Evolution and Human Behavior xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Evolution and Human Behavior

journal homepage: www.ehbonline.org

Formidability assessment mechanisms: Examining their speed and automaticity

Patrick K. Durkee *,[1], Aaron T. Goetz *,[1], Aaron W. Lukaszewski

Department of Psychology, California State University, Fullerton, United States

ARTICLE INFO

Article history:
Initial receipt 26 September 2017
Final revision received 20 December 2017
Available online xxxx

Keywords:
Formidability
Assessment mechanisms
Automatic
Evolution

ABSTRACT

Throughout vertebrate evolution, asymmetries in the ability to inflict costs on others (i.e., formidability) have determined the outcomes of contests over limited resources. Therefore, natural selection would have favored mechanisms designed to efficiently and accurately estimate the formidability of conspecifics. Although previous research has provided evidence for the existence of adaptations for formidability assessment, the design features of these mechanisms have not been fully examined. In the current study, participants underwent a battery of tasks to test hypotheses regarding the speed and automaticity of formidability assessment mechanisms. Results suggest that formidability is automatically and rapidly tracked and assessed from visual cues. With a few interesting exceptions, characteristics of the raters (N = 187) and targets (N = 64) did not influence these assessments. Additionally, we present eye–tracking data to highlight the salience of upper–body musculature as a cue to physical strength. Taken together, these findings bolster and extend evidence for formidability assessment mechanisms in humans.

1. Introduction

All animals must collect and decipher a bewildering array of information to survive and reproduce. Accordingly, natural selection has shaped sensory organs and brains to extract and analyze fitness–relevant information from the environment. Over the course of human evolution, some of the most pertinent information likely concerned asymmetries in the abilities and intentions of conspecifics, as many of these differences would have directly affected differential access to resources. One such asymmetry relates to individual differences in formidability—defined as the ability to impose physical costs on others (Sell, Hone, & Pound, 2012).

Physical formidability varies across individuals and is a major source of asymmetry in physical conflict (Maynard Smith & Parker, 1976). While physical competition brings the potential for large fitness gains if victorious (e.g., increased access to resources), there are also high potential costs to losing (e.g., resource loss, injury, death). Formidability is a key factor in deciding the outcome of physical combat; all else equal, less formidable individuals are more likely to suffer the costs while more formidable individuals are more likely to reap the benefits (Huntingford & Turner, 1987). Individuals that could efficiently and accurately assess the relative formidability of potential combatants would be better able to estimate the outcome of altercations, and thereby better equipped to make fitness–enhancing decisions about whether to withdraw from or escalate contests (Parker, 1974). Consequently, adaptations for assessing species–specific cues of formidability (e.g. body size, claw size) and guiding conflict decisions are common across taxa (see Arnott & Elwood, 2009).

To the degree that physical contests have regulated access to resources over human evolutionary history, these considerations suggest that humans should possess adaptations designed for estimating the formidability of conspecifics. This paper extends existing evidence supporting the hypothesis that humans possess psychological mechanisms designed to track and assess the formidability of others.

1.1. A violent past and the importance of formidability

Sell et al. (2012) previously detailed the ancestral importance of formidability, and several papers have summarized the evidence of violence in the evolutionary past (e.g., Buss & Shackelford, 1997; Goetz, 2010). We echo these arguments here, as they are crucial to understanding why the human mind might possess mechanisms designed to assess formidability.

Selection pressures related to formidability assessment are widespread across the animal kingdom, where violence and aggressive conflict has likely been a nearly universal occurrence throughout much of vertebrate evolution. Further, the archeological, anthropological, and historical record provide compelling evidence that violence and aggression were a common threat to ancestral humans, especially young men (Daly & Wilson, 1988; Keeley, 1996; Puts, 2010; Walker, 2001). Circumventing the costs of physical conflict would have been essential to the survival and reproductive potential of ancestral humans.

* Corresponding authors at: Department of Psychology, California State University, Fullerton, P.O. Box 6846, Fullerton, CA 92834-6846, United States.
E-mail addresses: pdurkee@utexas.edu (P.K. Durkee), agoetz@fullerton.edu (A.T. Goetz).

[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.evolhumbehav.2017.12.006
1090-5138/

The fingerprints of combat–related selection pressures are evident in the differences between men's and women's bodies. Due to the demands of physical competition—such as the need to withstand blunt force trauma—men's bodies (more so than women's) have been molded by natural selection to withstand the demands of aggressive conflict. For example, men have roughly 90% more upper–body strength, 11% higher muscle–to–fat ratio, nearly double the skeletal density by the end of puberty, and tend to recover faster from concussions (Covassin, Elbin, Harris, Parker, & Kontos, 2012; Kirchengast, 2010; Lassek & Gaulin, 2009; Miller, MacDougall, Tarnopolsky, & Sale, 1993; Wells, 2007; see also Sell et al., 2012). There is, however, considerably more individual variation in formidability among men than among women (Lindle et al., 1997), and thus, a statistically higher chance that there will be an asymmetry between two men drawn at random than between two women. This, coupled with higher rates of physical aggression in men (Archer, 2004), emphasizes the ancestral importance of tracking men's formidability to monitor relevant asymmetries and avoid potentially damaging conflicts.

The prominence of men's formidability in determining contest outcomes has a variety of additional downstream consequences for social decision–making. Alliance dynamics are crucial determinants of conflict outcomes among humans (Pietraszewski, 2016), and formidable individuals would have been differentially valuable as social partners, by virtue of their ability to cost–effectively elicit support and deference from various social actors with whom one might have conflict. Indeed, formidable men are preferred by women as mates (Fink, Neave, & Seydel, 2007; Frederick & Haselton, 2007; Snyder et al., 2011), and members of both sexes as cooperative partners (Eisenbruch, Grillot, Maestripieri, & Roney, 2016; von Rueden, Gurven, & Kaplan, 2008) and leaders (Lukaszewski, Simmons, Anderson, & Roney, 2016).

The formidability of others also contains predictive information about their likely behavioral characteristics. Because formidability influences men's social value in the eyes of others, it alters the cost–benefit ratios along various behavioral dimensions; for example, strong men are better positioned to capture the benefits, and avoid the costs, of social strategies involving assertive bargaining tactics, status competition, gregarious networking, and coalitional aggression (Lukaszewski & Roney, 2011). Consistent with this, men's physical strength has been found to associate positively with their levels of entitlement, anger proneness, aggression, extraversion, and prosocial leadership orientation (Archer & Thanzami, 2007; Archer & Thanzami, 2009; Durkee & Goetz, 2016; Lukaszewski, 2013; Lukaszewski & Roney, 2011; Petersen, Sznycer, Sell, Cosmides, & Tooby, 2013; Sell, Tooby, & Cosmides, 2009; Tooby & Cosmides, 1990a, 1990b; von Rueden, Lukaszewski, & Gurven, 2015).

These considerations suggest that, among human ancestors, relative formidability would have served as a proxy to estimate a man's (i) ability to win a given physical contest, (ii) value as a partner across a variety of social domains, and (iii) adoption of various behavioral strategies. Information about formidability would therefore have been relevant in solving a multitude of adaptive problems. Consequently, selection should have shaped formidability assessment mechanisms to detect and collect information that can be used to track and accurately assess the formidability of male conspecifics upon encounter.

1.2. Existing evidence for assessments of formidability

The first evidence that people can accurately assess formidability from visual cues was provided by Sell, Cosmides, et al. (2009). They showed that ratings of strength accurately tracked the actual strength of targets, regardless of whether the rater was shown only a target's face, only a target's body, or both. Accurate assessments of formidability can be made from auditory cues as well (Sell et al., 2010). These studies further demonstrated that the ability to accurately assess formidability held across stimuli from a range of populations (i.e., United States and Romanian college students, Bolivian horticulturists, and Andean pastoralists).

More evidence for visual assessments of formidability comes from two separate meta–analyses establishing that people track facial cues (i.e., facial width–to–height ratio) associated with aggressiveness (Geniole, Denson, Dixson, Carré, & McCormick, 2015; Haselhuhn, Ormiston, & Wong, 2015). These same cues have been shown to be predictive of actual fighting ability (Little, Trebicky, Havlicek, Roberts, & Kleisner, 2015; Třebický et al., 2015; Zilioli et al., 2015). Together, these studies provide evidence that people can and do make accurate judgments of formidability from visual cues.

Self–reports provide additional evidence of formidability assessment. Men engage in assessments of formidability more than women (Fox, 1997 as cited in Buss, 2015) and both sexes report consciously thinking about the strength and toughness of men more often than women (Durkee, Romero, Park, & Goetz, 2015). Men also report a desire to become more muscular and commonly cite intimidation and self–protection as a motivating factor (Frederick et al., 2007). These findings suggest that men, and to some degree women, assess formidability in everyday life.

1.3. The current study

Although there exists a cogent theoretical rationale and empirical support for the existence of evolved formidability assessment mechanisms, little is known about their specific design–features. Given that formidability was relevant across social domains, assessment mechanisms should be designed to automatically and rapidly track and assess others' formidability to calculate the costs and benefits of potential behavioral outputs at any time and across different contexts. No extant research has directly examined this theoretical possibility. To address this gap, we expand on previous research and examine the speed and automaticity of formidability assessment mechanisms.

1.3.1. Hypotheses

Due to the ancestral importance of formidability across social domains, we hypothesize that (*H1*) formidability assessment mechanisms should be designed to automatically assess formidability any time the proper cues (i.e., male conspecifics) are present—assessments of formidability should be spontaneous. For the purposes of this study, we predict that participants ratings of strength will track targets' actual strength in a surprise recall task where they are not explicitly asked to make this assessment before seeing the target individual.

If people assess the formidability of individuals, then it follows that they should be able to estimate the relative formidability of groups as well. Given the importance of formidability in coalitional conflict (Tooby & Cosmides, 1988), we hypothesize that (*H2*) the design features of the formidability assessment mechanism should automatically form estimates of group formidability. Specifically, we predict that in another surprise recall task groups with greater combined formidability—more strong/muscular men—will be endorsed more often as the stronger group than the groups with lesser combined formidability.

Many decisions that use others' formidability as an input, such as whether to escalate conflict or defer, must be made quickly. Because of this, we hypothesize that (*H3*) assessments of formidability happen rapidly upon exposure to male conspecifics. We predict that even with brief exposure to a target (i.e., 500 ms, 300 ms, 100 ms, 33 ms) participants' ratings of strength will track target strength. However, in the absence of strong theory to predict the minimum amount of exposure needed (aside from limitations of the visual system), we have no strong expectations about the minimum exposure needed.

Considering that information regarding the formidability of others would have been crucial to attend to, we hypothesize that (*H4*) assessment mechanisms have been designed to automatically detect and attend to formidable individuals. As a result, formidability should be attentionally salient. For the purposes of this study, we predict that when two men are presented on opposing sides of the visual field,

Please cite this article as: Durkee, P.K., et al., Formidability assessment mechanisms: Examining their speed and automaticity, *Evolution and Human Behavior* (2017), https://doi.org/10.1016/j.evolhumbehav.2017.12.006

participants should be more likely to attend to a more formidable man than a less formidable man.

Upper–body strength was likely the most relevant and reliable cue to ancestral fighting ability and assessment mechanisms should use relevant and reliable cues as input (Sell, Cosmides, et al. 2009; Sell et al., 2010). Given this, we hypothesize that (*H5*) assessors use upper–body musculature as a cue to formidability when making assessments. Specifically, we predict that participants will spend the most time looking at the chest and shoulders of targets, followed by the face, abdomen, and legs, respectively.

Finally, we explore characteristics of assessors and targets that influence assessments of formidability. Due to the stronger formidability–related selection pressures faced by ancestral men than women, men may be more likely to make assessments of formidability than women. By similar reasoning, men's assessments may be more accurate than women's assessments; however, previous research did not find differences in accuracy between men and women (Sell, Cosmides, et al., 2009). In the current study, we further examine the influence of sex on formidability estimation. We also examine additional characteristics of assessors that might influence assessments of formidability. Some research suggests that stronger men might underestimate the formidability of others (e.g., Fessler, Holbrook, & Gervais, 2014), but it is also plausible that the mechanism is designed to be impervious to such potentially confounding influences. In the current study, we further investigate whether and how proxies of assessor–formidability influence formidability assessments.

2. Materials and method

2.1. Participants

In total, 187 students (91 women) from a state university in Southern California participated in the study. We recruited at least 40 raters (20 of each sex) in each of the four stimuli–presentation conditions (see *H3*). The mean age of participants was 19.9 ($SD = 3.70$). A majority of participants identified as Hispanic (44%), followed by Asian/Pacific Islander (23%), Caucasian/White (23%), Other/Mixed (6%), and African American/Black (4%). Compensation was provided as partial fulfillment of course credit. We further incentivized participants throughout the experiment with tickets to win a $20 gift card to the campus store, ostensibly for their attention and the accuracy of their ratings (in reality, all participants were given equal chance to win). This study was approved by our university's Institutional Review Board.

2.2. Materials

2.2.1. PsychoPy

We created the experiment in PsychoPy, a free, open–source, experiment–building program based on the Python coding language (Peirce, 2007). Participants completed the experiment on a lab computer running Windows 7 (screen resolution: 1440 × 900).

2.2.2. Stimuli

The stimuli for strength assessments consisted of 64 photos of male students at Oklahoma State University (hereafter referred to as the OK Photoset) taken as part of an ongoing project (Lukaszewski, 2016). This photo set contains full body images of men in black gym shorts with no shirt. Each photo (hereafter referred to as targets) is associated with a standardized composite grip– and chest–strength measurement. Computer generated images of men's bodies from the UCLA Body Matrices II (Frederick & Peplau, 2007; Gray & Frederick, 2012) were used to test the group assessment hypothesis (*H2*) and the attentional adhesion hypothesis (*H4*); the specific images used for each task are detailed in the Procedures section.

2.2.3. Eye tracker

Gazepoint GP3 Eye Tracker and Remote Calibration software was used to track, record, and analyze participants' eye gaze. The eye tracker was mounted below the computer's monitor. Eye–tracking data collection took place on the same computer as the rest of the study.

2.2.4. Self–perceived formidability scale

The first author developed the Self–Perceived Formidability Scale (SPFS) as a measure of participants' self–perceived formidability. The psychometric properties of the SPFS were initially examined using Item Response Theory in a community sample ($N = 227$). The original pool of 21 items was reduced to nine items—each with four response categories (1 = strongly disagree; 4 = strongly agree)—based on scale modification criteria (e.g., clarity, redundancy, unidimensionality, local dependency). Overall, the final nine–item scale exhibited acceptable internal validity as evidenced by an exploratory factor analysis and good fit indices under the Nominal Response Model ($RMSEA = 0.04$, $NNFI = 0.97$). As evidence of external validity, theta scores for the SPFS were moderately correlated with Sell, Tooby, and Cosmides (2009) anger–proneness measure ($r_{\text{men}} = 0.32$, $r_{\text{women}} = 0.14$, p's < 0.03). In the current study, we present raw scores rather than theta scores for simplicity. We provide the psychometric properties of the SPFS as assessed by internal consistency and total reliability in Table 1.

2.2.5. Hand dynamometer

Participants' left– and right–hand grip strength were measured using a dynamometer (Camry digital hand dynamometer); an average of the left and right grip strength (in pounds) was computed for use in analyses ($M_{men} = 91.07$, $SD_{men} = 22.43$; $M_{women} = 60.80$, $SD_{women} = 11.72$). As shown in Fig. 1, participants' self–perceived formidability was moderately correlated with averaged grip strength for men, $r(95) = 0.42$, 95% CI [0.23, 0.57], $p = 3.4^{-5}$, and women, $r(91) = 0.58$, 95% CI [0.42, 0.70], $p = 0.00001$.

2.3. Procedures

Fig. 2 provides a visual summary of the procedures. Participants were asked to review and sign a consent form, which informed them that the study was about perceptions of others and that they would be asked questions about people while their eye–movements were recorded by an eye tracker. After obtaining consent, the research assistant opened the experiment in PsychoPy, which guided the participants through the majority of the study. Participants were not given any additional information about the tasks before starting; they were simply reminded that their accuracy and attention would be rewarded with chances to win an opportunity drawing and were instructed to follow the on–screen instructions to proceed. Subsequently, participants completed several different tasks as part of the study: each is explained below along with instructions for the participant.

Table 1
Self–Perceived Formidability Scale (SPFS) items and inter–item correlations.

Items	Inter–item correlations
1. I am a physically strong person.	0.76
2. I am physically stronger than most people my age and sex.	0.77
3. If I had to, I could beat most people my age and sex in a fight.	0.76
4. I am weaker than the majority of people my age and sex. [*R*]	0.71
5. I could win in hand–to–hand combat against most people my age and sex.	0.75
6. Most people my age and sex could take me in a fight. [*R*]	0.58
7. I could probably hold my own in a fight.	0.68
8. I am physically tough.	0.80
9. If I had to, I could probably use my strength and toughness to get my way.	0.59

Note. Inter–item correlations are corrected for item overlap and scale reliability (*Cronbach's alpha* = 0.90). Item 4 and Item 6 are reverse coded [*R*].

Please cite this article as: Durkee, P.K., et al., Formidability assessment mechanisms: Examining their speed and automaticity, *Evolution and Human Behavior* (2017), https://doi.org/10.1016/j.evolhumbehav.2017.12.006

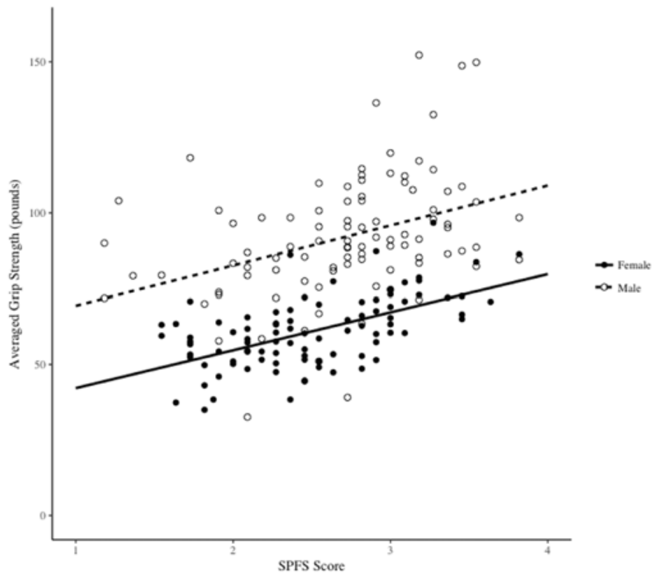

Fig. 1. Scatterplot depicting the relationship between self–perceived formidability and averaged grip strength for male and female participants.

2.3.1. Spontaneous assessment task

The first hypothesis—that formidability assessments are spontaneous (*H1*)—was necessarily tested by the first task of the experiment. Once the experiment program started, introductory instructions appeared and remained on the screen until the participant chose to proceed by clicking the ‘space’ bar. Next, a fixation cross (+) appeared on the screen for 1000 milliseconds (ms) followed by a target image, randomly selected from the OK photo set, for 2000 ms. After target removal, the participants used the mouse to rate the physical strength of the target using a 7–point scale anchored by “Very weak” and “Very strong” (as per Sell, Cosmides, et al., 2009).

Any claim of *spontaneous* assessments is dependent on the participant not being explicitly prompted to assess formidability before seeing targets. Therefore, we presented participants with distractor ratings (e.g., “How smart is this person?”, “How physically attractive is this person?”) following the next two targets so that participants were not expecting to assess formidability each time. The rating order, which was constant across participants, started with a strength rating, followed by an intelligence and attractiveness rating, two subsequent strength ratings, an intelligence rating, an attractiveness rating, and a strength rating. This order appeared random, and the trials did not repeat enough times for participants to identify a pattern. Moreover, after each rating, we presented participants with a distractor task meant to prevent them from knowing when the next image would appear.

2.3.2. Distractor task

Following a fixation cross, participants saw a colored square and were instructed to indicate if it was orange or not. The participant completed 20 to 30 trials of this tedious distractor task before another target appeared on the screen—without warning—for 2000 ms. The participant was again asked to rate the target's physical strength or a distractor rating (e.g., “How smart is this person?”). This cycle (i.e., distractor tasks followed by a target and then a rating prompt) repeated nine times in succession.

2.3.3. Spontaneous group assessment

To test the hypothesis that the formidability assessment mechanism should automatically form estimates of group formidability (*H2*), we presented participants with two, 2×2 arrays of pictures from the UCLA Body Matrices II (Frederick & Peplau, 2007). One array consisted of three muscular men (Image #18) and one weaker man (Image #15), and another consisted of three weaker men and one muscular man. Each pair was presented for 2000 ms in random order. All participants saw the same groups. Importantly, this task followed after a distractor task, and the participants did not know what the goal in the task was before seeing the stimuli. After both stimuli were presented and removed, participants indicated which group consisted of stronger men (the first or second group) by clicking a corresponding button on the computer screen.

2.3.4. Speed assessment task

We tested the hypothesis that assessments of formidability happen rapidly (*H3*) by quickly presenting targets from the OK photo set followed by a rating prompt. Upon starting the task, a fixation cross appeared for 1000 ms, followed by a randomly selected target presented for either 33, 100, 350, or 500 ms. Presentation time was constant within participants. Roughly equal numbers of men and women (around 20 of each sex) were assigned to each of the four stimuli presentation–time conditions. After the target was flashed and masked with a 100 ms visual noise patch, the participant rated the target's strength on a seven–point scale (1 = “Very Weak”; 7 = “Very strong”). There were 64 trials in total; participants saw every target once.

2.3.5. Attentional–adhesiveness task

We tested the hypothesis that assessment mechanisms automatically detect and attend to formidable individuals (*H4*) using the dot–probe paradigm, a common method for testing attentional–bias hypotheses (e.g., Bryant & Harvey, 1997). Participants read instructions before starting the task. Upon starting the task, a fixation cross appeared for 1000 ms followed by a screen with both a formidable person and a less–formidable person from the OK photo set (we determined pairings by taking one target photo from the highest quartile on the physical strength composite and one from the lowest quartile) or two images from different ends of the muscularity continuum of the UCLA Body Matrices II (image #15 was always paired with image #18; image #20 was always paired with image #22). The images were presented for 1000 ms, removed, and two small white dots appeared where one of the previous images had been (counter–balanced across trials). The participant then used the left and right arrow keys to indicate which side of the screen the dot was on and their reaction time (in milliseconds) was recorded by the computer, ending the first trial. This task consisted of seven practice trials and 60 counterbalanced experimental trials.

2.3.6. Eye–tracking ratings

After completing the PsychoPy portion of the study, participants underwent eye–tracker calibration where they followed a moving dot with their eyes through nine points; participants were asked to repeat the process if accuracy was poor. Once successfully calibrated, participants saw each target again in a randomized order that was fixed across participants. Participants again rated each target's physical strength on a 7–point scale, this time by saying a number to a research assistant who

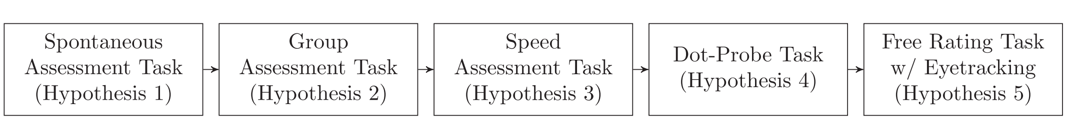

Fig. 2. Sequence of laboratory tasks completed by participants in relation to focal hypotheses.

Please cite this article as: Durkee, P.K., et al., Formidability assessment mechanisms: Examining their speed and automaticity, *Evolution and Human Behavior* (2017), https://doi.org/10.1016/j.evolhumbehav.2017.12.006

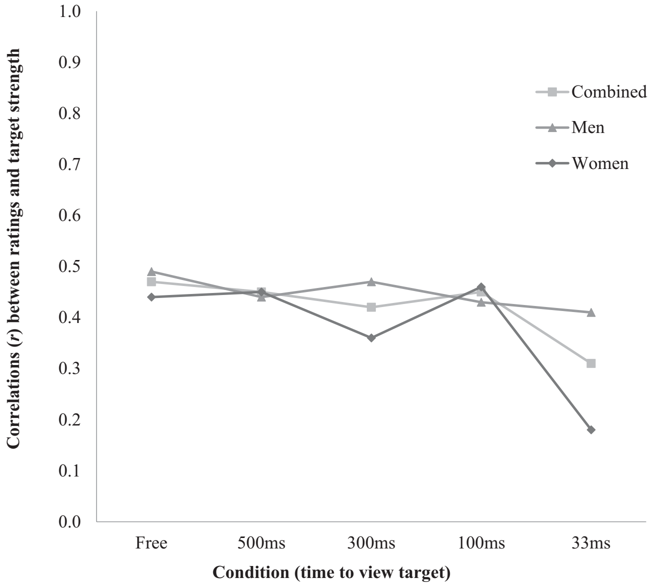

Fig. 3. The relationship between aggregated ratings of strength and actual strength across different exposure times for men's ratings only, women's ratings only, and combined ratings. All correlations are based on N = 63 and all $p < 0.02$, with the exception of women's ratings at 33 ms, where $p = 0.17$.

recorded the ratings into a spreadsheet while the participant's gaze was tracked and recorded by the eye–tracking program. This portion of the experiment tested the hypothesis that upper–body musculature is used as a cue to formidability (*H5*).

2.3.7. SPFS administration and grip strength measurement

After completing the experimental tasks above, participants answered basic demographic questions (e.g., height, weight, age, sex) and completed the Self–Perceived Formidability Scale (SPFS). Finally, participants' grip–strength was measured and recorded by a research assistant: While seated, participants were asked to hold the Camry digital hand dynamometer at their side, squeeze as hard as they could for three seconds with their dominant hand, and release; the research assistant recorded the resulting measurement and followed the same procedure with the participant's non–dominant hand. Upon completion of the study, participants were thanked for their time and asked not to share information about the tasks in the study with others.

3. Results

3.1. Data considerations

Data were screened and analyzed using R and SAS (R Core Team, 2012; Sas & Guide, 2008). Strength ratings and actual strength were standardized, as in Sell, Cosmides, et al., 2009. Unless otherwise noted, assumptions of each analysis were reasonably met. Less than 2% of the data were missing at the rater and target levels, so we used list–wise deletion for each analysis. One target was consistently a multivariate outlier and was excluded from all rating–analyses, thus, analyses at the target level (i.e., individuals from the OK Photoset) are based on $N = 63$ unless otherwise noted.

3.2. Confirmatory analyses

3.2.1. Are assessments of formidability spontaneous?

Yes. The prediction that assessments of formidability would be accurate even when raters were not primed to assess formidability (*H1*) is supported by a positive correlation between targets' composite strength and aggregated ratings ($N = 187$) of spontaneous assessments, $r(60)^2 = 0.31$, 95% CI [0.06, 0.52], $p = 0.02$. Further, assuming that the distractor task worked as intended (i.e., that the successive assessments were also spontaneous) adds additional support for the hypothesis; the aggregated subsequent ratings were positively and statistically correlated with targets' actual strength, $r(61) = 0.45$, 95% CI [0.23, 0.63], $p < 0.001$. Finally, in support of the prediction that participants could spontaneously assess group formidability (*H2*), 82% of the participants identified the more formidable group in the first trial (i.e., when prior knowledge of the task was absent), which is statistically greater than chance guessing (50%) as indicated by a one–tailed binomial test, 95% CI [0.77, 1.00], $p = 2.2^{-16}$.

Table 2
Correlations between aggregated ratings of strength by stimulus presentation time.

Condition	1	2	3	4	5	6
1. No time constraint (free ratings)	0.99					
2. 2000 ms (spontaneous ratings)	0.91	0.90				
3. 500 ms	0.97	0.92	0.97			
4. 300 ms	0.95	0.91	0.98	0.96		
5. 100 ms	0.96	0.90	0.97	0.98	0.96	
6. 33 ms	0.85	0.80	0.92	0.92	0.89	0.93

Note: All correlations are based on $N = 63$, p's < 0.001. Inter-rater reliability for each condition is listed on the diagonal.

3.2.2. Are assessments of formidability fast?

Yes. As shown in Fig. 3, the prediction that participants would be able to make accurate ratings of formidability even when stimulus presentation time is reduced (*H3*), is supported by positive statistical correlations of medium effect size between aggregated ratings and targets' actual strength at 500 ms ($r = 0.45$, 95% CI [0.23, 0.63], $p = 2.15^{-4}$), 300 ms ($r = 0.42$, 95% CI [0.20, 0.61], $p = 6.11^{-4}$), and 100 ms ($r = 0.45$, 95% CI [0.23, 0.63], $p = 2.15^{-4}$). Even when presentation time was 33 ms, aggregated strength ratings ($N = 66$; 23 female) were correlated with the targets' strength, $r = 0.31$, 95% CI [0.07, 0.52], $p = 0.01$. Interrater reliability was high for all conditions (ICC's = 0.90–0.99) and aggregated ratings at different presentation times were all highly correlated (Table 2).

3.2.2.1. Individual accuracy. To examine if ratings of strength are accurate within individuals, we used a cross–classified multilevel linear model (MLM). Cross–classified models allow ratings to be nested within both the targets and raters and the intercepts to vary independently for each. Modeling effects in this manner helps maximize power for designs with multiple random factors (see Judd, Westfall, & Kenny, 2017; Westfall, Kenny, & Judd, 2014).

Controlling for rater sex, individual ratings of strength were positively associated with target strength across the set presentation times ($\gamma = 0.24$, $p = 0.0005$) and when raters had unlimited time to view targets ($\gamma = 0.28$, $p = 0.0012$). Every standard deviation increase in target strength was associated, on average, with at least a 0.21 increase in individual ratings of that target, as would be predicted if ratings are tracking actual strength.

3.2.2.2. Differences by condition. The correlation between averaged ratings and target strength is somewhat lower at the fastest stimulus presentation (i.e., 33 ms) than when raters had no time limit to view the targets ($r = 0.47$, 95% CI [0.24, 0.65], $p < 0.001$). A dependent–samples Z test[3] showed that the difference between the two correlations—0.31

[2] One target was never shown in the spontaneous rating condition due to the randomization of stimuli across participants, so spontaneous ratings are based on $N = 62$.

[3] The dependent samples Z–statistic was computed using an online calculator (Hoerger, 2013), which uses three correlations to assess the statistical difference between two dependent samples correlations that share a common measure (i.e., strength) based on Steiger's (1980) formula. The third correlation, which restricts the level of deviation between the two correlations, is the correlation between the ratings being compared, in this case $r_3 = 0.91$ (Table 1).

Please cite this article as: Durkee, P.K., et al., Formidability assessment mechanisms: Examining their speed and automaticity, *Evolution and Human Behavior* (2017), https://doi.org/10.1016/j.evolhumbehav.2017.12.006

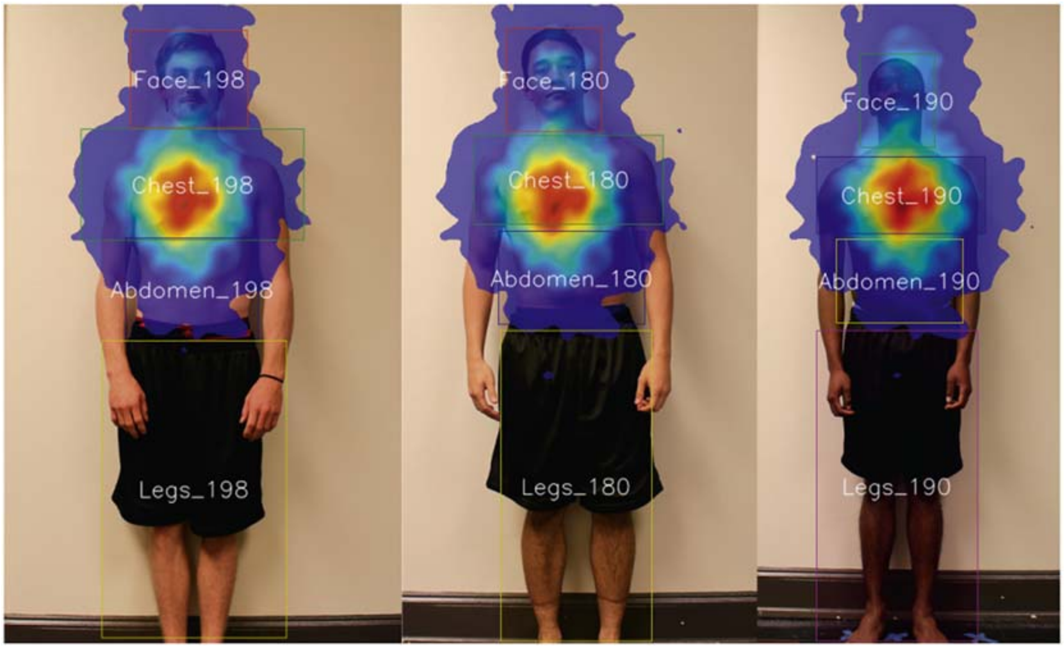

Fig. 4. Heat maps depicting aggregate time participants spent viewing different body areas for a representative subset of targets.

versus 0.47—was statistically different ($Z_H = 2.48$, $p = 0.0013$), suggesting that accuracy differs at varying stimulus presentation times for men's and women's ratings combined.

To further examine potential differences in accuracy across the set presentation conditions (i.e., the 500, 300, 100, and 33 ms conditions), we included presentation time in the MLM model as a covariate with strength and sex; we left presentation time unstandardized as milliseconds for ease of interpretation. Presentation time negatively predicted ratings ($\gamma = -0.00036$, $p = 0.04$), suggesting that every millisecond decrease in presentation time is associated with a slight overall increase in ratings. Importantly, target strength was still a robust predictor of ratings holding presentation time constant ($\gamma = 0.24$, $p = 0.0005$).

3.2.2.3. Sex differences in ratings. As shown in Fig. 3, when participants had unlimited time to view targets (i.e., in the free condition) and at exposure times of 500, 300, and 100 ms, correlations between target strength and aggregated ratings did not markedly differ by sex. With only 33 ms of exposure, women's ratings were no longer statistically correlated with target strength ($r = 0.18$, 95% CI [−0.07, 0.41], $p = 0.17$), while men's ratings were still moderately and statistically correlated with target strength ($r = 0.41$, 95% CI [0.25, 0.60], $p = 8.47^{-4}$).

We further examined the influence of sex on ratings by including the sex of raters (men = 0, women = 1) as a dummy–coded covariate in an MLM. In the free rating condition, women generally made higher ratings ($\gamma = 0.16$, $p = 0.03$); this addition to the model did not appreciably improve model fit (see ESM), suggesting that the influence of sex on the accuracy of ratings is negligible. Reproducing the model with the time–constrained ratings did not reveal a statistical influence of sex on ratings, ($\gamma = -0.02$, $p = 0.78$). In sum, there does not seem to be a meaningful difference in accuracy between male and female raters.

3.2.2.4. Is formidability attentionally adhesive?. Arguably, yes. We found partial support for our prediction that participants would pay more attention to formidable men than less formidable men in a dot–probe task (*H4*). When the dot–probe stimuli were photos from the OK photo set (i.e., real people), a paired samples t-test revealed no statistical difference in mean reaction time (in milliseconds) between the congruent conditions (i.e., probe on same side as the more formidable individual; $M = 478.21$, $SD = 66.50$) and incongruent conditions (i.e., probe opposite the more formidable individual; $M = 477.90$, $SD = 67.34$), $t(184) = 0.17$, $p = 0.87$). However, in conditions where the images from the UCLA Body Matrices II are used as the stimuli—where formidability is the only characteristic that differs—reaction times of congruent conditions ($M = 454.11$, $SD = 64.41$) were, on average, statistically faster than those in incongruent conditions ($M = 465.85$, $SD = 66.37$), $t(184) = 5.81$, $p = 2.69^{-8}$, $d = 0.42$. This effect[4] is small–to–moderate (Cohen, 1988), but suggests that, all else equal, formidability is indeed attentionally adhesive.

3.2.3. Is the upper–body a salient cue of formidability?

Yes. We found support for the hypothesis that assessors rely on upper–body musculature when assessing formidability (*H5*). A repeated–measures ANOVA with targets ($N = 64$) as subjects revealed an overall statistical difference in the number of seconds spent viewing different body parts within targets, $F_{greenhouse-geisser}(1.78, 115.86) = 771.67$, $p = 1.07^{-63}$. Post–hoc tests using Bonferroni–adjusted alpha levels revealed that, on average, participants spent statistically more time looking at the chest of targets ($M = 1.33$, $SD = 0.37$) in comparison to the face ($M = 0.44$, $SD = 0.29$), abdomen ($M = 0.35$, $SD = 0.23$), and legs ($M = 0.04$, $SD = 0.05$), p's $< 20^{-15}$, d's > 2.68. Further, as illustrated in Fig. 4, participants gave the face and abdomen roughly equal amounts of attention ($p = 0.053$) but gave statistically less attention to the legs of targets than the upper body (p's $< 6.4^{-11}$). The upper body, especially the chest, is given a disproportionate amount of attention in assessments of formidability.

3.3. Exploratory analyses

To explore individual traits of raters and targets that may influence assessments of formidability, we again employed cross–classified multilevel models treating both raters and targets as random effects. Statistically significant covariates are discussed in sections to follow. See the electronic Supplementary materials (ESM) for the full set of exploratory models.

[4] Effect size calculated using Morris and DeShon's (2002) equation eight, which corrects for dependence between means using the correlation between the two means, $r(185) = 0.91$.

Please cite this article as: Durkee, P.K., et al., Formidability assessment mechanisms: Examining their speed and automaticity, *Evolution and Human Behavior* (2017), https://doi.org/10.1016/j.evolhumbehav.2017.12.006

3.3.1. Target–level covariates of strength ratings

As in Sell, Cosmides, et al. (2009), we examined if target–level characteristics, such as height, weight, attractiveness, and age moderated ratings of strength. None of these variables were statistical contributors in the model after controlling for target strength, sex of the rater, and target presentation time (see ESM). The same was true in the model exploring target–level covariates of ratings based on free viewing (i.e., no time constraints) of the targets (see ESM). No target–level covariates influenced time spent looking at any of the targets in the eye–tracking task (see ESM). Target attractiveness (z–scored) was associated with ratings of strength in both the time–constrained and unconstrained models ($\gamma = 0.39, \gamma = 0.41$, p's < 0.001), even though attractiveness ratings are not correlated with target strength ($r = -0.04$, $p = 0.75$); this may reflect a more general halo–effect of attractiveness on rater perceptions. Target strength still predicted strength ratings even when attractiveness was included in the model, $\gamma = 0.15$–0.18, p's < 0.02.

3.3.2. Rater–level covariates of strength ratings

We also explored the influence of rater–level traits ratings of strength. We entered the rater's self–perceived formidability (SPF), averaged grip strength (GS), height, weight, age into the model with the aforementioned target–level traits, rater sex, and target strength (see ESM). Controlling for these covariates, self–perceived formidability (z–scored) was negatively related to ratings of strength in the unconstrained condition, $\gamma = -0.10$, $p = 0.02$. Each standard deviation increase in raters' self–perceived formidability was, on average, associated with a 0.10 standard deviation decrease in ratings of strength. This relationship was not found in the constrained conditions ($\gamma = -0.006$, $p = 0.88$) suggesting that rater–level traits may only influence assessments of formidability when exposure time is >500 ms. However, self–perceived formidability was negatively associated with gaze time in the eye–tracking task ($\gamma = -0.06$, $p = 0.03$), where raters with higher self–perceived formidability spent less time looking at targets while assessing formidability. No other rater–level traits under examination (GS, height, weight, age) were associated with ratings of strength in a statistically meaningful way (see ESM).

4. Discussion

These results offer support for the hypotheses that assessment mechanisms (*H1* and *H2*) spontaneously assess both individual and group formidability; (*H3*) quickly assess formidability; (*H4*) automatically attend to formidable men; and (*H5*) use cues contained in men's upper–body musculature. Our results suggest that assessments are only minimally affected by potentially confounding characteristics of the assessor (e.g., sex) and target (e.g., attractiveness). While previous research provided evidence of formidability assessment mechanisms by showing that people can accurately assess strength from visual cues, our findings extend existing evidence in several important ways.

To our knowledge, this study provides the first experimental evidence that formidability is spontaneously assessed upon encountering male conspecifics. Results of the dot–probe task further suggest that (all else equal) the mind automatically attends to more formidable individuals. Formidability appears to be an evolutionarily salient characteristic that the mind has been designed to automatically track and assess—much like sex, age, attractiveness, kinship, and group membership (Kurzban, Tooby, & Cosmides, 2001; Lieberman, Oum, & Kurzban, 2008; Maner, Gailliot, & DeWall, 2007; Stangor, Lynch, Duan, & Glas, 1992). Our results suggest that men and women are generally accurate, reflecting the significance of formidability assessment to both sexes: men would have benefited by avoiding costly conflicts and detecting potential coalitional allies, and women from identifying mates better able to acquire and maintain resources, defend offspring, and pass on genes related to formidability—especially in dangerous environments (c.f., Snyder et al., 2011).

Additionally, these results are the first to document the duration of exposure needed for accurate formidability assessment. Assessments tracked target strength even at extremely fast presentation times (i.e., 33 ms); that accuracy did not improve substantially after 100 milliseconds suggests that the necessary visual information about formidability is encoded within that time. Because information about formidability is likely a factor in the computation of welfare tradeoff ratios (Sell, Cosmides, et al., 2009) and calculating cost–benefit structure of different behavioral outputs (e.g., aggression, deference), our findings imply that formidability may begin to influence the course of social interactions almost as soon as they have begun.

Our results also highlight the potential for bias in assessment mechanisms, as assessments of target strength increased as exposure time decreased (although relative strength still tracked equally well). This overestimation is in accordance with the logic of Error Management Theory (Haselton & Buss, 2000). In the case of formidability assessment, individuals might overestimate the formidability of others under uncertainty because of the large potential fitness costs associated with losing a physical altercation. By similar logic, individuals who are relatively more formidable than average might underestimate the formidability of others under uncertain conditions due to the higher baseline odds of victory; our finding of a negative relationship between self–perceived formidability and strength ratings suggests that this may indeed be the case.

Finally, this is the first study to use eye–tracking data to examine the visual cues employed in formidability assessments. Our results show that more attention is paid to the chest and shoulders of targets, which emphasizes the salience of the upper–body musculature as a cue to formidability. This finding comports well with proposals that upper–body strength was the most ancestrally reliable predictor of formidability in human males (Sell, Cosmides, et al., 2009) and a characteristic under intrasexual selection via contest competition throughout evolutionary history (Puts, 2010; Sell et al., 2012). That raters' self–perceived formidability was negatively related to the amount of time they spent viewing each target provides further evidence of the potential for bias in formidability assessment.

4.1. Future directions

Future research could elucidate additional nuances in the design features of formidability assessment mechanisms. The developmental trajectory of formidability assessment should be studied to ascertain how these mechanisms are calibrated. Systematic examination of the modulating effects of contextual cues (e.g., current state of assessor, number of allies, past outcomes, imminence of combat) on assessments is also needed. Whether and how the outputs of rapid formidability assessments influence the course of downstream social outcomes is a fruitful avenue for future investigation as well.

4.2. Conclusion

The current study bolsters and extends formidability assessment research in important ways by identifying hypothesized design features of this crucial psychological adaptation. Our findings suggest that accurate formidability assessments are spontaneous, quick, and based on attentionally adhesive cues contained in sexually dimorphic upper–body musculature. The existence of this collection of design features attests to the ancestral importance of beginning social encounters with estimates of others' formidability in mind. Formidability assessment mechanisms likely helped our ancestors solve a raft of adaptive problems and, consequently, still play a role in human social cognition.

Data availability

All data associated with this study are available online through Open Science Framework (link: osf.io/qc6nk).

Acknowledgements

We are grateful to the Goetz Lab for their assistance with data collection and design input. We thank John Patton, Elizabeth Pillsworth, and members of CPB for their feedback throughout this project, and Jessica Tessler for statistical advice. We also thank the Buss Lab for helpful editorial suggestions and two anonymous reviewers for their valuable suggestions.

Funding

This work was supported by the California State University, Fullerton Associated Students Incorporated (ASI) Student Research Grant [#16–079].

Competing interests statement

The authors certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Appendix A. Supplementary tables

Supplementary tables to this article can be found online at https://doi.org/10.1016/j.evolhumbehav.2017.12.006.

References

Archer, J. (2004). Sex differences in aggression in real–world settings: A meta–analytic review. *Review of General Psychology*, *8*(4), 291–322.

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

Archer, J., & Thanzami, V. (2009). The relation between mate value, entitlement, physical aggression, size and strength among a sample of young Indian men. *Evolution and Human Behavior*, *30*(5), 315–321.

Arnott, G., & Elwood, R. W. (2009). Assessment of fighting ability in animal contests. *Animal Behaviour*, *77*(5), 991–1004.

Bryant, R. A., & Harvey, A. G. (1997). Attentional bias in posttraumatic stress disorder. *Journal of Traumatic Stress*, *10*(4), 635–644.

Buss, D. (2015). *Evolutionary psychology: The new science of the mind*. Psychology Press.

Buss, D. M., & Shackelford, T. K. (1997). Human aggression in evolutionary psychological perspective. *Clinical Psychology Review*, *17*(6), 605–619.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Covassin, T., Elbin, R. J., Harris, W., Parker, T., & Kontos, A. (2012). The role of age and sex in symptoms, neurocognitive performance, and postural stability in athletes after concussion. *The American Journal of Sports Medicine*, *40*(6), 1303–1312.

Daly, M., & Wilson, M. (1988). *Homicide*. Transaction Publishers.

Durkee, P., & Goetz, A. T. (2016, June). The relationship between formidability and aggression: A meta–analytic review. *Poster presented at 28th annual meeting of the Human Behavior and Evolution Society, Vancouver, Canada*.

Durkee, P., Romero, G. A., Park, J., & Goetz, A. T. (2015, April). Spontaneous assessments of formidability by men. *Poster presented at 95th annual Western Psychological Association Conference, Las Vegas, NV*.

Eisenbruch, A. B., Grillot, R. L., Maestripieri, D., & Roney, J. R. (2016). Evidence of partner choice heuristics in a one–shot bargaining game. *Evolution and Human Behavior*, *37*(6), 429–439.

Fessler, D. M., Holbrook, C., & Gervais, M. M. (2014). Men's physical strength moderates conceptualizations of prospective foes in two disparate societies. *Human Nature*, *25*(3), 393–409.

Fink, B., Neave, N., & Seydel, H. (2007). Male facial appearance signals physical strength to women. *American Journal of Human Biology*, *19*(1), 82–87.

Frederick, D. A., Buchanan, G. M., Sadeghi-Azar, L., Peplau, L. A., Haselton, M. G., Berezovskaya, A., & Lipinski, R. E. (2007). Desiring the muscular ideal: Men's body satisfaction in the United States, Ukraine, and Ghana. *Psychology of Men & Masculinity*, *8*, 103–117. https://doi.org/10.1037/1524–9220.8.2.103.

Frederick, D. A., & Haselton, M. G. (2007). *Why is muscularity sexy? Tests of the fitness indicator hypothesis*. Personality and Social Psychology Bulletin.

Frederick, D. A., & Peplau, L. A. (2007). The UCLA Body Matrices II: Computer–generated images of men and women varying in body fat and muscularity/breast size to assess body satisfaction and preferences. *Proceedings of the Eighth Annual Meeting of the Society for Personality and Social Psychology, Memphis, TN*.

Geniole, S. N., Denson, T. F., Dixson, B. J., Carré, J. M., & McCormick, C. M. (2015). Evidence from meta–analyses of the facial width–to–height ratio as an evolved cue of threat. *PLoS One*, *10*(7), e0132726.

Goetz, A. T. (2010). The evolutionary psychology of violence. *Psicothema*, *22*(1), 15–21.

Gray, P. B., & Frederick, D. A. (2012). Body image and body type preferences in St. Kitts, Caribbean: A cross–cultural comparison with US samples regarding attitudes towards muscularity, body fat, and breast size. *Evolutionary Psychology*, *10*(3).

Haselhuhn, M. P., Ormiston, M. E., & Wong, E. M. (2015). Men's facial width–to–height ratio predicts aggression: A meta–analysis. *PLoS One*, *10*(4), e0122637.

Haselton, M. G., & Buss, D. M. (2000). Error management theory: A new perspective on biases in cross–sex mind reading. *Journal of Personality and Social Psychology*, *78*(1), 81.

Hoerger, M. (2013). Z_H: An updated version of Steiger's Z and web–based calculator for testing the statistical significance of the difference between dependent correlations. Retrieved from http://www.psychmike.com/dependent_correlations.php.

Huntingford, F. A., & Turner, A. K. (1987). Animal conflict. *Animal behavior series*. Chapman & Hall Ltd.

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology*, *68*, 601–625.

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

Kirchengast, S. (2010). Gender differences in body composition from childhood to old age: An evolutionary point of view. *J. Life Sci.*, *2*(1), 1–10.

Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences*, *98*(26), 15387–15392.

Lassek, W. D., & Gaulin, S. J. (2009). Costs and benefits of fat–free muscle mass in men: Relationship to mating success, dietary requirements, and native immunity. *Evolution and Human Behavior*, *30*(5), 322–328.

Lieberman, D., Oum, R., & Kurzban, R. (2008). The family of fundamental social categories includes kinship: Evidence from the memory confusion paradigm. *European Journal of Social Psychology*, *38*(6), 998–1012.

Lindle, R. S., Metter, E. J., Lynch, N. A., Fleg, J. L., Fozard, J. L., Tobin, J., ... Hurley, B. F. (1997). Age and gender comparisons of muscle strength in 654 women and men aged 20–93 yr. *Journal of Applied Physiology*, *83*(5), 1581–1587.

Little, A., Trebicky, V., Havlicek, J., Roberts, S. C., & Kleisner, K. (2015). Human perception of fighting ability: Facial cues predict winners in Mixed Martial Arts fights. *Behavioral Ecology*, *26*(6), 1470–1475.

Lukaszewski, A. W. (2013). Testing an adaptationist theory of trait covariation: Relative bargaining power as a common calibrator of an interpersonal syndrome. *European Journal of Personality*, *27*(4), 328–345.

Lukaszewski, A. W. (2016). *Oklahoma photoset and associated data*. (Unpublished photoset and associated data).

Lukaszewski, A. W., & Roney, J. R. (2011). The origins of extraversion: Joint effects of facultative calibration and genetic polymorphism. *Personality and Social Psychology Bulletin*, *37*(3), 409–421.

Lukaszewski, A. W., Simmons, Z. L., Anderson, C., & Roney, J. R. (2016). The role of physical formidability in human social status allocation. *Journal of Personality and Social Psychology*, *110*(3), 385–406.

Maner, J. K., Gailliot, M. T., & DeWall, C. N. (2007). Adaptive attentional attunement: Evidence for mating–related perceptual bias. *Evolution and Human Behavior*, *28*(1), 28–36.

Maynard Smith, J., & Parker, G. A. (1976). The logic of asymmetric contests. *Animal Behaviour*, *24*(1), 159–175.

Miller, A. E. J., MacDougall, J. D., Tarnopolsky, M. A., & Sale, D. G. (1993). Gender differences in strength and muscle fiber characteristics. *European Journal of Applied Physiology and Occupational Physiology*, *66*(3), 254–262.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta–analysis with repeated measures and independent–groups designs. *Psychological Methods*, *7*(1), 105.

Parker, G. A. (1974). Assessment strategy and the evolution of fighting behaviour. *Journal of Theoretical Biology*, *47*(1), 223–243.

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1), 8–13.

Petersen, M. B., Sznycer, D., Sell, A., Cosmides, L., & Tooby, J. (2013). The ancestral logic of politics: Upper–body strength regulates men's assertion of self–interest over economic redistribution. *Psychological Science*, *24*(7), 1098–1103.

Pietraszewski, D. (2016). How the mind sees coalitional and group conflict: The evolutionary invariances of n–person conflict dynamics. *Evolution and Human Behavior*, *37*(6), 470–480.

Puts, D. A. (2010). Beauty and the beast: Mechanisms of sexual selection in humans. *Evolution and Human Behavior*, *31*(3), 157–175.

R Core Team (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing ISBN 3–900051–07–0, URL, http://www.R–project.org/.

Sas, S., & Guide, S. U. S. (2008). *SAS Institute Inc. Cary, NC, USA*.

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

Sell, A., Cosmides, L., Tooby, J., Sznycer, D., von Rueden, C., & Gurven, M. (2009). Human adaptations for the visual assessment of strength and fighting ability from the body and face. *Proceedings of the Royal Society of London B: Biological Sciences*, *276*(1656), 575–584.

Sell, A., Hone, L. S., & Pound, N. (2012). The importance of physical strength to human males. *Human Nature*, *23*(1), 30–44.

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

Snyder, J. K., Fessler, D. M. T., Tiokhin, L., Frederick, D. A., Lee, S. W., & Navarette, C. D. (2011). Trade–offs in a dangerous world: Women's fear of crime predicts preferences for aggressive and formidable mates. *Evolution and Human Behavior*, *32*, 127–137. https://doi.org/10.1016/j.evolhumbehav.2010.08.007.

Stangor, C., Lynch, L., Duan, C., & Glas, B. (1992). Categorization of individuals on the basis of multiple social features. *Journal of Personality and Social Psychology*, *62*(2), 207.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*, 245–251.

Tooby, J., & Cosmides, L. (1988). The evolution of war and its cognitive foundations. *Institute for evolutionary studies technical report. 88(1)*. (pp. 1–15).

Tooby, J., & Cosmides, L. (1990a). The past explains the present: Emotional adaptations and the structure of ancestral environments. *Ethology and Sociobiology*, *11*(4–5), 375–424.

Tooby, J., & Cosmides, L. (1990b). On the universality of human nature and the uniqueness of the individual: The role of genetics and adaptation. *Journal of Personality*, *58*(1), 17–67.

Třebický, V., Fialová, J., Kleisner, K., Roberts, S. C., Little, A. C., & Havlíček, J. (2015). Further evidence for links between facial width-to-height ratio and fighting success: commentary on Zilioli et al. (2014). *Aggressive Behavior*, *41*(4), 331–334.

von Rueden, C., Gurven, M., & Kaplan, H. (2008). The multiple dimensions of male social status in an Amazonian society. *Evolution and Human Behavior*, *29*(6), 402–415.

von Rueden, C. R., Lukaszewski, A. W., & Gurven, M. (2015). Adaptive personality calibration in a human society: Effects of embodied capital on prosocial traits. *Behavioral Ecology*, *26*(4), 1071–1082.

Walker, P. L. (2001). A bioarchaeological perspective on the history of violence. *Annual Review of Anthropology*, 573–596.

Wells, J. C. (2007). Sexual dimorphism of body composition. *Best Practice & Research Clinical Endocrinology & Metabolism*, *21*(3), 415–430.

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, *143*(5), 2020.

Zilioli, S., Sell, A. N., Stirrat, M., Jagore, J., Vickerman, W., & Watson, N. V. (2015). Face of a fighter: Bizygomatic width as a cue of formidability. *Aggressive Behavior*, *41*(4), 322–330.

Please cite this article as: Durkee, P.K., et al., Formidability assessment mechanisms: Examining their speed and automaticity, *Evolution and Human Behavior* (2017), https://doi.org/10.1016/j.evolhumbehav.2017.12.006