Electronic Supplementary Materials

Study 1

Body Measurements. To find proxies of lifting strength that could later be used in laboratory settings without cumbersome weight-lifting equipment, a variety of body measurements were taken, including height, weight, hip circumference, waist circumference, chest circumference (inside the arms with lungs deflated), unflexed biceps circumference, flexed biceps circumference, and neck circumference. Biceps measurements were taken on the dominant arm.

Strength measures. Subjects were led by the experimenter through four upper body exercises in random order: arm curl (biceps), abdominal crunch (abdominal muscles, e.g. rectus abdominus), chest press (pectorals), and super long pull (latissimus dorsi; "lats," deltoids). For each weight lifting exercise, the subject was asked to estimate his maximum 10-repetition lifting strength on that exercise. Then he was asked to lift 10 repetitions of the following weights: 50% of maximum estimated weight; 75%; 100%; maximum estimated weight + 5 pounds; +10 pounds; +15 pounds; and so on. The most weight the subject could successfully lift ten times was recorded for each of the four exercises. These amounts were then converted to z-scores and averaged together to create a single composite score representing the upper body strength of each subject. The reliability of these four weight-lifting measures was .92 (Cronbach's α). Unless otherwise specified, "lifting strength" refers to this upper body measure.

Performance on a leg press machine was used as a measure of leg strength; order was randomized with the four upper body machines.

History of fighting. The questionnaire filled out by the stimulus subjects contained one question relevant to the analyses reported: "I have been in ___ fights in the last four years. (Fights include shoving matches, fistfights, wrestling, and anything physical beyond yelling)." They were asked to not count sporting matches (e.g., wrestling, martial arts). Answers ranged from 0 (42% of subjects) to 11 fights (2 subjects). 61% had been in 0-1 fights, 70% in 0-2 fights. A square root transformation was applied to this data to normalize the distribution for analyses.

What were the best proxies of upper body strength?
Lifting strength measured on standardized weight-lifting machines may be as close as one can get to a gold standard for measuring strength. Access to weight lifting machines severely limits the populations one can test, however. To successfully explore the role that strength and formidability play in the social behavior of our species, researchers will require accurate yet portable measures of physical strength.

Because the upper body strength of men in Study 1 had been measured on weight lifting machines, we were able to determine which scales and physical measurements are well-correlated with lifting strength. Of all the physical measurements taken, flexed biceps circumference was the single best predictor of lifting strength: $r = .74$ ($p = 10^{-11}$). It accounted for 55% of the variance in lifting strength. Weight ($r = .48$, $p = .0001$), chest circumference ($r = .57$, $p = 10^{-6}$), and neck circumference ($r = .37$, $p = .003$) were all significant predictors of lifting strength, and height was marginally predictive ($r = .23$, $p = .073$). When all body measurements (i.e. height, weight, hip circumference, waist circumference, chest circumference, flexed and unflexed biceps circumference, and neck circumference) were put into a simultaneous regression analysis predicting lifting strength, only flexed biceps circumference accounted for unique variance, $\beta = .84$, $p = 10^{-6}$.

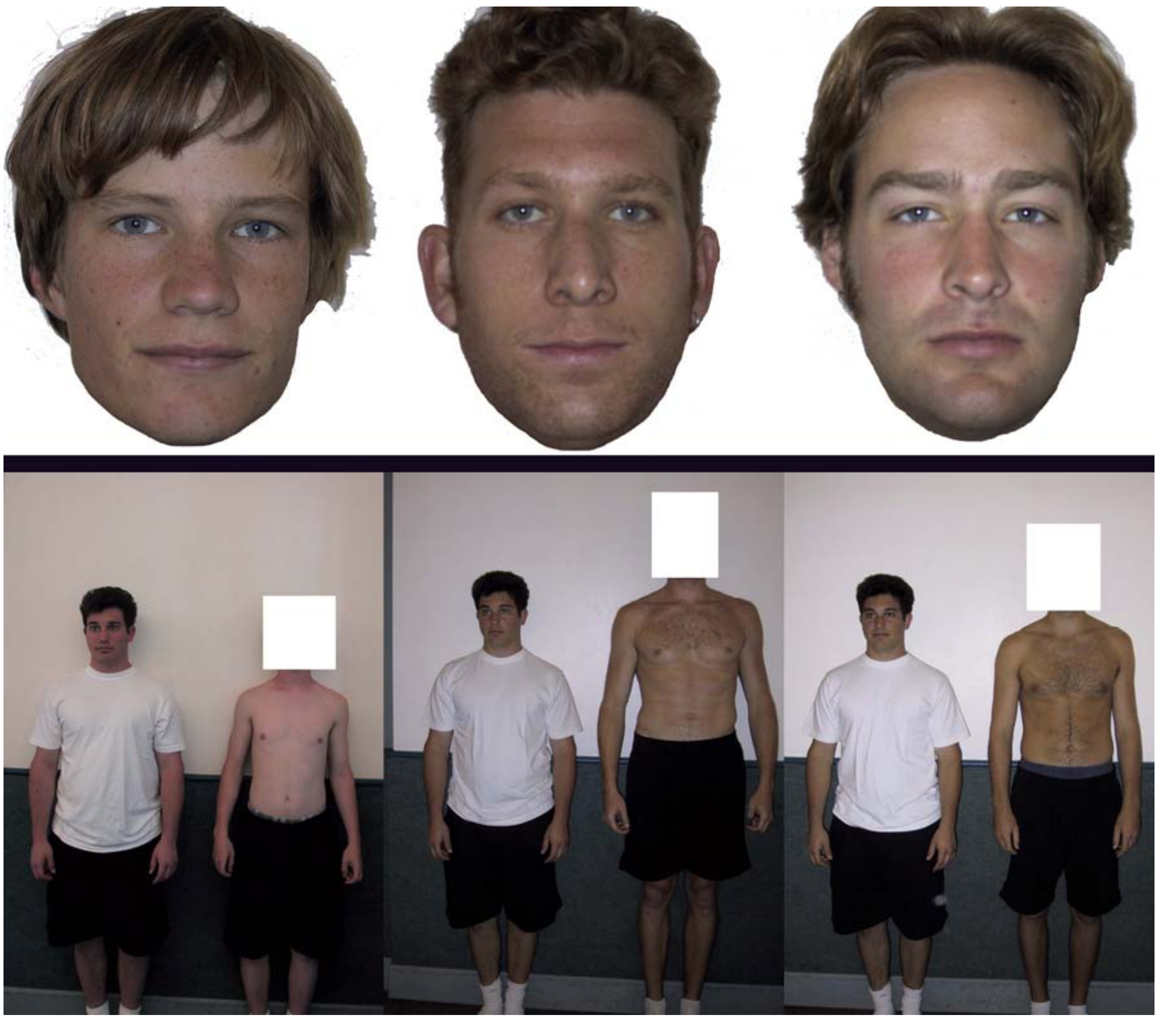
Results reported in the body of this article show that ratings of full person photographs correlated highly with weight-lifting strength ($r = .71$), accounting for 50% of the variance; to avoid circularity, we did not use this as a proxy measure of strength in our studies (which focus on the ability to rate strength from photos). It is, however, available to researchers who need a strength measure for other purposes. Finally, a self-report measure included in the questionnaire was a significant predictor of lifting strength as well, $r = .66$, $p = 10^{-8}$. The exact question asked was, "I am stronger than ____% of other men."

These three variables together—flexed biceps circumference, photographic ratings, and self-report—accounted for 66% of the variance in men's upper body lifting strength (adjusted R^2 = 64%). Importantly, each predictor accounted for unique variance in lifting strength: for flexed biceps circumference, $\beta = .41$, $p = 0003$; for full person photo ratings, $\beta = .30$, $p = .001$; for self-report, $\beta = .22$, $p = .05$. Prediction is not significantly improved by including the other body measures ($\Delta R^2 = .05$, $p = .24$). By creating a standard score for each of these three measures and adding them together, future researchers will have a useful proxy of lifting strength.

<u>HLM analyses</u>. Those interested in detailed reports of the HLM analyses will find these summarized for each study in tables at the end of this file. The HLMs estimating individual accuracy controlled for sex of rater, and included an error term that modeled the variability in accuracy among subjects. This is the variance component (with associated χ^2 statistic) reported in the tables at the end of this file.

Sample faces and bodies from Study 1. For those interested, a range of photographs used in Study 1 are presented in Figure ESM-1.

Figure ESM-1



Study 2

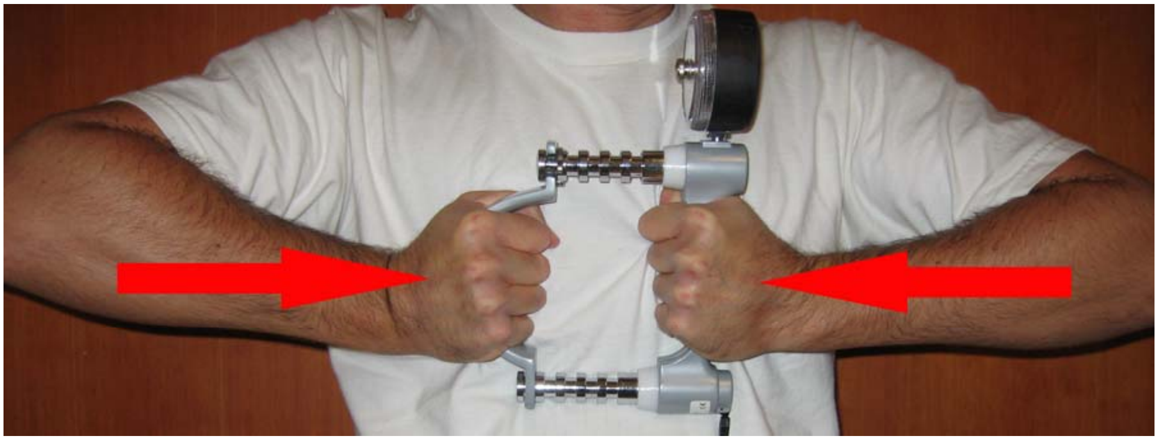
Strength measures. Study 2 included one direct measure of strength, and two proxy measures that were validated in Study 1 (see above, "What were the best proxies of upper body strength?").

1. Chest/arm strength. A measure of chest/arm strength was obtained by using a Rolyan Hydraulic hand dynamometer with its handles inverted (manufacturer: Smith & Nephew Rehabilitation). After grasping the inverted handles, the subject was instructed to hold the device to his or her chest with elbows extended and press in as hard as possible. Each subject was recorded twice and the higher score was used. See Figure ESM-2.

This is a new way of using the dynamometer. The motion required by this measure uses the pectoral, lattisimus dorsi, brachioradialis, and biceps muscles, so it is a direct measure of upper body strength (as are the weight-lifting machines). The measure derived from it correlates with flexed biceps circumference (r = .53) and with self-reported strength (r = .54), both of which have themselves been

shown to predict upper body lifting strength as measured on weight lifting machines (see above). Based on an independent sample of 14 men who did the four upper body weight-lifting exercises described in Study 1, we found that the inverted dynamometer chest/arm strength measure correlates at $r = .59$ ($p = .027$) with the measure of upper body lifting strength used in Study 1. (For those interested, handgrip, as measured by the same instrument in its standard use, correlated at $r = .49$ ($p = .07$) with the four machine weight-lifting measure of upper body strength in this sample of men; see below for measures involving the Tsimane).
2. Flexed biceps circumference (see above).
3. Self-report. Subjects filled in the following question, "I am physically stronger than ____% of others of my sex" (see above).

Figure ESM-2



Study 3

<u>Strength measures among the Tsimane</u>. There were six measures of strength among the Tsimane including exercises designed to measure both upper and lower body strength.

Upper body measures:
1. Chest strength. The subject pressed a Lafayette Manual Muscle Tester between the palms of his hands, with the elbows perpendicular to the body at mid-chest height.
2. Shoulder strength. The experimenter held the Lafayette Manual 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.
3. Handgrip strength. A Smedley III handgrip strength dynamometer was used to measure grip strength.
4. Flexed biceps circumference (see Study 1). Among the Tsimane, this indirect measure of strength correlated highly with the standardized average of the previous three direct measures, ($r = .54$, $p = .00004$). This shows that flexed biceps circumference continues to track strength in a non-industrial population without access to weight-lifting equipment.

Lower body measures:
5. Leg strength, measure 1. While the subject was laying on his side, the experiment held the Lafayette Manual Muscle Tester on his ankle. The subject then lifted his leg against the experimenter's resistance.
6. Leg strength, measure 2. While the subject was sitting with his legs bent, the experimenter held the Lafayette Manual Muscle Tester on his thigh just above the knee. The subject then lifted his leg against the experimenter's resistance.

There has been a great deal of research on handgrip strength on the assumption that it tracks upper body strength. However, it has not previously been shown that handgrip strength is highly correlated with upper body strength (although see Gallup et al. 2007 for links to body morphology). The Tsimane data demonstrate that handgrip strength is highly correlated with the two upper body strength measures taken: chest strength (r = .58) and shoulder strength (r = .67).

Do perceptions of men's strength reflect upper body strength more than leg strength?

In Study 1, an HLM estimating the raters use of upper and lower body strength in their rating showed that upper body strength was tracked at the expense of lower body strength. This was true using a composite of all four upper body weight-lifting exercises against the leg press machine. (These HLMs controlled for sex of rater, which never reached significance.)

Ratings	**Strength measures**	γ	s.e.	*t*	*df*	*p*	**Variance Component**	*df*	χ^2	*p*
Face only, strength	Upper body	**.31**	.030	10.17	34	10^{-11}	.0005	34	32.65	.53
	Leg	-.09	.030	-3.25	34	.003	.002	34	35.21	.41
Body only, strength	Upper body	**.44**	.022	20.05	32	10^{-19}	.0001	32	19.25	.96
	Leg	.007	.028	0.24	32	.81	.0007	32	27.07	.71
Full person, strength	Upper body	**.41**	.035	11.72	33	10^{-12}	.02	33	54.74	.01
	Leg	.06	.028	2.33	33	.03	.003	33	31.72	.53
Full person, fighting ability	Upper body	**.42**	.025	16.82	34	10^{-17}	.00018	34	25.45	.85
	Leg	.07	.029	2.37	34	.02	.00010	34	29.70	.68

One might think this pattern occurred because the upper body measure is based on four exercises and is therefore more reliable, but the same pattern is found for HLMs that pit the leg press measure against a single upper body exercise, the chest press, which is the upper body exercise most comparable in motion to the leg press:

Ratings	**Strength measures**	γ	s.e.	*t*	*df*	*p*	**Variance Component**	*df*	χ^2	*p*
Face only, strength	Chest press	**.29**	.03	10.84	34	10^{-11}	.0002	34	24.87	.87
	Leg press	-.07	.03	-2.50	34	.02	.002	34	29.67	.68
Body only, strength	Chest press	**.36**	.020	17.42	32	10^{-17}	.0001	32	17.00	.99
	Leg press	.10	.027	3.68	32	.001	.0005	32	26.92	.72
Full person, strength	Chest press	**.32**	.027	11.68	33	10^{-12}	.009	33	35.02	.37
	Leg press	.16	.027	5.69	33	10^{-5}	.009	33	34.03	.42
Full person, fighting ability	Chest press	**.34**	.019	17.78	34	10^{-18}	.00005	34	17.35	.37
	Leg press	.16	.025	6.48	34	10^{-6}	.0001	34	25.43	.42

There were no significant differences among raters for 15 of the 16 models summarized in the two tables above. The only exception was for a sample that included two persons who appeared to have inverted the strength scale; see **note on the Study 1 table below.

Detailed statistics from HLM analyses reported in the main text.

Study 1	Relationship between ratings and strength				Are there individual differences in the relationship between ratings and measured strength?		
Rated fighting ability	γ	s.e.	*t(34)*	*p*	Variance Component	χ^2	*P*
Measured strength	**.52**	0.017	30.19	10^{-25}	**0.00007**	24.41	.89
Sex of rater*	.0069	0.017	0.40	0.69			
Rated Strength							
Full person photos	γ	s.e.	*t(33)*	*p*			
Measured strength	**.50**	0.038	13.38	10^{-14}	**0.037****	110.53	10^{-9}**
Sex of rater	-.038	0.038	-1.02	.32			
Body alone photos	γ	s.e.	*t(32)*	*p*			
Measured strength	**.49**	.018	26.65	10^{-22}	**.00008**	23.82	.85
Sex of rater	-.015	.018	-0.85	.41			
Face alone photos	γ	s.e.	*t(34)*	*p*			
Measured strength	**.27**	.023	11.62	10^{-12}	**.00038**	32.20	.56
Sex of rater	-.036	.023	-1.56	.13			

* Sex of rater interactions test whether there are sex differences in accuracy. male raters = -1, female raters = +1

** The finding of significant individual differences in accuracy for full person photos was driven by two raters who appeared to invert their use of the scale (their correlations were negative; that is, they judged stronger men to be weaker than weaker men). There are no significant differences in accuracy when they are removed from the analysis.

The table above showed that there were few differences among raters in how accurately lifting strength predicted their judgments of strength and fighting ability. The table below shows the extent to which strength, height, weight and age predict ratings of fighting ability and strength when all four predictors are entered simultaneously into the analysis (i.e., the estimate for each predictor controls for the effects of the other three predictors). On average, ratings of strength and fighting ability reflect lifting strength more than height, weight or age. Controlling for the three other variables, individual differences among raters emerge in the extent to which strength tracks these ratings (the same is true for height, weight, and age). It is possible that some individuals might use height or weight more than strength in making their assessments, but this cannot be inferred from the fact that variability in use of a predictor emerges once other predictors are controlled for.

Study 1	Independent effects of target measurements on ratings*				Are there individual differences in the relationship between measurements and ratings?		
Rated fighting ability	γ	s.e.	*t(34)*	*P*	Variance Component	χ^2	*p*
strength	**.49**	0.029	16.81	10^{-17}	**.013**	49.32	.043
height	.29	0.37	7.81	10^{-8}	.028	68.52	.001
weight	-.06	0.054	-1.12	.27	.077	111.60	10^{-9}
age	.05	0.016	3.01	.005	.0017	30.40	.64
Rated Strength							
Full person photos	γ	s.e.	*t(33)*	*P*			
strength	**.50**	0.038	13.13	10^{-14}	**.038**	102.62	10^{-8}
height	.36	0.046	7.68	10^{-8}	.060	111.82	10^{-9}
weight	-.16	0.050	-3.10	.004	.061	96.73	10^{-7}
age	.06	0.012	4.79	10^{-4}	.0003	14.92	.99
Body alone photos	γ	s.e.	*t(32)*	*P*			
strength	**.52**	0.024	21.18	10^{-19}	**.007**	35.07	.32
height	.36	0.042	8.49	10^{-9}	.038	78.89	10^{-5}
weight	-.21	0.059	-3.55	.001	.093	121.51	10^{-11}
age	.10	0.014	6.82	10^{-7}	.0005	19.13	.96
Face alone photos	γ	s.e.	*t(34)*	*P*			
strength	**.22**	0.027	8.27	10^{-8}	**.002**	33.62	.49
height	.05	0.032	1.70	.10	.004	36.59	.35
weight	.07	0.034	2.05	.05	.004	33.40	.50
age	.04	0.025	1.70	.10	.008	50.58	.03

*controlling for sex of rater; with one exception, men and women did not significantly differ in their use of strength, height, weight, or age.

Study 2 **Male targets:** **Rated Strength**	Relationship between strength ratings and measured strength				Are there individual differences in the relationship between strength ratings and measured strength?		
Body alone photos	γ	s.e.	*t(32)*	*p*	Variance Component	χ^2	*p*
Measured strength	**.43**	0.042	10.17	10^{-10}	**.05**	173.77	10^{-21}
Sex of rater*	-.03	0.042	-.71	.48			
Face alone photos	γ	s.e.	*t(34)*	*p*			
Measured strength	**.22**	0.019	11.66	10^{-12}	**.001**	36.94	.33
Sex of rater	.02	0.019	0.82	.42			

*Sex of rater interactions test whether there are sex differences in accuracy. male raters = -1, female raters = +1

Study 2 Male Targets: Rated Strength	Independent effects of target measurements on strength ratings*				Are there individual differences in the relationship between measurements and strength ratings?		
Body alone photos	γ	s.e.	*t(32)*	*p*	Variance Component	χ^2	*p*
strength	**.46**	0.042	11.13	10^{-11}	**.039**	97.96	10^{-8}
height	.11	0.025	4.44	.0001	.014	80.18	10^{-5}
weight	-.11	0.041	-2.47	.02	.049	144.65	10^{-16}
age	.08	0.014	5.56	10^{-5}	.0007	30.59	.54
Face alone photos	γ	s.e.	*t(34)*	*p*			
strength	**.13**	0.032	3.99	.0003	**.0071**	43.54	.13
height	-.05	0.025	-1.96	.06	.012	72.77	.0001
weight	.14	0.039	3.56	.001	.038	112.37	10^{-9}
age	-.01	0.014	-0.77	.45	.0002	24.34	.89

*controlling for sex of rater; men and women did not significantly differ in their use of strength, height, weight, or age.

Study 2 Female targets: Rated Strength	Relationship between strength ratings and measured strength				Are there individual differences in the relationship between strength ratings and measured strength?		
Body alone photos	γ	s.e.	*t(30)*	*p*	Variance Component	χ^2	*P*
Measured strength	**.27**	0.048	5.62	10^{-5}	**.07**	226.77	10^{-31}
Sex of rater*	-.10	0.048	-2.11	.04			
Face alone photos	γ	s.e.	*t(28)*	*p*			
Measured strength	**.14**	0.018	7.96	10^{-7}	**.0004**	27.21	.61
Sex of rater	-.05	0.018	-2.79	.01			

*Sex of rater interactions test whether there are sex differences in accuracy. male raters = -1, female raters = +1

Study 2 Female Targets: Rated Strength	Independent effects of target measurements on ratings*				Are there individual differences in the relationship between measurements and strength ratings?		
Body alone photos	γ	s.e.	*t(30)*	*p*	Variance Component	χ^2	*p*
strength	**.15**	0.021	7.29	10^{-7}	**.0027**	27.06	.62
height	.17	0.039	4.45	.0001	.042	214.99	10^{-29}
weight	.05	0.058	0.92	.37	.105	329.32	10^{-51}
age	-.05	0.016	-2.98	.006	.003	45.65	.033
Face alone photos	γ	s.e.	*t(28)*	*p*			
strength	.04	0.031	1.42	.17	**.011**	42.28	.041
height	.05	0.019	2.49	.02	.004	41.89	.044
weight	.09	0.030	2.88	.008	.014	58.59	.001
age	-.02	0.013	-1.28	.21	.0007	21.89	.79

*controlling for sex of rater; men and women did not significantly differ in their use of strength, height, weight, or age.

Study 3 Tsimane men: Rated Strength	Relationship between strength ratings and measured strength				Are there individual differences in the relationship between strength ratings and measured strength?		
Face alone photos	γ	s.e.	*t(30)*	*p*	Variance Component	χ^2	*p*
Measured strength	**.30**	0.034	8.88	10^{-9}	**.01**	40.90	.09
Sex of rater*	.03	0.034	-.87	.39			

*Sex of rater interactions test whether there are sex differences in accuracy. male raters = -1, female raters = +1

Study 3 Tsimane men: Rated Strength	Independent effects of target measurements on strength ratings*				Are there individual differences in the relationship between measurements and ratings?		
Face alone photos	γ	s.e.	*t(30)*	*p*	Variance Component	χ^2	*p*
strength	**.18**	0.047	3.77	.001	**.011**	30.97	.42
height	.07	0.034	1.94	.06	.009	39.87	.11
weight	.06	0.039	1.64	.11	.013	40.52	.10
age	-.05	0.042	-1.28	.21	.034	63.55	.001

*controlling for sex of rater; men and women did not significantly differ in their use of strength, height, weight, or age.

Study 4 Andean men: Rated Strength	Relationship between strength ratings and measured strength				Are there individual differences in the relationship between strength ratings and measured strength?		
Face alone photos	γ	s.e.	*t(26)*	*p*	Variance Component	χ^2	*p*
Measured strength	**.29**	0.037	7.77	10^{-7}	**.0004**	24.00	.58
Sex of rater*	-.04	0.037	-1.11	.28			

*Sex of rater interactions test whether there are sex differences in accuracy. male raters = -1, female raters = +1

Study 4 Andean men: Rated Strength	Independent effects of target measurements on strength ratings*				Are there individual differences in the relationship between measurements and ratings?		
Face alone photos	γ	s.e.	*t(26)*	*p*	Variance Component	χ^2	*p*
strength	**.43**	0.073	5.94	10^{-5}	**.053**	30.52	.25
height	.05	0.052	0.95	.35	.044	40.71	.03
weight	.01	0.11	0.05	.96	.22	77.54	10^{-6}
age	-.22	0.049	-4.40	.0002	.049	53.92	.001

*controlling for sex of rater; men and women did not significantly differ in their use of strength, height, weight, or age.