Human adaptations for the visual assessment of strength and fighting ability from the body and face

Aaron Sell[1,2,*], Leda Cosmides[1,2], John Tooby[1,3], Daniel Sznycer[1,3], Christopher von Rueden[1,3] 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*

Selection in species with aggressive social interactions favours the evolution of cognitive mechanisms for assessing physical formidability (fighting ability or resource-holding potential). The ability to accurately assess formidability in conspecifics has been documented in a number of non-human species, but has not been demonstrated in humans. Here, we report tests supporting the hypothesis that the human cognitive architecture includes mechanisms that assess fighting ability—mechanisms that focus on correlates of upper-body strength. Across diverse samples of targets that included US college students, Bolivian horticulturalists and Andean pastoralists, subjects in the US were able to accurately estimate the physical strength of male targets from photos of their bodies and faces. Hierarchical linear modelling shows that subjects were extracting cues of strength that were largely independent of height, weight and age, and that corresponded most strongly to objective measures of upper-body strength—even when the face was all that was available for inspection. Estimates of women's strength were less accurate, but still significant. These studies are the first empirical demonstration that, for humans, judgements of strength and judgements of fighting ability not only track each other, but accurately track actual upper-body strength.

Keywords: RHP; formidability; physical strength; face; evolutionary psychology; aggression

1. INTRODUCTION

The negotiating position of an organism is, in large part, a function of the magnitude of the costs that it can inflict on competitors—i.e. its resource-holding potential (Parker 1974) or formidability. Consequently, in social species such as humans, natural selection typically organizes adaptations designed to enhance the organism's capacity to inflict damage. In order to make advantageous decisions about when to persevere or defer in conflicts, members of social species benefit from being able to make accurate assessments of individual differences in aggressive formidability. Indeed, the so-called ritualized animal contests are widely interpreted as joint advertisements of formidability, during which animals demonstrate (and exaggerate) cues of their fighting ability, as well as use their observations of others to modulate their subsequent actions cost-effectively (for reviews see Huntingford & Turner 1987; Archer 1988; Krebs & Davies 1993). Obviously, methods of assessment that can be carried out at a distance and prior to combat—such as through visual inspection—are more advantageous than methods that entail the risk of damage inherent in direct physical contact.

A growing body of evidence supports the view that the selection pressures shaping non-human conflict also applied to ancestral humans—a conclusion supported by the existence of a number of evolved parallels between humans and comparable species. These include signals of imminent aggression such as facial expressions (Ekman *et al.* 1987), vocal changes (Scherer *et al.* 2001), and body postures (Duclos *et al.* 1989), signals of submission (Keltner & Buswell 1997), patterns of escalation with the most violent outcomes reserved for situations in which both combatants are evenly matched (Luckenbill 1977; Daly & Wilson 1988), and greater investment in adaptations for aggression in the sex that is most reproductively limited by access to mates (Daly & Wilson 1983). Of equal importance, widespread palaeoanthropological evidence indicates that aggressive conflict among our foraging ancestors was substantial enough to constitute a major selection pressure, especially on males (Manson & Wrangham 1991; Keeley 1996). It is suggestive that in some ethnographically investigated small-scale societies where actual rates can be measured, a third of the adult males are reported to die violently (Keeley 1996), with rates going as high as 59 per cent, reported for the Achuar (Bennett Ross 1984).

Evidence of aggression among our ancestors, as well as its ubiquitous role in the social lives of related species (Symons 1978; Smuts *et al.* 1987), leads to the expectation that the human cognitive architecture includes mechanisms that are well designed for extracting information about formidability from cues that were typically available in ancestral environments. Despite the fact than an individual's formidability can be substantially augmented by allies—a process of great importance in humans—the assessment of individual formidability remains critical in social decision-making, because adversaries are often encountered alone or without allies, conflicts emerge among individuals within alliances, and

(in small-scale societies) the formidability of an alliance must be computed in part from the individuals that constitute it.

Anatomical evidence supports the view that, for ancestral humans, the single most important factor driving the differential ability to inflict costs was upper-body strength. In humans, the view that upper-body strength is more relevant for fighting than lower-body strength is empirically supported by the considerable sexual dimorphism in human upper-body size and strength (for review see Lassek & Gaulin in preparation). Men, for example, have approximately 75 per cent more muscle mass than women in the arms, but only 50 per cent more muscle mass in legs. Although ancestral humans were zoologically unusual in their use of tools in some types of aggression, the force driving the weapon remains largely a function of upper-body strength (Brues 1959). Moreover, given the persistence of weapons-free fights (at least within social groups) after the emergence of tool use and continuing into the present, it seems likely that the neurocomputational assessment specializations that evolved during the tens of millions of years prior to tool use would remain useful and be maintained. Skill in fighting and weapon use, as well as neurologically driven differences such as rapid reflexes or courage, are obviously also relevant to formidability, but require narrowly restricted observational circumstances for their assessment. Because human males are substantially stronger than human females (Lassek & Gaulin in preparation), and deploy physically aggressive strategies more often (Daly & Wilson 1988; Campbell 1999; Archer 2004), the cognitive specializations for strength assessment are expected to be better engineered for evaluating males than females.

Finally, a number of factors suggest that selection should have tailored strength assessment specializations to use information present in the face alone, if there are cues in the face that reliably predict strength. Under ancestral conditions, the upper body would sometimes have been obscured by clothing, carried objects, other people, vegetation and other obstructions. If the face also manifested cues of strength, then this would have provided a separate channel when direct assessment of the relevant musculature was not possible. The brain is known to contain neurocomputational specializations designed to extract dynamic and static social information from the face, including identity, eye direction, emotional state, sex, age, attractiveness and long-term testosterone exposure (Bruce & Young 1986; Baron-Cohen 1995; Sugiyama 2005). Given the documented array of social face-processing competences, the hypothesis that strength detection coevolved with the rest of face processing seems worthy of empirical testing. Indeed, because aggression-related decisions are often made under severe time constraints, special efficiencies may arise from the ability to interrelate data streams from parallel face-processing mechanisms.

In short, the studies reported here were designed to test the hypothesis that the human neurocognitive architecture includes mechanisms that are well designed to visually assess individual formidability, especially in males, through accurately assessing their upper-body strength from cues present in the body and face.

(a) Predictions

In the four studies presented below, we tested the following predictions.

— Humans should be good at assessing others' actual physical strength from exposure to visual cues in the body—i.e. perceived physical strength should track actual physical strength.
— Because the use of physical aggression was far more common and consequential for males than females in ancestral environments, the assessment of physical strength should be more accurate for male targets than for female targets.
— Because strength is the primary variable underlying the ability to inflict costs physically, assessments of men's ability to win fights should be most strongly correlated with assessments of men's physical strength, rather than their height, weight, age or other cues.
— Assessments of strength (and fighting ability) should disproportionately reflect upper-body strength or its correlates, compared with other components of strength, such as leg strength, or other cues, such as size. This should be true if strength judgements are produced by adaptations designed to assess fighting ability, because upper-body strength is more closely causally tied to the capacity to inflict damage through aggression.
— Because the face is a frequent locus of attention and the least occluded body area, humans should be able to assess physical strength using only information available in the face.
— Face-derived strength assessments should also correspond most closely to upper-body strength, the variable that most strongly predicts formidability.
— If strength assessment is the output of a species-typical adaptation, and is based on species-typical cues in the body and face, then individuals should be able to accurately assess strength from members of other cultures.

2. STUDY 1

The purpose of study 1 was to determine whether people can assess men's strength based on photos of the face alone, the body alone and the full person. In addition, study 1 explicitly tests the hypothesis that judgements of physical strength are related to judgements of aggressive formidability. The stimuli were photos of American men whose strength had been measured using weight-lifting machines at a gymnasium.

(a) Stimulus subjects

To create stimuli, 59 male undergraduates were recruited from a campus gym at the University of California, Santa Barbara (UCSB) and paid $10 for their participation (mean age: 21.1, s.d. 2.4, range 18–32; 62% Euro-American, 15% Asian-American, 5% African-American, 2% Middle Eastern, 5% Hispanic, 11% other, with no significant differences in strength as a function of ethnicity). Each man was assessed individually. Each completed a brief questionnaire (not reported), after which his body measurements (height, weight, etc.; see electronic supplementary materials (ESM)) and photographs were taken. Finally, his physical strength was assessed on five weight-lifting machines.
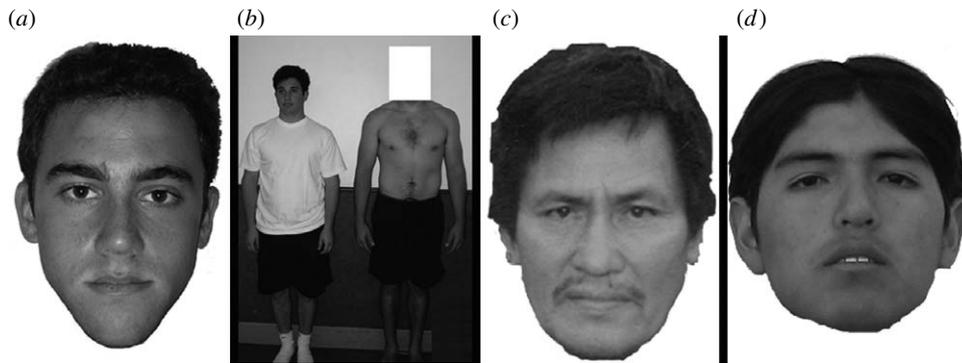
Figure 1. (*a*) US face, (*b*) US body, (*c*) Bolivian face and (*d*) Andean face photographs.

(i) *Photographs*

Stimulus subjects were dressed in standard black gym shorts and no shirts (to show upper-body musculature), and asked to keep a neutral expression. They posed for two colour photographs: (i) face only, facing forward; and (ii) full body, facing forward. *Face photos*: The face photographs were cropped below the jaw so that subjects' necks were not visible; this was to get a pure facial measure, unaffected by assessments of neck muscles (figure 1*a*). To maximize visibility of features, the faces were magnified to fill a standard-size box; a limitation of this choice is that it obscures information about relative head size. The second photograph was used for a full-person and a body-only photograph. *Full-person photo*: Each stimulus subject stood next to a male experimenter; this provided a benchmark allowing raters to determine the subject's relative height. *Body-only photo*: Using Photoshop v.8.0, the full person photos were edited to remove the subject's face, as shown in figure 1*b*.

(ii) *Strength measurement*

Upper-body strength was assessed on four weight-lifting machines in random order: arm curl, abdominal crunch, chest press and super long pull (see ESM for details). The most weight the subject could successfully lift 10 times on each machine was converted to z-scores and averaged together to create a single composite score representing the upper-body strength of each subject (reliability = 0.92 (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 for 47 of the 59 stimulus subjects; order was randomized with the four upper-body machines (leg-press data could not be collected from the last 12 subjects due to a machine malfunction).

(b) *Face and body ratings*

An additional 142 UCSB undergraduates were paid \$5 each to rate the three sets of 59 photographs. Each subject rated only one set (full person, body only or face only). In all studies, photos were presented on a computer screen (resolution: 1024×768). Order of presentation was randomized across raters.

To obtain strength judgements, raters were asked 'Please rate the following [men/bodies/faces] on how physically strong you think the man is compared to other men of his age' using a 7-point Likert scale (1 = very weak; 7 = very strong). The same instructions were used in studies 2–4. For full-person photos, $n=35$ (19 female); for body-only photos, $n=34$ (18 female); for face-only photos, $n=36$ (22 female).

To determine whether judgements of physical strength are indeed related to judgements of aggressive formidability, an additional 37 subjects (25 female) were asked to rate the full-person photos on aggressive formidability, with the instruction 'Please look at the following photographs of men and rate them on how tough each would be in a physical fight—how likely he would be to beat his opponent', with a 7-point Likert scale (1 = not tough; 7 = very tough).

Reliability among raters was high (full person: Cronbach's $\alpha=0.95$ (strength), 0.96 (toughness); body only (strength): $\alpha=0.96$; face only (strength): $\alpha=0.90$). Strength ratings for each target photo were averaged to find the average perceived strength of each target individual, and these were correlated with the targets' actual lifting strength.

We were also interested in the average accuracy of individual raters. Each subject rated multiple targets, so ratings are nested under individuals; given this nested structure, hierarchical linear modelling (HLM) is the most appropriate analytic tool. For each study, one hierarchical linear model was used to estimate the average accuracy of individual raters, controlling for their sex (there were no effects of rater's sex unless otherwise noted). A second HLM was used to estimate the effects of lifting strength, height, weight and age on raters' strength judgements. Prior to being entered into the HLM, all variables were z-scored (ratings of strength and ratings of toughness were standardized within rater). This makes the HLM's gamma coefficients (γ) equivalent to standardized βs in multiple regressions (and the first level intercepts zero for all analyses).

The results of all studies are summarized in table 1, with details in the electronic supplementary material.

(c) *Results of study 1*
(i) *Can people assess men's strength from static visual images?*

Yes. The correlation between the average perceived strength score for a target male and his actual upper-body lifting strength was $r=0.71$ ($p=10^{-10}$) for photographs of the full person. It was almost as large—$r=0.66$ ($p=10^{-8}$)—for photographs of the body alone. When subjects saw only a man's face, they were still able to assess his strength: $r=0.45$ ($p=0.0003$). This robust correlation

Table 1. Summary of results.

	full person		body only			face only				
	fighting ratings	strength ratings	strength ratings			strength ratings				
	men	men	men	men	women	men	men	Tsimane	Andean	women
	US-1	US-1	US-1	US-2	US-2	US-1	US-2			US-2
relationship between ratings and upper body strength										
averaged across raters r (p)	0.69 (10^{-9})	0.71 (10^{-10})	0.66 (10^{-8})	0.57 (10^{-8})	0.51 (10^{-11})	0.45 (0.0003)	0.39 (10^{-5})	0.52 (0.0001)	0.47 (0.01)	0.21 (0.01)
average individual accuracy γ (p)	0.52 (10^{-25})	0.50 (10^{-14})	0.49 (10^{-22})	0.43 (10^{-10})	0.27 (10^{-5})	0.27 (10^{-12})	0.22 (10^{-12})	0.30 (10^{-9})	0.29 (10^{-7})	0.14 (10^{-7})
independent effects of target measurements on ratings of strength or fighting ability										
measured strength γ (p)	0.49 (10^{-17})	0.50 (10^{-14})	0.52 (10^{-19})	0.46 (10^{-11})	0.15 (10^{-7})	0.22 (10^{-8})	0.13 (0.0003)	0.18 (0.001)	0.43 (10^{-5})	0.04 n.s.
height γ (p)	0.29 (10^{-8})	0.36 (10^{-8})	0.36 (10^{-9})	0.11 (0.0001)	0.17 (0.0001)	0.05 n.s.	−0.05 n.s.	0.07 n.s.	0.05 n.s.	0.05 (0.02)
weight γ (p)	−0.06 n.s.	−0.16 (0.004)	−0.21 (0.001)	−0.11 (0.02)	0.05 n.s.	0.07 n.s.	0.14 (0.001)	0.07 n.s.	0.01 n.s.	0.09 (0.008)
age γ (p)	0.05 (0.005)	0.06 (0.0001)	0.09 (10^{-7})	0.08 (10^{-5})	−0.05 (0.006)	0.04 n.s.	−0.01 n.s.	−0.05 n.s.	−0.22 (0.0002)	−0.02 n.s.

shows that men's faces contain cues from which strength can be accurately assessed, and that human minds are tuned to pick up on these cues.

To confirm that the accurate assessment of strength emerges not only when data are aggregated across raters, but also exists at the individual level, HLM was used to determine the average accuracy of individuals and the extent to which raters differed in their accuracy. Individual accuracy was good, estimated as $\gamma = 0.50$ ($p = 10^{-14}$) for full-person photos, 0.49 ($p = 10^{-22}$)—almost identical—for body-alone photos and 0.27 ($p = 10^{-12}$) for photos of the face alone. (Given that all variables have been standardized, the magnitude of each gamma can be interpreted in the same way as a standardized regression coefficient.) For the body-alone and face-alone photos, there were no appreciable differences across raters in their accuracy (as reflected in small and non-significant variance components; see the electronic supplementary material).

For this and all other studies involving male targets, there were no significant differences in the accuracy of male and female raters.

(ii) Are perceptions of fighting ability related to perceptions of physical strength?

Yes. Ratings of toughness in a physical fight, defined as how likely the man would be to beat an opponent, were averaged for each target to obtain a measure of his perceived fighting ability. These perceptions of the targets' fighting ability were almost perfectly correlated with perceptions of their physical strength: $r = 0.96$ ($p = 10^{-32}$). This striking relationship exists even though fighting ability and strength were rated by different groups of subjects. Furthermore, perceptions of a man's fighting ability and perceptions of his strength were both predicted by his upper-body lifting strength to the same degree: $r = 0.69$ ($p = 10^{-9}$) for fighting ability and 0.71 for strength. The same pattern emerges when ratings for a target are disaggregated so that the individual rater is the unit of analysis: the assessments that individual raters make are predicted by the targets' actual lifting strength to the same degree, whether they are asked to assess fighting ability ($\gamma = 0.52$, $p = 10^{-25}$) or strength (0.50). Across raters, there were no significant differences in the extent to which lifting strength predicted their ratings of fighting ability. Taken together, these results support the hypothesis that judgements of physical strength and judgements of fighting ability are based on the same underlying computations, at least when they are made in response to stationary visual stimuli.

Although a tournament to determine the targets' actual fighting ability would be neither ethical nor practical, we know the number of fights target males reported to have had during the last 4 years (electronic supplementary material). If more formidable men are more likely to initiate fights or less likely to avoid them (because they are more likely to win), then the number of fights a man has been in would be a rough index of his actual formidability. Indeed, men who were seen as tougher had been in more fights: there was a correlation of $r = 0.30$ ($p = 0.02$) between average toughness scores and the targets' actual fighting behaviour. This means that perceptions of a man's fighting ability track a real world behaviour that is a plausible index of his actual formidability.

3. STUDY 2

In study 1, all the targets were men who work out at a gym. If strength training with machines modifies the musculature of the face and body in ways uncharacteristic of normal exertion, the ability to assess strength could be restricted to this population. So the goal of study 2 was to determine whether the ability to visually assess strength generalizes to men drawn from a broader population. Study 2 also extends the analysis by including female targets. Because ratings based on the body alone were almost as accurate as those based on the full person, and because we are interested in which cues are used to infer strength, ratings were based on the body alone or the face alone.

(a) Stimulus subjects

To create stimuli, 109 male and 146 female students were recruited from a building at UCSB frequented by undergraduates of all majors, which houses food kiosks and a bookstore (mean age: 19.4 years, s.d. =1.67; 58% Euro-American, 16% Asian-American, 13% Hispanic, 4% Middle Eastern, 5% African-American, 3% other, with no significant differences in strength as a function of ethnicity). Target subjects were paid $10 for their participation.

(i) Photographs

After taking body measurements, target subjects were posed for a single full-person photograph against a constant background. Men removed their shirts for the photographs; for cultural reasons, women could not be photographed shirtless, and were instead given a white T-shirt to wear over their shirts to standardize style of dress. Face-alone and body-alone versions of each photo were created using PHOTOSHOP v.8.0, as in study 1.

(ii) Strength measurement

Chest/arm strength was obtained by using a Rolyan Hydraulic hand dynamometer with its handles inverted (the electronic supplementary material). Also taken were two proxy measures of strength that had been validated in study 1 (the electronic supplementary material), which together account for 60 per cent of the variance in upper-body strength as measured by the weight-lifting machines: flexed biceps circumference and a self-reported question ('I am physically stronger than _____% of others of my sex'). The reliability of the three measures (Cronbach's α) was 0.78 for male targets and 0.66 for female targets. Each strength measure was converted to a z-score, and the three z-scores for a target were averaged to form a single composite measure of physical strength.

(b) Face and body ratings

An additional 132 UCSB undergraduates (76 female) were paid $5 to rate the photographs. Owing to the large number of targets, each rater saw only faces or bodies of one sex. Inter-rater reliability was high (male targets: bodies $\alpha=0.94$, faces $\alpha=0.89$; female targets: bodies $\alpha=0.86$, faces $\alpha=0.81$).

(c) Results of study 2

(i) Does the ability to assess strength in men generalize to a broader population of male targets?

It does. *Body alone.* Even though the male targets had been recruited from a popular campus location having no

relationship to a gym or weight-training, average perceived strength assessed from the body alone was robustly correlated with measured strength, $r=0.57$ ($p=10^{-8}$). The average accuracy of individual raters was estimated as $\gamma=0.43$ ($p=10^{-10}$).

Face alone. Average perceived strength scores based on the face alone were correlated with measured strength at $r=0.39$ ($p=10^{-5}$), similar to the value of 0.45 found in study 1. The average accuracy of individuals was $\gamma=0.22$ ($p=10^{-12}$), similar to the γ of 0.27 in study 1. Across raters, there were no appreciable differences in accuracy (see the electronic supplementary material).

(ii) Can people assess strength from women's bodies?

Yes. Average ratings of strength from photographs of women's bodies correlated with their measured strength at levels similar to those for men's bodies, $r=0.51, p=10^{-11}$. The average accuracy of individuals was estimated as $\gamma=0.27$ ($p=10^{-5}$) by an HLM that controlled for sex of rater. Sex of rater was also significant ($\gamma=-0.10, p=0.043$), with male raters being more accurate than female raters (γs=0.37 versus 0.17).

(iii) Can people assess strength from women's faces?

Yes, but not very well. Average perceived strength judged from the face alone was correlated with women's measured strength at $r=0.21$ ($p=0.01$), half the effect size found for male faces. The average accuracy of individuals was low, estimated by HLM as $\gamma=0.14$ ($p=10^{-7}$), and showed that male raters performed better than female raters ($p=0.01$; γs=0.19 versus 0.09). There were no appreciable differences in individual accuracy beyond this sex difference (see the electronic supplementary material).

4. STUDY 3

If there are species-typical cues of strength in men, and natural selection has designed mechanisms that use these cues for assessing strength, then these mechanisms should successfully assess strength even when operating on photographs of men from unfamiliar cultures. In study 3, we tested this prediction using the faces of men from an indigenous Amerindian group. The targets were Tsimane men from the Ton'tumsi village in Bolivia whose physical strength had been measured; the raters were undergraduates at UCSB.

(a) Stimulus subjects

Fifty-three adult Tsimane men had strength measurements and photographs taken as part of a larger project (see von Rueden *et al.* 2008; ages 19–77, mean age= 37.0 years, s.d. =14.5). The Tsimane are forager–horticulturalists who inhabit lowland Bolivia along the Maniqui River and in adjacent forests. Tsimane families may spend weeks or months on hunting or fishing trips away from settled villages; however, the Tsimane are semi-sedentary and live in communities ranging from 30 to 500 individuals.

(i) Photographs

Using PHOTOSHOP, photographs of men's faces were cropped below the jaw line and around the head to completely remove the background as in study 1

(figure 1*c*). Because the photos used in studies 3 and 4 were taken for field purposes, the poses are slightly less standardized than those in studies 1 and 2, and the resolution is lower.

(ii) *Strength measurement*
Direct measurements of chest, shoulder and handgrip strength, along with flexed biceps circumference, were taken (reliability: $\alpha = 0.81$); they were *z*-scored and combined to form a composite measure of upper body strength. The same was done for two direct measures of leg strength ($\alpha = 0.75$; see the electronic supplementary material for measurement procedures). Height and weight measures were also taken.

(b) *Face ratings*
Thirty-two UCSB undergraduates (17 female) were paid $5 to rate the strength of Tsimane men based on their faces. Inter-rater agreement was high (Cronbach's $\alpha = 0.86$).

(c) *Results of study 3*
Ratings of physical strength based on photographs of the face alone were highly correlated with the targets' upper-body strength: $r = 0.52$, $p = 0.0001$. For individual raters, average accuracy was estimated as $\gamma = 0.30$ ($p = 10^{-9}$), with no significant difference among raters in their accuracy (see the electronic supplementary material). These correlations are as high as, or higher than, those for the faces of men from the raters' own culture (table 1).

5. STUDY 4
Study 4 had the same goal as study 3: to find whether the ability to assess men's strength from their faces generalizes to faces drawn from unfamiliar populations. In study 4, the faces were drawn from an Andean pastoralist population, herder–horticulturalists from the villages of Gobernador Solá and Ingeniero Maury in the province of Salta, Argentina.

(a) *Stimulus subjects*
Twenty-eight adult men had strength measurements taken and were photographed as part of a larger project (ages 15–71, mean age = 34.3, s.d. = 17.37). The villagers have a mixed economy of goat herding, horticulture and market exchange (see Raffino 1972; Tarragó 2000).

(i) *Photographs*
Photographs of the Andean men were taken against a neutral background. Face-alone photos were created using PHOTOSHOP, as described in study 1 (figure 1*d*).

(ii) *Strength measurement*
Flexed-biceps circumference and a direct measure of chest/arm strength (used in study 2; see the electronic supplementary material) were taken ($\alpha = 0.61$); they were *z*-scored and averaged into a single composite measure of upper-body strength. Height and weight were also obtained for 20 of the 28 subjects.

(b) *Face ratings*
Twenty-eight UCSB undergraduates (19 females) were paid $5 to rate strength from the Andean men's faces. Instructions for rating physical strength based on the faces

were the same as in the previous studies. Inter-rater agreement was high ($\alpha = 0.88$).

(c) *Results of study 4*
Based on the face alone, average perceived strength scores were well correlated with men's measured physical strength, $r = 0.47$ ($p = 0.01$). Accuracy of individual raters was estimated as $\gamma = 0.29$ ($p = 10^{-7}$), with no significant difference among raters in their accuracy (see the electronic supplementary material). As with the Tsimane, these correlations are as high as, or higher than, those found for the faces of culturally familiar men.

6. DO PERCEPTIONS OF MEN'S STRENGTH REFLECT UPPER BODY STRENGTH MORE THAN LEG STRENGTH?
Fighting ability is disproportionately a function of upper-body strength; if strength assessments are made by a mechanism that is specialized for judging fighting ability, then they should also disproportionately reflect upper-body strength. To determine whether this was the case, hierarchical linear models were made of ratings of men's strength from study 1, in which the two predictor variables were upper-body strength and leg strength as measured on the leg press; these models controlled for sex of rater (n.s.). Conceptually, ratings based on the full person and body alone are less interesting because the men were wearing gym shorts that partly obscured their thighs; this might bias subjects towards relying more heavily on the upper body for theoretically uninteresting reasons. Most probative are ratings based on the face alone, because neither the upper body nor the legs were visible when subjects were making these judgements.

Strength ratings based on the face alone were positively associated with upper-body strength, $\gamma = 0.31$ ($p = 10^{-11}$), but not with leg strength, which yielded a small negative effect ($\gamma = -0.09$, $p = 0.003$). For ratings of strength and fighting ability from photos of the full person and body alone, the effect sizes for upper-body strength were positive and robust, ranging from $\gamma = 0.41$ to 0.44, whereas those for leg strength were small, ranging from $\gamma = 0.007$ to 0.07. 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 the chest press, the single upper-body exercise that is most comparable in motion to the leg press. In general, there were no appreciable differences among raters in these analyses (for details, see the electronic supplementary material).

The other population for which measures of both leg and upper-body strength were available was the Tsimane (study 3). An HLM of facial ratings of Tsimane men's strength, with upper-body and leg strength entered as predictor variables, showed that the effect size for upper-body strength, $\gamma = 0.22$ ($p = 10^{-5}$), was about 10 times larger than the (non-significant) effect size for leg strength ($\gamma = 0.023$; no significant difference among raters). Using faces from an unfamiliar indigenous Amerindian group, the results of study 1 were replicated.

An adaptation specialized for assessing fighting ability should disproportionately weight upper-body strength;

people's ratings of physical strength show precisely this bias, even when they are based on the face alone.

7. MUSCULARITY OR BODY SIZE?

Upper-body lifting strength is correlated with weight and height (see ESM: study 1), raising the possibility that people's judgements of strength are tracking overall body size, with no effects attributable to cues such as muscularity. To gain insight into which physical features were being used by subjects in rating strength and fighting ability, HLM was used to estimate the extent to which subjects' strength judgements reflected the targets' measured strength, height, weight and age in each study. The results of these analyses are shown in the bottom half of table 1; each γ represents the effect size for that predictor variable, controlling for all the others.

(a) *Male targets*

Whenever subjects could see men's bodies (full-person and body-alone photos), measured strength was always a significant independent predictor of both their strength ratings and their ratings of the men's fighting ability ($10^{-11} \leq p\text{s} \leq 10^{-19}$). The effect sizes for measured strength were reliable and robust ($0.46 \leq \gamma\text{s} \leq 0.52$); moreover, they were virtually identical for ratings of fighting ability and ratings of strength ($0.49, 0.50$). As table 1 shows, the effect sizes for measured strength were always considerably larger than those for any other body size variable. Because these large effects of measured strength are found in analyses that control for effects of body size, the cognitive mechanisms whereby people assess strength must be using visual cues of lifting strength, such as muscularity, that are independent of overall body size. The results further suggest that, when actual strength is held constant, men are seen as stronger and better fighters when they are taller ($0.11 \leq \gamma\text{s} \leq 0.36$) and leaner ($-0.06 \leq \gamma\text{s} \leq -0.21$). However, these effects were smaller than the effects of measured strength, and less consistent, dropping in importance for targets drawn from outside the gym (study 2). There was a tiny effect of age ($0.05 \leq \gamma\text{s} \leq 0.09$), consistent with the fact that men continue to physically mature during this age span.

When subjects could see nothing but a man's face, measured strength was always a significant predictor of their strength ratings ($0.001 \leq p\text{s} \leq 10^{-8}$), with effect sizes two (or more) times larger than those for any other predictor in three of the four studies ($0.13 \leq \gamma\text{s} \leq 0.43$). No other variables yielded effects that were consistent across populations (we note, however, that in study 2, the effect size for weight was positive and similar to that for measured strength).

(b) *Female targets*

As judges of women's strength, male raters were more accurate than female raters, but an HLM testing for interactions showed no difference in the extent to which measured strength, height, weight and age were affecting their judgements, whether these were based on photos of women's bodies or just their faces.

Unlike the results for male targets, an HLM on ratings of strength from women's bodies showed that height was the best predictor of individual ratings of women's strength ($\gamma = 0.17$), followed closely by measured strength

($\gamma = 0.15$). There was a small effect of age ($\gamma = -0.05$), with younger women being judged a bit stronger than older women, and no significant effect of weight. Because more of the upper body was covered for female than male targets, the fact that measured strength was a better predictor of strength ratings for male than for female bodies should be interpreted cautiously. Face photographs, however, do not have this limitation. Not only was individual accuracy for women's faces low ($\gamma = 0.14$), but measured strength was not a significant independent predictor of perceived strength for female faces ($\gamma = 0.04$); instead, weight was the strongest independent predictor. This result is in sharp contrast to the four studies of male faces, where measured upper-body strength was always a significant independent predictor of perceived strength, usually with the largest effect size. Controlling for body size, the effect of upper-body strength on perceived strength was 3–5 times larger for male faces than for female faces in the US samples.

8. GENERAL DISCUSSION AND CONCLUSIONS

Decisions have larger pay-offs when uncertainty can be reduced and the outcomes of alternative courses of action can be better predicted. Evolutionary analyses indicate that one of the key variables governing social interactions in species like humans should be formidability, the relative ability to inflict costs. These studies explore the hypothesis that the human cognitive architecture is well engineered to detect formidability in others visually. The results of studies 1–4 are summarized in table 1. All predictions were supported. When asked to rate the strength of men from photographs of full persons or the body alone, people were able to do so accurately, even though the stimuli were small and static, and hence substantially degraded compared with real visual exposure to others in normally experienced social environments. Tellingly, when asked about strength, the subjects supplied estimates that disproportionately track upper-body strength, the strength component most relevant to the ability to win fights, over all other measures. Even more striking, when asked to rate each man's ability to win fights, their perceptions of the men's fighting ability were almost perfectly correlated with their perceptions of the men's strength. Not only was average perceived strength strongly correlated with actual strength, but performance was robust when analysed at the individual level as well, suggesting a capacity that reliably develops across individuals. Taken together, these results imply that the cognitive abilities underlying strength perception and representation were specifically shaped by selection for accurate formidability assessment.

Humans are also good at assessing strength based on the face alone. Even though no part of the men's bodies was available for inspection in these photos, the subjects were able to successfully perceive strength. Indeed, in our data, upper-body strength independently predicted facial ratings of strength, while leg strength did not. This means that the cues the strength detection system is using to judge a man's strength from his face are ones that disproportionately weight the component of strength that is most relevant to fighting ability. This supports

the hypothesis that social face processing includes mechanisms designed for formidability detection.

It is often assumed that fighting ability, and judgements of fighting ability, are primarily a function of body size. Our findings indicate otherwise. Men's upper-body lifting strength robustly predicted their perceived strength and fighting ability, even when controlling for their body size and age; when pitted against each other, measured strength always predicted ratings of men's strength and fighting ability better than height, weight and age did when the body was visible, and it was usually the strongest predictor even when raters could see only the face. This means that people are tracking cues of upper-body strength, such as muscularity, that are independent of body size. All else equal, taller, leaner men were seen as stronger when their bodies were visible, but these effects were smaller and less consistent than those for upper-body strength. Height does nevertheless predict strength judgements, and it also predicts reach, which is likely to be an independent contributor to formidability. Along with sex, height is also easy to perceive at a distance, so sex and height might provide the earliest and fastest formidability assessment, to be revised on closer encounter.

Male and female distributions in upper-body strength overlap by less than 10 per cent, with over 99 per cent of women below the male mean (Lassek & Gaulin in preparation). This renders sex itself a powerful cue in formidability detection, and underscores why in human sociality males tend to monopolize the use of force. For this reason, individual differences in female formidability might not be as urgent to discriminate. Although both men and women can judge female strength from the body and face, as expected, they perform substantially less well than they do for men. Because women's upper bodies were (unlike men's) clothed, the decrement in assessing female strength from the body alone could be an artefact; sexual dimorphism in fat deposition may also obscure critical cues. However, these problems are absent from the studies exploring strength judgements from the face. Based on the face alone, the accuracy of people's strength judgements was, on average, twice as high for male than female faces, suggesting that superior assessment of male formidability is a genuine characteristic of the system.

Although not designed to test questions of ontogeny, these studies supply some limited insights. If formidability assessment were derived from a developmental history of actual conflicts, one might expect men to be better judges of male strength than women are, given that males engage in far more rough-and-tumble play with each other during development (Boulton & Smith 1992). Yet men and women both were accurate judges of men's strength and fighting ability. Analogously, many anthropologists might expect that humans would learn to exploit culturally specific cues through exposure. However, our subjects were just as good at judging strength from the faces of men of other cultures as from their own. That is, thousands of times more experience with members of one's local culture had no effect on the accuracy of the system.

These results suggest an alternative explanation for research on female preferences for 'masculine' or 'dominant' faces. A number of researchers have used real or composite faces to extract ratings of 'masculinity', 'dominance' or, more rarely, 'maturity' from male faces. There appears to be cross-cultural consistency in which faces are rated as masculine (Keating *et al.* 1981; McArthur & Berry 1987) and the rated masculinity of male faces has been shown to correlate positively with ratings of attractiveness by some women (Zuckerman *et al.* 1995; Thornhill & Gangestad 1999; Penton-Voak & Perrett 2000; Johnston *et al.* 2001), though this general effect has not been entirely reliable (Perrett *et al.* 1998; Swaddle & Reierson 2002; DeBruine *et al.* 2006). Some studies show that ratings of dominance or masculinity correlate with testosterone levels (Penton-Voak & Chen 2004; Roney *et al.* 2006) and with handgrip strength (Fink *et al.* 2007).

These preferences have been hypothesized to result from mate-selection mechanisms that are designed to detect cues of circulating testosterone and thus a genetically high-quality immune system (Fink & Penton-Voak 2002; Penton-Voak & Chen 2004). However, an alternative interpretation would be that the features in the face that are perceived as masculine or dominant are cues of physical strength and hence formidability, characteristics that are inherently desirable for women to have in a mate. Formidability in males should be an important part of mate selection for women, with substantial direct benefits (e.g. Ellis 1992; Sell 2006; Fink *et al.* 2007; Frederick & Haselton 2007). If so, this could explain why women were as accurate as men at rating men's strength.

The formidability preference hypothesis and the testosterone preference hypothesis overlap substantially. In humans, as in similar primate species, sexual dimorphism in strength plausibly reflects an evolutionary history of male–male competition. Accordingly, a substantial subset of the long-term developmental effects that testosterone has on the body can be theoretically understood as sexually selected design for aggression. This is consistent with the fact that, in humans, male muscularity is directly related to testosterone levels and develops during puberty as testosterone levels rise (Griggs *et al.* 1989). Moreover, testosterone and aggression have been linked empirically (Archer 1998, 2006; Mazur & Booth 1998), as have been aggression and strength (Sell 2006; Archer & Thanzami 2007; Gallup *et al* 2007). In consequence, cumulative long-term testosterone levels (and their observable effects) will be associated with strength. It is known that testosterone affects facial morphology, specifically thickening the brow ridge, squaring the jaw and increasing the face's width relative to its length (Thornhill & Gangestad 1999; Verdonck *et al.* 1999; Schaefer *et al.* 2005; Carre & McCormick 2008); indeed, the brow ridge and jaw are the areas of the face used to distinguish male from female skulls (Buikstra & Ubelaker 1994). These effects of testosterone on the face should covary with the effects testosterone has on the body, including increased physical strength (Bhasin *et al.* 1996, 2001). Hence, it is unclear whether the adaptive benefits of preferring masculinity are strength, immune competence or both—or, indeed, whether masculinity perception is strength perception rather than testosterone detection.

In sum, both theoretical analyses and evidence from other species suggest that selection strongly favours the evolution of cognitive specializations engineered to accurately assess fighting ability. Using what can be considered a gold standard for measuring strength—lifting

strength as measured on standardized weight-lifting machines—these studies provide the first direct evidence that both men and women can accurately assess adult men's physical strength, that these assessments perfectly track assessments of men's fighting ability, and that the cues employed are not solely by-products of size but instead track visual correlates of upper body strength, such as muscularity. The overall pattern of results supports the hypothesis that the human cognitive architecture contains specializations for formidability assessment.

REFERENCES

Archer, J. 1988 *The behavioural biology of aggression.* Cambridge: Cambridge University Press.

Archer, J. 1998 The association between testosterone and aggression among young men: empirical findings and a meta-analysis. *Aggr. Behav.* **24**, 411–420. (doi:10.1002/(SICI)1098-2337(1998)24:6<411::AID-AB2>3.0.CO;2-9)

Archer, J. 2004 Sex differences in aggression in real-world settings: a meta-analytic review. *Rev. Gen. Psychol.* **4**, 291–322. (doi:10.1037/1089-2680.8.4.291)

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

Baron-Cohen, S. 1995 *Mindblindness: an essay on autism and theory of mind.* Cambridge, MA: MIT Press.

Bennett Ross, J. 1984 Effects of contact on revenge hostilities among the Achuara Jivaro. In *Warfare culture, and environment* (ed. R. B. Ferguson), pp. 83–124. Orlando, FL: Academic Press.

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

Bhasin, S., Woodhouse, L. & Storer, T. 2001 Proof of the effect of testosterone on skeletal muscle. *J. Endocrinol.* **170**, 27–38. (doi:10.1677/joe.0.1700027)

Boulton, M. J. & Smith, P. K. 1992 The social nature of play fighting and play chasing. In *The adapted mind* (eds J. Barkow, L. Cosmides & J. Tooby). New York: Oxford University Press.

Bruce, V. & Young, A. 1986 Understanding face recognition. *Br. J. Psychol.* **77**, 305–327.

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)

Buikstra, J. & Ubelaker, D. 1994 *Standards for data collection from human skeletal remains: proceedings of a seminar at the field museum of natural history.* Arkansas Archeological Survey Research Series No. 44, Fayetteville, AR

Campbell, A. 1999 Staying alive: evolution, culture, and women's intrasexual aggression. *Behav. Brain Sci.* **22**, 203–214. (doi:10.1017/S0140525X99001818)

Carre, J. & McCormick, C. 2008 In your face: facial metrics predict aggressive behavior in the laboratory and in varsity and professional hockey players. *Proc. R. Soc. B* **275**, 2651–2656. (doi:10.1098/rspb.2008.0873)

Daly, M. & Wilson, M. 1983 *Sex, evolution, and behavior*, 2nd edn. Belmont, CA: Wadsworth Publishing Company.

Daly, M. & Wilson, M. 1988 *Homicide.* New York: de Gruyter.

DeBruine, L. *et al.* 2006 Correlated preferences for facial masculinity and ideal or actual partner's masculinity. *Proc. R. Soc. B* **273**, 1355–1360. (doi:10.1098/rspb.2005.3445)

Duclos, S., Laird, J., Schneider, E., Sexter, M., Stern, L. & Van Lighten, O. 1989 Emotion-specific effects of facial expressions and postures on emotional experience. *J. Pers. Soc. Psychol.* **57**, 100–108. (doi:10.1037/0022-3514.57.1.100)

Ekman, P. *et al.* 1987 Universals and cultural differences in the judgments of facial expressions of emotion. *J. Pers. Soc. Psychol.* **53**, 712–717. (doi:10.1037/0022-3514.53.4.712)

Ellis, B. 1992 The evolution of sexual attraction: evaluative mechanisms in women. In *The adapted mind* (eds J. Barkow, L. Cosmides & J. Tooby), pp. 267–288. New York: Oxford University Press.

Fink, B. & Penton-Voak, I. 2002 Evolutionary psychology of facial attractiveness. *Curr. Dir. Psychol. Sci.* **11**, 154–158. (doi:10.1002/ajhb.20583)

Fink, B., Neave, N. & Seydel, H. 2007 Male facial appearance signals physical strength to women. *Am. J. Hum. Biol.* **19**, 82–87. (doi:10.1177/0146167207303022)

Frederick, D. & Haselton, M. 2007 Why is muscularity sexy? Tests of the fitness indicator hypothesis. *Pers. Soc. Psychol. Bull.* **33**, 1167–1183. (doi:10.1111/1467-8721.00190)

Gallup, A., White, D. & Gallup, 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)

Griggs, R. C., Kingston, W., Josefowics, R. F., Herr, B. E., Forbes, G. & Halliday, D. 1989 Effect of testosterone on muscle mass and muscle protein synthesis. *J. Appl. Physiol.* **66**, 498–503.

Huntingford, F. A. & Turner, A. K. 1987 *Animal conflict.* New York: Chapman & Hall.

Johnston, V., Hagel, R., Franklin, M., Fink, B. & Grammer, K. 2001 Male facial attractiveness: evidence for hormone-mediated adaptive design. *Evol. Hum. Behav.* **22**, 251–267. (doi:10.1016/S1090-5138(01)00066-6)

Keating, C., Mazur, A. & Segall, M. 1981 A cross-cultural exploration of physiognomic traits of dominance and happiness. *Ethol. Sociobiol.* **2**, 41–48. (doi:10.1016/0162-3095(81)90021-2)

Keeley, L. H. 1996 *War before civilization.* Oxford: Oxford University Press.

Keltner, D. & Buswell, B. N. 1997 Embarrassment: its distinct form and appeasement functions. *Psychol. Bull.* **122**, 250–270. (doi:10.1037/0033-2909.122.3.250)

Krebs, J. & Davies, N. 1993 *An introduction to behavioural ecology*, 3rd edn. Oxford: Blackwell Scientific Publications.

Lassek, W. & Gaulin, S. In preparation. *Costs and benefits of fat-free muscle mass in men: relationship to mating success, dietary requirements, and natural immunity.*

Luckenbill, D. 1977 Criminal homicide as a situated transaction. *Soc. Probl.* **25**, 176–186. (doi:10.1525/sp.1977.25.2.03a00050)

Manson, J. & Wrangham, R. 1991 Intergroup aggression in chimpanzees and humans. *Curr. Anthropol.* **32**, 369–390. (doi:10.1086/203974)

Mazur, A. & Booth, A. 1998 Testosterone and dominance in men. *Behav. Brain Sci.* **21**, 353–397.

McArthur, L. Z. & Berry 1987 Cross-cultural agreement in perceptions of babyfaced adults. *J. Cross Cult. Psychol.* **18**, 165–192. (doi:10.1177/0022002187018002003)

Parker, G. A. 1974 Assessment strategy and the evolution of fighting behaviour. *J. Theor. Biol.* **47**, 223–243. (doi:10.1016/0022-5193(74)90111-8)

Penton-Voak, I. & Chen, J. 2004 High salivary testosterone is linked to masculine male facial appearance in humans. *Evol. Hum. Behav.* **25**, 229–241. (doi:10.1016/S1090-5138(99)00033-1)

Penton-Voak, I. & Perrett, D. 2000 Female preference for male faces changes cyclically: further evidence. *Evol. Hum. Behav.* **21**, 39–48. (doi:10.1016/j.evolhumbehav.2004.04.003)

Perrett, D. I., Lee, K., Penton-Voak, I., Rowland, D., Yoshikawa, S., Burt, D. M., Henzi, S. P., Castles, D. L. & Akamatsu, S. 1998 Effects of sexual dimorphism on facial attractiveness. *Nature* **394**, 884–887. (doi:10.1038/29772)

Raffino, R. 1972 Las sociedades agrícolas del período tardío en la Quebrada del Toro y aledaños. *Revista del Museo de La Plata, Nueva Serie* **45**, 157–210.

Roney, J., Hanson, K., Durante, K. & Maestripieri, D. 2006 Reading men's faces: women's mate attractiveness judgments track men's testosterone and interest in infants. *Proc. R. Soc. B* **273**, 2169–2175. (doi:10.1098/rspb.2006.3569)

Schaefer, K., Fink, B., Mitteroecker, P., Neave, N. & Bookstein, F. 2005 Visualizing facial shape regression upon 2nd to 4th digit ratio and testosterone. *Coll. Antropol.* **29**, 415–419.

Scherer, K., Banse, R. & Wallbott, H. 2001 Emotion inferences from vocal expression correlate across languages and cultures. *J. Cross Cult. Psychol.* **32**, 76–92. (doi:10.1177/0022022101032001009)

Sell, A. 2006 Regulating welfare tradeoff ratios: three tests of an evolutionary-computational model of human anger. *Diss. Abs. Int. B* **66**, 4516.

Smuts, B., Cheney, D., Seyfarth, R., Struhsaker, T. & Wrangham, R. 1987 *Primate societies*. Chicago, IL: University of Chicago Press.

Sugiyama, L. 2005 Physical attractiveness in adaptationist perspective. In *The handbook of evolutionary psychology* (ed. D. M. Buss), pp. 292–343. Hoboken, NJ: Wiley.

Swaddle, J. & Reierson, G. 2002 Testosterone increases perceived dominance but not attractiveness in human males. *Proc. R. Soc. B* **269**, 2285–2289. (doi:10.1098/rspb.2002.2165)

Symons, D. 1978 *Play and aggression: a study of rhesus monkeys*. New York: Columbia University Press.

Tarragó, M. N. 2000 *Los pueblos originarios y la conquista, Nueva Historia Argentina*. Barcelona: Sudamericana.

Thornhill, R. & Gangestad, S. 1999 Facial attractiveness. *Trends Cogn. Sci.* **3**, 452–460. (doi:10.1016/S1364-6613(99)01403-5)

Verdonck, A., Gaethofs, M., Carels, C. & de Zegher, F. 1999 Effect of low-dose testosterone treatment on craniofacial growth in boys with delayed puberty. *Eur. J. Orthod.* **21**, 137–143. (doi:10.1093/ejo/21.2.137)

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. (doi:10.1016/j.evolhumbehav.2008.05.001)

Zuckerman, M., Miyake, K. & Elkin, C. S. 1995 Effects of attractiveness and maturity of face and voice on interpersonal impressions. *J. Res. Pers.* **29**, 253–272. (doi:10.1006/jrpe.1995.1015)