

Research



Cite this article: Green PA, Patek SN. 2018 Mutual assessment during ritualized fighting in mantis shrimp (Stomatopoda). *Proc. R. Soc. B* **285**: 20172542.
<http://dx.doi.org/10.1098/rspb.2017.2542>

Received: 13 November 2017

Accepted: 14 December 2017

Subject Category:

Behaviour

Subject Areas:

behaviour, ecology, biomechanics

Keywords:

contests, assessment, resource holding potential, resource ownership, animal weapons, mantis shrimp

Author for correspondence:

P. A. Green

e-mail: pag16@duke.edu

Electronic supplementary material is available online at <https://dx.doi.org/10.6084/m9.figshare.c.3965835>.

Mutual assessment during ritualized fighting in mantis shrimp (Stomatopoda)

P. A. Green and S. N. Patek

Biology Department, Duke University, Durham, NC, USA

PAG, 0000-0002-2434-8795

Safe and effective conflict resolution is critical for survival and reproduction. Theoretical models describe how animals resolve conflict by assessing their own and/or their opponent's ability (resource holding potential, RHP), yet experimental tests of these models are often inconclusive. Recent reviews have suggested this uncertainty could be alleviated by using multiple approaches to test assessment models. The mantis shrimp *Neogonodactylus bredini* presents visual displays and ritualistically exchanges high-force strikes during territorial contests. We tested how *N. bredini* contest dynamics were explained by any of three assessment models—pure self-assessment, cumulative assessment and mutual assessment—using correlations and a novel, network analysis-based sequential behavioural analysis. We staged dyadic contests over burrow access between competitors matched either randomly or based on body size. In both randomly and size-matched contests, the best metric of RHP was body mass. Burrow residency interacted with mass to predict outcome. Correlations between contest costs and RHP rejected pure self-assessment, but could not fully differentiate between cumulative and mutual assessment. The sequential behavioural analysis ruled out cumulative assessment and supported mutual assessment. Our results demonstrate how multiple analyses provide strong inference to tests of assessment models and illuminate how individual behaviours constitute an assessment strategy.

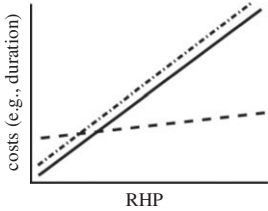
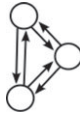
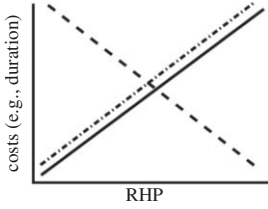
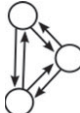
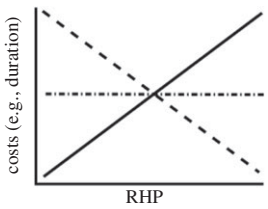

1. Introduction

Competing animals face the fundamental challenge of resolving conflict while minimizing the costs involved. To resolve contests efficiently, animals are thought to assess their own and/or their competitor's resource holding potential (RHP; defined as an individual's absolute competitive ability, [1]). Theoretical and experimental studies have established three primary strategies, originally developed from game-theoretical models, by which animals assess RHP to resolve contests: pure self-assessment, cumulative assessment and mutual assessment (reviewed in [2]). We term these 'assessment models' following [2] and note that they focus on assessment of competitors as opposed to mates or other individuals.

Assessment models provide an important framework in studies of animal contests, because they establish if individuals use information about (i.e. assess) their own and/or their competitor's RHP, how contest costs accrue, and how losers decide to give up the contest (table 1). In pure self- and cumulative assessment [3–5], individuals assess only their own RHP, giving up the contest when accrued costs reach a pre-set threshold. In these models, costs can be self-imposed by displays, or, in cumulative assessment only, imposed by a competitor (e.g. through injury). In mutual assessment [6], individuals assess both their own and their competitor's RHP; the loser gives up the contest when it has assessed that it is competitively inferior.

Assessment models can be differentiated by testing how behaviours progress throughout contests and how variation in contest costs correlates with variation in competitor RHP [8,9]. Competitors using mutual assessment assess both their own and each other's ability; therefore, contest behaviours

Table 1. Assessment models differ in the function of behaviours and determinants of outcome. (Models predict different directionality and strength of relationships between contest costs and RHP for losers (solid line) and winners (dashed line) of randomly matched contests, and for the averaged RHP of competitors in RHP-matched contests (dash-dot line) (following [2]). Models also predict different trends in contest behaviours, such as the presence of physical contact, the directionality of behavioural sequences, and how behaviours escalate and/or de-escalate (circles represent behaviours; arrows represent transitions between behaviours). Additional details and definitions are provided in the main text.)

model	model description	relationships between contest costs and RHP	trends in contest behaviours	references
pure self-assessment	impose energetic costs to self only loser reaches own cost threshold first		 no physical contact; behaviours progress in any direction; escalation and de-escalation	[2–4]
cumulative assessment	impose energetic costs to self and other costs (e.g. energetic, injury) to competitor loser reaches own cost threshold first		 physical contact, injuries likely; behaviours progress in any direction; escalation and de-escalation	[2,5]
mutual assessment	assess both self and competitor RHP loser retreats upon assessing it is the inferior competitor		 physical contact, injuries rare; behaviours progress unidirectionally via phases; escalation without de-escalation	[2,6,7]

are predicted to progress in phases of escalating intensity that give increasingly accurate information about RHP [7]. Competitors should not de-escalate to previously used (and therefore uninformative) behavioural phases. As behaviours in pure self- and cumulative assessment function only to impose costs, not to compare ability, behaviours can progress in any sequence and without phases [3,5]. Correlations between contest costs (e.g. duration) and RHP can also be used to differentiate among assessment models, especially correlations between costs and winner RHP in randomly matched contests and costs and averaged competitor RHP in RHP-matched contests. See table 1 for a summary of the logic and predictions of assessment models.

While experimental tests have validated assessment models in certain systems (e.g. [7,10]), results from other systems do not clearly align with any particular model (reviewed in [11], e.g. [12,13]). This incongruence is often because of the difficulty of acquiring the breadth of experimental data needed to test aspects related to both behavioural progressions and cost-RHP correlations [2,11]. To best differentiate among assessment models, many researchers have recommended a combined approach that incorporates both tests of behavioural progressions and correlations of costs and RHP [9,11]. However, while the techniques for correlational analyses are well-established and widely used [9], relatively fewer studies employ tests of behavioural progressions.

Testing behavioural progressions requires rigorous analysis and visualization of behavioural sequences; one technique that can be used to achieve this goal is sequential behavioural

analysis (also termed ‘sequential analysis’; [14–16]). While sequential analysis is common in the social sciences [16], it has rarely been applied to contest dynamics [17–19]. Sequential analysis can quantify the sequences of behaviours used by competitors, which assist in later tests against theoretical predictions (table 1). Additionally, this technique can identify how individual behaviours connect to a given assessment strategy [8]. For example, a sequential analysis of stalk-eyed fly (*Teleopsis dalmanni*) contest behaviour revealed progressions that supported mutual assessment, resolving confusion from prior studies that vacillated among assessment models [17]. Furthermore, the analysis found that competitors lined up their exaggerated eye stalks frequently (32% of all contest behaviours) and early in contests, suggesting that this behaviour facilitates assessment of relative ability [17].

Here, we examine assessment, using both correlational and sequential analyses, in the mantis shrimp *Neogonodactylus bredini* (Stomatopoda: Crustacea). *Neogonodactylus bredini* competes over access to territories (burrows) using powerful striking weapons. Previous research suggests that mantis shrimp may use either or both cumulative or mutual assessment. However, most research in *N. bredini* contest behaviour was conducted before these assessment models—and their predictions—were formalized [5,7]. This unresolved assessment strategy in an animal with dangerous weaponry makes *N. bredini* a useful system for testing among assessment models.

In cumulative assessment, competitors deliver energetic and/or injury costs to each other via contest behaviours

(although competitors do not assess each other's RHP directly, [5]). Cumulative assessment may be an appropriate model for describing *N. bredini* contest behaviour because individuals possess potentially deadly weapons that may deliver high costs. Both sexes of *N. bredini* use their second maxillipeds (termed raptorial appendages) to strike competitors [20,21]. These strikes can reach bullet-like accelerations ([22] and references therein), delivering high peak forces [23]. Competitors can also unfurl their raptorial appendages to stab each other, causing puncture wounds [20,24].

Despite the potential danger of striking and stabbing, *N. bredini* contest behaviours may allow for cost minimization via assessment of relative ability, which would support mutual assessment [6]. *Neogonodactylus bredini* use behaviours that may minimize costs; for example, competitors ritualistically and non-lethally exchange strikes on each other's armoured tailplates [20] in a behaviour termed 'telson sparing' [25] and, in our experience, potentially dangerous stabbing behaviours are rare [25]. Additionally, *N. bredini* exhibits individual competitor recognition (reviewed in [26]) and analyses of non-territorial contests found a behavioural stereotypy like that expected by mutual assessment [21,27]. Overall, the evidence suggesting cumulative and mutual assessment in *N. bredini* calls for a thorough test of assessment models in this system. The use of ritualized striking also presents an opportunity to examine how animals with deadly weapons use those weapons to inflict costs or as a means of assessing relative competitive ability.

In this study, we perform correlational and sequential analyses of *N. bredini* contest dynamics with the goals of: (i) identifying the assessment strategy used by *N. bredini*, (ii) explaining how individual behaviours function in assessment in *N. bredini*, and (iii) establishing if and how using multiple analysis techniques can conclusively support one assessment model by rejecting alternatives. We established a metric of RHP and then tested how RHP correlates with two measures of costs in dyadic, sex-matched contests over territory where competitors were either randomly matched or matched for body size (within 5% total body length). We then used a novel application of social network techniques [28,29] to conduct a sequential analysis. We matched the behavioural dynamics from this sequential analysis to predictions of assessment models and identified how specific behaviours fit within the broader progression of contest behaviours. Our results support recent recommendations [11] by showing how testing multiple predictions can substantiate one assessment model. We also reveal how the use of deadly weapons, and other contest behaviours, functions in assessment.

2. Methods

Our dataset from body length-matched contests is a more detailed analysis of contests previously studied in [25]. See the electronic supplementary material for additional methodological details.

(a) Collection and measurement

We collected *N. bredini* individuals from coral rubble burrows in *Thalassia* spp. seagrass habitats on the Atlantic coast of Panamá. In *N. bredini*, burrows are a valuable resource that function as sites for feeding, mating and egg-brooding and as refuges from predation; burrow availability is the main limitation on population size [30]. Individuals that had moulted or were found

with eggs after collection were not used in experiments for at least 7 days (moulted and brooding individuals differ in contest behaviours as compared to the rest of the population; [31–33]). Individuals were housed separately in perforated bottles in flow-through seawater tanks until their use in a contest.

We measured each individual's body length (mm) within one week of collection and body mass (g) either on the morning of (body length-matched contests) or the evening after (randomly matched contests) each individual's use in a contest. We used the mean of three replicate measurements in all analyses.

(b) Contest protocol

Contests were conducted in clear plastic arenas with sandy substrate and a laminated piece of centimetre graph paper on one of the broad sides. We observed both male–male and female–female contests; both sexes compete over burrows and show no differences in contest behaviour [25,34]. Each individual was used in only one contest to avoid effects of prior experience on contest behaviour [35].

We allowed one individual (the 'resident') to establish residency in a single-entrance artificial burrow made of clear vinyl tubing. The burrow was wrapped in black vinyl tape except for a clear area at the top and was matched for resident body length [30]. After several hours (3–6, body length-matched contests; 8–18, randomly matched contests; differences in residency time had no effect on results), we introduced either a randomly matched or a body length-matched competitor (the 'intruder') behind an opaque barrier. After 10 min, we raised the barrier and videotaped the contest using two orthogonally placed GoPro cameras.

If the resident did not establish residency in the burrow before the contest began, it was removed and used as an intruder in another, randomly selected contest; we then selected another individual and allowed it to establish residency in the burrow for a subsequent day's contest. If competitors did not interact with each other for 20 min, we removed both to be selected for use in subsequent days' contests. If one individual interacted, but the other did not, we discarded the contest data and did not use either individual in subsequent contests. We discarded a total of 10 randomly matched (approx. 20% of initiated trials) and seven body length-matched contests (approx. 17% of initiated trials) owing to a lack of interaction or the presence of one-sided interactions.

(c) Analyses

A contest began when individuals first made eye contact or when one individual approached the other; a contest ended when one individual made a clear, directed movement away from its competitor and towards the edge of the contest arena (see ethogram, electronic supplementary material, table S1). The winner was the individual that resided in the burrow after its competitor's retreat.

We conducted separate analyses for body length-matched and randomly matched datasets. We removed four total outliers—two from the randomly matched and two from the body length-matched datasets—that had contest duration greater than three standard deviations from the mean (`uv.outliers` function, [36]). The overall trends were the same with and without inclusion of the outliers (electronic supplementary material, tables S7 and S8). All data were analysed using R v. 3.0.1 [37].

(d) Resource holding potential and resource ownership

We tested between body length and body mass as measures of RHP, following methods outlined by [8]. We randomly selected one 'focal' individual from each contest and created a metric of focal body length or mass relative to opponent body length or

mass. We tested the effects of relative mass, relative length and their interaction on focal individual contest outcome (win or lose) using a binomial generalized linear model (GLM) with a log link function [38]. We ran all possible models including and excluding these terms and identified which models had the lowest Akaike information criterion (AIC) scores (within 2 Δ AIC, [39]). We chose which variable appeared most frequently in the best-supported (lowest AIC) models as our metric of RHP.

After establishing a metric of RHP, we tested how burrow residency interacts with RHP to affect outcome. We first ran a binomial GLM with focal outcome as a function of relative RHP, focal role (resident/intruder) and their interaction, and identified significant terms using *p*-values and *z*-scores from the full model results [40]. We also tested for the benefit, in RHP units, of residency [41,42]. Here, we ran a binomial GLM with the probability of an intruder winning the contest as a function of intruder minus resident RHP. We calculated the RHP difference value at which the intruder had a 50% chance of winning the contest (the inflection point of the binomial function). With no residency advantage, the intruder would have a 50% chance of winning the contest at an intruder minus resident RHP value of 0. With a residency advantage, an intruder would have a 50% chance of winning only if its RHP were greater than the resident's [42].

(e) Correlational tests of assessment models

Following methods established by [9], we tested the effects of RHP and residency on two measures of contest costs: total contest duration (in seconds) and the total number of raptorial appendage strikes exchanged during a contest. Both contest duration and number of strikes were log-transformed [\log_{10} (duration) and $\log_{10}(1 + \text{number of strikes})$] to meet assumptions of normality. We ran two multiple regressions, each with one cost variable as the dependent variable. For randomly matched contests, our independent variables included winner RHP, winner role (resident or intruder), loser RHP, loser role, the interaction of winner and loser RHP and the interaction of winner and loser role (use of interaction terms following [43]). For body length-matched contests, the independent variables were averaged competitor RHP and competitor role. We used the direction, strength and statistical significance of the full model results [40] to differentiate among assessment models (table 1 [2,9]). We visualized our results using simple linear regression, but note that the statistical results of simple linear regression are similar, but not equal to, multiple regression, as simple regression does not take interaction terms into account.

(f) Sequential behavioural analysis

Assessment models differ in their predictions regarding (among other predictions) if behaviours occur in phases and if and how behaviours escalate during contests (table 1 [3,5–7]). We used a sequential analysis to test how *N. bredini* behaviours match these predictions.

We coded all contest behaviours using JWATCHEr software [44], following an ethogram similar to that of Dingle & Caldwell [21]. Our ethogram consisted of 14 mutually exclusive contest behaviours (electronic supplementary material, table S1). We combined behavioural sequences for all body length-matched competitors and, separately, all randomly matched competitors.

Full details on our sequential analysis technique and an R code for its execution are available in the electronic supplementary material. Briefly, we used the igraph network analysis package [28] to summarize our behavioural sequence data into adjacency matrices of all randomly matched and, separately, body length matched, contest behaviours. Each row and column in an adjacency matrix represented one of 14 contest behaviours (14×14 matrix). Each cell in a matrix represented the number

of times, across the dataset, that one behaviour from an individual (row) transitioned to (i.e. was followed immediately by) a subsequent behaviour from that individual (column). Therefore, the adjacency matrices summarized all intra-individual behavioural transitions that occurred in contests.

To identify patterns in *N. bredini* behaviours and match them to assessment model predictions, we isolated which transitions were more frequent than expected by chance using permutation procedures common in sequential behavioural analysis ([15]; see the electronic supplementary material). We resampled the second column of our two-column behavioural transitions dataset 10 000 times, keeping the relative frequency of behaviours but randomizing the transitions between behaviours. After each resampling iteration, we saved the resulting adjacency matrix. This technique resulted in a distribution of the expected number of transitions between behaviours if behaviours transitioned randomly (i.e. a null distribution), constrained only by the relative frequency of each behaviour. We extracted the 95% quantile of the null distribution for each cell (i.e. each transition) in the randomized adjacency matrix and compared our observed values to these quantiles. Any transitions in the observed dataset that were more frequent than their respective 95% null quantile were more common than expected. We term these 'significant' transitions.

After isolating only significant transitions for both randomly matched and body length-matched contests, we plotted the resulting adjacency matrices as network graphs in igraph. Individual behaviours are represented as network vertices (circles), and significant transitions between behaviours are represented as directed network edges (arrows). Vertex size was proportional to five categories of scaled degree—the percentage of total contest behaviours made up by one behavioural state. Edge width was proportional to five categories of transitional probability—the number of times a transition occurred from one behaviour to another divided by the total number of transitions from that behaviour to all other behaviours (higher values are more likely transitions). We used transitional probability values from the original observed dataset (pre-simplification), and present tables including observed values, 95% null quantile values and transitional probability values in the electronic supplementary material, tables S10 and S11.

We used the network graphs to identify the presence or absence of phases following the definitions of [7]. We defined a phase as beginning with the onset of (i.e. a transition to) a new subset of behaviours. Multiple behaviours could occur within a phase, in which case these behaviours should be used with relatively equal frequency. We defined a phase as ending with a transition to a new subset of behaviours (i.e. a new phase), after which previously used behaviours were unlikely to re-occur (i.e. there were few transitions to behaviours from previous phases). If *N. bredini* contests did not progress in phases, we expected to see no subsets of repeatedly used behaviours and that any behaviour could occur at any point within the behavioural sequence [5].

We also used the behavioural sequence data to identify patterns of escalation. Contest behaviours can fall along a continuum from low to high escalation, where high escalation behaviours typically involve costlier physical contact [6]. For example, *Servavaea incana* jumping spiders escalate from no physical contact, to 'sparring', to biting with the chelicerae [10]. Assessment models differ in predicted patterns of escalation: in mutual assessment, behaviours escalate with rare de-escalation, while behaviours in pure self- and cumulative assessment can both escalate and de-escalate (table 1; reviewed in [2,17]). We identified which behaviour(s) involved physical contact and examined whether behaviours progressed from these escalated behaviours to de-escalated behaviours. We caution that defining degree of escalation is a relatively subjective choice by the experimenter—what a human observer defines as 'costly' may be less so to the animal [2].

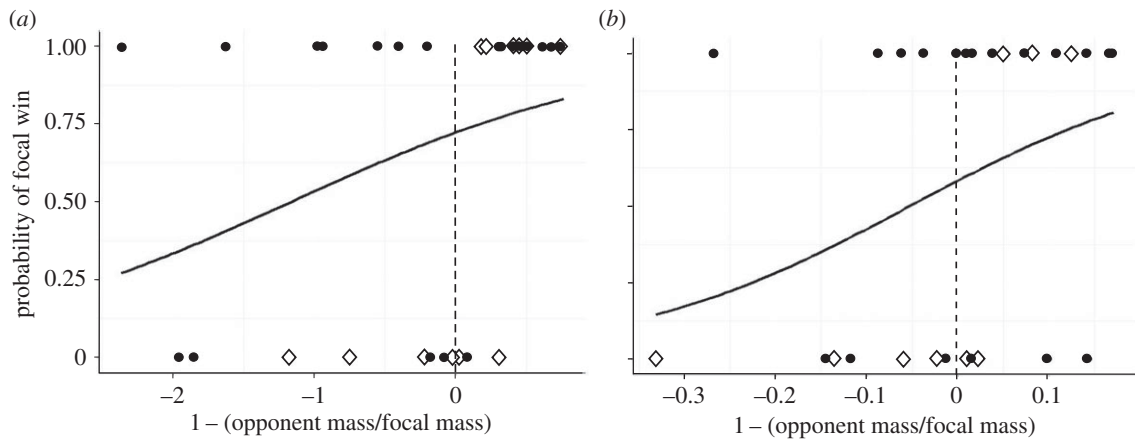


Figure 1. In (a) randomly matched and (b) body length-matched contests, the probability that a focal individual won a contest increased with increasing relative body mass, and residents (closed circles) were more likely to win contests than intruders (open diamonds). Vertical dashed lines in (a) and (b) represent relative mass difference of 0. The solid black line represents the curve fit of a GLM predicting a focal individual's probability of winning from relative body mass, burrow residency and the mass : residency interaction (see Methods). The y -axes are equivalent in (a) and (b).

3. Results

We collected RHP, residency and contest cost data from 35 randomly matched and 29 body length-matched contests. From these, we collected behavioural sequence data from 35 randomly matched and 28 body length-matched contests. In isolation, either the correlational or the sequential analysis results could not conclusively support one assessment model by rejecting the others. The two approaches combined ruled out both pure self- and cumulative assessment and supported mutual assessment in *N. bredini*. Summary statistics of contest variables and behaviours are in the electronic supplementary material, table S9.

(a) Resource holding potential and resource ownership

Body mass occurred most frequently in the best-supported models predicting contest outcome (electronic supplementary material, table S2) and was therefore the best metric of RHP. Using body length as a measure of RHP did not change the results of the correlations between contest costs and RHP (electronic supplementary material, tables S5 and S6).

RHP (body mass) and resource ownership (burrow residency) interacted to predict contest outcome (figure 1). In randomly matched contests, residency ($B = 3.5$, $z_{34} = 2.0$, $p = 0.045$) was a better predictor of contest outcome than mass ($B = 11.0$, $z_{34} = 1.62$, $p = 0.11$) or the mass : residency interaction ($B = -10.3$, $z_{34} = 1.5$, $p = 0.13$). The full GLM for body length-matched contests did not have a high enough sample size to converge (d.f. = 25). By testing models with only individual effects, we found that residency ($B = 1.54$, $z_{27} = 1.79$, $p = 0.07$) had a similar effect on contest outcome as mass ($B = 6.34$, $z_{27} = 1.71$, $p = 0.09$), and both these effects were greater than that of the mass : residency interaction (intruder: $B = 37.5$, $z_{26} = 1.36$, $p = 0.17$, resident: $B = 2.83$, $z = 0.70$, $p = 0.49$).

Intruders needed an average of 0.51 g greater mass (49.6% of mean competitor mass) to overcome the residency advantage in randomly matched contests (binomial GLM: intercept = -1.13 , $B = 2.23$, $z_{33} = 2.56$, $p = 0.01$) and 0.16 g greater mass (9.7% of mean competitor mass) in body length-matched contests (intercept = -0.72 , $B = 4.66$, $z_{27} = 1.67$, $p = 0.09$) (electronic supplementary material, figure S1).

(b) Correlational tests of assessment models

The correlations of randomly matched contest costs and RHP ruled out pure self-assessment, but could not differentiate between mutual and cumulative assessment (full model results: $F_{6,28} = 2.49$, $R^2 = 0.21$, $p = 0.047$; figure 2; electronic supplementary material, table S3). As predicted by all three assessment models, there was a positive correlation between contest duration and the interaction of loser mass and loser residency (figure 2a). We also found a negative correlation between contest duration and winner mass, which supported both mutual and cumulative assessment but not pure self-assessment (figure 2b). Residency allowed eventual losers to compete for longer; this effect increased with increasing body mass (figure 2a). Winner mass alone correlated negatively with contest duration, but this effect was driven by winners who were residents (figure 2b). There was a negative correlation between total number of contest strikes and winner mass, but no correlation with loser mass or residency (full model results: $F_{6,28} = 3.76$, $R^2 = 0.33$, $p = 0.007$; electronic supplementary material, table S3). Losers who were residents exchanged more strikes than losers who were intruders; conversely, winners who were residents exchanged fewer strikes than winners who were intruders, an effect that decreased with increasing body mass (electronic supplementary material, table S3).

The correlations of body length-matched contest costs and RHP ruled out cumulative assessment and supported mutual assessment. There was no relationship between contest duration and average competitor mass, which matches the predictions of mutual assessment but not cumulative assessment ($F_{2,26} = 2.12$, $R^2 = 0.07$, $p = 0.14$; figure 2c; electronic supplementary material, table S4). Contests were shorter when losers were intruders and winners were residents. There was also no correlation between total number of strikes and average competitor mass, but the total number of strikes did correlate with residency ($F_{2,26} = 6.04$, $R^2 = 0.27$, $p < 0.01$; electronic supplementary material, table S4): competitors exchanged fewer strikes when losers were intruders and winners were residents. Importantly, these results ruled out cumulative assessment according to the criteria of Arnott & Elwood [2], but not those of Fawcett & Mowles [45] (see Discussion). Therefore, we could not fully reject cumulative assessment using correlations alone.

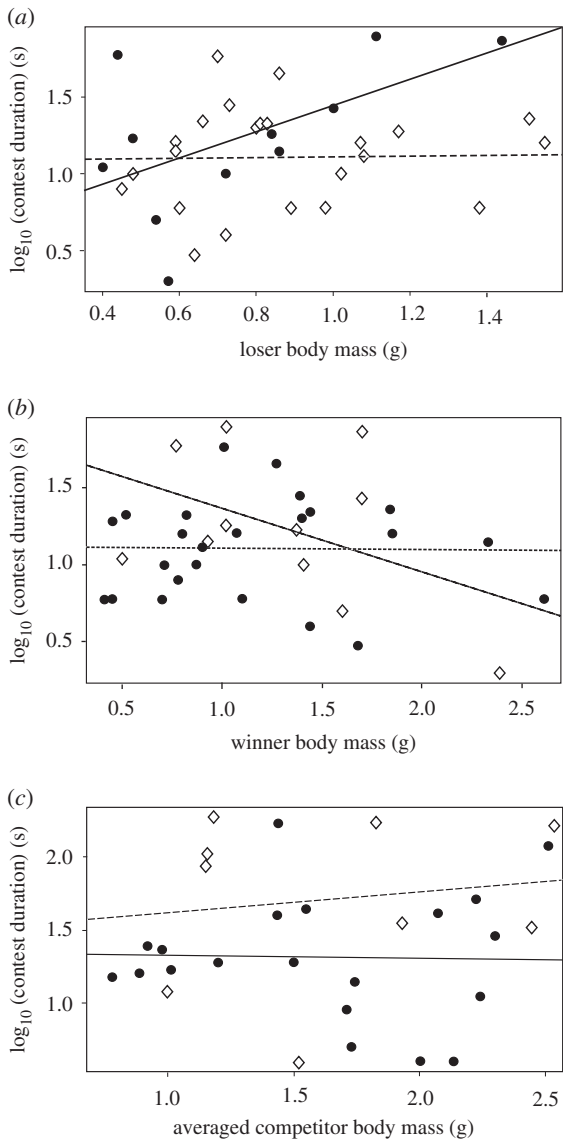


Figure 2. Correlations of log-corrected contest duration as predicted by log-corrected (a) loser body mass and (b) winner body mass in randomly matched contests ruled out pure self-assessment, but could not differentiate between cumulative and mutual assessment. The correlation of contest duration as predicted by log-corrected (c) averaged competitor body mass in body length-matched contests ruled out cumulative assessment according to [2]. Correlation lines depict differences between residents (closed circles, solid lines) and intruders (open diamonds, dashed lines; icons and lines in c represent winner roles only), but do not necessarily depict statistical relationships (statistical results from multiple regressions are reported in results and the electronic supplementary material, tables S3 and S4).

(c) Sequential behavioural analysis

The sequential analysis revealed behavioural progressions that matched the predictions of mutual assessment, supporting the correlational results. In both randomly matched and body length-matched contests, behaviours progressed in phases from (1) visual tracking or approaching, to (2) visual meral spread displays or chemosensory antennular flicking, to (3) ritualized striking via telson sparring and finally to (4) contest resolution (figure 3). The frequency of use of behaviours within a phase was relatively similar; for example, the meral spread and antennular flick behaviours each represented 0–5% and 5.1–10% of total contest behaviours in body length-matched and randomly matched contests, respectively (figure 3). Competitors were also likely to

transition among behaviours within a phase (e.g. between antennular flick and meral spread in body length-matched contests; among lunge, strike, coil in randomly matched and body length-matched contests); however, we found few significant transitions from behaviours in one phase to behaviours in previously occurring phases (figure 3; electronic supplementary material, tables S10 and S11).

Contests showed evidence of escalation to physical contact via telson sparring, specifically the ‘strike’ behaviour. Strikes represented 7.9% and 16.7% of total contest behaviours in randomly matched and body length-matched contests, respectively. Once contests escalated to sparring, we saw little evidence of de-escalation to behaviours without physical contact; instead, contests frequently transitioned within the sparring phase or to contest resolution via a retreat (figure 3; electronic supplementary material, tables S10 and S11). This escalation to physical contact (sparring) with rare de-escalation does not match predictions of pure self- or cumulative assessment and supports those of mutual assessment. We saw the winner stab the loser in two body length-matched contests, only as the loser was already retreating (see also [25]). We saw no overt evidence of injury owing to striking.

The sequential analysis also revealed how the use of behaviours varied within and between contest types. Competitors almost doubled their use of sparring (lunge, strike, coil) in body length-matched (64.1% of total contest behaviours) as compared to randomly matched contests (37.5%). Within both body length-matched and randomly matched contests, sparring usually began with one competitor’s lunge and was strongly predicted by an antennular flick or meral spread behaviour (electronic supplementary material, tables S10 and S11). One sparring behaviour, the telson coil, was a predictor of a retreat in both contest types (electronic supplementary material, tables S10 and S11).

4. Discussion

Using correlational tests of contest costs as a function of RHP and residency in randomly matched and body length-matched contests, we ruled out pure self-assessment and cumulative assessment according to [2]; however, owing to disagreement in the literature [45], we could not fully differentiate between cumulative and mutual assessment based solely on correlations. By incorporating a sequential analysis, we identified behavioural phases and escalation without de-escalation, which supported mutual assessment—specifically, the sequential assessment model [6,7]—as the most likely assessment strategy used by *N. bredini*. The sequential analysis also found that telson sparring plays a prominent role in contest assessment and resolution. Our results reveal that animals with deadly weapons can use them to assess relative ability and resolve conflict safely. We show that complementary analyses of contest dynamics enable strong inference of assessment models. Furthermore, our use of sequential analysis identifies how individual behaviours—such as telson sparring—fit within broader patterns of assessment.

(a) Resource holding potential and resource ownership

In both randomly matched and body length-matched contests, body mass predicted contest outcome (figure 1). Body mass may be a particularly relevant measure of RHP in

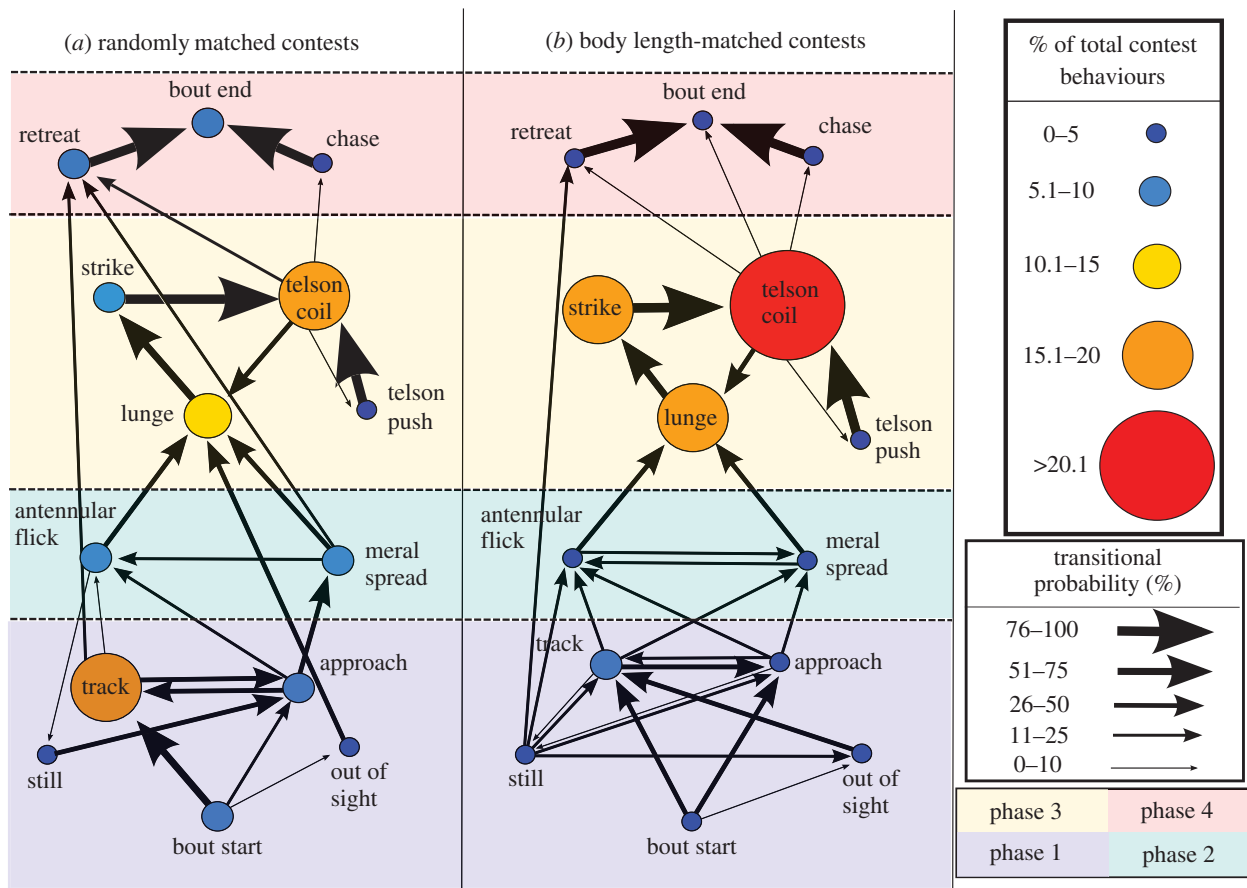


Figure 3. Sequential analysis of contest behaviours in (a) randomly matched and (b) body length-matched contests ruled out pure self- and cumulative assessment and supported mutual assessment. Contest behaviours progressed in phases that we identified following [7]: track or approach (phase 1), antennular flick or meral spread (phase 2), telson sparring (phase 3) and contest resolution (phase 4). Transitions to behaviours from previous phases were unlikely, as was de-escalation from behaviours involving physical contact (phase 3) to behaviours without contact (phases 1 and 2). Circles represent behaviours; circle size and colour is scaled to the percentage of total contest behaviours. Arrows represent significant transitions between behaviours; arrow width is scaled to transitional probability. Phases are colour-coded for visualization purposes. See Methods for details.

N. bredini contests, especially in the context of telson sparring. In a closely related species, *Neogonodactylus wemmerae*, the telson (but not nearby abdominal exoskeleton) dissipates the energy of a strike proportionally to body mass—more massive individuals have telsons that dissipate a greater amount of energy [46]. Telson sparring was common in both randomly matched and body length-matched *N. bredini* contests (figure 3). Competitors may use telson sparring to assess body mass as a metric of RHP, in addition to reducing the costs of receiving strikes. Tests of the biomechanics of telson sparring may further elucidate the mechanical role of sparring in assessment, if and how competitors vary strike performance during sparring, and how the *N. bredini* telson withstands sparring strikes.

Burrow residency also played a key role in determining contest outcome. While the effect of residency was greater than that of body mass, the degree of this effect differed between randomly matched and body length-matched contests. In randomly matched contests, residents had an effective RHP advantage equal to 49.6% greater than the average competitor RHP; this advantage was only 9.7% in body length-matched contests. While differences in experimental design did not appear to affect contest dynamics (see the electronic supplementary material), factors such as the possibility of allometric scaling of RHP with body size [47] could explain some of the differences we found in residency effects. Future work could identify the multiple

factors that contribute to RHP, their relationships with body size and their effects on contest outcome and dynamics.

(b) Tests of assessment models

We found a negative correlation between winner RHP and contest costs in randomly matched contests, which allowed us to rule out pure self-assessment, but this correlation could not differentiate between cumulative and mutual assessment [2,9]. By finding that average competitor RHP was not correlated with contest costs in body length-matched contests, we rejected cumulative assessment and supported mutual assessment according to Arnott & Elwood [2]. However, Fawcett & Mowles [45] recently suggested that both mutual and cumulative assessment should predict no relationship between costs and averaged competitor RHP. While our correlational tests suggested mutual assessment as the most likely model by rejecting alternatives, discrepancies in the literature precluded fully supporting a single assessment model using correlational data alone.

The sequential analysis identified behavioural phases and escalation with rare de-escalation, matching the predictions of mutual assessment [6,7]. We defined phases and escalation following Enquist & Leimar ([7]; see Methods): phases are subsets of behaviours that occur with relatively equal frequency and once a new phase begins, behaviours from previous phases are unlikely to re-occur; escalated

behaviours are those that involve presumably costlier physical contact. In *N. bredini*, contest behaviours progressed in subsets (i.e. phases) from (1) eye tracking and approaching, to (2) visual or chemosensory behaviours, to (3) telson sparring and finally to (4) contest resolution (figure 3). Competitors transitioned among behaviours within a phase with relatively equal frequency. Once competitors transitioned to a new phase, they were more likely to transition within that phase or to progress to the next phase than they were to transition to behaviours that occurred previously. Competitors occasionally ‘skipped’ phases; however, these transitions were rarely more common than expected, except for when competitors transitioned directly to a retreat behaviour (figure 3; electronic supplementary material, tables S10 and S11). Competitors also escalated from behaviours that did not involve physical contact (phases 1 and 2) to behaviours that did involve contact (phase 3). We found no evidence of de-escalation; that is, there were no significant transitions from phases involving physical contact to those without physical contact. Finally, competitors in body length-matched contests used sparring behaviours more frequently than those in randomly matched contests, which matches a prediction of the sequential assessment model of mutual assessment [6,7]: closely matched contestants should use costlier behaviours to determine RHP differences. Overall, our behavioural results ruled out pure self- and cumulative assessment and supported mutual assessment in *N. bredini*.

In addition to supporting mutual assessment as an overall strategy, the sequential analysis points towards future work testing if and how assessment strategies change within a contest. For example, within the phase of telson sparring, competitors may show behavioural escalation (e.g. to striking) and de-escalation (e.g. to telson coiling). Therefore, even though the overall contest dynamics support mutual assessment (as do correlations using the number of strikes as a cost variable, see Results), competitors may use cumulative assessment while sparring. While this question is beyond the scope of the present study, recent models that allow for assessment strategy switching (e.g. [48,49]) could generate predictions for behavioural progressions to be empirically tested using sequential analysis.

References

- Parker GA. 1974 Assessment strategy and the evolution of fighting behavior. *J. Theor. Biol.* **47**, 223–243. (doi:10.1016/0022-5193(74)90111-8)
- Arnott G, Elwood RW. 2009 Assessment of fighting ability in animal contests. *Anim. Behav.* **77**, 991–1004. (doi:10.1016/j.anbehav.2009.02.010)
- Payne RJH, Pagel M. 1996 Escalation and time costs in displays of endurance. *J. Theor. Biol.* **183**, 185–193. (doi:10.1006/jtbi.1996.0212)
- Mesterton-Gibbons M, Marden JH, Dugatkin LA. 1996 On wars of attrition without assessment. *J. Theor. Biol.* **181**, 65–83. (doi:10.1006/jtbi.1996.0115)
- Payne RJH. 1988 Gradually escalating fights and displays: the cumulative assessment model. *Anim. Behav.* **56**, 651–662. (doi:10.1006/anbe.1998.0835)
- Enquist M, Leimar O. 1983 Evolution of fighting behaviour: decision rules and assessment of relative strength. *J. Theor. Biol.* **102**, 387–410. (doi:10.1016/0022-5193(83)90376-4)
- Enquist M, Leimar O, Ljungberg T, Mallner Y, Segerdahl N. 1990 A test of the sequential assessment game: fighting in the cichlid fish *Nannacara anomala*. *Anim. Behav.* **40**, 1–14. (doi:10.1016/S0003-3472(05)80660-8)
- Briffa M, Hardy ICW, Gammell MP, Jennings DJ, Clarke DD, Goubault M. 2013 Analysis of animal contest data. In *Animal contests* (eds ICW Hardy, M Briffa), pp. 47–85. Cambridge, UK: Cambridge University Press.
- Taylor PW, Elwood RW. 2003 The mismeasure of animal contests. *Anim. Behav.* **65**, 1195–1202. (doi:10.1006/anbe.2003.2169)
- McGinley RH, Prenter J, Taylor PW. 2015 Assessment strategies and decision making in male–male contests of *Servaea incana* jumping spiders. *Anim. Behav.* **101**, 89–95. (doi:10.1016/j.anbehav.2014.12.014)
- Briffa M, Elwood RW. 2009 Difficulties remain in distinguishing between mutual and self-assessment in animal contests. *Anim. Behav.* **77**, 759–762. (doi:10.1016/j.anbehav.2008.11.010)
- Briffa M. 2008 Decisions during fights in the house cricket, *Acheta domestica*: mutual or self assessment of energy, weapons and size? *Anim.*

5. Conclusion

The field of assessment has had a strong theoretical foundation since its first game-theory models were established [50]. However, inconclusive experimental results have sometimes led to uncertainty as to whether current theory is sufficient to explain assessment in many species [47–49,51], or whether experiments need to incorporate a broader array of approaches [11,52]. We found that, in *N. bredini*, a combination of correlational and sequential analyses ruled out alternative models and supported mutual assessment [6,7]. Our analyses resolve the frequently encountered challenge in contest studies, which is that correlational and sequential approaches on their own are often inconclusive: correlations cannot easily differentiate between cumulative and mutual assessment, nor can sequential analysis clearly distinguish between pure self- and cumulative assessment. Our study gives a pathway for strong empirical tests of both current and future assessment models.

Ethics. All methods complied with legal requirements of the Panamanian Ministerio de Ambiente and of Duke University. After the conclusion of all trials in Panama, some animals were released at the same sites where they were collected, while others were transported to Duke University for use in other research (ANAM export permits no.SEX/A-23-14 and no.SEX/A-106-15).

Data accessibility. The datasets and R code supporting this article have been uploaded as part of the electronic supplementary material.

Authors’ contributions. P.A.G. conceived of the study and collected and analysed the data. P.A.G. and S.N.P. wrote the paper and gave final approval for publication.

Competing interests. We have no competing interests.

Funding. This study was supported by a National Science Foundation grant to S.N.P. (NSF IOS 1149748), and a STRI Short-Term Fellowship, a Duke University Biology Fellowship, Grant in Aid of Research, and Summer Research Fellowship, and a University of Massachusetts, Amherst, Natural History Collections grant to P.A.G.

Acknowledgements. We thank R. Cohn and C. Schanche-Perret Gentil for help collecting behavioural data and in initial analyses. C. Kuo, R. Crane, M. deVries and the staff at STRI helped in the collection and transport of mantis shrimp, and in facilitating the study. M. Briffa, J. Christy, R. Fitak, M. Hughes, S. Johnsen, R. McGinley, S. Nowicki, A.V. Palaoro, D. Pfennig, V.L. Roth, D. Shizuka, and three anonymous reviewers gave helpful feedback on design, analysis and writing.

- Behav.* **75**, 1053–1062. (doi:10.1016/j.anbehav.2007.08.016)
13. Jennings DJ, Gammell MP, Payne RJH, Hayden TJ. 2005 An investigation of assessment games during fallow deer fights. *Ethology* **111**, 511–525. (doi:10.1111/j.1439-0310.2005.01068.x)
 14. Bakeman R, Quera V. 1995 Log-linear approaches to lag-sequential analysis when consecutive codes can and cannot repeat. *Psychol. Bull.* **118**, 272–284. (doi:10.1037/0033-2909.118.2.272)
 15. Bakeman R, Robinson BF, Quera V. 1996 Testing sequential association: estimating exact *p* values using sampled permutations. *Psychol. Methods* **1**, 4–15. (doi:10.1037/1082-989X.1.1.4)
 16. Bakeman R, Quera V. 2011 *Sequential analysis and observational methods for the behavioral sciences*. Cambridge, UK: Cambridge University Press.
 17. Egge AR, Brandt Y, Swallow JG. 2010 Sequential analysis of aggressive interactions in the stalk-eyed fly *Teleopsis dalmanni*. *Behav. Ecol. Sociobiol.* **65**, 369–379. (doi:10.1007/s00265-010-1054-5)
 18. Chen S, Lee AY, Bowens NM, Huber R, Kravitz EA. 2002 Fighting fruit flies: a model system for the study of aggression. *Proc. Natl Acad. Sci. USA* **99**, 5664–5668. (doi:10.1073/pnas.082102599)
 19. Clutton-Brock T, Albon SD, Gibson RM, Guinness FE. 1979 The logical stag: adaptive aspects of fighting in red deer (*Cervus elaphus* L.). *Anim. Behav.* **27**, 211–225. (doi:10.1016/0003-3472(79)90141-6)
 20. Caldwell RL. 1987 Assessment strategies in stomatopods. *Bull. Mar. Sci.* **41**, 135–150.
 21. Dingle H, Caldwell RL. 1969 The aggressive and territorial behaviour of the mantis shrimp *Gonodactylus bredini* Manning (Crustacea: Stomatopoda). *Behaviour* **33**, 115–136. (doi:10.1163/156853969X00341)
 22. McHenry MJ, Anderson PS, VanWassenbergh S., Matthews DG, Summers AP, Patek SN. 2016 The comparative hydrodynamics of rapid rotation by predatory appendages. *J. Exp. Biol.* **219**, 3399–3411. (doi:10.1242/jeb.140590)
 23. Patek SN, Caldwell RL. 2005 Extreme impact and cavitation forces of a biological hammer: strike forces of the peacock mantis shrimp *Odontodactylus scyllarus*. *J. Exp. Biol.* **208**, 3655–3664. (doi:10.1242/jeb.01831)
 24. Berzins IK, Caldwell RL. 1983 The effect of injury on the agonistic behavior of the stomatopod, *Gonodactylus bredini* (Manning). *Mar. Behav. Physiol.* **10**, 83–96. (doi:10.1080/10236248309378609)
 25. Green PA, Patek SN. 2015 Contests with deadly weapons: telson sparring in mantis shrimp (Stomatopoda). *Biol. Lett.* **11**, 20150558. (doi:10.1098/rsbl.2015.0558)
 26. Vetter KM, Caldwell RL. 2015 Individual recognition in stomatopods. In *Social recognition in invertebrates: the knowns and the unknowns* (ed. LT Aquiloni), pp. 17–36. Cham, Switzerland: Springer.
 27. Dingle H. 1969 A statistical and information analysis of aggressive communication in mantis shrimp *Gonodactylus bredini* Manning. *Anim. Behav.* **17**, 561–575. (doi:10.1016/0003-3472(69)90165-1)
 28. Csardi G, Nepusz T. 2006 The igraph software package for complex network research. *InterJournal, Complex System* 1695.
 29. Kolaczyk ED, Csardi G. 2014 *Statistical analysis of network data with R*. New York, NY: Springer.
 30. Steger R. 1987 Effects of refuges and recruitment on gonodactylid stomatopods, a guild of mobile prey. *Ecology* **68**, 1520–1533. (doi:10.2307/1939236)
 31. Adams ES, Caldwell RL. 1990 Deceptive communication in asymmetric fights of the stomatopod crustacean *Gonodactylus bredini*. *Anim. Behav.* **39**, 706–716. (doi:10.1016/S0003-3472(05)80382-3)
 32. Montgomery EL, Caldwell RL. 1984 Aggressive brood defense by females in the stomatopod *Gonodactylus bredini*. *Behav. Ecol. Sociobiol.* **14**, 247–251. (doi:10.1007/BF00299495)
 33. Steger R, Caldwell RL. 1983 Intraspecific deception by bluffing: a defense strategy of newly molted stomatopods (Arthropoda: Crustacea). *Science* **221**, 558–560. (doi:10.1126/science.221.4610.558)
 34. Caldwell RL, Dingle H. 1975 Ecology and evolution of agonistic behavior in stomatopods. *Naturwissenschaften* **62**, 214–222. (doi:10.1007/BF00603166)
 35. Caldwell RL. 1979 Cavity occupation and defensive behaviour in the stomatopod *Gonodactylus festai*: evidence for chemically mediated individual recognition. *Anim. Behav.* **27**, 194–201. (doi:10.1016/0003-3472(79)90139-8)
 36. McGarigal K. 2013 BIOMSTATS R package. See https://www.umass.edu/landeco/teaching/ecodata/labs/ecodata_labs.html.
 37. R Core Team. 2008 *R: a language and environment for statistical computing*, 3.0.1 edn. Vienna, Austria: R Foundation for Statistical Computing.
 38. Crawley MJ. 2007 *The R book*. New York, NY: Wiley.
 39. Burnham KP, Anderson DR, Huyvaert KP. 2010 AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behav. Ecol. Sociobiol.* **65**, 23–35. (doi:10.1007/s00265-010-1029-6)
 40. Mundry R, Nunn CL. 2009 Stepwise model fitting and statistical inference: turning noise into signal pollution. *Am. Nat.* **173**, 119–123. (doi:10.1086/593303)
 41. Kasumovic MM, Mason AC, Andrade MCB, Elias DO. 2010 The relative importance of RHP and resource quality in contests with ownership asymmetries. *Behav. Ecol.* **22**, 39–45. (doi:10.1093/beheco/arq161)
 42. Kokko H, Lopez-Sepulcre A, Morrell LJ. 2006 From hawks and doves to self-consistent games of territorial behavior. *Am. Nat.* **167**, 901–912.
 43. Elias DO, Kasumovic MM, Punzalan D., Andrade MC, Mason AC. 2008 Assessment during aggressive contests between male jumping spiders. *Anim. Behav.* **76**, 901–910. (doi:10.1016/j.anbehav.2008.01.032)
 44. Blumstein D, Daniel J, Evans C. 2007 JWatcher. See <http://www.jwatcher.ucla.edu/>.
 45. Fawcett TW, Mowles SL. 2013 Assessments of fighting ability need not be cognitively complex. *Anim. Behav.* **86**, e1–e7. (doi:10.1016/j.anbehav.2013.05.033)
 46. Taylor JR, Patek SN. 2010 Ritualized fighting and biological armor: the impact mechanics of the mantis shrimp's telson. *J. Exp. Biol.* **213**, 3496–3504. (doi:10.1242/jeb.047233)
 47. Palaoro AV, Briffa M. 2017 Weaponry and defenses in fighting animals: how allometry can alter predictions from contest theory. *Behav. Ecol.* **28**, 328–336. (doi:10.1093/beheco/arw163)
 48. Mesterton-Gibbons M, Heap SM. 2014 Variation between self- and mutual assessment in animal contests. *Am. Nat.* **183**, 199–213. (doi:10.1086/674443)
 49. Prenter J, Elwood RW, Taylor PW. 2006 Self-assessment by males during energetically costly contests over precopula females in amphipods. *Anim. Behav.* **72**, 861–868. (doi:10.1016/j.anbehav.2006.01.023)
 50. Maynard SJ, Price GR. 1973 The logic of animal conflict. *Nature* **246**, 15–18. (doi:10.1038/246015a0)
 51. Lane SM, Briffa M. 2017 The price of attack: rethinking damage costs in animal contests. *Anim. Behav.* **126**, 23–29. (doi:10.1016/j.anbehav.2017.01.015)
 52. Vieira MC, Peixoto PEC. 2013 Winners and losers: a meta-analysis of functional determinants of fighting ability in arthropod contests. *Funct. Ecol.* **27**, 305–313. (doi:10.1111/1365-2435.12051)