and not how strong the subject believed himself to be. This is the opposite of what would be predicted if men who believed themselves strong were modulating their voice and that this relationship was responsible for the relationship between physical strength and ratings of strength from the voice. This is not to say that there are not vocal cues of confidence or lack of fear that may be produced during aggressive interactions (we believe there are), but those cues are not responsible for the ability to assess strength from normal speaking voices.

(g) *Did F0 account for variance in physical strength and judgments of physical strength?*

F0 is the acoustic correlate of perceptual pitch and is the result of vocal fold vibration during speech. Studies generally show no relationship between F0 and physical size in adult males (see Evans *et al.* (2006) for a notable exception), but experiments that manipulate F0 suggest that humans might use F0 as a cue to strength. For example, men show more deference to a voice with lower F0 (Puts *et al.* 2006). Rendall *et al.* (2007) suggested three possibilities for people's misguided use of F0 in body size judgments. One is that listeners might be generalizing the between-speaker differences that do exist (e.g. women and children, on average, have higher pitched voices than adolescent and adult males). Second, listeners might be generalizing from other negative pitch–size relationships (i.e. larger things produce lower frequencies on average in the world). Last, listeners might be letting other F0 correspondences (e.g. F0 is related to testosterone exposure) intrude on pitch–body size judgments (see also Collins (2000) for related discussion).

To address these questions, all voice samples were analysed for F0 and correlated with average voice ratings and actual body ratings. Results are reported in table 2. F0 failed to reliability correlate with measures of strength. While F0 did tend to affect raters' estimations of strength

Table 2. Fundamental frequency (F0) correlates with voice ratings but not actual body measurements. (*p*-values have not been corrected for family wise error.)

	correlation between F0 and body measurements			correlation between F0 and voice ratings		
	strength	height	weight	strength	height	weight
sample 1: US males	−0.14, $p = 0.26$	−0.24, $p = 0.077$	0.01, $p = 0.96$	−0.37, $p = 0.003$	−0.33, $p = 0.009$	−0.51, $p = 0.00002$
sample 2: US males	−0.12, $p = 0.40$	−0.17, $p = 0.24$	−0.02, $p = 0.89$	−0.55, $p = 0.00004$	−0.45, $p = 0.001$	−0.60, $p = 10^{-6}$
sample 3: Tsimane males	0.20, $p = 0.18$	−0.091, $p = 0.55$	0.19, $p = 0.21$	−0.14, $p = 0.36$	−0.47, $p = 0.001$	−0.46, $p = 0.002$
sample 4: Andean males	−0.35, $p = 0.14$	−0.19, $p = 0.42$	−0.13, $p = 0.58$	−0.35, $p = 0.13$	−0.22, $p = 0.35$	−0.019, $p = 0.94$
sample 5: US females	−0.19, $p = 0.19$	−0.13, $p = 0.36$	−0.23, $p = 0.11$	−0.40, $p = 0.004$	−0.40, $p = 0.004$	−0.31, $p = 0.027$
sample 8: US males—vowels	0.01, $p = 0.96$	0.11, $p = 0.41$	0.18, $p = 0.19$	−0.25, $p = 0.07$	−0.24, $p = 0.08$	−0.17, $p = 0.22$

(i.e. lower pitch correlated with perceptions of greater strength), this percept was erroneous. Recent work has explored this 'perceptual pull' that F0 exerts on judgments of size, and the results are consistent with ours (Rendall *et al.* 2007).

(h) *Did Df account for variance in physical strength and judgments of physical strength?*

Df is the most reliable known acoustic index of body size in humans, as well as other primates. Df is the averaged distance between adjacent resonance frequencies resulting from the vocal source filtering of the superlaryngeal vocal tract. Fitch (1997) proposed that because vocal tract length (VTL) positively correlates with height, the distance between adjacent frequencies should provide a reliable index (i.e. Df should negatively correlate with VTL). There is only limited evidence that listeners can detect body size from this cue however (Rendall *et al.* 2007). Rendall and his collaborators found that judges could discriminate between speakers with different Df but only when the actual height difference between presented pairs exceeded 10 cm. These data are consistent with other work showing that VTL just noticeable differences are approximately 4–7% in synthesized speech (Ives *et al.* 2005; Smith *et al.* 2005). Detecting formant spatial relationships is clearly difficult even for a system that probably has related specializations for perceiving formant information in other domains (e.g. formant transitions in vowel sounds). Nevertheless, there is a distinct possibility that humans respond to Df as a cue of physical strength.

High-quality acoustic recordings of vowels representing a range of vocal tract configurations are needed for accurate measurement of Df. Sample 8 was designed primarily to assess the relationship between Df and formidability and ratings of strength. The voice samples consisted of vowels produced by subjects in a sound studio. Df did not correlate with physical strength or weight (strength: $r = -0.08$, $p = 0.58$; weight: $r = 0.04$, $p = 0.77$), but did show a significant negative relationship with height ($r = -0.36$, $p = 0.01$). Like F0, Df showed significant negative relationships with voice ratings as well (strength: $r = -0.45$, $p = 0.001$; height: $r = -0.51$, $p = 0.0001$; weight: $r = -0.37$, $p = 0.01$). However, Df cannot be responsible for the accuracy of strength assessment as it is not correlated with strength. Additionally, when controlling for Df, ratings of strength are still accurate (partial $r = 0.30$, $p = 0.03$). Ratings of height become only marginally significant when controlling for Df (partial $r = 0.25$, $p = 0.07$), suggesting raters may be using Df in their height assessments. In summary, Df is a genuine cue of height that raters might be using when assessing height, but Df is not related to physical strength and is not the cue raters are using to accurately assess strength.

For the moment, the acoustic variables that people use to assess physical strength are unknown.

4. CONCLUSION

These findings support the hypothesis that the human voice—especially the male voice—contains cues of physical strength, and that natural selection built specializations into the human neurocognitive architecture designed for assessing fighting ability based on these cues. These specializations extract information not available from the visual channel alone and cues that are uniquely predictive of upper-body strength rather than height and weight. These results complement and support earlier work, indicating that fighting assessment from visual channels focuses on predicting upper-body strength in men (Sell *et al.* 2009*a*).

These results offer an explanation as to why a 30 year research programme examining the ability of raters to detect height and weight from the voice has returned such inconsistent findings. From an evolutionary point of view, determining a conspecific's height or weight had no clear function in itself. On the other hand, determining physical strength was vitally important. Consistent with this, ratings tracked physical strength independent of both height and weight across all samples. However raters were not able to consistently extract unique cues of weight or height.

Both theoretical analyses and evidence from other species indicate that natural selection would favour cognitive mechanisms that accurately assess fighting ability. The studies presented here provide direct evidence that both men and women can accurately assess adult men's physical strength from cues in the speaking voice. Furthermore, the cues are not solely cues of size but instead appear to track correlates of upper-body strength more directly and in ways that do not require experience with a particular language. Taken together, the results strongly support the hypothesis that the human cognitive architecture contains specializations for formidability assessment through auditory channels.

The studies presented here in were approved by the Human Subjects Committee ethics board at the University of California, Santa Barbara.

We thank Christina Larson, June Betancourt, Mahsa and Golsa Afsharpour, Lauren Click, Andy Delton, Max Krasnow, Jim Roney, Roxana Toma, Silviu Nisu and Phillip Smith for their valuable insights and assistance. We would also like to thank Professor P. Petrescu of the University of Timisoara, Norma Naharro, Mirta Santoni, the head of the Museo de Antropología de Salta, and the villagers of Gobernador Solá and Ingeniero Maury as well as the Tsimane people. We thank Kevin Kelly for donating his time and expertise with the UCSB sound studio. Financial support for this project was provided by an NIH Director's Pioneer Award to L.C. and the UCSB Pre-Dissertation Field Site Visit Grant awarded to D.S.

REFERENCES

Archer, J. 2006 Testosterone and human aggression: an evaluation of the challenge hypothesis. *Neurosci. Biobehav. Rev.* **30**, 319–345. (doi:10.1016/j.neubiorev.2004.12.007)

Archer, J. 2009 Does sexual selection explain human sex differences in aggression? *Behav. Brain Sci.* **32**, 249–311. (doi:10.1017/S0140525X09990951)

Archer, J. & Thanzami, V. 2007 The relation between physical aggression, size and strength, among a sample of young Indian men. *Pers. Indiv. Differ.* **43**, 627–633. (doi:10.1016/j.paid.2007.01.005)

Balikci, A. 1970 *The Netsilik eskimo.* Garden City, NY: The Natural History Press.

Bhasin, S. *et al.* 1996 The effects of supraphysiologic doses of testosterone on muscle size and strength in normal men. *New Engl. J. Med.* **335**, 1–7. (doi:10.1056/NEJM199607043350101)

Boehm, C. 1984 *Blood revenge. The anthropology of feuding in Montenegro and other tribal societies*. Lawrence, KS: University Press of Kansas.
Brues, A. 1959 The Spearman and the Archer: an essay on selection in body build. *Am. Anthropol.* **61**, 457–469. (doi:10.1525/aa.1959.61.3.02a00080)
Buss, A. H. & Perry, M. 1992 The aggression questionnaire. *J. Pers. Soc. Psychol.* **63**, 452–459. (doi:10.1037/0022-3514.63.3.452)
Chagnon, N. 1983 *Yanomamo: the fierce people*, 3rd edn. New York, NY: Holt, Rinehart and Winston.
Clutton-Brock, T. & Albon, S. 1979 The roaring of red deer and the evolution of honest advertisement. *Behaviour* **69**, 145–170. (doi:10.1163/156853979X00449)
Collins, S. 2000 Men's voices and women's choices. *Anim. Behav.* **60**, 773–780. (doi:10.1006/anbe.2000.1523)
Dabbs, J. & Mallinger, A. 1999 High testosterone levels predict low voice pitch among men. *Pers. Indiv. Differ.* **27**, 801–804. (doi:10.1016/S0191-8869(98)00272-4)
Daly, M. & Wilson, M. 1988 *Homicide*. New York, NY: de Gruyter.
Davies, N. & Halliday, T. 1978 Deep croaks and fighting assessment in toads *Bufo bufo*. *Nature* **274**, 683–685. (doi:10.1038/274683a02000)
Deliyski, D. D., Shaw, H. S. & Evans, M. K. 2005 Influence of sampling rate on accuracy and reliability of acoustic voice analysis. *Logoped. Phoniatr. Vocol.* **30**, 55–62. (doi:10.1080/1401543051006721)
Evans, S., Neave, N. & Wakelin, D. 2006 Relationships between vocal characteristics and body size and shape in human males: an evolutionary explanation for a deep male voice. *Biol. Psychol.* **72**, 160–163. (doi:10.1016/j.biopsycho.2005.09.003)
Fairbanks, G. 1960 *Voice and articulation drillbook*, 2nd edn. New York, NY: Harper & Row.
Fitch, W. T. 1997 Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *J. Acoust. Soc. Am.* **102**, 1213–1222. (doi:10.1121/1.421048)
Frederick, D. & Haselton, M. 2007 Why is muscularity sexy? Tests of the fitness indicator hypothesis. *Pers. Soc. Psychol. Bull.* **33**, 1167–1183. (doi:10.1177/0146167207303022)
Freeman, D. 1983 *Margaret Mead and Samoa: the making and unmaking of an anthropological myth*. Cambridge, MA: Harvard University Press.
Gallup, A., White, D. & Gallup Jr, G. 2007 Handgrip strength predicts sexual behavior, body morphology, and aggression in male college students. *Evol. Hum. Behav.* **28**, 423–429. (doi:10.1016/j.evolhumbehav.2007.07.001)
Garn, S. & Clark, L. 1953 The sex difference in the basal metabolic rate. *Child Dev.* **24**, 215–224.
Gonzalez, J. & Oliver, J. C. 2004 Percepción a través de la voz de las características físicas del hablante: identificación de la estatura a partir de una frase o una vocal. *Revista de Psicología General y Aplicada* **57**, 21–33.
Gonzalez, J., Cervera, T. & Llau, M. J. 2003 Acoustic analysis of pathological voices compressed with MPEG system. *J. Voice* **17**, 126–139.
Hardouin, L., Reby, D., Bavoux, C., Burneleau, G. & Bretagnolle, V. 2007 Communication of male quality in owl hoots. *Am. Nat.* **169**, 552–562. (doi:10.1086/512136)
Hughes, S. & Gallup Jr, G. 2004 Ratings of voice attractiveness predict sexual behavior and body configuration. *Evol. Hum. Behav.* **25**, 295–304. (doi:10.1016/j.evolhumbehav.2004.06.001)
Huntingford, F. A. & Turner, A. K. 1987 *Animal conflict*. New York, NY: Chapman & Hall.
Ives, D. T., Smith, D. R. R. & Patterson, R. D. 2005 Discrimination of speaker size from syllable phrases. *J. Acoust. Soc. Am.* **118**, 3816–3822. (doi:10.1121/1.2118427)
Kitchen, D. M., Seyfarth, R. M., Fischer, J. & Cheney, D. L. 2003 Loud calls as indicators of dominance in male baboons (*Papio cynocephalus ursinus*). *Behav. Ecol. Sociobiol.* **53**, 374–384.
Krebs, J. & Davies, N. 1993 *An introduction to behavioural ecology*, 3rd edn. Oxford, UK: Blackwell Scientific Publications.
Ladich, F. 1998 Sound characteristics and outcome of contests in male croaking gouramis (Teleostei). *Ethology* **104**, 517–529. (doi:10.1111/j.1439-0310.1998.tb00087.x)
Lass, N. J., Barry, P. J., Reed, R. A., Walsh, J. M. & Amuso, T. A. 1979*a* The effect of temporal speech alterations on Speaker height and weight identification. *Lang. Speech* **22**, 163–171.
Lass, N. J., DiCola, G. A., Beverly, A. S., Barbera, C., Henry, G. H. & Badali, M. K. 1979*b* The effect of phonetic complexity on speaker height and weight identification. *Lang. Speech* **22**, 297–309.
Lassek, W. & Gaulin, S. 2009 Costs and benefits of fat-free muscle mass in men: relationship to mating success, dietary requirements, and natural immunity. *Evol. Hum. Behav.* **30**, 322–328.
Lewis, O. 1961 *The children of Sanchez: autobiography of a Mexican family*. New York, NY: Random House.
Mager, J., Walcott, C. & Piper, W. 2007 Male common loons, *Gavia immer*, communicate body mass and condition through dominant frequencies of territorial yodels. *Anim. Behav.* **73**, 683–690.
Newman, S., Butler, J., Hammond, E. H. & Gray, S. D. 2000 Preliminary report on hormone receptors in the human vocal fold. *J. Voice* **14**, 72–81. (doi:10.1016/S0892-1997(00)80096-X)
Patton, J. 2000 Reciprocal altruism and warfare: a case from the Ecuadorian Amazon. In *Adaptation and human behavior: an anthropological perspective* (eds L. Cronk, N. Chagnon & W. Irons), pp. 417–436. Hawthorne, NY: Aldine de Gruyter.
Plavcan, M. & Van Schaik, C. 1997 Interpreting hominid behavior on the basis of sexual dimorphism. *J. Hum. Evol.* **32**, 345–374. (doi:10.1006/jhev.1996.0096)
Puts, D. A. 2005 Mating context and menstrual phase affect female preferences for male voice pitch. *Evol. Hum. Behav.* **26**, 388–397. (doi:10.1016/j.evolhumbehav.2005.03.001)
Puts, D., Gaulin, S. & Verdolini, K. 2006 Dominance and the evolution of sexual dimorphism in human voice pitch. *Evol. Hum. Behav.* **27**, 283–296. (doi:10.1016/j.evolhumbehav.2005.11.003)
Raudenbush, S. W. & Bryk, A. S. 2002 *Hierarchical linear models: applications and data analysis methods*, 2nd edn. Newbury Park, CA: Sage.
Rendall, D., Vokey, J. R. & Nemeth, C. 2007 Lifting the curtain on the Wizard of Oz: biased voice-based impressions of speaker size. *J. Exp. Psychol. Hum. Perc. Perform.* **33**, 1208–1219. (doi:10.1037/0096-1523.33.5.1208)
Sargent, W. 1974 *People of the valley*. New York, NY: Random House.
Sell, A. 2006 Regulating welfare tradeoff ratios: three tests of an evolutionary-computational model of human anger. *Dissert. Abs. Int. B Sci. Eng.* **66**, 4516.
Sell, A., Cosmides, L., Tooby, J., Sznycer, D., von Rueden, C. & Gurven, M. 2009*a* Human adaptations for the visual assessment of strength and fighting ability from the body and face. *Proc. R. Soc. B* **276**, 575–584. (doi:10.1098/rspb.2008.1177)

Sell, A., Tooby, J. & Cosmides, L. 2009*b* Formidability and the logic of anger. *Proc. Natl Acad. Sci. USA* **106**, 15 073–15 078. (doi:10.1073/pnas.0904312106)

Smith, D. R. R., Patterson, R. D., Turner, R., Kawahara, H. & Irino, T. 2005 The processing and perception of size information in speech sounds. *J. Acoust. Soc. Am.* **117**, 305–318. (doi:10.1121/1.1828637)

Titze, I. R. 1994 *Principles of voice production.* Englewood Cliffs, NJ: Prentice Hall.

Toch, H. 1969 *Violent men.* Chicago, IL: Aldine.

Tooby, J., Cosmides, L., Sell, A., Lieberman, D. & Sznycer, D. 2008 Internal regulatory variables and the design of human motivation: a computational and evolutionary approach. In *Handbook of approach and avoidance motivation* (ed. J. Andrew Elliot), pp. 251–271. New York, NY: Psychology Press.

von Rueden, C., Gurven, M. & Kaplan, H. 2008 The multiple dimensions of male social status in an Amazonian society. *Evol. Hum. Behav.* **29**, 402–415.