TO FIGHT OR NOT TO FIGHT:
DOES CONSPECIFIC STRENGTH INFLUENCE DEFENSIVE SIGNALING?

By

David J. Johnson

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Psychology – Master of Arts

2014

ABSTRACT

TO FIGHT OR NOT TO FIGHT:
DOES CONSPECIFIC STRENGTH INFLUENCE DEFENSIVE SIGNALING?

By

David J. Johnson

Humans and social animals show similar responses to defensive threats such as the presence of predators or rival conspecifics (Blanchard et al., 2001). The current work tested two extensions of this research: first, whether humans show similar assessment processes compared to non-human animals including dynamically updating their assessments based on new information, and second, whether humans send different signals (i.e., willingness to escalate or submission) based on differences in physical formidability and whether those signals have behavioral consequences. Using an experimental procedure where randomly paired same-sex naïve participants competed against one another in a physical task, the current experiment revealed evidence consistent with assessment; participants became more accurate in their judgments of strength after gaining information from a physical contest. In contrast, participants did not send different signals based on differences in formidability, insofar as those signals were broadcasted by changes in strength. Implications of using animal models to predict human defensive behaviors are discussed, as well as relevant connections to game theory.

TABLE OF CONTENTS

LIST OF TABLES

INTRODUCTION

How do individuals signal for interactions with physically formidable others? Research on non-human animals characterizes defensive behavior as driven by an interaction between ability and context given the presence of a threat (D. C. Blanchard, 1997). In this view, defensive behaviors are the product of an evolved computational process that takes into account contextual variables in order to prepare an animal to act in ways that will afford successful defense from a threat (D. C. Blanchard, Griebel, Pobbe, & R. J. Blanchard, 2011; Gawronski & Cesario, 2013).

In a competitive context, certain actions will be more or less appropriate given the unique discrepancy in ability between individuals. Game theory (Maynard Smith, 1974; Maynard Smith & Parker, 1976; Maynard Smith & Price, 1973; Parker, 1974) provides a useful model for predicting what behaviors will occur in such situations. According to this framework animals should be sensitive to differences in physical formidability, or resource holding potential (RHP; Parker, 1974). When discrepancies in RHP are large, more formidable opponents should signal willingness to escalate conflict and weaker opponents should signal submission or withdraw from conflict.

As a part of a growing body of work demonstrating commonalities between human and animal defensive behaviors (e.g., D. C. Blanchard, Hynd, Minke, Minemoto, & R. J. Blanchard, 2001; Sell et al., 2009) I propose that the influence a person has on another's defensive signaling is partially dependent on the discrepancy in formidability between the two. In competitive contexts, stronger individuals will signal readiness to escalate conflict while weaker individuals will signal submission. Critically, these signals will be indexed through changes in strength relative to strength measured when the individual is alone. Weaker individuals will show

1

submission by demonstrating less strength, while stronger individuals will display willingness to engage in conflict by demonstrating greater strength.

Animal Assessment and Defensive Behaviors

Theoretical accounts.

Animals in social species such as humans have historically faced recurrent threats from both predators and rival conspecifics (same-species organisms). Although there are important differences between threats from predators and conspecifics (e.g., predator-prey relationships are typically characterized by large asymmetries in formidability), one commonality across these domains is that aggression should only occur in limited circumstances because of the high costs attached to it, namely risk of injury or death (Parker, 1974). These costs are elevated when an individual is faced with a more formidable conspecific (or predator). An ability to accurately gauge individual differences in formidability would provide an advantage in determining whether to escalate or withdraw from conflicts (Gawronski & Cesario, 2013; Griskevicius et al., 2009; Sell et al., 2009). In favor of this hypothesis, the existence of reliable cues to physical formidability, including (but not limited to) size, weight, and weaponry, are well documented amongst non-human animals (cf. Arnott & Elwood, 2010).

From a game theoretical perspective (Maynard Smith, 1974; Maynard Smith & Parker, 1976; Maynard Smith & Price, 1973) every organism has a specific level of fighting ability, or resource holding potential (RHP; Parker, 1974). RHP is influenced by several factors including size, strength, weaponry, group size, and experience. Each factor serves as a partial cue to the individual's absolute fighting ability. All other things equal, an organism with greater RHP than his or her opponent has a higher probability of winning a physical fight. According to Parker (1974), when discrepancies in RHP are large, more formidable opponents should signal to

escalate and weaker opponents to withdraw. Fights should only occur where both opponents escalate; this is more likely to occur as differences in RHP approach zero.

It is important to note that all that is meant by assessment is that an organism reacts differentially based on the magnitude of a given threat (Parker, 1974). Conscious intent is not required; rather, actions prepared by the organism are those favored by selection. Historically, organisms that successfully defended themselves from interpersonal threats had better reproductive success. These behavioral tendencies were inherited by their offspring, over time developing into neural mechanisms that process information about the magnitude of threats. These mechanisms activate a set of behaviors appropriate to that unique interaction (Maynard Smith & Parker, 1976; Parker, 1974), which result in differential responding partially based on relative threat level (i.e., differences in RHP).

How long should an organism assess a threat before preparing an appropriate behavioral response? In betta fish, lateral displays of body size are given for up to several minutes until one opponent withdraws (Simpson, 1968). Similarly, red deer typically engage in roaring contests (where vocalizations accurately index fighting ability) for several minutes before decisions to engage or withdraw are made (Clutton-Brock & Albon, 1979). Although game theory predicts that extended bouts of assessment give increasingly accurate estimations of RHP, they are also more costly (Dawkins & Guilford, 1991; Maynard Smith & Price, 1973). Animals that signal for long periods of time or engage in prolonged assessment of others' signals are at higher risk of incurring injury from opponents or predators drawn to the displays. In addition, animals also lose valuable time to pursue other desirable resources or mates. Consider the case of a bird that sings to deter others from invading his or her territory. The song puts not only the signaler at risk for predation, but also increases risk for those close enough to hear the warning song.

Often there is a trade-off between assessment length and accuracy. When costs of losing a conflict are high and the costs assessment are low, prolonged assessment may be beneficial (Dawkins & Guilford, 1991; Maynard Smith, 1982). When observing a rival from a distance, the only costs of assessment would likely be the time and energy lost in pursuit of other activities. These low costs might encourage protracted assessment. In contrast, approaching a rival might increase assessment accuracy beyond what could be observed from a distance, but at larger risk of potential injury, especially the longer the animal remains close to the threat. However, these situations are likely to vary widely across species, and are dependent on assessment accuracy. In some cases (e.g., close proximity, high cost situations) even if extensive assessments are more accurate, it may be more cost effective to make quicker, but less accurate decisions about the formidability of conspecifics. The value of such assessments would increase as their accuracy approaches that of prolonged assessments.

Similarly, the degree to which assessment is accurate is a key determinant of what actions are selected in response to threats. As assessment accuracy decreases, fighting is more likely to occur between conspecifics (Maynard Smith & Parker, 1976; Parker, 1974). Recall that when there is a discrepancy in RHP the more formidable opponent should signal to escalate and the weaker one to withdraw. Only when the difference is small (or nonexistent), is it likely that both opponents will determine that they are more formidable, resulting in mutual escalation and physical conflict. However, when assessment is inaccurate, it is harder to determine disparities in RHP. This increases the margin of error for determining when escalation is appropriate, resulting in increased conflict. In contrast, when assessment is accurate, conflicts occur less often, as it is easier to gauge the probability of winning a bout (Maynard Smith & Parker, 1976; Parker, 1974).

Experimental evidence.

The purpose of assessment is to determine what kind of action will most likely to lead to a successful defense from a threat (D. C. Blanchard et al., 2011). Assessment is only useful if the actions prepared in response to the threat benefit the assessor. Therefore, one should expect different responses to threats in non-human animals based on the magnitude of the threat and the context in which the threat is encountered. Indeed, hyenas and other social species respond differently to threats based on their magnitude (Benson-Amram, Heinen, Dryer, & Holekamp, 2011). When exposed to calls from unfamiliar conspecifics, hyena behavior varied flexibly based on the ratio of intruders to allies. They demonstrated more vigilance to larger groups of outsiders, and approach behavior was contingent upon having a numerical advantage. Hyenas demonstrate a sophisticated system of threat assessment that takes into account the formidability of both allies and intruders and chooses appropriate actions accordingly.

Similarly, research on rodents also provides support for differential responding based on individual ability and context. The presence of an ambiguous threat (a threat of unclear RHP) encourages orientation towards, and investigation of, the threat. In contrast, the presence of a clear threat (e.g., a predator with much higher RHP) elicits different responses based on the physical context. At a large distance threats elicit flight if escape is possible and freezing if it is not. At closer distances rodents give defensive "threats" (e.g., displays of weaponry), and at even closer distances, they engage in defensive attack (D. C. Blanchard et al., 1997; D. C. Blanchard, 2011). When at a clear disadvantage, rodents seek to withdraw from conflict, and only resort to escalation if all other options have been exhausted.

In addition to the defensive behaviors highlighted by D. C. and R. J. Blanchard (D. C. Blanchard, 1997; D.C. Blanchard & R. J. Blanchard, 2003), submissive behaviors are also common in social animals such as primates (Bernstein & Gordon, 1980; D. C. Blanchard et al.,

2011). These signals are most likely to occur when the value gained from continuing a contest is lower than the potential costs of continuing the fight (Matsumura & Hayden, 2006). Typically, this occurs when discrepancy in RHP is slight: accepting submission is advantageous because opponents are likely to retaliate if attacks are continued. For this reason, submissive signals are only evolutionarily stable when assessment is accurate enough to distinguish subtle differences in formidability. In contrast, submissive signals are likely to be ignored when the difference between opponents is large, because the benefits of continued attack (e.g., asserting dominance) outweigh the risk of injury (Matsumura & Hayden, 2006).

Submission is directly related to dominance and the formation of hierarchies in animals. Many animals form dominance relationships, including insects, fish, birds and mammals (cf. Chase & Seitz, 2011). Although a review of dominance hierarchies is beyond the scope of this proposal, certain aspects are relevant to understanding submissive behaviors. In particular, primates have a complex system of hierarchy largely based on a rank order of submission; the most formidable primate submits to no one and the weakest to everyone. Rhesus monkeys simultaneously introduced to each other will compete for dominance but typically settle into a stable hierarchy within an hour (Bernstein & Gordon, 1980; Bernstein, Gordon, & Rose, 1974), suggesting that assessment is a relatively quick process. Interestingly, lone male rhesus monkeys introduced to an established group immediately give submissive signals and assume the lowest position in a hierarchy (Bernstein & Gordon, 1980). This may seem contradictory if the intruder is more formidable than some members of the group. However, chimpanzees are known to solicit and receive help from conspecifics during conflicts (de Waal & Hoekstra, 1980). Therefore, in any given group, the combined RHP of all or multiple members will almost always outnumber

the intruder's RHP by several magnitudes, making submission a reasonable behavior (Bernstein & Gordon, 1980; Matsumura & Hayden, 2006).

In sum, many non-human animals show an ability to accurately assess formidability in both predators and conspecifics. Longer assessments may be more accurate, but have higher costs. Selection should favor decision rules that flexibly execute accurate or quick assessments depending on contextual contingencies, such as the magnitude of the threat. All other things equal, accurate assessment limits physical conflict to opponents with similar RHP. When discrepancies in RHP are large, weaker animals should signal withdraw by fleeing or submitting, whereas stronger animals should signal willingness to fight. Critically, the neural mechanisms that orchestrate assessment should respond to contextual differences, including the formidability of opponents, the presence of allies, and the physical situation where the encounter occurs.

Human Assessment and Defensive Behaviors

Until recently defensive behaviors in animals have typically been characterized by psychologists as inflexible and innate (Gawronski & Cesario, 2013), whereas human aggression was thought of as a learned behavior unique to the species (D. C. Blanchard et al., 2001). This distinction is partially due to the constrained meaning of aggression typically applied to humans versus the more general definition used in the biological sciences. In humans, aggression is typically thought of as a learned hostile behavior with the intent to inflict pain. However, a more general definition, applicable across species, regards aggression as behavior associated with physical attack or escalation towards attack, without regards to intent (Sell, Hone, & Pound, 2012; van Staaden, Searcy, & Hanlon, 2011). According to the former definition, aggression is a uniquely human phenomenon requiring conscious intent and hostile motives. According to the latter, aggression just one of many defensive behaviors employed across species. Although not

ignoring the likely possibility that certain aspects of aggression are unique to humans, this definition embraces the possibility that similar selection pressures shaped the defensive behaviors of humans and non-human animals. From this perspective, the same computational processes of assessment that occur in non-human animals provide valuable hypotheses about the effect of context on human defensive behavior.

For the computational models of threat assessment to apply to human defensive behavior, humans must first demonstrate an ability to accurately gauge their own formidability as well as that of others. As with non-human animals, both the accuracy and speed of assessment is critical for understanding what actions humans will signal in conflict. Paralleling findings in the animal literature, humans are able to both quickly and accurately assess physical formidability. In a series of experiments, Sell and colleagues (2009, 2010) demonstrated that humans were able to assess formidability (as defined by upper body strength) from mere facial photos or voice recordings. Raters were more accurate at determining male strength than female strength, which is at least partially due to lower variability in female strength due to sexual dimorphism (Sell et al., 2012). These results have also been replicated by other researchers (Archer & Thanzami, 2009). Sell and colleagues results suggest that humans can quickly and accurately assess formidability, even from relatively impoverished information like pictures or voice recordings.

Although it may seem obvious that humans prepare different actions based on the magnitude of a threat and the context in which the threat occurs, it is important to question if the behaviors exhibited by non-human animals under threat match human behaviors under similar conditions. In an experimental study, D. C. Blanchard and colleagues (2001) asked individuals to report how they would respond in several threatening situations, which systematically varied the level of threat and ability to escape. Human threat responses closely paralleled rodent threat

responses. When threats were ambiguous (e.g., hearing a suspicious noise) the most reported behavioral choice was to investigate the threat. In contrast, when threats were clear (e.g., being attacked) defensive responses were most common. These defensive responses did vary somewhat across the sexes, such that males were more prone to report defensive attack in the context of a clear threat, while females were more likely to report defensive threat (i.e. screaming). These could reflect a higher percentage of males assessing that they were more formidable than the threat and thus a retaliatory attack would be most advantageous. Interestingly, D. C. Blanchard and colleagues (2001) noted that one limitation to this research was not including a response option for submissive or pleading behaviors, noting while that these responses are unlikely in rodents[1] they would be in the range of possibility for animals such as primates or humans.

Due to the historical focus of studying aggression as a learned phenomenon, little research has investigated the role of RHP disparities on defensive signaling. Moreover, the studies that have done so have typically relied upon hypothetical conflict situations (Archer & Benson, 2008; D. C. Blanchard et al., 2001). Although studies examining responses to hypothetical threats are advantageous in that they can manipulate threat in ways that would not be ethical to examine otherwise, they are limited by relying on prospective self-report, which may not accurately reflect actual behaviors chosen in a conflict situation. For example, males may report being less likely to scream than females when threatened due to self-presentation concerns, rather than actual differences in behavior.

A stronger test of whether RHP influences defensive signaling would be to actually set up dyadic interactions between individuals of varying RHP and observe how differences in

---

[1] The definition of submission used in this article excludes behaviors such as "playing dead" which are assumed to be qualitatively different than submission, as they may be used by animals to evade predation but also to lure opponents closer.

formidability influence these signals. Changes in upper body strength in response to a partner could serve as an index of defensive signaling. Although relatively stable, upper body strength can be reliably increased or decreased by a variety of factors, including motivation and threat (Ikai & Steinhaus, 1961). Insofar as submission is indexed by showing weakness (i.e., less strength) and escalation is broadcasted by exaggerating those cues (i.e., more strength), individuals may change their strength dependent on the unique relationship between partners, specifically the degree in which their RHP differs.

Pilot Study

The aforementioned hypotheses were pilot tested in an unpublished study conducted by Cesario and Johnson (2013). Specifically, male undergraduates were randomly paired with one of five male confederates of varying physical formidability. Confederates were blind to study hypotheses. Participants were told that they would compete against each other at a later point in time. They were instructed to stand while strength was measured, with the idea being that this might encourage natural assessment of their partners' formidability. The experimenter measured participants' upper body strength using an inverted hand dynamometer. Measurements were taken with the participant's partner turned around, so participants could not observe their partner's measured strength. However, participants could to see their own measured strength.

To initially examine men's defensive signaling as a function of confederate strength, a single factor ANOVA was first conducted with confederate as predictor of participant upper body strength. This yielded a significant effect for confederate $F(4, 116) = 2.74$, $p = .032$, $\eta^2 = .086$. Participant upper body strength was lower in the presence of a stronger confederate, and higher in the presence of a weaker confederate. Although these results are promising, they ignore the dyadic nature of the data. Because the hypotheses concerned whether one's own ability and

the confederate's ability would influence strength, the data were analyzed with a one-with-many model (OWM; Marcus, Kashy, & Baldwin, 2009).

The OWM design is a variance decomposition model that can be used with hierarchically structured data in which multiple individuals (the many; participants in this case) interact with or are tied to the same partner (the one; confederates in this case). The model separates confederate variance from participant variance. In a reciprocal OWM design both the confederates and the participants have an outcome score for each dyadic combination (i.e., strength), and the percentage of variance at the confederate and participant levels can be estimated for both. Confederate-level variance for the measure of the confederate's strength estimates the extent to which a confederate is consistently strong (or weak) across participants, and confederate-level variance for the measure of the participants' strength estimates the extent to which all participants paired with the same confederate are consistently strong or weak. The correlation between these two confederate-level effects measures generalized reciprocity or the extent to which participants paired with stronger confederates consistently show strong stronger (or weaker) behavior than do participants who are tied to weaker confederates.

Multilevel modeling with restricted maximum likelihood was used to estimate a OWM model assessing the reciprocal effects of confederate grip strength on participant grip strength and vice versa. The key finding was that the generalized reciprocity correlation was significant, substantial, and negative ($r = -.83$, $p = .008$). The presence of a strong confederate tended to attenuate participants' strength, such that the mean strength for participants paired with that confederate was lower than the mean strength across all confederates. Conversely, the presence of a weak confederate tended to enhance participants strength, such that the mean participant strength for that confederate was higher than the mean strength across all confederates.

11

The model also revealed that a considerable amount of the variance in confederate strength was at the participant or dyad level, which accounted for 38% ($p < .001$) of the variance in confederates' strength. This indicates that confederates' strength changed depending on the participant they interacted with. Combined with the significant generalized reciprocity correlation, these results replicate the findings from the single factor ANOVA while demonstrating that participants also influence confederate's strength. These findings are consistent with the hypothesis that humans assess the formidability of conspecifics and automatically signal actions that will lead to successful defense (e.g., escalation, submission).

The Present Research

Drawing on research from both human and non-human animals, as well as the pilot study described above, the current research tested whether physical formidability influences defensive signaling in humans. Consistent with prior research (Sell et al., 2010; 2009), I predicted that (H_1) humans would be able to accurately gauge the strength of people they interact with. However, because there is less variation in upper body strength between females (Sell et al., 2012), the relationship between formidability and defensive signaling would be stronger for males rather than females.

Second, assessment will alert individuals to the degree of RHP discrepancy between them and the other competitor. Interacting with a competitor who is much stronger (weaker) would elicit signals of submission (willingness to escalate conflict; H_2). By signaling I refer to changes in upper body strength from baseline, when the other competitor is not present. As evidenced in the pilot study, signaling should not require explicit comparisons of formidability (e.g., displaying scores on upper body strength measures), although such a manipulation might certainly enhance the effect.

Finally, changes in signaled strength should predict success in a physical competitive task over and above any effects due to baseline differences in strength (H_3). Individuals who signal submission should lose more than individuals who signal willingness to escalate conflict.

METHOD

Same-sex pairs of participants reported separately to the lab and had baseline measurements of strength taken by a same-sex confederate. Participants were then brought together and had their strength measured again. Finally they competed against each other in contest requiring upper body strength. Throughout the experiment participants answered a variety of questions related to perceptions of self and partner strength.

Participants

Participants were 398 Michigan State undergraduate students (196 women; 24.7% non-White) who participated together in dyads (98 female dyads, 101 male dyads). The experiment was advertised as two separate experiments to avoid participants signing up with friends and to ensure random sampling.

Measurements

Participants' left hand, right hand, and upper body strength were measured via handgrip dynamometer. Each strength measurement was taken three times, and the average of these scores was used in analyses. Height and arm length were measured in centimeters. Because variation in bicep circumference was smaller than the former two, it was measured in millimeters.

Procedure

The first participant arrived for an experimental session fifteen minutes prior to the second participant. Participants were told that the current experiment was interested with how physical strength related to personality. The experimenter then took baseline measurements of upper body strength and other body measurements. Measurements were not shared with participants. The participant was then escorted to an individual suite and given instructions to complete measures related to anger and formidability (Sell et al., 2009) unrelated to the current

experiment. After leaving the first participant, the experimenter escorted the second participant into a different suite, where he or she repeated the same procedures.

When both participants finished the measures, the experimenter brought them together into a large room in the lab and repeated the experimental cover story, elaborating that the two participants would later compete against each other in a physical task. Before the competition, the experimenter took the strength measurements again, this time with the other participant in the room. Measurements were taken individually with the hand dynamometer while their competitor faced away. Experimenters measured participants in the order that they arrived.

After measuring the strength of both participants, participants were escorted to two side-by-side computers. They then completed several measures concerned with relative judgments of strength (e.g., how strong are you relative to the other participant?) and predicted fight outcomes (e.g., how likely do you think you would be to win a physical fight against your partner?).

After both participants completed these measures they competed against each other in an arm wrestling contest (156 dyads) or a mercy contest (43 dyads).[2] Each participant was given three tickets into a raffle for a $50 gift card. For each round they won, participants were given one of their competitor's tickets. In the arm wrestling task, participants used their right hand to arm wrestle against their opponent. Participants won matches if they managed to push their competitors arm down against the table. In the mercy contest, participants interlocked the fingers on both their hands with their competitor's fingers. They then squeezed as hard as could in order to try to get the other participant to give in. In both contests, if a winner was not determined in less than 30 seconds, the match ended in a draw. Each dyad competed in up to three rounds,

---

[2] The competitive task was changed from a mercy task to an arm wrestling task because the former did not effectively differentiate between strong and weak participants; most matches (84.2%) ended in draws.

unless one participant forfeited. If a participant forfeited at any time, all their tickets were given to their competitor.

Participants then again completed measures asking about their strength relative to their partner and predicted fight outcomes. If participants consented, pictures of each were taken in order to facilitate rated judgments of strength and attractiveness. Finally, participants were fully debriefed and dismissed.

RESULTS

Planned Comparisons

Signaled strength.

To examine the effect competitors have on signaled strength, I first created composite variables of strength based on the average of the upper body, left hand, and right hand strength measures at each time point separately (αs = .924 and .925, respectively). The composite variable was strongly correlated between time points (r = .972, p < .001), as was the upper body strength measure (r = .920, p < .001), the left hand measure (r = .954, p < .001) and the right hand measure (r = .954, p < .001). This substantial collinearity presents problems with data analysis and will be addressed further in the discussion.

As expected and shown in Table 1, men were substantially stronger than women on all strength measures. In particular, the mean score for men on the standardized composite variable (0.70) was almost one standard deviation higher than the mean score for women (-0.76). When examining all the strength variables, in over 99% of cases, the mean score for men exceeded the maximum score for women. Variability in strength also differed by gender, with men having almost twice as much variability compared to women on all strength measures.

These composite strength variables were entered into a multilevel model using restricted maximum likelihood in order to estimate an indistinguishable actor-partner interdependence model (APIM; Kenny, Kashy, & Cook, 2006) assessing the effects of baseline strength on strength when a competitor was present (hereafter referred to as signaled strength). If participants adaptively signal strength depending on competitor strength, being paired with a stronger individual should result in less signaled strength (a negative partner effect). However, this effect should be moderated by the individual's own level of strength, such that decreases in strength

signaling should be greater when participants are weak (a negative actor-partner interaction). Finally, baseline strength should strongly predict signaled strength (a positive actor effect), such that strength should largely be stable across time points.

The APIM conducted on the composite strength measure revealed no evidence that signaled strength was influenced by a competitor, $\beta = .021$, $t(260) = 1.30$, $p = .196$. However, the actor-partner strength interaction $\beta = -.032$, $t(185) = -2.11$, $p = .036$ was significant and negative, indicating that weaker participants paired with stronger partners tended to signal less strength. Thus, as predicted, participants signaled strength changed based on who they were competing against. Finally, a large and positive actor effect was found $\beta = .945$, $t(259) = 59.25$, $p < .001$, indicating high stability between strength measurements. Participant sex did not influence signaled strength, nor did it interact with the strength variables, (ts < 2, ps $> .200$).

The above analysis was also conducted separately on each of the three strength measures: upper body strength, left arm strength, and right arm strength. While in all cases there was stability between baseline strength measurements (all ts > 30, ps $< .001$), no actor-partner strength interactions reached significance (all ts < 2, ps $> .080$). Additionally, in the analysis of the right arm strength measure the presence of the confederate actually led to greater strength ($b = .038$, $t(281) = 2.11$, $p = .036$, opposite of what was predicted.

Examination of the intraclass correlations for each of the strength measures revealed that the data were nonindependent, intraclass $r = .55 - .65$, ps $< .001$. That is, scores within dyads were more similar than scores between dyads. Although this ostensibly suggests that participants' strength was influenced by whom they competed against, the null results obtained for the partner effects are inconsistent with this interpretation. Furthermore, rerunning the analyses with only baseline strength as a predictor revealed that it was sufficient to explain the

nonindependence in the data; all intraclass correlations were reduced to nonsignficance. Thus, it is likely that the similarity in the data is artifactual. Recall that participants were paired with same-sex partners. Because men's scores were more similar to other men rather than to women, and women's scores were more similar to other women than to men (see Table 1), the nonindependence might simply be due to experimental design. Indeed, as Table 2 shows, when examining the similarity between competitors separately by sex, the nonindependence was reduced to nonsignificance in all but one case. Thus, the intraclass correlation is inflated by experimental design, giving little evidence to suggest that participants' strength was influenced by their opponent.

Diagnostic analyses.

One reason why the current study may have found no evidence for changes in strength signaling is due to the fact that most dyads were evenly matched in terms of strength, compared to the pilot study where strength differences were manipulated via confederates. When examining the upper body strength measure standardized across sex, 152 dyads (81.2%) had individuals within one standard deviation of each other. Even when standardized within sex, 117 dyads (62.5%) still had individuals within one standard deviation of each other. For comparison, in the pilot study of 121 men, paired with five confederates, only 45 dyads (37.2%) had individuals within one standard deviation of each other. Recall that in the pilot study, post-hoc tests revealed that decreases in upper body strength were most pronounced for participants paired with the two strongest confederates. Given the smaller discrepancies in strength in the current study, the finding that competitors do not influence personal strength is ambiguous.

To reiterate, I hypothesized that individuals are sensitive to discrepancies in RHP, which lead to different signals (submission or willingness to escalate) indexed by strength changes. The

larger the discrepancy, the more pronounced the signal. This issue is complicated by the fact that assessments are not perfect; while they are strongly correlated with measurements of physical strength ($r = .66$ for men and $r = .51$ for women; Sell et al. 2009), there is still a substantial degree of noise. Because judgments are not perfect, when RHP between competitors is roughly equal, they may both assess themselves as stronger (Parker, 1974). In this case, both individuals should signal willingness to fight. Given that a large percentage of participants were of roughly equal RHP, this may have obscured the relationship between differences in formidability and greater signaling of submission.

One way to test this hypothesis directly is to split the dyads into two groups, one where there the discrepancy in strength between competitors is large, and another where differences between competitors is small. This was accomplished by using the composite measure of strength (standardized within sex)[3] to split the 199 dyads into two groups, one group where the difference in strength between competitors was less than one standard deviation (the similar strength subset, n = 112) and the other where the difference was more than one standard deviation (the different strength subset, n = 77). If similarity in strength was obscuring the relationship between formidability and (submissive) signaling, examining the different strength group should reveal the expected pattern. The similar strength group should signal increased strength as the difference between competitor formidability decreases, above and beyond any effects due to personal strength and competitor strength.

However, running the APIM with the different strength dyads did not change the pattern of results. Upper body strength as measured with a competitor present was still strongly

---

[3] As there was more variability in strength in the male dyads, measurements were standardized within sex to prevent imbalance in the number of male and female dyads selected into the similar and dissimilar groups. Standardizing across sex did not change the pattern of results.

20

influenced by baseline strength ($\beta = .952$, $t(127) = 54.58$, $p < .001$), but the effect of competitor strength did not reach significance ($\beta = .014$, $t(127) = -0.76$, $p = .428$), nor did their interaction ($\beta = -.023$, $t(74) = -0.95$, $p = .348$). Similarly, running the APIM with the same strength dyads also did not change the pattern of results. Upper body strength as measured with a competitor present was still strongly influenced by baseline strength ($\beta = .893$, $t(120) = 20.78$, $p < .001$), but the effect of competitor strength did not reach significance ($\beta = .066$, $t(120) = 1.53$, $p = .128$), nor did the absolute magnitude of their difference ($\beta = -.022$, $t(74) = -0.95$, $p = .348$). The coefficient represents the degree to which greater (or lesser) discrepancy in baseline strength is related to greater strength in the presence of the competitor.

In sum, we did not find evidence for strength signaling, even when splitting the dyads into two groups, one where the discrepancy in strength was large, and the other where the discrepancy was low. Implications of this are further addressed in the discussion.

Competition outcomes.

To examine how baseline and signaled strength influence competitive outcomes, arm wrestling outcomes were regressed on strength measures at the dyad level. If signaled strength influences competition outcomes above and beyond baseline differences in strength, its inclusion should improve the predictive validity of the model when controlling for any differences due to baseline strength. That is, does strength signaling influence the likelihood that an individual will win a physical contest such as arm wrestling?

Because competition outcomes between competitors are completely dependent (i.e., winning a match necessarily means that the other competitor loses), analyses were conducted at the dyad level. As assignment to Person 1 and Person 2 is arbitrary, in all analyses intercepts were suppressed. Predictors were determined by taking the difference of the two competitors

scores (e.g., $X_1 - X_2$). The difference in outcomes (range: -3 to 3) was then regressed on the difference between predictors. Thus, positive coefficients represent that greater asymmetry in the predictor (e.g., upper body strength) predicts competitive success.

To control for individual differences not (directly) related to upper body strength, hand dominance and arm length were first entered into the model. The overall model was significant, $R^2 = .059$, $F(2, 140) = 4.36$, $p = .015$. Handedness did not predict success in the arm wrestling competition, $b = .090$, $t(140) = .540$, $p = .590$, and was removed from the model. Greater arm length significantly predicted success, $b = .081$, $t(140) = 2.933$, $p = .004$, and was retained. In the next step, baseline right arm strength (all matches were right handed) and upper body strength were added as predictors. As can be seen in Table 3, the overall model was again significant, $\Delta R^2 = .203$, $F(2, 139) = 19.05$, $p < .001$. Greater upper body strength ($b = .045$, $t(139) = 3.00$, $p = .003$) and right arm strength ($b = .060$, $t(139) = 3.17$, $p = .002$) both significantly predicted success. In the last step of the model, signaled upper body and right arm strength were added as predictors. However, adding in these variables did not significantly increase the predictive power of the model $\Delta R^2 = .008$, $F(2, 137) = 0.78$, $p = .460$. Thus, Hypothesis 3 was not supported, in that signaled strength did not influence the likelihood of winning a physical contest.

Exploratory Analyses

Updating strength assessments.

To further examine whether participants were accurately assessing formidability of conspecifics (Hypothesis 1), I investigated participants' relative strength assessments. Recall that participants were asked to report their strength relative to their partner on a 7-point scale (1 = I am much weaker than the other participant, 7 = I am much stronger than the other participant) at two time points, first just after meeting the other participant and second after the competition. I

predicted participants would use their own strength as well as their competitor's strength when judging who was stronger. However, participants would gain additional information about strength from the competition that would influence their relative strength assessments such that participants who won would reassess themselves as stronger, while participants who lost would reassess themselves as weaker.

To test this hypothesis, I examined whether or not participants updated their relative strength assessments based on information gained from the arm wrestling competition. To control for nonindependence in the data, I ran a multilevel regression model where the second relative strength assessment was regressed onto the initial relative strength assessment, as well as the outcome of the competition, participant sex, and the sex by competitive outcome product term. A positive coefficient for the competition outcome reflects that controlling for the initial relative strength assessment, participants who won (lost) the competition were more likely to rate themselves as stronger (weaker) than their competitor. Table 4 lists the model coefficients. As predicted, the coefficient for the competitive outcome was significant and positive, $b = .547$, $t(166) = 19.38$, $p < .001$, indicating that controlling for their initial strength assessment, participants adjusted their strength ratings based on the competitive task. This effect was moderated by a significant sex by competitive outcome interaction, $b = -.079$, $t(140) = -3.01$, $p < .001$, indicating that men updated their strength ratings less than women. The overall model accounted for the vast majority the variance in relative strength assessments, pseudo-$R^2 = .734$.

Accuracy of updated assessments.

Although updating strength assessments after a physical competition is consistent with an adaptive reappraisal of strength differences, it is unclear whether changes in assessment better reflect reality. Consider the case where two individuals meet and initially consider themselves of

23

equal strength. With some difficulty, one participant edges out the other in the competition and updates his or her relative strength assessment to suggest that he or she is stronger. The issue is whether this updated assessment more accurately reflects the difference in strength between the two individuals. It is possible that participants might overcorrect, resulting in assessments that less accurately track strength differences than those made before the competition. However, if participants are accurately using the information from the competition to inform their strength assessments, these posterior assessments should more accurately reflect strength differences.

A preliminary test revealed that, as expected, initial assessments of strength were nonindependent, intraclass $r = -.268$, p < .001, indicating that the more (less) individuals rated themselves as stronger relative to their competitor, the less (more) their partner rated themselves as stronger relative to their competitor. More relevant, strength assessments were more correlated after the arm wrestling competition, intraclass $r = -.783$, $Z = 5.66$, $p < .001$. This increase in nonindependence is expected because as both participants' accuracy increases, so should the correlation between their assessments.

Because relative assessments of strength are nonindependent, they were entered into a multilevel model using restricted maximum likelihood in order to estimate an indistinguishable over-time APIM assessing the effects of baseline strength (composite measure) on relative strength assessments. I predicted participants' relative strength assessments would be strongly influenced by their own strength (a positive actor effect) as well as the strength of their partner (a negative partner effect) at both time points, but particularly after the arm wrestling competition. That is, not only would participants reassess strength judgments after the physical competition, but these judgments would be more accurate (i.e., better predicted by baseline strength).

As predicted and displayed in Table 5, the over-time APIM revealed that relative strength assessments were positively related to greater personal baseline strength, $b = .971$, $t(211) = 8.33$, $p < .001$, and negatively related to the competitor's baseline strength, $b = -.661$, $t(208) = -5.71$, $p < .001$. Thus, participants' strength assessment was influenced by both their own strength and their competitor's strength. More importantly, the predicted actor strength by time ($b = .226$, $t(169) = 3.56$, $p < .001$) and partner strength by time ($b = -.190$, $t(169) = -3.00$, $p = .003$) interactions were significant, indicating participant's assessments after the competition better tracked baseline strength differences. That is, participants' relative strength assessments became more accurate after the competition. There was also a main effect of participant sex, such that men were more likely to rate their competitor as less strong than women, $b = -.254$, $t(144) = -2.87$, $p = .005$. However, sex did not interact with any other variable (ts < 2, ps > .10). In sum, 32.3% of the variance in participants' strength assessments were explained by differences in physical strength.

A similar analysis was conducted on judgments of predicted fight outcomes. As with relative strength assessments, participants were asked to report on a 7-point scale how likely they would be to win a physical fight with their partner as well as what the outcome of the fight would be, where higher numbers represent better predicted success. These judgments were also reported before and after the arm wrestling competition (αs = .930 and .959, respectively). Paralleling assessments of physical strength, the over-time APIM revealed that predicted fight outcomes were positively related to greater personal baseline strength, $b = .797$, $t(237) = 6.14$, $p < .001$, and negatively related to the competitor's baseline strength, $b = -.465$, $t(234) = -3.61$, $p < .001$, see Table 6. Thus, participants' predicted fight outcomes were influenced by both their own strength and their competitor's strength. More importantly, the predicted actor strength by time

(b = .125, $t(184)$ = 3.22, p = .002) and partner strength by time (b = -.100, $t(184)$ = -2.57, p = .011) interactions were significant, indicating participant's judgments after the competition better tracked baseline strength differences. That is, participants' predicted fight outcomes more accurately tracked differences in physical strength after the competition. Unlike the relative strength measure, there was no main effect of participant sex, b = -.187, $t(145)$ = -1.62, p = .102, nor did it interact with any other variable (ts < 2, ps > .05). In sum, 15.3% of the variance in participants' predicted fight outcomes were explained by differences in physical strength.

DISCUSSION

This research emphasizes that humans can accurately assess differences in physical formidability and update these assessments based on information gained during an interaction. Individuals took into account both their own strength and the strength of a competitor when making initial assessments about who was stronger as well as who would win in a physical fight. In addition, participants updated their assessments after competing against the competitor in a physical task. Like non-human animals, humans appear to have psychological mechanisms that accurately track cues to physical formidability, and they dynamically update these cues based on relevant information.

Assessment Accuracy

Individuals' relative assessment of strength accurately tracked their own physical formidability as well as that of the other competitor. Although their initial estimate tracked strength differences, it was improved by competing against the other individual in a contest determined by physical strength. Improvement was obtained regardless of whether dyads were male or female, indicating that both men and women gained information from the competition.

Improvement was not due to mere exposure to the other participant. Recall that a portion of the dyads (38 out of 199) engaged in a different competitive task (i.e., a mercy task) before making their second assessment of strength. This task was not very sensitive to differences in physical formidability; 84.2% of all matches ended in draws compared to only 32.9% in the arm wrestling contest, $\chi^2(1) = 32.18$, $p < .001$. In those dyads, accuracy in strength assessments actually became worse after the competition (time by actor/partner effects were in the opposite direction). Similarly, participants predicted fight outcomes did not more accurately track differences in physical formidability after the competition (time by actor/partner effects were

27

nonsignificant). Thus, completing a task that is a poor indicator of physical formidability, or unrelated to physical formidability at all, is unlikely to improve assessment accuracy.

Defensive Signaling

Based on the results of a pilot study, I predicted that individuals would modify their strength based on the strength of their competitor. Specifically, when an individual was weaker than their competitor they would signal submission, resulting in decreased strength relative to baseline. In contrast, when an individual was stronger than their competitor they would signal willingness to escalate conflict, resulting in increased strength relative to baseline. While the data did not support this hypothesis, this null result may be due to experimental design limitations.

The issue is that the current study did not manipulate strength via pairing participants with confederates chosen based on physical formidability. Instead, all individuals were naïve participants whose strength varied based on individual differences. As discussed in the results section, this led to considerably less dyads where there was a large discrepancy in strength. In line with prior theorizing (e.g., Parker, 1974), because strength assessment is not perfect, when individuals are similar in strength, they may both signal willingness to escalate conflict. This could have created noise in the data that obscured the relevant differences in signaling.

However, dividing dyads into two groups where the difference in strength was either similar or dissimilar (i.e., less or greater than one standard deviation, respectively) did not change the pattern of pattern of results. Personal strength was not influenced by the presence of a competitor. If the prediction is correct, this null funding may simply be due to low power. However, if signaling changes do occur when individuals interact with competitors, the data suggest the size of the effect is small. To ascertain if the effect is reliable as well as to narrow down the confidence interval of the effect, it will be necessary to conduct additional research that

manipulates the discrepancy in strength between individuals. A highly powered test of this hypothesis is necessary to replicate the original effect and help rule out the possibility of a false positive. It may also be the case that the effect driving strength changes in the pilot study was confounded with strength amongst the confederates selected. To rule out these possibilities, more research with additional confederates is needed.

Because there were no differences in signaled strength, it is therefore unsurprising that I did not find support for the hypothesis that defensive signaling influenced competitive outcomes. Given well-known problems with including highly correlated variables as predictors in a regression model (Cronbach, 1987), it not unusual that including the second upper body strength ($r = .920$) and right arm strength ($r = .954$) measures would not significantly improve model fit. Because this hypothesis is contingent upon finding evidence for strength differences based on the presence of a competitor, this hypothesis cannot directly be evaluated with the current data.

Implications For Game Theory

The finding that humans can accurately assess the formidability of conspecifics and update those assessments based on information gained during contests supports predictions derived from game theoretical principles. As Parker and colleagues (Maynard Smith & Parker, 1976; Parker & Rubenstein, 1981) predicted, "it is likely that good information, particularly about RHP, can be acquired only during a contest itself" (Parker & Rubenstein, 1981, p. 288). That is, individuals initially might not know (perfectly) the asymmetry in RHP between them and their competitor, but gain such information through physical competition. Insofar as this process is available to conscious introspection, it should be reflected through changes in assessments of relative strength and fighting ability, as it was in the current experiment.

Typically, reassessment of relative strength is determined not by conscious changes in relative strength, but is inferred from the decision to withdraw from a contest. However, low rates of forfeiting in the current study (4.0%) prevented conclusive analyses of this. This is likely due to the structure of the competitive tasks used: participants were told that forfeiting at any point would result in a loss of all their raffle entries. This design encouraged participants to keep competing, even if they were likely to lose, because the opportunity cost of not competing was always higher than the cost of competing. When combined with minimal risk of physical injury (i.e. low damage costs), low rates of forfeiture are consistent with game theoretical predictions. This interpretation is bolstered by evidence that participants did update their relative strength assessments based on the match, even if they did not often forfeit.

This logic suggests that changing the nature of the contest would change the duration of time before the weaker individual withdraws. Instead of a winner-take-all style competition, if participants are given the option to save their resources in addition to risking them for potential reward, the decision to withdraw should be contingent upon how great the potential loses are for a given round. All else equal, participants for whom potential loses are greater should be quicker to withdraw after discovering that their competitor has higher RHP. This non-fixed framework would also allow inferences as to how the duration of a contest influences assessment accuracy. The current design does not allow for a test of whether contest duration influences accuracy because the vast majority of individuals engaged in three rounds of the physical competition. However, it seems likely that additional bouts would have an asymptotic relationship with RHP assessment; given little information about a competitor (i.e., few rounds of competition) an additional round should help accuracy, however, after many rounds, an additional round is

unlikely to improve accuracy. Future research should test this prediction in order to provide additional support for the predictive validity of game theoretical predictions to human conflict.

Similarities to Animal Assessment

These results share several similarities with studies on competition between human as well as non-human animals. First, these results add to the growing literature demonstrating that humans have psychological mechanisms that accurately determine the strength of conspecifics (e.g., Archer & Thanzami, 2009; Sell et al., 2009). Additionally, they extend this literature by demonstrating that assessments of predicted fight outcomes take into account both personal strength as well as competitor strength. While there is considerable debate about how to best model RHP assessment in non-human animals (for a review, see Taylor & Elwood, 2003), our results provide strong evidence that humans engage in mutual assessment. Though predicted by Arnott and Elwood (2010), no other study has directly tested how strength differences between competitors predict expected conflict outcomes.

More generally, these results parallel findings in the non-human animal literature demonstrating that assessment improves accuracy of RHP. For example, during the mating season, red deer stags attempt to guard fertile hinds from other stags (Clutton-Brock & Albon, 1979). A small proportion of approaches between stags actually result in fights; typically when discrepancy in fighting ability is visible, the stag with lower RHP will withdraw. Only when RHP is roughly equal do stags engage in roaring contests, where vocalizations accurately reflect fighting ability. The majority of fights are also preceded by a parallel walks, where stags display their full body size by walking alongside each other. In both cases displays provide additional information that influences the decision to engage or withdraw. Similarly, in the current study, individuals have initial impressions of the strength of their competitor. The accuracy of these

31

impressions is improved based on several rounds of a competition that reflects fighting ability. In many cases where discrepancies in RHP are not clear, competition provides additional evidence as to the magnitude of those discrepancies.

Conclusion

I found evidence that humans can accurately assess differences in physical formidability and that these assessments are dynamically updated based on information gained from competitive bouts. These processes show remarkable similarities to assessment in non-human animals and suggest that animal models are useful for predicting human behaviors in competitive situations. Similarly, these results are largely consistent with predictions derived from a game theoretical framework.

While the prediction that individuals would broadcast different signals depending on the relative difference in strength between them and their competitor was not confirmed, this null effect may be due to the small degree of differences between individuals in dyads. Similarly, the validity of the hypothesis that conflict outcomes would be influenced based on signaling submission or willingness to escalate conflict cannot be ascertained without successfully manipulating differential signaling. More work will need to be done to replicate the original effect and help rule out the possibility of a false positive.

APPENDIX

Table 1

Descriptive Statistics for Baseline Strength Measures

	Strength Measure	Minimum	Maximum	*M*	*SD*
Women	Upper Body	1.67	36.00	19.20	6.10
	Right Arm	14.33	44.33	27.11	5.15
	Left Arm	12.33	47.33	25.68	5.14
	Composite	-1.67	0.39	-0.76	0.39
Men	Upper Body	4.67	69.33	39.84	12.33
	Right Arm	20.67	71.33	42.76	9.13
	Left Arm	14.33	64.67	42.76	8.98
	Composite	-1.03	2.60	0.70	0.73

Note. Women $N = 188$, Men $N = 197$. Upper body, right arm, and left arm strength are all

measured in kilograms. Composite is the average of the three standardized strength measures.

Table 2

Intraclass Correlations Between Strength Measures

Strength Measure	All Ps (*N* = 374)	Women (*N* = 184)	Men (*N* = 190)
Upper Body	.592***	.088	.204*
Left Arm	.597***	.137	.069
Right Arm	.546***	.081	.006
Composite	.647***	.097	.098

Note. P = participant. Higher correlations indicate greater nonindependence, i.e., that strength measurements within dyads are more similar than strength measurements between dyads.

*p < .05, ***p < .001

Table 3

Multiple Regression Results Predicting Success in Arm Wrestling Competition From Strength Measures

Strength Measure	Model 1		Model 2		Model 3	
	b	β	b	β	b	β
Arm length	.080**	.238	$.043^{\dagger}$	.128	$.047^{\dagger}$	.139
Upper Body Strength (T_1)	--	--	.045**	.254	.032*	.182
Right Arm Strength (T_1)	--	--	.060**	.280	.088	.408
Upper Body Strength (T_2)	--	--	--	--	.017	.109
Right Arm Strength (T_2)	--	--	--	--	-.036	-.166
R^2	.057		.260		.268	
F	8.47**		16.25***		10.32	
ΔR^2	--		.203		.008	
ΔF	--		19.05		0.78	

Note. T_1 = baseline strength measurement. T_2 = strength measurement with competitor present. Model 1 *df* = 141, Model 2 *df* = 139, Model 3 *df* = 137. All variables were grand-mean centered prior to analysis. Regression was conducted at the dyad level, so the intercept was suppressed.

†p < .10 *p < .05, **p < .01, ***p < .001.

Table 4

Multilevel Multiple Regression Results Predicting Updated Strength Assessments From Past

Strength Assessments and Competition Outcomes

	b	*β*	*t*	*df*
Intercept	3.085	--	--	--
Strength Judgment (T_1)	.251	.215	6.42***	263
Competition Outcome	.547	.748	19.38***	166
Sex	.048	.032	1.24	138
Sex*Competition Outcome	-.079	-.107	-3.01**	140

Note. T_1 = relative strength judgment prior to competition. All continuous variables were grand-

mean centered prior to analysis; sex was effects coded (-1 = women, 1 = men).

**$p < .01$, ***$p < .001$.

Table 5

Over-time APIM Model Predicting Relative Strength Assessment From Strength Measurements

Before and After Competition

	b	β	*t*	*df*
Intercept	3.877	--	--	--
Actor Strength	.971	.660	8.33***	211
Partner Strength	-.661	-.449	-5.71***	208
Sex	-.254	-.181	-2.87**	144
Time	.131	.093	4.28***	143
Actor Strength*Time	.226	.153	3.56***	169
Partner Strength*Time	-.190	-.129	-3.00**	169

Note. Strength refers to the composite strength measure taken at baseline. All continuous

variables were grand-mean centered prior to analysis; sex was effects coded (-1 = women, 1 =

men) and time was effects coded (-1 = initial assessment, 1 = reassessment after competition).

**p < .01, ***p < .001.

Table 6

Over-time APIM Model Predicting Fight Outcomes From Strength Measurements Before and

After Competition

	b	β	t	df
Intercept	4.079	--	--	--
Actor Strength	.797	.557	6.14***	237
Partner Strength	-.465	-.325	-3.61***	234
Sex	-.187	-.137	-1.62	145
Time	.082	.059	3.47**	144
Actor Strength*Time	.125	.088	3.21**	184
Partner Strength*Time	-.100	-.070	-2.57**	184

Note. Strength refers to the composite strength measure taken at baseline. All continuous

variables were grand-mean centered prior to analysis; sex was effects coded (-1 = women, 1 =

men) and time was effects coded (-1 = initial judgment, 1 = judgment after competition).

**$p < .01$, ***$p < .001$.

REFERENCES

REFERENCES

Archer, J., & Benson, D. (2008). Physical aggression as a function of perceived fighting ability and provocation: An experimental investigation. *Aggressive Behavior, 34*(1), 9-24. doi: 10.1002/ab.20179

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. doi: 10.1016/j.evolhumbehav.2009.03.003

Arnott, G., & Elwood, R. W. (2010). Signal residuals and hermit crab displays: flaunt it if you have it! *Animal Behaviour, 79*(1), 137-143. doi: 10.1016/j.anbehav.2009.10.011

Benson-Amram, S., Heinen, V. K., Dryer, S. L., & Holekamp, K. E. (2011). Numerical assessment and individual call discrimination by wild spotted hyaenas, Crocuta crocuta. *Animal Behaviour, 82*(4), 743-752. doi: 10.1016/j.anbehav.2011.07.004

Bernstein, I. S., & Gordon, T. P. (1980). The social component of dominance relationships in rhesus monkeys (Macaca mulatta). *Animal Behaviour, 28*(4), 1033-1039. doi: 10.1016/S0003-3472(80)80092-3

Bernstein, I. S., Gordon, T. P., & Rose, R. M. (1974). Factors influencing the expression of aggression during introductions to rhesus monkey groups. *Primate Aggression, Territoriality and Xenophobia*, 211-240.

Blanchard, D. C. (1997). Stimulus, environmental, and pharmacological control of defensive behaviors. In M. E. Bouton & M. S. Fanselow (Eds.), *Learning, motivation, and cognition: The functional behaviorism of Robert C. Bolles* (pp. 283-303). Washington, DC, US: American Psychological Association.

Blanchard, D. C., & Blanchard, R. J. (2003). What can animal aggression research tell us about human aggression? *Hormones and Behavior, 44*(3), 171-177. doi: 10.1016/S0018-506X(03)00133-8

Blanchard, D. C., Griebel, G., Pobbe, R., & Blanchard, R. J. (2011). Risk assessment as an evolved threat detection and analysis process. *Neuroscience & Biobehavioral Reviews, 35*(4), 991-998. doi: 10.1016/j.neubiorev.2010.10.016

Blanchard, D. C., Hynd, A. L., Minke, K. A., Minemoto, T., & Blanchard, R. J. (2001). Human defensive behaviors to threat scenarios show parallels to fear- and anxiety-related defense patterns of non-human mammals. *Neuroscience & Biobehavioral Reviews, 25*(7–8), 761-770. doi: 10.1016/S0149-7634(01)00056-2

Cesario, J., & Johnson, D. J. (2013). *Partner formidability influences personal strength*. Manuscript in preparation.

Chase, I. D., & Seitz, K. (2011). Self-Structuring Properties of Dominance Hierarchies: A New Perspective. In D. L. B. Robert Huber & B. Patricia (Eds.), *Advances in Genetics* (Vol. Volume 75, pp. 51-81): Academic.

Clutton-Brock, T. H., & Albon, S. D. (1979). The roaring of red deer and the evolution of honest advertisement. *Behaviour*, 145-170.

Cronbach, L. J. (1987). Statistical tests for moderator variables: Flaws in analyses recently proposed. *Psychological Bulletin, 102*(3), 414-417.

Dawkins, M. S., & Guilford, T. (1991). The corruption of honest signalling. *Animal Behaviour, 41*(5), 865-873. doi: 10.1016/S0003-3472(05)80353-7

de Waal, F. B. M., & Hoekstra, J. A. (1980). Contexts and predictability of aggression in chimpanzees. *Animal Behaviour, 28*(3), 929-937. doi: 10.1016/S0003-3472(80)80155-2

Gawronski, B., & Cesario, J. (2013). Of mice and men: What animal research can tell us about context effects on automatic responses in humans. *Personality and Social Psychology Review, 17*(2), 187-215. doi: 10.1177/1088868313480096

Griskevicius, V., Tybur, J. M., Gangestad, S. W., Perea, E. F., Shapiro, J. R., & Kenrick, D. T. (2009). Aggress to impress: Hostility as an evolved context-dependent strategy. *Journal of Personality and Social Psychology, 96*(5), 980-994. doi: 10.1037/a0013907

Ikai, M., & Steinhaus, A. H. (1961). Some factors modifying the expression of human strength. *Journal of Applied Physiology, 16*(1), 157-163.

Kenny, D. A., Kashy, D., & Cook, W. L. (2006). *Dyadic data analysis*. New York, NY: Guilford.

Marcus, D. K., Kashy, D. A., & Baldwin, S. A. (2009). Studying psychotherapy using the one-with-many design: The therapeutic alliance as an exemplar. *Journal of Counseling Psychology, 56*(4), 537.

Matsumura, S., & Hayden, T. J. (2006). When should signals of submission be given?–A game theory model. *Journal of Theoretical Biology, 240*(3), 425-433. doi: 10.1016/j.jtbi.2005.10.002

Maynard Smith, J. (1974). The theory of games and the evolution of animal conflicts. *Journal of Theoretical Biology, 47*(1), 209-221.

Maynard Smith, J. (1982). Do animals convey information about their intentions? *Journal of Theoretical Biology, 97*(1), 1-5. doi: 10.1016/0022-5193(82)90271-5

Maynard Smith, J., & Parker, G. A. (1976). The logic of asymmetric contests. *Animal Behaviour, 24*(1), 159-175. doi: 10.1016/S0003-3472(76)80110-8

Maynard Smith, J., & Price, G. R. (1973). The Logic of Animal Conflict. *Nature, 246*(5427), 15-18. doi: 10.1038/246015a0

Parker, G. A. (1974). Assessment strategy and the evolution of fighting behaviour. *Journal of Theoretical Biology, 47*(1), 223-243. doi: 10.1016/0022-5193(74)90111-8

Parker, G. A., & Rubenstein, D. I. (1981). Role assessment, reserve strategy, and acquisition of information in asymmetric animal conflicts. *Animal Behaviour, 29*(1), 221-240. doi: 10.1016/S0003-3472(81)80170-4

Sell, A., Bryant, G. A., Cosmides, L., Tooby, J., Sznycer, D., von Rueden, C., et al. (2010). Adaptations in humans for assessing physical strength from the voice. *Proceedings of the Royal Society B: Biological Sciences, 277*(1699), 3509-3518. doi: 10.1098/rspb.2010.0769

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 B: Biological Sciences, 276*(1656), 575-584. doi: 10.1098/rspb.2008.1177

Sell, A., Hone, L. S., & Pound, N. (2012). The importance of physical strength to human males. *Human Nature: An Interdisciplinary Biosocial Perspective, 23*(1), 30-44. doi: 10.1007/s12110-012-9131-2

Simpson, M. J. A. (1968). The display of the Siamese fighting fish, Betta splendens. *Animal Behaviour Monographs, 1*(1).

Taylor, P. W., & Elwood, R. W. (2003). The mismeasure of animal contests. *Animal Behaviour, 65*(6), 1195-1202. doi: 10.1006/anbe.2003.2169

van Staaden, M. J., Searcy, W. A., & Hanlon, R. T. (2011). Signaling aggression. In D. L. B. Robert Huber & B. Patricia (Eds.), *Advances in Genetics* (Vol. 75, pp. 23-49). New York, NY: Academic.