The Behavior and Energetics of Ritualized Weapon Use in Mantis Shrimp

(Stomatopoda)

by

Patrick Andrew Green

Department of Biology
Duke University

Date:_______________________
Approved:

___________________________
Sheila Patek, Supervisor

___________________________
Sönke Johnsen

___________________________
Stephen Nowicki

___________________________
David Pfennig

___________________________
V. Louise Roth

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of
Biology in the Graduate School
of Duke University

2018
ABSTRACT

The Behavior and Energetics of Ritualized Weapon Use in Mantis Shrimp

(Stomatopoda)

by

Patrick Andrew Green

Department of Biology
Duke University

Date:_______________________
Approved:

___________________________
Sheila Patek, Supervisor

___________________________
Sönke Johnsen

___________________________
Stephen Nowicki

___________________________
David Pfennig

___________________________
V. Louise Roth

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biology in the Graduate School of Duke University
2018
Abstract

Contests are essential parts of an animal’s life history, as they dictate access to critical resources like mates, food, or territory. Studying how animals efficiently assess competitive ability to resolve contests is a central goal of research in animal behavior. Additionally, studies of how animals use traits like signals and weapons in contests lends insight to the evolution of those traits. In this thesis, I study assessment and resolution of territorial contests – as well as the function of signals and weapons in contests – in the mantis shrimp *Neogonodactylus bredini* (Stomatopoda: Crustacea).

Behavioral theory predicts that animals may use visual or other displays to communicate reliable information on ability, resolving contests without the use of potentially costlier combat, such as biting, grasping, or striking with weapons. In Chapter 2, I show that *N. bredini* do not match these predictions – the size of structures presented during visual weapon displays did not correlate with strike performance, and almost all contests involved weapon use *via* high-force striking. Because most strikes were exchanged on the armored telson (tailplate), I hypothesized that the ritualized “telson sparring” behavior helps competitors avoid contest costs and functions as a signal, instead of dangerous combat.

Studies of assessment help show what information competitors use to make decisions during contests and can reveal the role specific behaviors play. In Chapter 3, I
show that *N. bredini* use mutual assessment during both size-matched and non size-matched contests; that is, competitors gather information about both themselves and their opponent. I also show the role telson sparring and other behaviors play during this assessment.

Finally, in Chapter 4, I test how the energetic cost of delivering sparring strikes scales with body size. I find that larger competitors used proportionally more energy when striking, that this positive scaling of energy resulted in constant scaling of velocity across size, and that these results matched predictions from a mathematical model of the strike mechanism. Furthermore, I show that these scaling dynamics are different from those of strikes delivered in another behavioral context: feeding on hard-shelled prey.

Overall, this thesis shows that the use of deadly weapons in contests should not be assumed as dangerous combat; instead, I show how ritualized behaviors allow for weapon use to function in assessment. The approaches used and conclusions made from this thesis can inform work in contest behavior, functional morphology, and biomechanics.
Dedication

“The power which resides in him is new in nature, and none but he knows what that is which he can do, nor does he know until he has tried.”

Emerson, Self-Reliance
Contents

Abstract .................................................................................................................................................. iv
Dedication ............................................................................................................................................... vi
List of Tables ......................................................................................................................................... xi
List of Figures ........................................................................................................................................ xiii
Acknowledgements ............................................................................................................................. xiv
1. Introduction .......................................................................................................................................... 1
2. Contests with deadly weapons: telson sparring in mantis shrimp (Stomatopoda) ...... 6
   2.1 Introduction ...................................................................................................................................... 6
   2.2 Methods .......................................................................................................................................... 8
   2.3 Results .......................................................................................................................................... 10
   2.4 Discussion ..................................................................................................................................... 14
3. Mutual assessment during ritualised fighting in mantis shrimp (Stomatopoda) ....... 16
   3.1 Introduction ...................................................................................................................................... 16
   3.2 Methods .......................................................................................................................................... 24
      3.2.1 Collection and Measurement ................................................................................................. 24
      3.2.2 Contest Protocol ...................................................................................................................... 25
      3.2.3 Analyses .................................................................................................................................. 26
      3.2.4 RHP and Resource Ownership ................................................................................................. 27
      3.2.5 Correlational Tests of Assessment Models .............................................................................. 28
      3.2.6 Sequential Behavioural Analysis .............................................................................................. 29
3.3 Results .......................................................................................................................... 32
  3.3.1 RHP and Resource Ownership ................................................................................. 32
  3.3.2 Correlational Tests of Assessment Models ............................................................... 34
  3.3.3 Sequential Behavioural Analysis .............................................................................. 36
3.4 Discussion ..................................................................................................................... 39
  3.4.1 RHP and Resource Ownership ................................................................................. 40
  3.4.2 Tests of Assessment Models .................................................................................... 41
  3.4.3 Conclusions ............................................................................................................ 44
4. Context-Dependent Scaling of Weapon Energetics and Kinematics in Mantis Shrimp (Stomatopoda) ........................................................................................................... 45
  4.1 Introduction .................................................................................................................. 45
  4.2 Methods ....................................................................................................................... 49
    4.2.1 Simulations ............................................................................................................ 49
    4.2.2 Experimental Measurements ................................................................................ 51
      4.2.2.1 Animal Collection and Care .......................................................................... 51
      4.2.2.2 Filming Protocol ........................................................................................... 52
      4.2.2.3 Morphological Measurements ...................................................................... 54
      4.2.2.4 Digitizing, Kinematic, and Energetic Analyses ........................................... 57
    4.2.3. Statistical Analyses ............................................................................................. 60
  4.3 Results .......................................................................................................................... 63
    4.3.1 Simulation Data ...................................................................................................... 66
    4.3.2 Experimental Data ................................................................................................. 66
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.3.2.1 Energetics</td>
<td>66</td>
</tr>
<tr>
<td>4.3.2.2 Kinematics</td>
<td>66</td>
</tr>
<tr>
<td>4.3.3 Matching Simulation and Experimental Datasets</td>
<td>68</td>
</tr>
<tr>
<td>4.3.4 Post-hoc test: The effects of relative size on sparring energetics and kinematics</td>
<td>69</td>
</tr>
<tr>
<td>4.4 Discussion</td>
<td>72</td>
</tr>
<tr>
<td>4.4.1 Sparring</td>
<td>73</td>
</tr>
<tr>
<td>4.4.2 Feeding</td>
<td>74</td>
</tr>
<tr>
<td>4.4.3 Simulations</td>
<td>75</td>
</tr>
<tr>
<td>4.4.4 Measuring animal performance in context</td>
<td>76</td>
</tr>
<tr>
<td>4.4.5 Conclusion</td>
<td>77</td>
</tr>
<tr>
<td>Chapter 5 – Conclusions</td>
<td>78</td>
</tr>
<tr>
<td>Appendix A – Supplementary Methods, Figures, and Tables for Ch. 2.</td>
<td>80</td>
</tr>
<tr>
<td>A.1 Methods</td>
<td>80</td>
</tr>
<tr>
<td>A.1.1 Raptorial appendage morphology</td>
<td>80</td>
</tr>
<tr>
<td>A.1.2 Body length, body mass, and maximum strike force measurements</td>
<td>81</td>
</tr>
<tr>
<td>A.1.3 Contest behaviour</td>
<td>82</td>
</tr>
<tr>
<td>A.1.4 Statistical analyses</td>
<td>84</td>
</tr>
<tr>
<td>A.1.4.1 Morphology and force correlations</td>
<td>84</td>
</tr>
<tr>
<td>A.1.4.2 Contest behaviour and correlates of contest success</td>
<td>86</td>
</tr>
<tr>
<td>A.2 Results</td>
<td>88</td>
</tr>
<tr>
<td>A.2.1 GLM and residual correlation results</td>
<td>88</td>
</tr>
</tbody>
</table>
Appendix B: Supplementary Methods, Tables, and Figures for Chapter 3

B.1 Methods

B.1.1 Animal Collection and Maintenance

B.1.2 Morphological Measurements

B.1.3 Contest Protocol

B.1.3.1 Randomly-matched Contest Protocol

B.1.3.2 Body length-matched contest protocol

B.1.3.3 Differences between randomly-matched and body length-matched contest protocols

B.1.4 Contest Data

B.1.5 Analysis Techniques

B.1.5.1 RHP Analysis

B.1.5.2 Effects of RHP and Residency on Contest Outcome

B.1.5.3 Correlational tests of assessment models

B.1.5.6 Sequential Behavioural Analysis

B.1.6 Ethics Note

B.2 Supplementary Tables and Figures

References

Biography
List of Tables

Table 1: Winners (W) of contests do not have greater strike force than losers (L); instead, winners strike a greater number of times during contests. ................................................................. 13

Table 2: Assessment models differ in the function of behaviours and determinants of outcome. ......................................................................................................................... 18

Table 3: Summary statistics. ......................................................................................................................... 63

Table 4: Summary of statistical results ........................................................................................................... 64

Table 5: Residual values from LMMs testing the effect of relative competitor size on the scaling of sparring strike potential energy and maximum angular velocity. ................. 71

Table 6: Appendage morphology does not strongly correlate with maximum strike force. ........................................................................................................................................... 89

Table 7: Ethogram of contest behaviours. All behaviours are mutually exclusive. Further definitions are provided in Methods ........................................................................................................... 107

Table 8: In randomly-matched contests, body mass was a better predictor of contest outcome than body length (BL) .................................................................................................................. 108

Table 9: Pure self-assessment was ruled out in randomly-matched contests ......................... 109

Table 10: Cumulative assessment was ruled out according to the criteria of (Arnott & Elwood, 2009) but not (Fawcett & Mowles, 2013), in body length-matched contests.....110

Table 11: In randomly-matched contests, correlations of contest costs as a function of body length and residency ........................................................................................................ 111

Table 12: In body length-matched contests, correlations of contest costs as a function of body length and residency ........................................................................................................ 112

Table 13: Previously-removed outliers were added to the dataset and the effects of RHP and residency on randomly-matched contest costs were re-analyzed here ...................... 113

Table 14: Previously-removed outliers were added to the dataset and the effects of RHP and residency on body length-matched contest costs were re-analyzed.............................. 114
Table 15: Summary statistics ........................................................... 115

Table 16: Full transitional data matrix for randomly-matched contests. .................. 116

Table 17: Full transitional data matrix for body length-matched contests. ................. 118
List of Figures

Figure 1: During aggressive contests, mantis shrimp often present meral spread displays (A)........................................................................................................................................7

Figure 2: Winners of contests do not have greater strike force than losers; instead, winners strike a greater number of times during contests.................................................................12

Figure 3: In (a) randomly-matched and (b) body length-matched contests, the likelihood that a focal individual won a contest increased with increasing relative body mass ......33

Figure 4: Correlations of log-corrected contest duration as predicted by log-corrected (a) loser body mass and (b) winner body mass...........................................................35

Figure 5: Sequential analysis of contest behaviours in (a) randomly-matched and (b) body length-matched contests .....................................................................................................38

Figure 6: Summary figure of methods; see main text for full details. ..............................................56

Figure 7: Experimentally-measured strike potential energy exhibits distinct scaling relationships across behavioral contexts ..............................................................................................67

Figure 8: Strike maximum angular velocity exhibits different scaling relationships between behavioral contexts..................................................................................................................68

Figure 9: Residuals of maximum strike force on body length show a positive correlation with residuals of saddle length.................................................................90

Figure 10: To have an equal chance of winning a contest.................................................................121

Figure 11: The distribution of contest costs for randomly-matched competitors within 5% of total body length matches closely to that of body length-matched competitors .......122
Acknowledgements

My dissertation has been supported by people who believe in me, at many times more strongly than I believe in myself.

Sheila Patek, my advisor, patiently guided me through innumerable ideas, hypotheses, and changes of interest. She challenged me to avoid complacency and to try just a little bit harder, both because she knew I could and because that’s what good scientists do. She is a role model for me in how hard she works and how passionately she sticks to her convictions. I’m very proud to have been her student.

A number of faculty have mentored me throughout my career. Jeff Podos at UMass, Amherst showed me how to have fun while doing good work and remains my template for writing style. My thesis committee at Duke – Sheila Patek, David Pfennig, V. Louise Roth, Sönke Johnsen, and Steve Nowicki – have given sage advice and timely support. Faculty at Duke, especially Mohamed Noor, Sönke Johnsen, and Steve Nowicki, have shown a trust and faith in me that I appreciate immensely. Mohamed has supported every professional idea I’ve thought up and has taken me in on his own ventures. Sönke has shown me how to take care of my personal life in what often seems like an overwhelming career. Steve took me under his wing and has taught me to be a careful, thorough behaviorist; he and Susan Peters model what I hope to be as a member of a scientific partnership.
Grad students at UMass, Amherst and in Duke Biology have been my best friends during my dissertation. They’ve supported me through the trials of my mid- to late-20s and have been sources of incredible fun. I want to thank members of the Patek Lab for their advice and collaboration. Ryan Campbell and Matthew Zipple have been constant companions, and I am incredibly grateful for their friendship.

My family has not always understood what I am doing, but they’ve never doubted my ability to do it. My grandparents taught me how to live my life with dignity, grace, and humor. My brother has shown me how to persevere through hardship; I truly admire his ability to continually grow and learn. My parents have shown me how to try hard at everything I do and have been there for me whenever I needed them, even from 3000 miles away. They have an inherent belief in my abilities that I cannot appreciate enough.

Finally, Eleanor Caves remains my biggest believer, strongest supporter, and best friend. I am constantly in awe of how hard she tries to be the best person she can be. She has contributed just as much to my career and my love of science as she has to my personality and who I am as a man, and I am grateful for all of it.
1. Introduction

Resources like mates, territory, and food are essential to an animal’s survival and reproduction; however, these resources are often limited. To determine who can access limited resources, animals compete. According to animal behavior theory, and our own intuition, animals should compete efficiently; that is, competitors with greater ability should win contests and the process of contest resolution should involve minimal costs (reviewed in: Hardy & Briffa, 2013; Maynard Smith, 1974; Maynard Smith & Price, 1973; Parker, 1974). To accomplish this efficient resolution, animals assess, or gather information, about themselves and/or their opponents (reviewed in: Arnott & Elwood, 2009). Assessment can be facilitated by the behaviors and structures used in contests, such as signals and animal weapons (Emlen, 2008; Searcy & Nowicki, 2005).

A comprehensive understanding of a contest system requires answering questions related to ability, assessment, and the function of behaviors and weaponry (Briffa et al., 2013; Hardy & Briffa, 2013). These questions include: (1) what is ability (what differentiates winners from losers); (2) how is ability assessed (what information do competitors gather when making competitive decisions); (3) how do behaviors and weapons function in assessment (when and in which ways are behaviors and weapons used); and (4) how do behaviors or weapons allow for reliable assessment (what information is communicated through behaviors or weapon use)?
The classic system of red deer (*Cervus elaphus*) is a good example of a comprehensively-studied contest system. During the rut (mating season), male red deer compete over access to harems of females – males who win contests gain reproductive access to most females in a harem, while unsuccessful males go without mating opportunities. Clutton-Brock and colleagues showed that males between the ages of 7 and 10 are most likely to win contests (Clutton-Brock, Albon, Gibson, & Guinness, 1979); therefore ability is related to age (question 1, above). Males with a greater number of antler points are also more likely to win (question 1), though this effect is weaker than that of age (Clutton-Brock et al., 1979). During contests, competitors gather information on both their own ability and that of their competitors (question 2) by progressing in a behavioral sequence from vocal “roar” displays to visual “parallel walks” and, if necessary, to weapon use *via* “antler sparring” (question 3; Clutton-Brock & Albon, 1979). The acoustic properties of a male’s roar are constrained by his body size, such that the roar is thought to communicate reliable information on size (question 4; Reby & McComb, 2003). Finally, roars may also communicate an individual’s energetic state (question 4): males have been observed to collapse from exhaustion after roaring contests (Clutton-Brock & Albon, 1979). Work in red deer shows how males parse out access to a limited resource (access to reproduction) and reveals that the full understanding of a contest system requires integration across fields like behavior,
functional morphology, and biomechanics. However, similar comprehensive studies of contest systems are rare, and most well-studied animal weapon systems involve similar types of weapons. Testing the classic questions of this field in systems with weapons that function differently can reveal new information or reinforce our common knowledge about how animals compete.

In this thesis, I study the territorial contests of *Neogonodactylus bredini* (Stomatopoda: Crustacea) mantis shrimp by testing (1) what determines competitive ability, (2) how competitors assess, (3) how behaviors and weapons function in assessment, and (4) how energetics may facilitate assessment. Mantis shrimp are a good system for studies of contest behavior and weaponry. During contests, many species of mantis shrimp present visual weapon displays (termed “meral spreads”) and deliver high-force strikes to each other with power-amplified raptorial appendages (Caldwell, 1987; Caldwell & Dingle, 1975). Mantis shrimp strikes occur over timescales that are orders of magnitude faster than commonly-studied weapon systems, making them a good comparison point for testing classic predictions. A history of research studying mantis shrimp contest behavior (e.g., Adams & Caldwell, 1990; Caldwell, 1987; Caldwell & Dingle, 1975; Dingle & Caldwell, 1969) and the biomechanical mechanism powering the appendage strike (e.g., M.J. McHenry et al., 2016; S. N. Patek & Caldwell, 2005; S.N. Patek, Korff, & Caldwell, 2004; S. N. Patek, Nowroozi, Baio, Caldwell, & Summers, 2007)
sets up a strong foundation in this system. However, the more recently-discovered knowledge about the biomechanics of the appendage strike have yet to be combined with behavioral approaches in a comprehensive study of mantis shrimp contest behavior and weapon use.

I begin by establishing the relative roles of meral spread visual displays and ritualized striking in contests where competitors are matched for size (Chapter 2; P. A. Green & Patek, 2015). Contrary to predictions based on common findings in other contest systems, I show that visual displays were not used to communicate information on strike performance and did not resolve contests before striking. Appendage morphology did not correlate with strike force, and almost all contests involved striking, even if visual displays were used beforehand. I also show that an overwhelming majority of strikes were delivered onto a competitor’s coiled tailplate – the telson – and that the number of strikes delivered was a predictor of competitive success. I term the ritualized exchange of strikes “telson sparring” and propose that it functions to facilitate safe conflict resolution.

Next, using a detailed behavioral analysis, I show that N. bredini uses mutual assessment during contests; that is, competitors assess both their own ability and that of their opponent (Chapter 3; P. A. Green & Patek, 2018). I show that both body mass and residency in a protective burrow were good measures of ability – these two factors
strongly predicted competitive success in contests where competitors were both matched for size and randomly-matched. Furthermore, by using network analysis to visualize how telson sparring fits in the sequence of contest behaviors, I show how it (and, therefore, weapon use) played a crucial role in overall contest dynamics.

Finally, I test how the energy used to power sparring strikes may facilitate the assessment that occurs during telson sparring (Chapter 4). I show that the energy of strikes (the work that powers the forward motion of the appendage, and of surrounding water, during a strike) scaled positively with body mass: larger animals had proportionally more energetically-costly strikes. The ability to withstand the energetic costs of delivering strikes may differentiate low- from high-quality competitors. I also show that these energy dynamics were reflected in strike velocity – velocity was constant across a size range – and that the scaling of energy with size was likely achieved by increased compression of the spring that powers the strike. Finally, I show that the energy dynamics of sparring differed from those of another behavioral context – feeding – suggesting that individuals modify the use of their spring-powered weapons depending on context.

By studying weapons that function over vastly different timescales from commonly-studied weapons and by developing new approaches, this work informs classic theories of animal contests and animal weaponry.
2. Contests with deadly weapons: telson sparring in mantis shrimp (Stomatopoda)


2.1 Introduction

Game theoretical models of aggressive behaviour (reviewed in Enquist & Leimar, 1990; Searcy & Nowicki, 2005; van Staaden, Searcy, & Hanlon, 2011) and empirical tests in both vertebrates (Barrette & Vandal, 1990; Hof & Podos, 2013) and invertebrates (Egge, Brandt, & Swallow, 2010; Mowles & Briffa, 2012) find that competitors use signals to resolve contests before escalating to dangerous combat. Some mantis shrimp species (Stomatopoda) use their raptorial appendages to crack and kill hard-shelled prey with strikes that deliver forces exceeding their body weight by a thousand times or more (S. N. Patek & Caldwell, 2005; Weaver et al., 2012). During territorial contests, both sexes also use their appendages to strike competitors (Caldwell & Dingle, 1975). These conspecific strikes may be as lethal as those delivered to prey; thus, mantis shrimp should resolve contests by signalling before escalating to combat.

Some mantis shrimp species use a visual display and the ritualized exchange of strikes during territorial conflicts (reviewed in Caldwell & Dingle, 1975). During the “meral spread” visual display, the raptorial appendages are spread laterally and ventrally
such that several parts of the appendage are presented to the competitor and the individual displaying the meral spread is biomechanically unable to strike (Figure 2.1) (Caldwell & Dingle, 1975). The meral spread is considered a signal of aggressive motivation (Adams & Caldwell, 1990; Caldwell & Dingle, 1975; Steger & Caldwell, 1983) and possibly performance (Searcy & Nowicki, 2005). In addition, competitors exchange strikes using a ritualized “telson coil” behaviour (Caldwell, 1987; Caldwell & Dingle, 1975), in which the receiver of a strike coils its tailplate, or telson, in front of its body to receive the blow (Figure 1). How these behaviours actually resolve conflicts has yet to be examined in this system. More broadly, studying processes of conflict resolution is essential for understanding the evolution of deadly weapons for both inflicting and circumventing injury.

Figure 1: During aggressive contests, mantis shrimp often present meral spread displays (A). When receiving a strike during a ritualized fight (B), an individual coils its tailplate in front of its body in a “telson coil” posture. Eyes (dotted lines), one appendage (solid lines), and telson (arrow) are labeled. (A) Neogonodactylus wennerae, (B) Neogonodactylus oerstedii. Images courtesy of Dr. Roy Caldwell.
Here we test how *N. bredini* competitors use meral spreads and telson strikes to resolve sex- and size-matched contests. We predicted that few contests would escalate to striking, and instead most contests would be resolved by meral spread displays. We also predicted that appendage morphological components presented during meral spreads act as index signals (Maynard Smith & Harper, 2003) of strike performance (sensu Arnold, 1983) via a strong correlation between the size of appendage components and maximum strike force (e.g., Lappin, Brandt, Husak, Macedonia, & Kemp, 2006). Finally, based on other studies of weapon performance in size-matched contests (Lappin et al., 2006; Sneddon, Huntingford, Taylor, & Orr, 2000; Wilson, Angilletta, James, Navas, & Seebacher, 2007), we predicted that winners of contests would strike with greater maximum force than losers.

### 2.2 Methods

See Appendix A for additional details. We collected individuals from burrows in coral rubble. We measured body length, body mass, and three morphological components visible during meral spreads that are biomechanically important for storing elastic energy and delivering strikes. Strike performance was defined as the maximum peak force from 10 strikes, following standard organismal performance methodology (Lappin et al., 2006; Sneddon et al., 2000; Wilson et al., 2007).
We paired 68 unique individuals in 34 sex- and body-length matched (mean difference 2.0%±1.2%; range <0.01-4.0%) contests over access to a body length-matched artificial burrow (Steger, 1987) following commonly-used methods in contest behaviour research (Adams & Caldwell, 1990; Caldwell, 1979; Mowles & Briffa, 2012; Steger & Caldwell, 1983; Wilson et al., 2007). After giving the burrow resident 3-6 hours to acclimate, we acclimated a second individual (the “intruder”) behind an opaque barrier for 10 minutes. We then lifted the barrier and recorded contest behaviours until one competitor was deemed the winner. The winner was defined as the individual that resided in the burrow after the other made a clear, directed retreat toward the edge of the arena. Burrow residents have been found to win contests more often (Adams & Caldwell, 1990; Caldwell, 1987; Steger & Caldwell, 1983), yet we saw no resident advantage (perhaps due to brief residency time) and did not use residency as a factor in our analyses.

To test whether meral spreads occurred at the beginning of contests to avoid escalation to striking, we compared the number of contests that began with competitors displaying meral spreads to those that began with other behaviours, including striking, visually tracking competitors, and flicking antennae. We also asked whether meral spreads were used to resolve conflicts by comparing the number of contests resolved by meral spreads to the number that escalated to striking.
When contests escalated to striking, we tested how telson strikes were used in conflict resolution. We analysed contest videos and counted the number of strikes received on a competitor’s telson compared to other parts of the body. We also tested if winners and losers of contests differed in strike force, the number of strikes delivered during a contest, body length (within the size-matched range), and body mass.

We tested whether meral spreads displayed index signals of strike force by correlating maximum strike force with body length, body mass, and the size of three appendage morphological components (generalized linear model: gamma error distribution, log-link function; reduced using Akaike Information Criterion simplification (Crawley, 2007)). We calculated partial correlation coefficients (partial $r^2$) between maximum strike force and the independent effect of each morphological component. A high amount of variation (low partial $r^2$) between a morphological component’s individual effect and maximum strike force suggests that the size of that component, independent of the size of other components or body size, does not reliably signal strike force (Hughes, 2000). Alternatively, a high partial $r^2$ suggests component size may reliably signal strike force (Lappin et al., 2006).

**2.3 Results**

Meral spreads were not used to avoid escalation to strikes. Meral spreads were not more common than other behaviours at the beginning of contests (one-sided proportions
test: $\chi^2 = 0.266, p=0.246$), and 33 out of 34 contests escalated to striking, even after competitors presented meral spreads (17/18 contests).

Meral spreads also did not display index signals of strike force: there was only a weak correlation between appendage morphological components and strike force. Saddle length, the component with the highest partial $r^2$, accounted for only 5.9% of variation in strike force (Appendix Table 6, Figure 9). In comparison, other studies have described potential index signals with $r^2$ values greater than 0.40 (Hughes, 1996; Lappin et al., 2006).
Figure 2: Winners of contests do not have greater strike force than losers; instead, winners strike a greater number of times during contests. Each variable was z-score transformed. The y-axis shows the winner minus loser z-score. Dark solid lines represent median z-score, box ends represent 25% and 75% quartiles, whiskers represent the most extreme values within 1.5x the interquartile range, and open circles represent values greater than 3 standard deviations from the mean.

Winners of contests did not have greater maximum strike force than losers; instead, winners struck a greater number of times during contests (Table 1, Fig. 2).

Winners also had greater body mass than losers, although this variable was collinear ($r^2 = 0.31$) with body length.
Table 1: Winners (W) of contests do not have greater strike force than losers (L); instead, winners strike a greater number of times during contests. The mean difference between winner and loser values is indicated for each variable followed by the statistical tests of these differences. Bold values represent significant differences (p < 0.05). Body length was correlated with body mass (see Results). See Appendix for explanation of statistical analyses.

<table>
<thead>
<tr>
<th>Variable</th>
<th>W-L Mean +/- SE</th>
<th>p-value</th>
<th>Statistical analysis</th>
<th>Hedge's g</th>
<th>95% CI of Hedge's g</th>
</tr>
</thead>
<tbody>
<tr>
<td>Body length</td>
<td>0.31 +/- 0.16 mm</td>
<td>0.058</td>
<td>Two-sided t-test</td>
<td>0.330</td>
<td>-0.023, 0.713</td>
</tr>
<tr>
<td>Body mass</td>
<td>0.08 +/- 0.03 g</td>
<td><strong>0.020</strong></td>
<td>Two-sided t-test</td>
<td>0.409</td>
<td>0.057, 0.735</td>
</tr>
<tr>
<td>Maximum strike force</td>
<td>-2.48 +/- 9.31 N</td>
<td>0.604</td>
<td>One-sided t-test</td>
<td>-0.045</td>
<td>-0.392, 0.288</td>
</tr>
<tr>
<td>Number of strikes</td>
<td>0.97 +/- 0.35 strikes</td>
<td><strong>0.005</strong></td>
<td>Two-sided Wilcoxon</td>
<td>0.469</td>
<td>0.201, 0.659</td>
</tr>
</tbody>
</table>

When contests escalated to striking, almost every strike (214/227, 94.3%, one-sided proportions test: $\chi^2 = 176.2, \ p < 0.001$) was received on the competitor’s telson in a telson coil posture. In rare cases, as the loser was retreating the winner escalated past striking to stabbing. During stabbing, the appendage was used to puncture the competitor’s abdomen, resulting in significant injury.
2.4 Discussion

Contrary to our predictions, *N. bredini* competitors did not use meral spreads to resolve size-matched contests by presenting index signals of strike force. Instead, most contests escalated to the ritualized exchange of strikes, and almost all strikes were received on the telson. These results evoke parallels to sparring in mammals (Barrette & Vandal, 1990; Chelliah & Sukumar, 2013), in which weapons are used in a ritualized manner and damage is unlikely. Thus, we introduce the term “telson sparring” to describe the function of ritualized strikes in mantis shrimp contests. While the signalling function of sparring is often difficult to quantify, our finding that winners deliver more strikes than losers suggests that telson sparring may communicate aggressive persistence or physiological endurance. Alternatively, or additionally, sparring may signal performance other than peak strike force (Byers, Hebets, & Podos, 2010).

The close match between body and contested burrow size in the field (Steger, 1987) makes size-matched contests likely; however, future work should investigate the progression of behaviours in non-size-matched contests. One study found variation in the number of strikes and meral spreads given to larger or smaller competitors, but did not measure the progression of these behaviours throughout a contest (Adams & Caldwell, 1990), which is essential information for theoretical models of escalation (Arnott & Elwood, 2009).
Telson sparring illuminates how animals with lethal weapons potentially lower their risk of fatality by using ritualized fighting to signal persistence, endurance, or performance (Enquist & Leimar, 1990). Indeed, the telson’s mechanical behaviour suggests that impact dynamics may inform competitor assessment. The telson dissipates a high percentage of impact energy and the amount of energy returned to the striker correlates with the size of the receiving animal (J. R. Taylor & Patek, 2010). Since body size can be difficult to visually assess in burrows, telson sparring may communicate a competitor’s size. Telson sparring thus reveals multiple potential signalling functions for high-force strikes in mantis shrimp. These findings encourage new perspectives on other well-studied weapon systems, in which peak force has been the key measured variable, yet which may be using sparring strategies to reduce the risk of damage during contests.
3. Mutual assessment during ritualised fighting in mantis shrimp (Stomatopoda)


3.1 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, Parker, 1974). 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 Arnott & Elwood, 2009). We term these “assessment models” following (Arnott & Elwood, 2009) 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 2). In pure self- and cumulative assessment (M Mesterton-Gibbons, Marden, & Dugatkin, 1996; Payne, 1988; Payne & Pagel, 1996), 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 (Enquist & Leimar, 1983), 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.
Table 2: 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 Arnott & Elwood, 2009). 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.
<table>
<thead>
<tr>
<th>Model</th>
<th>Model description</th>
<th>Relationships between contest costs and RHP</th>
<th>Trends in contest behaviours</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pure self-assessment</td>
<td>Impose energetic costs to self only.</td>
<td><img src="image1" alt="Graph" /></td>
<td>No physical contact; behaviours progress in any direction; escalation and de-escalation.</td>
<td>(Arnott &amp; Elwood, 2009; M Mesterton-Gibbons et al., 1996; Payne &amp; Pagel, 1996)</td>
</tr>
<tr>
<td></td>
<td>Loser reaches own cost threshold first.</td>
<td><img src="image2" alt="Graph" /></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cumulative</td>
<td>Impose energetic costs to self and other costs (e.g., energetic, injury) to</td>
<td><img src="image3" alt="Graph" /></td>
<td>Physical contact, injuries likely; behaviours progress in any direction; escalation and de-escalation.</td>
<td>(Arnott &amp; Elwood, 2009; Payne, 1988)</td>
</tr>
<tr>
<td>assessment</td>
<td>competitor.</td>
<td><img src="image4" alt="Graph" /></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Loser reaches own cost threshold first.</td>
<td><img src="image5" alt="Graph" /></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mutual</td>
<td>Assess both self and competitor RHP.</td>
<td><img src="image6" alt="Graph" /></td>
<td>Physical contact, injuries rare; behaviours progress unidirectionally via phases; escalation without de-escalation.</td>
<td>(Arnott &amp; Elwood, 2009; Enquist &amp; Leimar, 1983; Enquist, Leimar, Ljungberg, Mallner, &amp; Segerdahl, 1990)</td>
</tr>
<tr>
<td>assessment</td>
<td>Loser retreats upon assessing it is the inferior competitor.</td>
<td><img src="image7" alt="Graph" /></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Assessment models can be differentiated by testing how behaviours progress throughout contests and how variation in contest costs correlates with variation in competitor RHP (Briffa et al., 2013; P. W. Taylor & Elwood, 2003). Competitors using mutual assessment assess both their own and each other’s ability; therefore, contest behaviours are predicted to progress in phases of escalating intensity that give increasingly accurate information about RHP (Enquist et al., 1990). 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 (Payne, 1988; Payne & Pagel, 1996). 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., Enquist et al., 1990; McGinley, Prenter, & Taylor, 2015), results from other systems do not clearly align with any particular model (e.g., Briffa, 2008; reviewed in Briffa & Elwood, 2009; Jennings, Gammell, Payne, & Hayden, 2005). This incongruence is often due to the difficulty of acquiring the breadth of experimental data needed to test aspects related to both behavioural progressions and cost-RHP correlations (Arnott & Elwood, 2009; Briffa & Elwood, 2009). 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 (Briffa & Elwood, 2009; P. W. Taylor & Elwood, 2003). However, while the techniques for correlational analyses are well-established and widely-used (P. W. Taylor & Elwood, 2003), relatively fewer studies employ tests of behavioural progressions.

Testing behavioural progressions requires rigorous analysis and visualisation of behavioural sequences; one technique that can be used to achieve this goal is sequential behavioural analysis (also termed "sequential analysis"; Bakeman & Quera, 1995, 2011; Bakeman, Robinson, & Quera, 1996). While sequential analysis is common in the social sciences (Bakeman & Quera, 2011), it has rarely been applied to contest dynamics (Chen, Lee, Bowens, Huber, & Kravitz, 2002; Clutton-Brock et al., 1979; Egge et al., 2010). Sequential analysis can quantify the sequences of behaviours used by competitors, which assist in later tests against theoretical predictions (Table 2). Additionally, this technique can identify how individual behaviours connect to a given assessment strategy (Briffa et al., 2013). 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 (Egge et al., 2010). 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 (Egge et al., 2010).
Here we examine assessment, using both correlational and sequential analyses, in the mantis shrimp *Neogonodactylus bredini* (Stomatopoda: Crustacea). *N. 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 (Enquist et al., 1990; Payne, 1988). 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, Payne, 1988). 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 (Caldwell, 1987; Dingle & Caldwell, 1969). These strikes can reach bullet-like accelerations (M.J. McHenry et al., 2016 and references therein), delivering high peak forces (S. N. Patek & Caldwell, 2005). Competitors can also unfurl their raptorial appendages to stab each other, causing puncture wounds (Berzins & Caldwell, 1983; Caldwell, 1987).

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 (Enquist & Leimar, 1983). *N. bredini* use behaviours
that may minimize costs; for example, competitors ritualistically and non-lethally exchange strikes on each other’s armoured tailplates (Caldwell, 1987) in a behaviour termed “telson sparring” (P. A. Green & Patek, 2015) and, in our experience, potentially dangerous stabbing behaviours are rare (P. A. Green & Patek, 2015). Additionally, *N. bredini* exhibits individual competitor recognition (reviewed in Vetter & Caldwell, 2015) and analyses of non-territorial contests found a behavioural stereotypy like that expected by mutual assessment (Dingle, 1969; Dingle & Caldwell, 1969). 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 ritualised 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 (1) identifying the assessment strategy used by *N. bredini*, (2) explaining how individual behaviours function in assessment in *N. bredini*, and (3) 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 (Csardi & Nepusz, 2006; Kolaczyk & Csardi, 2014) 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 (Briffa & Elwood, 2009) 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.

3.2 Methods

Our dataset from body length-matched contests is a more detailed analysis of contests previously studied in (P. A. Green & Patek, 2015). See the Appendix B for additional methodological details.

3.2.1 Collection and Measurement

We collected Neogonodactylus 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 (Steger, 1987). 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; Adams & Caldwell, 1990; Montgomery & Caldwell, 1984; Steger & Caldwell, 1983). 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.

### 3.2.2 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 (Caldwell & Dingle, 1975; P. A. Green & Patek, 2015). Each individual was used in only one contest to avoid effects of prior experience on contest behaviour (Caldwell, 1979).

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 (Steger, 1987). 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 minutes, 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 minutes, 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 (~20% of initiated trials) and 7 body length-matched contests (~17% of initiated trials) due to a lack of interaction or the presence of one-sided interactions.

3.2.3 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 toward the edge of the contest arena (see ethogram, Table 7). 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, McGarigal, 2013). The overall trends were the same with and without inclusion of the outliers (Tables 13 & 14). All data were analysed using R Version 3.0.1 (Core Team, 2008).
3.2.4 RHP and Resource Ownership

We tested between body length and body mass as measures of RHP, following methods outlined by (Briffa et al., 2013). 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 (Crawley, 2007). We ran all possible models including and excluding these terms and identified which models had the lowest AIC score (within 2 ∆AIC, Burnham, Anderson, & Huyvaert, 2010). 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 (Mundry & Nunn, 2009). We also tested for the benefit, in RHP units, of residency (Kasumovic, Mason, Andrade, & Elias, 2010; Kokko, Lopez-Sepulcre, & Morrell, 2006). 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 at an intruder minus resident RHP value of 0. With a residency advantage, an intruder would have a 50% chance of winning the contest only if its RHP were greater than the resident’s (Kokko et al., 2006).

3.2.5 Correlational Tests of Assessment Models

Following methods established by (P. W. Taylor & Elwood, 2003), 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 \([\log_{10}(\text{duration}) \text{ and } \log_{10}(1 + \text{number of strikes})]\) 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 Elias, Kasumovic, Punzalan, Andrade, & Mason, 2008). 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 (Mundry & Nunn, 2009) to differentiate among assessment models (Table 2; Arnott & Elwood, 2009; P. W. Taylor & Elwood, 2003). 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.
3.2.6 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 2; Enquist & Leimar, 1983; Enquist et al., 1990; Payne, 1988; Payne & Pagel, 1996). We used a sequential analysis to test how *N. bredini* behaviours match these predictions.

We coded all contest behaviours using JWatcher software (Blumstein, Daniel, & Evans, 2007), following an ethogram similar to that of Dingle & Caldwell (Dingle & Caldwell, 1969). Our ethogram consisted of 14 mutually exclusive contest behaviours (Table 7). 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 Appendix B. Briefly, we used the igraph network analysis package (Csardi & Nepusz, 2006) 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 x 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 (Bakeman et al., 1996; see Appendix B). 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 Tables 16 & 17.

We used the network graphs to identify the presence or absence of phases
following the definitions of (Enquist et al., 1990). 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 (Payne, 1988).

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 (Enquist & Leimar,
1983). For example, Servaea incana jumping spiders escalate from no physical contact, to
“sparring”, to biting with the chelicerae (McGinley et al., 2015). 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 2; reviewed in Arnott & Elwood, 2009; Egge et al., 2010). 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 (Arnott & Elwood, 2009).

**3.3 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 Table 15.

**3.3.1 RHP and Resource Ownership**

Body mass occurred most frequently in the best-supported models predicting contest outcome (Table 8) 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 (Tables 11 & 12).
RHP (body mass) and resource ownership (burrow residency) interacted to predict contest outcome (Figure 3). In randomly-matched contests, residency ($B=3.5, z_M=2.0, p=0.045$) was a better predictor of contest outcome than mass ($B=11.0, z_M=1.62, p=0.11$) or the residency : mass interaction ($B=-10.3, z_M=1.5, p=0.13$). The full GLM for body length-matched contests did not have a high enough sample size to converge ($df=25$). By testing models with only individual effects, we found that residency ($B=1.54, z_M=1.79, p=0.07$) had a similar effect on contest outcome as mass ($B=6.34, z_M=1.71, p=0.09$), and both these effects were greater than that of the mass : residency interaction (Intruder: $B=37.5, z_M=1.36, p=0.17$, Resident: $B=2.83, z=0.70, p=0.49$).

![Graph](image)

**Figure 3:** In (a) randomly-matched and (b) body length-matched contests, the likelihood 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 likelihood of winning from relative body mass, burrow residency, and the mass : residency interaction (see Methods). The y-axes are equivalent in (a) and (b).
Intruders needed an average of 0.51g 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=2.56, p=0.01) and 0.16g greater mass (9.7% of mean competitor mass) in body length-matched contests (Intercept = -0.72, B=4.66, z=1.87, p=0.09) (Figure 10).

3.3.2 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, R²=0.21, p=0.047; Figure 4; Table 9). As predicted by all three assessment models, there was a positive correlation between contest duration and the interaction of loser mass and loser residency (Fig. 4a). We also found a negative correlation between contest duration and winner mass, which supported both mutual and cumulative assessment but not pure self-assessment (Fig. 4b). Residency allowed eventual losers to compete for longer; this effect increased with increasing body mass (Fig. 4a). Winner mass alone correlated negatively with contest duration, but this effect was driven by winners who were residents (Fig. 4b). 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, R²=0.33, p=0.007; Table 9). 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 (Table 9).
Figure 4: 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 (Arnott & Elwood, 2009). Correlation lines depict differences between residents (closed circles, solid lines) and intruders (open diamonds, dashed lines; icons and lines in 2c represent winner roles only), but do not necessarily depict statistical relationships (statistical results from multiple regressions are reported in results and Tables 9 & 10).
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 4c; Table 10). 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 ($F_{2,26}=6.04$, $R^2=0.27$, $p<0.01$; Table 10): 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 (Arnott & Elwood, 2009), but not those of Fawcett & Mowles (Fawcett & Mowles, 2013) (see Discussion). Therefore, we could not fully reject cumulative assessment using correlations alone.

### 3.3.3 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 (Fig. 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 (Fig. 5; Tables 16 & 17).
Figure 5: 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 (Enquist et al., 1990): 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 & 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 visualisation purposes. See Methods for details.

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 (Fig. 5; Tables 16 & 17). 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 P. A. Green & Patek, 2015). We saw no overt evidence of injury due 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 (Tables 16 & 17). One sparring behaviour, the telson coil, was a predictor of a retreat in both contest types (Tables 16 & 17).

3.4 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 (Arnott & Elwood, 2009); however, due to disagreement in the literature (Fawcett & Mowles, 2013), 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 (Enquist & Leimar, 1983; Enquist et al., 1990) – 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.

### 3.4.1 RHP and Resource Ownership

In both randomly-matched and body length_matched contests, body mass predicted contest outcome (Fig. 3). Body mass may be a particularly relevant measure of RHP in *N. bredini* contests, especially in the context of telson sparring. In a closely-related species, *N. wennerae*, 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 (J. R. Taylor & Patek, 2010).

Telson sparring was common in both randomly-matched and body length-matched *N. bredini* contests (Fig. 5). 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 Appendix B), factors such as the possibility of allometric scaling of RHP
with body size (Palaoro & Briffa, 2017) 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.

### 3.4.2 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 (Arnott
& Elwood, 2009; P. W. Taylor & Elwood, 2003). 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 and
Elwood (Arnott & Elwood, 2009). However, Fawcett & Mowles (Fawcett & Mowles, 2013) 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 (Enquist & Leimar, 1983; Enquist et al., 1990). We defined phases and escalation following Enquist & Leimar (Enquist et al., 1990; 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 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 (Fig. 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 (Fig. 5; Tables 16 & 17). Competitors also escalated from behaviours
that did not involve physical contact (phases 1 & 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 (Enquist & Leimar, 1983; Enquist et al., 1990): 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 toward 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., M. Mesterton-Gibbons & Heap, 2014; Prenter, Elwood, & Taylor, 2006) could generate predictions for behavioural progressions to be empirically tested using sequential analysis.
3.4.3 Conclusions

The field of assessment has had a strong theoretical foundation since its first game-theory models were established (Maynard Smith & Price, 1973). However, inconclusive experimental results have sometimes led to uncertainty as to whether current theory is sufficient to explain assessment in many species (Lane & Briffa, 2017; M. Mesterton-Gibbons & Heap, 2014; Palaoro & Briffa, 2017; Prenter et al., 2006), or whether experiments need to incorporate a broader array of approaches (Briffa & Elwood, 2009; Vieira & Peixoto, 2013). We found that, in *N. bredini*, a combination of correlational and sequential analyses ruled out alternative models and supported mutual assessment (Enquist & Leimar, 1983; Enquist et al., 1990). 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.
4. Context-Dependent Scaling of Weapon Energetics and Kinematics in Mantis Shrimp (Stomatopoda)

4.1 Introduction

Energy is a central currency in biology that fuels how an animal accomplishes essential tasks (Nagy, 2005). Studies of energetic scaling – how energy use varies as body size varies – have established unifying parameters (Brown & West, 2000; Schmidt-Nielsen, 1984) and also discovered variation due to phylogenetic relationships and life history differences (Nagy, 2005). However, even though energy use is at the core of theories related to animal competition (Briffa & Sneddon, 2007) and feeding strategies (Pyke, 1984), few studies have examined energy use across behavioral contexts and how this energy use scales with size.

Energy plays a crucial role in animal contests. An animal’s energetic stores can limit its ability to compete (Payne, 1988; Payne & Pagel, 1996) or the vigor of its competitive displays (Briffa & Lane, 2017; Byers et al., 2010). Energetic costs thereby differentiate high- from low-quality competitors and can lead to efficient contest resolution (Briffa & Sneddon, 2007; Searcy & Nowicki, 2005).

Despite the importance of energy to contest theory, direct tests of contest energetics are rare, and tests that incorporate scaling are even less common. Using carbon dioxide production rate as a proxy for energy metabolism in sierra dome spiders
Neriene litigiosa, DeCarvalho et al. (2004) found that the energetic costs scaled positively with mass for displays used early in contests; however, once competitors escalated to physical grappling, energetic costs were constant across size (DeCarvalho et al., 2004). These results suggest that energy use reliably reflects ability (size) in some stages of contests, but not all (DeCarvalho et al., 2004). Though important to validating contest theory, tests like these are rare, in part because of the inherent challenges of measuring carbon dioxide production (Boisseau, Woods, & Goubault, 2017; DeCarvalho et al., 2004; Hack, 1997) or other proxies like lactate or glucose production (Briffa & Elwood, 2001; Copeland, Levay, Sivaraman, Beebe-Fugloni, & Earley, 2011).

In addition to measuring energetic costs during a single context like competition, testing energetic scaling across multiple contexts can explain how animals utilize limited energetic reserves to accomplish multiple tasks (Nagy, 2005). For example, Boisseau et al. (2017) found that fig wasp Eupelmus vuilleti contests over egg-deposition sites require less energy (i.e. lower CO₂ production rates) than drilling into the sites to deposit eggs, but that both activities are costlier than depositing eggs in a hole already drilled by a competitor. These animals benefit by avoiding contests and instead simply use a conspecific’s pre-drilled hole to deposit eggs (although they did not examine scaling in either context; Boisseau et al., 2017).

The goal of this study is to quantify energetic scaling across two behavioral contexts in the mantis shrimp Neogonodactylus bredini (Stomatopoda: Crustacea). N. bredini delivers high-force strikes to both competitors and to hard-shelled prey items.
Mantis shrimp strikes are powered by the storage and release of elastic potential energy.

In preparation for a strike, the lateral extensor muscle in the raptorial appendage contracts and compresses the exoskeleton which acts as a spring to store elastic energy. Simultaneously, flexor muscles engage latches (modified apodemes) and allow continued spring contraction while preventing output motion. Upon latch release, energy is transferred from the spring through a linkage mechanism and into work done on the striking body (propodus and dactyl) of the appendage to rotate through the water (for detailed descriptions of the strike mechanism, see: M.J. McHenry et al., 2016; M. J. McHenry, Claverie, Rosario, & Patek, 2012; S.N. Patek et al., 2004; S. N. Patek et al., 2007; S. N. Patek, Rosario, & Taylor, 2013; Zack, Claverie, & Patek, 2009). The fast movement of appendage strikes prevents individuals from altering kinematics after the latch is released; however, in advance of a strike, individuals can modify strike kinematics by contracting muscles to a different degree and thereby differentially compressing their exoskeletal spring (Kagaya & Patek, 2016).

Mantis shrimp strikes are central to both contests and foraging. During contests over access to limited resources (burrows in coral rubble; Steger, 1987), *N. bredini* competitors ritualistically exchange strikes on each other’s armored tailplates (Dingle & Caldwell, 1969) in a behavior termed “telson sparring” (P. A. Green & Patek, 2015). Telson sparring functions in mutual assessment (P. A. Green & Patek, 2018), letting individuals assess both their own ability and that of their opponent. Furthermore, this ability is related to size and the number of strikes delivered: larger individuals are more...
likely to win contests (P. A. Green & Patek, 2018) and in size-matched contests, winners deliver a greater number of strikes (P. A. Green & Patek, 2015). Outside of contests, mantis shrimp also smash hard-shelled prey with these same appendages. Snails, clams, and other hard-shelled prey make up a substantial portion of the N. bredini diet (devries, Stock, Christy, Goldsmith, & Dawson, 2016); indeed, N. bredini are considered voracious snail predators (Caldwell, Roderick, & Shuster, 1989).

We tested how the potential energy powering appendage strikes scales with size using a previously-validated mathematical model (M. J. McHenry et al., 2012). This model calculates the potential energy of strikes by incorporating prior experimental data on spring energy storage and linkage mechanics (S. N. Patek et al., 2013; Zack et al., 2009), as well as the dynamics of motion in fluids (M.J. McHenry et al., 2016). This math model allows us to simulate and, when combined with measured kinematics, closely estimate the energetics of only the strike itself. This approach is distinct from other techniques that measure whole-organism energetics (Briffa & Sneddon, 2007).

We first used simulations of the math model to establish a set of possible scaling relationships between elastic potential energy and appendage size. Given that individuals can manipulate elastic potential energy storage mainly through differential spring compression (Kagaya & Patek, 2016; S. N. Patek et al., 2013; Zack et al., 2009), we ran simulations across a range of appendage sizes (measured as the length of the striking body, see Methods) that varied only in the amount of spring compression. For each simulation, we measured the elastic potential energy stored in the exoskeletal
spring. This established a set of possible scaling relationships between elastic potential energy and size that was controlled only by variation in spring compression.

We then measured strike kinematics from high-speed video and combined these with the mathematical model to experimentally measure strike potential energy in the contexts of feeding and sparring. This allowed us to test (1) how energetic scaling varies across behavioral contexts and (2) how variation in experimentally-measured energetics matches simulated strike energetics. Finally, using our experimental dataset, we tested how the maximum angular velocity of strikes scales with body mass and tested if these scaling relationships differ across contexts.

4.2 Methods

4.2.1 Simulations

We used the mathematical model from McHenry et al (2012) to conduct simulations of mantis shrimp strikes. The model input requires (1) the link lengths of the mantis shrimp four-bar linkage system, which transmits spring motion into striking body movement (Anderson, Claverie, & Patek, 2014; S. N. Patek et al., 2007); (2) the stiffness \( k \) of the exoskeletal torsion spring, which incorporates the spring constant and other measurements related to spring materials testing (e.g., force and displacement; S. N. Patek et al., 2013; Zack et al., 2009); (3) the length of the striking body (“striking body length”), defined as the length from the carpus-propodus joint to the propodus-dactyl joint (M.J. McHenry et al., 2016; see Figure 6); and (4) the initial input angle \( \theta_{\text{start}} \) and final resting angle \( \theta_{\text{rest}} \) between two links of the four-bar linkage that control spring
compression. Subtracting $\theta_{\text{start}}$ from $\theta_{\text{rest}}$ gives $\theta_{\text{range}}$, the range of spring compression; a greater value of $\theta_{\text{range}}$ represents greater spring compression.

Following prior studies of strike biomechanics (S. N. Patek et al., 2013; Zack et al., 2009), we simulated mantis shrimp size variation as larger mantis shrimp having (1) larger striking bodies and (2) greater abilities to compress their exoskeletal spring. Across simulations, we kept spring stiffness ($k$) and link length constant at the mean value for *N. bredini* as measured in Patek et al. (2013) and Anderson et al. (2014). We changed striking body length in our simulations to span the range of measurements from the animals used in our experiments (0.005 m to 0.009 m, by 0.001 m increments). We kept $\theta_{\text{rest}}$ constant across simulations at the maximum value theoretically possible for the four-bar linkage geometry, given average *N. bredini* link lengths ($\theta_{\text{rest}} = 85$ degrees). For each striking body length, we then ran simulations for values of $\theta_{\text{start}}$ increasing from the minimum to maximum theoretically possible for the average *N. bredini* link lengths (67 degrees to 83 degrees, respectively). Therefore, $\theta_{\text{range}}$ varied by 1-degree increments between 2 and 18 degrees (17 total simulations for each striking body length measure). From each simulation, we extracted the maximum elastic potential energy of the system, calculated as spring energy: $EE = \frac{1}{2} k_{\text{spring}} (\theta_{\text{range}})^2$, where $k$ is the stiffness of the torsion spring (a function of linear spring stiffness and materials testing parameters, see M. J. McHenry et al., 2012). Spring energy is a measure of energy that is stored in the spring and that performs work on the appendage. Therefore, spring energy is equivalent
to strike potential energy, assuming that energy is transferred perfectly from the spring to striking body rotation and that no energy is lost during spring release due to dissipation, heat, or other factors.

4.2.2 Experimental Measurements

4.2.2.1 Animal Collection and Care

We collected *N. bredini* from burrows in coral rubble in *Thalassia* spp. seagrass habitats on the Atlantic coast of Panama (ANAM Collection Permits #SE/A-115-13; #SE/A-92-15; #SE/A-52-17). Most individuals were used in previously-published studies on contest behavior (P. A. Green & Patek, 2015; P. A. Green & Patek, 2018) before being transported to Duke University (ANAM Export Permits #SEX/A-23-14; SEX/A-106-15; SEX/A-48-17). At Duke, they were housed individually in clear plastic cubes (10cm x 10cm x 10cm, AMAC Plastics Corp., Sausalito, CA, USA) that were placed in groups of up to eight cubes in larger tanks in an aquarium system with circulating artificial seawater (27.2 deg C, 12h/12h light/dark schedule). Each individual was provided an artificial burrow made of opaque PVC tubing that had been cut longitudinally in half and secured to a corner of the cube with aquarium-safe sealant (i.e., the burrow had only one available opening and the animal was visible from the exterior of the cube). Individuals were fed approximately twice weekly with frozen krill and brine shrimp, or fresh snails.
4.2.2.2 Filming Protocol

4.2.2.2.1 Staged contests

To conduct a staged contest, we allowed one individual (the “resident”) to remain in its PVC burrow and introduced a competitor (the “intruder”) in a second, longitudinally-halved PVC burrow that was placed directly in front of opening of the resident’s burrow. We ensured both competitors had not molted in the past week, following previous studies (P. A. Green & Patek, 2015; P. A. Green & Patek, 2018). Competitors were matched visually by body length [body length percent difference, mean ± sd (min, max) = 4.7 ± 2.8 (0.6, 10.4); body mass percent difference, mean ± sd (min, max) = 14.0 ± 10.2 (0.8, 40.0)]. Competitors were separated by an opaque barrier until filming began. Before raising the barrier, we ensured both competitors were facing each other. Once the barrier was raised, opponents began interacting, and often sparring, almost immediately.

As competitors sparred, we opportunistically filmed their strikes with 30,000–40,000 frames per second (fps) high-speed video (HSV; Photron SA-Z or SA-X2; 10-15μsec shutter speed; pixel resolution: 1024 x 512 or 688, SA-Z, 896 x 496, SA-X2; Photron FastCam Viewer v3; Photron, San Diego, CA, USA). For many contests, we also simultaneously took 30 fps video using a Sony Handycam HDR CX-900 video camera (Sony Corp., Minato, Tokyo, Japan). These videos and our own observations suggested that behaviors in these staged contests were similar to behaviors we have observed in previous studies (P. A. Green & Patek, 2015; P. A. Green & Patek, 2018). We recorded up
to three strikes before stopping filming to save HSV data. While saving data, we separated the competitors to prevent them from further, un-filmed sparring. After saving data, we often re-introduced the same competitors (following the same protocol above) to film additional strikes. We used either competitor in up to approximately three “bouts” (i.e., three consecutive re-introductions), after which we did not use the animal in any contest trials for the rest of the day.

4.2.2.2 Feeding trials

We fed *N. bredini* individuals snails (*Tritia obsoleta*) that were collected from the Duke Marine Lab, Beaufort, NC, USA and maintained at Duke University in an aquarium system with circulating artificial seawater (24.0 deg C, 12h/12h light/dark schedule). Snails were fed frozen krill or brine shrimp twice weekly.

To film feeding strikes, we ensured *N. bredini* were in their PVC burrows and then introduced a snail into the burrow and placed an opaque barrier at the front of the burrow to ensure neither animal left the burrow. *N. bredini* prefer a certain size of snail when feeding (Full, Caldwell, & Chong, 1989); we visually assessed that snails were relatively matched for size with each mantis shrimp; that is, larger animals received larger snails. We introduced snails to all mantis shrimp at one time and then opportunistically filmed whichever *N. bredini* began striking its snail.

Once an individual began striking a snail, we filmed its strikes with 30,000 – 40,000 fps HSV, using the same camera settings as for sparring trials. As above, we stopped filming after three recorded strikes to save HSV data; however, we did not
remove the snail as we were saving data. Instead, we allowed the individual to continue striking and re-started filming as soon as possible.

4.2.2.3 Morphological Measurements

After filming strikes, we measured each *N. bredini* individual’s body mass (Denver Instruments APX-3202 balance; range: 0-3100g, readability: 0.01g, Sartorius AG, Goettingen, Germany) and body length (Mitutoyo Digimatic Caliper, range: 0-150mm, resolution: 0.01mm, Mitutoyo Corp., Kawasaki, Japan). We also took standardized photographs of the lateral and frontal surface of the propodus and dactyl (the “striking body”) using a 12 megapixel, digital SLR camera (Nikon D300; AF Micro-NIKKOR 60 mm f/2.8D or 105 mm f/2.8D macro lenses, Nikon Inc., Melville, NY) and an external light source (EM-140 DG macro-flash, Sigma Corp., Ronkonkoma, NY or Leica KL-300 LED, Leica Corp., Wetzlar, Germany). We took three replicates each of body mass, body length, and each photograph.

From the standardized photographs, we used Fiji software (Schindelin et al., 2012) to measure three aspects of the shape of the striking body as detailed in McHenry et al (2012): striking body length, average chord width, and average thickness (Figure 6b, 6c). These measurements are necessary for calculating energetics (see below). Striking body length was measured from the joint of the carpus and propodus to the joint of the propodus and dactyl (Figure 6b). Average chord width was measured using a blade-element approach: from lateral images, we measured the length of the anterior to posterior face of the striking body at approximately 21 points along the proximal to
distal striking body length (Figure 6b); we used the mean of these values for an average chord width. Average thickness was measured using a similar blade-element approach; however, average thickness was the mean of approximately 22 measurements along the lateral to medial distance from frontal images of the striking body (Figure 6c). The number of measurements along the striking body did not affect values for average chord width or thickness.
In brief, we digitized high speed video of sparring individuals (a) with two points on the striking body and two points on the merus. We used a blade-element approach to measure (b) striking body length and average chord width, and (c) average thickness (lines on appendage in b and c represent blade-element measurements). Using morphological input (b, c) and kinematic input from digitizing throughout the strike (representative still images d, e, f, g), the mathematical model output included (h) strike angular velocity, kinetic energy, and drag energy as a function of time. Solid vertical lines in (h) show the strike time for each still image (d – g). Scaling analyses used angular velocity and energy at the point of maximum angular velocity (vertical dotted line in h). In (a), (b), and (c) scale bar represents 5mm; scale from (a) is maintained in (d, e, f, g).
4.2.2.4 Digitizing, Kinematic, and Energetic Analyses

4.2.2.4.1 Digitizing
We only digitized videos that had good resolution and in which the appendage was visible and remained in focus throughout the strike, which suggested the strike was perpendicular to the plane of the camera. We digitized appendage motion from the beginning of appendage movement to several frames after contact using the MTrackJ plugin in Fiji (Meijering, Dzyubachyk, & Smal, 2012), following methods outlined in Kagaya and Patek (2016). We digitized two points on the striking body and two points on the merus; all points were natural color or brightness patterns on the exoskeleton. To measure the angular displacement of the striking body during a strike, we calculated the relative change in angle between (1) the line created by connecting the two striking body points and (2) the line created by connecting the two merus points (Kagaya & Patek, 2016). We calculated angular displacement from the raw digitizing data using an R code (Core Team, 2008) provided in Kagaya and Patek (2016).

4.2.2.4.2 Kinematics and Energetics
From the angular displacement data, we analyzed strike kinematics and energetics using the mathematical model published in McHenry et al. (2016; 2012). As stated above, the model was developed to simulate strike kinematics and energetics from data on appendage morphology. We adapted the model to allow us to input strike displacement from high speed video alongside measured morphological data; from these inputs, the model produced output kinematics and energetics data. Details of the model are in McHenry et al. (2012); we briefly outline the model below.
We first input morphological measurements (striking body length, average chord width, average thickness), HSV frame rate, and displacement data. We then fit a 7th, 8th, or 9th-order polynomial spline to the displacement data. To ensure an appropriate spline fit, the spline was truncated to end at the frame immediately before contact, as determined during digitizing. For each strike, we visually fit as high-order of a spline as possible, ensuring a proper measurement of velocity while minimizing noise. For further details on polynomial spline fitting technique, see deVries, Murphy, and Patek (2012) and Cox, Schmidt, Modarres-Sadeghi, and Patek (2014).

After fitting a spline to the displacement data, we generated a high-resolution time vector (5000 time points) and used this vector and the spline fit to generate high-resolution displacement data (5000 displacement points). We calculated strike velocity as the derivative of the high-resolution displacement data with respect to time. These velocity data were used in energetics measurements.

Measuring energetics from kinematics requires knowledge of both strike velocity and how mass is distributed along the striking body. McHenry et al. (2012) found that the striking body could be modeled as a uniform cylinder with little material density variation and established a scaling relationship for inferring striking body mass from striking body length for the morphologically-similar species Gonodactylus smithii (M. J. McHenry et al., 2012). We calculated our own scaling relationship for N. bredini by measuring the mass of the propodus and dactyl of several striking bodies dissected from naturally-deceased animals (XPE 56 microbalance, resolution: 10^-6g, Mettler Toledo).
Corp., Columbus, OH, USA). Our scaling relationship allowed us to infer striking body mass from striking body length.

From the strike kinematics data and our morphological data, we used the model to calculate strike kinetic energy, energy lost (due to drag), and total potential energy. Strike kinetic energy was calculated as \( KE = 0.5I\omega^2 \), where \( I \) is the summed moment of inertia of the striking body and the added mass from the acceleration reaction force and \( \omega \) is the angular velocity of the strike. Drag energy was calculated as \( DE = \int_{t_0}^{t_f} \tau_{drag} \), the integral of drag torque \( (\tau_{drag}) \) along the displacement of the strike; therefore, drag energy was a cumulative measure influenced by the displacement of the strike as well as the length and drag index of the striking body, the density of water, and strike velocity. Finally, we calculated strike potential energy as \( PE = (1.25KE_{max}) + DE_{KE_{max}} \); that is, potential energy was equal to 1.25 times the maximum strike kinetic energy, plus the drag energy at the point of maximum kinetic energy.

We extracted velocity and energetics data for each strike and plotted strike angular velocity, kinetic energy, and energy lost as a function of time, up to the point of strike contact (Figure 6h).

Hereafter, we refer to “energy” when discussing either the elastic potential energy of our simulations or the potential energy from our experimental measurements.
4.2.3. Statistical Analyses

Our simulations dataset produced 17 values of energy for each size value. Because our goal was to test how the simulated data matched scaling relationships between experimentally-measured energy and size, we first analyzed the scaling relationships from our experimental dataset.

We established scaling relationships for our experimental dataset using linear mixed models (LMMs) executed through the lmer function in the lme4 package in R version 3.0.1 (Bates, Maechler, Bolker, & Walker, 2015; Core Team, 2008). We first tested if several variables should be included as random effects in our final analysis by testing if they significantly improved the fit of models predicting strike energy from body mass. In all cases, we included Individual ID as a random effect to control for the fact that we took multiple measurements from each individual. We tested for random effects of (1) the date of HSV recording (as Julian date), (2) the fit of the polynomial spline for each strike (7th, 8th, or 9th-order), (3) the order the strike was delivered in the recording sequence (e.g., 1st, 2nd, 3rd…nth strike recorded from each individual in a given sparring or feeding bout), and (4) the order of the bout (e.g., 1st, 2nd, 3rd, time an individual was paired with a competitor). Spline fit, strike order, and bout order had no significant effect on the results and were therefore not included as random effects. Including recording date significantly improved model fit; therefore, we included recording date (1|date) and individual ID (1|ID) as random effects in all models.
To test if strike energy scaled with body mass for sparring and/or feeding strikes, our LMM structure was

$$\log_{10}(potential\ energy) \sim \log_{10}(body\ mass) \times strike\ type + (1|date) + (1|ID),$$

where strike type was either sparring or feeding. To test if maximum angular velocity scaled with body mass for sparring and/or feeding strikes, our LMM structure was

$$\log_{10}(maximum\ angular\ velocity) \sim \log_{10}(body\ mass) \times strike\ type + (1|date) + (1|ID)$$

For both models, we tested if scaling relationships differed between sparring and feeding strikes by comparing the fit of the full model to a model in which we did not include strike type (i.e., a model that grouped all strikes together). A better-fit model was one that had a $\Delta AICc$ of less than -2 (lower AIC scores were better fit; AICc function in MuMIn package in R). For each strike context, we also tested whether the scaling relationships between energy or maximum angular velocity and size were significantly different from zero. We subset our full dataset to two separate datasets: one for only sparring strikes and one for only feeding strikes. Then, for each of these datasets, we tested whether the fit of models that included size and random effects was better than models with only random effects. Since this analysis involved testing the hypothesis that the slope values were significantly different from zero (as opposed to simply comparing
models), use used the $\chi^2$ values and associated p-values from likelihood ratio tests (ANOVA function in R) to test for significance. After establishing the scaling relationships of our experimental dataset, we compared these slopes to our simulation data to find which values of spring compression resulted in similar scaling relationships. We first plotted strike energy as a function of size for both feeding and, separately, sparring strikes. These plots used striking body length as a measure of body size (instead of body mass), as we only could manipulate striking body length in the simulation. Striking body length is closely related to body mass in *N. bredini* ($\log_{10}-\log_{10}$ Pearson correlation $= 0.96$). After plotting experimental data, we superimposed simulation data onto the plot and visually assessed which values of spring compression most closely matched the linear regressions of experimentally-measured strike energy. We subset our simulation dataset to only these values and built (OLS) linear regressions of $\log_{10}$-transformed energy as predicted by $\log_{10}$-transformed striking body length. We built two models from our simulated data: one model for each of sparring and feeding strike data. We did not use LMMs for simulation data, as there were no random effects to include. We extracted the scaling slope and intercept values using the summary function in R. To test how closely the simulation slopes matched the experimental slopes, we used an OLS slope test *via* the slope.test function in R.
4.3 Results

Summary statistics are presented in Table 3. Statistical results from scaling regressions are presented in Table 4. In general, we found differential scaling of strike energetics across behavioral contexts. These scaling relationships matched simulation data that varied the amount of appendage spring compression and resulted in differential scaling of strike velocity across contexts.

Table 3: Summary statistics. Values for strikes / individual are median ± standard deviation (min, max). Values for max angular velocity, potential energy, and body mass are mean ± sd (min, max). The “both” strike context shows individuals for which we had data on both sparring and feeding strikes.

<table>
<thead>
<tr>
<th>Strike context</th>
<th>Number of individuals</th>
<th>Number of strikes</th>
<th>Strikes / individual</th>
<th>Max angular velocity (rad/s)</th>
<th>Potential energy (J)</th>
<th>Body mass (g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>13</td>
<td>104</td>
<td>7 ± 3 (4, 15)</td>
<td>4478 ± 1918.4 (1163, 10320)</td>
<td>0.0109 ± 0.0122 (0.0004, 0.0661)</td>
<td>1.86 ± 0.74 (0.56, 3.23)</td>
</tr>
<tr>
<td>Sparring</td>
<td>11</td>
<td>47</td>
<td>4 ± 1 (2, 7)</td>
<td>4819 ± 1963.2 (1648, 9732)</td>
<td>0.0163 ± 0.0154 (0.0004, 0.0661)</td>
<td>2.08 ± 0.75 (0.81, 3.23)</td>
</tr>
<tr>
<td>Feeding</td>
<td>9</td>
<td>57</td>
<td>6 ± 3 (2, 15)</td>
<td>4197 ± 1850.8 (1163, 10320)</td>
<td>0.0064 ± 0.0057 (0.0007, 0.0352)</td>
<td>1.69 ± 0.68 (0.56, 2.58)</td>
</tr>
<tr>
<td>Both</td>
<td>7</td>
<td>63</td>
<td>9 ± 2 (6, 13)</td>
<td>4740 ± 2039.4 (1586, 10320)</td>
<td>0.0105 ± 0.0129 (0.0004, 0.0661)</td>
<td>1.64 ± 0.63 (0.78, 2.58)</td>
</tr>
</tbody>
</table>
Table 4: Summary of statistical results. Each model is differentiated by data type (experimental or simulation) and the data used in the model (feeding, sparring, or all data combined). For “All” data models, we used AIC scores to compare the fit of a full model to a model without the effect of strike type. For “Feeding” or “Sparring” data models, we tested whether the slope was different from zero using $\chi^2$ values and p-values from likelihood ratio tests. For “Simulation” data, we tested whether the slope was different from zero using the F-value and p-value from a OLS model fit.
<table>
<thead>
<tr>
<th>Data type</th>
<th>Data</th>
<th>Dependent variable</th>
<th>Independent variable(s)</th>
<th>Random effects</th>
<th>Model AICc (random effects AICc)</th>
<th>$\chi^2$, df (p-value)</th>
<th>$\log_{10} - \log_{10}$ model equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>Log10(Potential Energy)</td>
<td>Log10(Body Mass)</td>
<td>Julian Date ID</td>
<td>106.80 (120.65)</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Feeding</td>
<td>Log10(Potential Energy)</td>
<td>Log10(Body Mass)</td>
<td>Julian Date ID</td>
<td>1.90, 1 (0.17)</td>
<td>y = 0.12x - 2.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sparring</td>
<td>Log10(Potential Energy)</td>
<td>Log10(Body Mass)</td>
<td>Julian Date ID</td>
<td>4.30, 1 (0.04)</td>
<td>y = 2.13x - 2.66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feeding</td>
<td>Log10(Potential Energy)</td>
<td>Log10(Striking Body Length)</td>
<td>Julian Date ID</td>
<td>1.56, 1 (0.21)</td>
<td>y = 1.29x + 0.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sparring</td>
<td>Log10(Potential Energy)</td>
<td>Log10(Striking Body Length)</td>
<td>Julian Date ID</td>
<td>6.12, 1 (0.01)</td>
<td>y = 4.68x + 7.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>Log10(Maximum Angular Velocity)</td>
<td>Log10(Body Mass)</td>
<td>Julian Date ID</td>
<td>-56.07 (-48.91)</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Feeding</td>
<td>Log10(Maximum Angular Velocity)</td>
<td>Log10(Body Mass)</td>
<td>Julian Date ID</td>
<td>7.57, 1 (&lt; 0.01)</td>
<td>y = -0.73 + 3.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sparring</td>
<td>Log10(Maximum Angular Velocity)</td>
<td>Log10(Body Mass)</td>
<td>Julian Date ID</td>
<td>0.26, 1 0.61</td>
<td>y = 0.13x + 3.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feeding</td>
<td>Log10(Elastic energy)</td>
<td>Log10(Striking Body Length)</td>
<td>NA</td>
<td>---</td>
<td>$F = 23.35$</td>
<td>$df = 1, 7$ (&lt; 0.01)</td>
<td>y = 1.34 + 0.57</td>
</tr>
<tr>
<td>Sparring</td>
<td>Log10(Elastic Energy)</td>
<td>Log10(Striking Body Length)</td>
<td>NA</td>
<td>---</td>
<td>$F = 142.7$</td>
<td>$Df = 1, 5$ (&lt; 0.01)</td>
<td>y = 4.87x + 8.37</td>
</tr>
</tbody>
</table>
4.3.1 Simulation Data

The model simulations resulted in 17 values of energy for each of the five values of striking body length (85 total values; Figure 7a). For each of the five striking body length values, energy as a function of spring compression were the same. This result is expected because the simulations only quantified the energy stored in the spring; they did not incorporate any metrics related to strike output. We discuss the fit between model simulations and experimental data in a later section.

4.3.2 Experimental Data

4.3.2.1 Energetics

The best fit model predicting strike energy included both body mass and strike type (sparring or feeding) as independent variables, with recording date and individual ID as random effects. Energy scaled positively with body mass for sparring strikes (slope = 2.13) and was constant across mass for feeding strikes (slope = 0.12; Table 4, Figure 7b).

4.3.2.2 Kinematics

The best fit model predicting strike maximum angular velocity included both body mass and strike type (sparring or feeding) as independent variables, with recording date and individual ID as random effects. Maximum angular velocity was constant across mass for sparring strikes (slope = 0.13) and scaled negatively with mass for feeding strikes (slope = -0.73; Table 4, Figure 8).
Figure 7: Experimentally-measured strike potential energy exhibits distinct scaling relationships across behavioral contexts; these results match simulations of elastic energy storage. (a) Simulations of elastic energy storage show increasing energy with increasing spring compression (open circles). Experimentally-measured energy of sparring (green squares) and feeding (orange triangles) strikes overlap with range of simulated data. (b) Scaling relationships of experimental data (sparring: green solid line; feeding: orange dot-dash line) are matched by simulation data for certain values of spring compression (sparring: green circles, green dotted line; feeding: orange circles, orange dotted line). Axes are log₁₀-transformed. Open triangles and squares represent individual strikes; filled triangles and squares
represent means for each value of striking body length. Lines are for visualization purposes; slopes calculated from LMM and OLS models given in main text.

![Graph showing maximum angular velocity vs. body mass for different strike contexts.](image)

**Figure 8:** Strike maximum angular velocity exhibits different scaling relationships between behavioral contexts. When sparring (green squares, solid green line), maximum angular velocity remains constant across mass; when feeding (orange triangles, dot-dash orange line), maximum angular velocity scales negatively with mass. Open triangles and squares represent individual strikes; filled triangles and squares represent means for each value of body mass. Axes are log10-corrected. Lines are for visualization purposes; slopes calculated from LMMs (see main text).

### 4.3.3 Matching Simulation and Experimental Datasets

We found scaling relationships between simulated energy and size (striking body length) that closely matched our experimentally-obtained scaling slopes and intercepts (Table 4, Figure 7b). The scaling relationships between simulated energy and size were not significantly different from those for either sparring (slope test $F = 2.25$, $p = 0.19$) or feeding (slope test $F = 0.04$, $p = 0.86$) strikes.
To match the scaling relationships between simulated and experimental data, sparring strikes required a greater simulated variation in spring compression than feeding strikes. As size increased, spring compression ($\theta_{\text{range}}$) needed to match our sparring strike data increased from 4 degrees (0.005 m striking body length, SBL) to 17 degrees (0.009 m SBL). However, over the same size range, the $\theta_{\text{range}}$ needed to match feeding strike data varied only from 6 (0.005 m SBL) to 9 degrees (0.009 m SBL).

4.3.4 Post-hoc test: The effects of relative size on sparring energetics and kinematics

During preliminary analyses, we found that including the relative size difference between the focal (striking) individual and its opponent significantly improved the fit of a model predicting sparring strike energy from body mass (recording date and individual ID as random effects). While we did not include the effect of relative size in our scaling comparisons between sparring and feeding, we conducted post-hoc tests to establish how sparring strike energetics and kinematics were affected by relative opponent size.

After subsetting the original dataset to only those sparring strikes for which we had data on opponent body mass (9 of 11 individuals), we built mixed models of the form:

$$\log_{10}(\text{energy [or maximum angular velocity]}) \sim \log_{10}(\text{body mass}) + \text{relative mass} + (1|\text{date}) + (1|\text{ID})$$
where relative mass is the mass of the focal individual relative to its opponent (Briffa et al., 2013): \( \text{relative mass} = 1 - \frac{\text{opponent mass}}{\text{focal mass}} \). Relative mass is a predictor of contest success in \( N. \text{bredini} \), such that individuals with higher relative mass are more likely to win contests (P. A. Green & Patek, 2018). We tested whether the full model (i.e., including relative mass) was a better predictor of strike energy or maximum angular velocity than models that did not include relative mass.

The full model, including the effect of relative mass, was the best predictor of sparring strike energy (AICc: 49.17; \( \Delta \text{AICc} \) from model without relative mass = -5.63; likelihood ratio test \( \chi^2 = 9.33, \text{df} = 1, p < 0.01 \)). The slope estimate of the relative mass effect (-2.23) showed that focal individuals that were relatively larger than their opponents used less energy in sparring strikes than focal individuals that were relatively smaller than their opponents.

The full model including relative mass was also the best predictor of sparring strike maximum angular velocity (AICc = -2.48; \( \Delta \text{AICc} \) from model without relative mass = -2.23, likelihood ratio test \( \chi^2 = 6.97, \text{df} = 1, p < 0.01 \)). The slope estimate of the relative mass effect (-0.76) showed that focal individuals that were relatively larger than their opponents struck with lower velocity than focal individuals that were relatively smaller than their opponents.
As an additional way of understanding this effect, we extracted the residual values of energy and maximum angular velocity from the above LMMs and grouped them by whether the focal individual was smaller or larger than its opponent. Relative to their opponents, larger individuals had lower residuals for both energy and velocity (Table 5). Due to small sample size, we did not test whether these groups were significantly different.

**Table 5: Residual values from LMMs testing the effect of relative competitor size on the scaling of sparring strike potential energy and maximum angular velocity.**

<table>
<thead>
<tr>
<th>Relative size difference between focal and opponent</th>
<th>Mean potential energy residual ± sd (min, max)</th>
<th>Mean maximum angular velocity residual ± sd (min, max)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focal smaller than opponent</td>
<td>0.04 ± 0.17 (-0.11, 0.3)</td>
<td>0.02 ± 0.06 (-0.06, 0.12)</td>
</tr>
<tr>
<td>Focal larger than opponent</td>
<td>-0.01 ± 0.03 (-0.05, 0.02)</td>
<td>± 0.08 (-0.07, 0.11)</td>
</tr>
</tbody>
</table>

To ensure that our results were not affected by incorrect matching (i.e., larger or smaller individuals were preferentially matched with relatively larger or smaller opponents), we tested whether relative mass was significantly affected by focal body mass. We fit a LMM of the form \( \text{relative mass} \sim \text{body mass} + (1|\text{date}) + (1|\text{ID}) \) and tested whether this model was a better fit than including only random effects. If larger or smaller individuals were preferentially matched with larger or smaller opponents, then the model including body mass would be a better fit than only the random effects.
We found that the full model was not a better predictor than the model including only random effects (full model AICc = -105.42, ΔAICc from model including only random effects = 5.77, likelihood ratio effects χ² = 0.06, df = 1, p = 0.81). Therefore, relative mass did not show significant directional variation over our variation in focal individual body size.

4.4 Discussion

We tested how energetic and kinematic scaling differs between behavioral contexts, and how this variation is explained by spring compression. As body mass increased, sparring strikes required proportionally greater amounts of energy, whereas feeding strikes required similar energy across mantis shrimp body sizes. These differences in energetic scaling are reflected in the scaling of strike maximum angular velocity: velocity was constant across mantis shrimp body mass for sparring strikes and scaled negatively with mass for feeding strikes. Finally, comparing these experimental results to our simulations revealed that these energetic differences are likely driven by differential compression of the exoskeletal spring that powers strike movement. Our results show (1) that sparring *N. bredini* may use strike energy to assess competitive ability (size), (2) that energy and velocity scaling relationships are altered according to behavioral context, and (3) how animals can achieve variation in energetic and kinematic scaling in spring-driven biomechanical systems. Below, we interpret the
relevance of these results to N. bredini life history and to the broader fields of contest energetics, animal performance, and power-amplified movements.

4.4.1 Sparring

Sparring strike energy scaled positively with body mass and strike velocity was not significantly correlated with mass. These results fit well with the current understanding of N. bredini contest dynamics and suggests a communicative function for energetics in this system. As part of broader contest dynamics, telson sparring functions in mutual assessment of ability that is related to size such that larger individuals have a competitive advantage (P. A. Green & Patek, 2018). Size assessment via striking may be advantageous in N. bredini, as visual assessment may be difficult: contests often occur in enclosed, dark burrows where competitors’ bodies are hidden by their coiled, protective telsons. The present study suggests that receivers of strikes could assess ability via strike energy, but size assessment may also be possible for the striking individual. The telson’s ability to dissipate strike energy scales with body mass (this result was not found for nearby abdominal exoskeleton, J. R. Taylor & Patek, 2010), such that striking individuals could potentially assess competitor size through this dissipation of strike energy (J. R. Taylor & Patek, 2010). Furthermore, our post-hoc tests showing that individuals used lower-energy, lower-velocity strikes against relatively smaller opponents (Results, Table 5) suggests that striking individuals may assess opponent size and alter strikes accordingly. These combined results suggest that energy flow – from the strike of the
“signaler” to the receiving of the strike by the “receiver” – may relay information about body size that helps individuals make competitive decisions.

4.4.2 Feeding

Feeding strike energy was not significantly correlated with body mass, while velocity scaled negatively with mass. In contrast with the presumably high motivation of competing *N. bredini*, these results may reflect a low motivation during feeding, especially for larger individuals. Feeding *N. bredini* take an average of approximately 73 strikes to open snails (Crane, Cox, Kisare, & Patek, in revision), whereas size-matched *N. bredini* contests involve an average of only 5 strikes (P. A. Green & Patek, 2018); it may be more important for competing *N. bredini* to use maximum energy during sparring, as the ultimate outcome (winning or losing) depends on fewer strikes. In our feeding experiments, individuals were also isolated in their protective burrows for as long as needed to process prey – conditions that may facilitate low-stakes, low-energy striking. These conditions may also occur when *N. bredini* is feeding in nature; however, individuals may be interrupted by intruding competitors or predators.

The scaling relationships for feeding showed that smaller individuals were using high-velocity strikes with potentially high energetic costs. We attempted to matched *N. bredini* size with snail size, but we did not directly measure snail size before feeding trials. It may be the smaller individuals were receiving proportionally-larger snails than larger individuals. Although many of our *N. bredini* did not strike snails during their
trials, we did not notice any trends between individual size and the likelihood of striking snails. Furthermore, a recent study (Crane et al., in revision) that matched snail size to *N. bredini* size found that 50% of *N. bredini* immediately consumed size-matched snails during feeding trials. This low likelihood of consuming even size-matched prey may support the idea that *N. bredini*, at least under these conditions, have low motivation. Assuming that the snails we gave our *N. bredini* were well size-matched, these scaling results suggest that processing snails may be costlier for smaller individuals; as they grow, they may be able to process prey using relatively lower-energy strikes. Future work should alter snail size relative to mantis shrimp size to test if energy use changes according to the size or armor of prey.

### 4.4.3 Simulations

Our math model simulations first generated a set of possible scaling relationships and then found a close match between simulated and experimental data by modifying only the range of spring compression. The parameters used in our model matched previous research that found a positive correlation between spring compression and body size, even though spring constant remains similar across the same size range (S. N. Patek et al., 2013; Zack et al., 2009). Although other aspects of the strike mechanism, such as the link lengths of the 4-bar linkage system (Anderson et al., 2014), may change across a size range, we kept these variables constant for simplicity.
Future work could test whether altering these and other parameters change scaling dynamics.

The finding that changing only spring compression can modify energetic scaling has relevance to behavior in this system and other spring-driven (or “power-amplified”) systems. Adjustment of spring compression prior to a strike is a key component of “behavioral control” in the mantis shrimp strike system. Individuals cannot alter strike kinematics once the appendage is released; however, before a strike, they can contract their muscles to achieve greater or lesser spring compression (Kagaya & Patek, 2016). Our results show that animals adjust their spring-driven system based on the behavioral context, in this case, sparring or feeding. Understanding how animals control spring-powered, power-amplified movements is an important topic in biomechanics (Patek et al in press); our results lend insight to both how (spring compression) and why (behavioral context) animals alter these movements.

4.4.4 Measuring animal performance in context

Comparisons across the contexts of feeding and sparring inform broader studies of animal performance. Performance research measures an animal’s ability to execute an ecologically-relevant task (Arnold, 1983), and often attempts to relate these findings to fitness. For example, a series of studies by Husak, Lappin and colleagues (Husak, Lappin, Fox, & Lemos-Espinal, 2006; Husak, Lappin, & Van Den Bussche, 2009; Lappin et al., 2006) identified how bite force in collared lizards predicts competitive and
reproductive success. Many performance measurements are conducted in lab conditions, for example by placing a force sensor inside the animal’s jaws or claws, or directly in front of the animal, until it bites, grasps, or strikes (P. A. Green & Patek, 2015; Husak et al., 2006; Lappin et al., 2006; Wilson et al., 2007). These measurement conditions are both motivationally and mechanically distinct from the scenarios in which the animal performs its ecologically-relevant task. Motivationally, performance is measured in a different behavioral context (e.g., while being handled by a researcher, not in a contest); mechanically, researchers ensure proper measurement technique (e.g., force plate alignment), which is likely not comparable to the conditions of the actual task. When we allowed mantis shrimp to compete or feed freely, we found that behavioral context dramatically altered output performance (strike energy or velocity). We support calls (Losos, Creer, & Shulte, 2002) to measure performance in the context(s) in which it functions, so as to gather more relevant performance data.

4.4.5 Conclusion

Studies of energetic scaling influence our understanding of how animals power a vast range of relevant tasks (Nagy, 2005) and can relate to central behavioral theories (Briffa & Sneddon, 2007). We found differential scaling of energy and velocity in a spring-powered system and showed how this scaling may be achieved by variation in spring compression. Future work testing if and how energetic scaling changes across contexts can help establish how animals accomplish the myriad tasks they face.
Chapter 5 – Conclusions

The results of this dissertation lend insight to our understanding of contest behavior and the role of animal weapons. In Chapter 2, I showed that visual weapon displays did not communicate performance-based information, and that weapon use was not dangerous “combat”, as is often presumed. Instead, ritualized weapon use can help competitors resolve conflicts safely. These results encourage researchers to more closely examine how competitors use their weapons in contests, as these interactions may be more than simply dangerous combat.

In Chapter 3, I conducted a comprehensive test of assessment models, using several pieces of evidence to show that *N. bredini* mutually assess competitive ability. This study shows that researchers studying assessment should use multiple routes to test among assessment models, while giving new analysis techniques to facilitate these tests. Finally, by showing the central role telson sparring plays in contests, I further support the notion from Chapter 2 that weapon use can facilitate efficient conflict resolution.

In Chapter 4, I show how the energy used to power sparring strikes is related to size. This result shows that energy use may facilitate the assessment of size that occurs during sparring: individuals delivering strikes vary, according to body size, in how much energy they put into a strike and therefore in strike velocity. These results reveal a central role of energetics in competitive assessment. By showing that energy dynamics
change between the behavioral contexts of sparring and feeding, I show how animals parse out limited resources to complete essential life history tasks. Finally, I show how energy dynamics relate to the input (spring compression) and output (strike velocity) of the raptorial appendage strike, lending insight to how animals achieve performance variation in power-amplified systems.

Within the mantis shrimp system, continued work is needed across levels of analysis to understand (among other questions): how resource value influences contest behavior, how natural contests (outside of staged arenas) progress and influence overall seagrass bed ecology, and how the telson receives and senses strike forces. These and other studies across taxa are needed to connect the strong theory of contest behavior with recent advances in behavioral and biomechanical analysis. Through these connections, we can lend insight to classic and still-open questions regarding the function and evolution of the myriad traits associated with animal contests.
Appendix A – Supplementary Methods, Figures, and Tables for Ch. 2.

A.1 Methods

We collected 150 individuals of *Neogonodactylus bredini* from the Galeta Marine Laboratory of the Smithsonian Tropical Research Institute (STRI) in Panama (ANAM Collection Permit #SE/A-115-13 and Export Permit #SEX/A-23-14; water temperature: 28.5°C).

A.1.1 Raptorial appendage morphology

To test if meral spreads presented index signals of strike force, we tested whether maximum strike force was correlated with three biomechanically-relevant measures of appendage morphology: merus length, saddle length, and dactyl heel width. The merus segment of the raptorial appendage houses the elastic mechanism that generates the strikes (Mendoza Blanco & Patek, 2014). We defined merus length as the distance between the proximal and distal end of the merus, not including the meral-v (Figure 9). This measurement is similar, though not identical, to that of Mendoza Blanco and Patek (Mendoza Blanco & Patek, 2014), which correlates strongly with muscle physiological cross-sectional area. The saddle is a saddle-shaped part of the merus exoskeleton (S.N. Patek et al., 2004) and is part of the elastic mechanism (Zack et al., 2009). We measured saddle length as the distance between two black spots that lie on the proximal and distal
edges of the saddle (Figure 9). The dactyl heel is the hammer-shaped part of the proximal dactyl that contacts competitors during a strike. We defined dactyl heel width as the lateral to medial width of the dactyl heel, measured at its widest part (Figure 9).

We took 3 replicates of each measurement from 142 total individuals within one week of collection (Mitutoyo Digimatic Caliper, range: 0-150mm, resolution: 0.01mm, Mitutoyo Corp., Kawasaki, Japan). We used the mean of three replicates of each measurement in all analyses. Measurement error was assessed using an ANOVA and was less than 5% of the total variance in lengths (Klingenberg & McIntyre, 1998; Warton, Wright, Falster, & Westoby, 2006).

A.1.2 Body length, body mass, and maximum strike force measurements

Body length was defined as the distance from the tip of the rostrum to the apex of the telson (Ahyong, 2001). On the day of each individual’s collection, we measured each individual’s body length.

We measured body mass on the day of each individual’s use in a contest. Between each of three replicates, the individual was removed from the scale, the scale was wiped dry and re-zeroed, and the individual was reweighed. Mass was measured using either a Mettler AE163 (range: 0-31g, readability: 0.00001g, Mettler-Toledo, LLC, OH, USA) or a Denver Instruments APX-3202 balance (range: 0-3100g, readability: 0.01g, Sartorius AG, Goettingen, Germany).
Maximum strike force was measured over several days between each individual’s collection date and its use in a contest. We presented each individual with a one-axis force sensor (model W200B02/002C10, force range 444.8N, upper frequency limit 75kHz, PCB Piezotronics, Depew, NY USA) mounted on a heavy metal rod and connected to a computer via a signal conditioner (Model 480C02, PCB Piezotronics) and data acquisition board (sample rate: 500,000 samples s\(^{-1}\), NIDAQ USB-6251, National Instruments, Austin, TX USA). We allowed each individual to strike the sensor repeatedly. If necessary, we presented the sensor over subsequent hours or days to collect at least 10 strikes from as many individuals as possible. Strike data were converted to peak force (Matlab R2013b 32-bit, The MathWorks, Inc., Natick, MA USA; R Version 3.0.1, R Foundation for Statistical Computing, Vienna, Austria). We used the maximum peak force of the first 10 strikes as a metric of maximum strike force. We collected 10 strikes from each of 81 individuals, and removed one individual from the dataset that seemed to have an injury that prevented it from striking at full force. Thus, the full dataset used for force and morphology correlations consisted of measurements for each of 80 individuals.

**A.1.3 Contest behaviour**

Behavioural tests were performed in Panama (Naos Marine Laboratory, STRI) and at Duke University. Each individual was sexed and checked for its moult condition and presence of eggs. Individuals who moulted or were found with eggs after collection
were not used in experiments for at least 7 days. Their morphology was re-measured after moulting.

At least one day after collecting final strike force data for each individual, two sex- and body length-matched competitors were paired in a contest. Contests were conducted in clear plastic arenas (32.5cm L x 20cm W x 11.5cm D). One individual was randomly chosen via a coin flip as the ‘resident’ and was placed in the arena in an artificial burrow made of clear vinyl tubing. The burrow had only one entrance, was covered in black vinyl tape except for a clear area at the top of the burrow, and was placed approximately 5cm from one end of the arena. Burrow volume was matched for individual body length within 10% (mean±sd=2.7 ± 2.2%, range=0.05-8.9%) following N. bredini-specific equations derived by Steger (Steger, 1987). Each resident was allowed to acclimate for approximately 3 to 6 hours in the burrow. There was no significant relationship between resident acclimation time and the probability that a resident won a contest (generalized linear model with binomial error distribution, df=31, p=0.992). After the resident acclimation period, a sex- and body length-matched individual (the ‘intruder’) was introduced into the arena behind a barrier made of gray laminated paper placed approximately 9 cm from the burrow entrance. The intruder was given 10 minutes to acclimate, after which the barrier was removed and the behavioural observation period began.
We documented each competitor’s behaviour during a contest using video and audio recording. In order to view the entire arena, we recorded contests using two video cameras (Hero 3+ Silver Edition, GoPro, Inc., USA). One camera was placed directly over the top of the arena, while a second camera was placed facing the broad side of the arena. Video recordings were made at 60 frames/sec and 1920 x 1080 px resolution. We also used a handheld audio recorder (ICD-PX312, Sony Electronics Inc., San Diego, CA USA) to dictate each individual’s contest behaviour. We documented each behaviour of both the resident and the intruder, following an ethogram similar to that of Dingle and Caldwell (Dingle & Caldwell, 1969). We discarded contests in which both competitors did not notice each other for 10 minutes, or one individual clearly did not interact with the other. Our observational dataset consisted of 34 sex- and body length-matched contests from 68 unique individuals.

A.1.4 Statistical analyses

In all statistical tests, we set our significance threshold at p=0.05. In tests involving multiple comparisons, we did not correct our significance levels using a Bonferroni correction; instead, we used measures of effect size as recommended by Garamszegi (Garamszegi, 2006) and Nakagawa (Nakagawa, 2004).

A.1.4.1 Morphology and force correlations

We tested whether maximum strike force was predicted by body length, body mass, merus length, saddle length, and dactyl heel width by regressing these variables
using a generalized linear model (GLM, stats package in R, version 3.0.1) (Crawley, 2007). In our model, we specified a Gamma error distribution because the response variable (maximum strike force) was right-skewed and could have only positive values (Zuur, Ieno, Walker, Saveliev, & Smith, 2009). We also used a log-link function to log-transform the strike force data when building the model (Crawley, 2007). After building the full model with all interactions, we used an Akaike Information Criterion (AIC) approach to sequentially reduce the model, removing non-significant terms and choosing the simplest, best-fit model as defined by the lowest AIC score (Crawley, 2007). We calculated partial correlation coefficients between maximum strike force and each morphological measurement (Table 6).

To visualize the relationships between raptorial appendage morphology (merus length, saddle length, dactyl heel width) and maximum strike force, and as a comparison to our GLM-based approach, we calculated the residuals of each variable from a log-transformed ordinary-least squares regression on body length. We then regressed these residuals against each other. This approach allowed us to plot relationships between variables after removing the effects of body size (Figure A.1). While the results of the residual approach matched those of the GLM-based approach, we used the GLM-based approach for hypothesis testing (Darlington & Smulders, 2001; Garcia-Berthou, 2001).
A.1.4.2 Contest behaviour and correlates of contest success

Previous research on *N. bredini* found that residents were more likely to win contests than intruders (Adams & Caldwell, 1990; Steger & Caldwell, 1983). In a preliminary test, we found no significant difference in the number of winners who were residents versus those who were intruders (resident winners: 22, intruder winners: 12, \( \chi^2=2.94, \text{df}=1, \text{p}=0.0864, \text{effect size} = 0.086 \) (S. B. Green & Salkind, 2003)). Thus, in our dataset, residents were no more likely to win contests than intruders and we did not use residency as a factor in our analyses. The lack of resident effect may be due to the 3-6 hour time we allowed residents to acclimate in the burrow; studies that acclimated residents for up to 12 hours did see a resident effect (Adams & Caldwell, 1990; Steger & Caldwell, 1983). Potentially, if we had acclimated residents for more than 6 hours, we may have seen a residency effect. Nonetheless, residents vigorously defended burrows during contests, behavioural evidence that suggests residency time was not a large factor in a resident’s motivation to defend its territory.

We tested if the proportion of contests that began with at least one competitor presenting a meral spread was significantly greater than the proportion that began with any other behaviour. We also compared the proportion of contests that were resolved by meral spreads to the proportion resolved by strikes. Finally, in contests that escalated to telson striking, we tested whether the proportion of strikes received on a competitor’s telson was greater than the proportion received anywhere else on the body. All
proportion tests were conducted using a one-sided test of proportions (prop.test function, stats package in R, version 3.0.1).

We also tested whether winners and losers of contests differed in body length, body mass, maximum strike force, and the number of strikes delivered during a contest. For each contest and each variable, we subtracted the loser’s value from the winner’s value, so our final dataset consisted of each contest’s winner minus loser value for each variable. Although competitors were body length-matched within 5%, we included body length in our analyses given that variation within this range may affect contest success.

Our null hypothesis was that winners and losers would not differ in their values for a given variable. Thus, our null hypothesis predicted that winner minus loser values should not differ significantly from zero. For maximum strike force, we predicted that winners would have greater maximum strike force than losers. Therefore, our alternative hypothesis for this variable was that winner minus loser maximum strike force would be significantly greater than zero. For all other measured variables, we only tested whether winner minus loser values were significantly different from zero.

We evaluated statistical significance using t-tests and Wilcoxon tests, as well as measures of effect size (Cohen, 1988) and their 95% confidence intervals (CIs). We used a one-sided t-test to test if winner minus loser strike force was significantly greater than zero, and we used two-sided t-tests to test if winner minus loser body length and body mass were significantly different from zero. Because strike number was not normally
distributed, we used a two-sided Wilcoxon test for this variable only. We also calculated the Hedge’s g effect size for the winner minus loser values of each variable (Hedges & Olkin, 1985). We created 95% CIs for effect sizes using nonparametric bootstrapping methods in the bootES package (Kirby & Gerlanc, 2013) in R (version 1.01, 10000 replicates, CI method:BCa). While recent studies have used generalized linear models to control for correlations between predictor variables (e.g. (Chelliah & Sukumar, 2013; McCullough & Emlen, 2013)), we were unable to use this method due to sample size constraints. Instead, we used correlation coefficients to identify potential multicollinearity, and instances of high correlation are mentioned in Results in the main text.

**A.2 Results**

**A.2.1 GLM and residual correlation results**

The results from the GLM analysis are presented in Table 6. No variable was a strong correlate of maximum strike force. The independent effect with the greatest partial r-squared value was saddle length, which predicted on 5.9% of variation in maximum strike force.

Relationships between residuals of appendage morphology and residuals of maximum strike force showed the same trends as those of the GLM approach (Figure 9). There was a positive relationship between residual strike force and residual saddle length, and no relationship between residual strike force and residual merus length or
residual dactyl heel width. Despite the statistical significance of the relationship between residual saddle length and residual strike force, the high amount of variation indicates that, for a given saddle size, strike force is highly variable and thus saddle size is unlikely to be a reliable signal of strike force (Hughes, 2000).

Table 6: Appendage morphology does not strongly correlate with maximum strike force. Retained explanatory variables and interaction terms from a simplified GLM correlating strike force with body length, body mass, merus length, dactyl heel width, and saddle length (full model AIC = 879.92, simplified model AIC = 853.75).

<table>
<thead>
<tr>
<th>Explanatory variable</th>
<th>Estimate</th>
<th>Standard error</th>
<th>t-value</th>
<th>Partial r-squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-5.32</td>
<td>13.94</td>
<td>-0.38</td>
<td>0.002</td>
</tr>
<tr>
<td>Body length</td>
<td>1.15</td>
<td>0.83</td>
<td>1.381</td>
<td>0.028</td>
</tr>
<tr>
<td>Body mass</td>
<td>14.26</td>
<td>11.98</td>
<td>1.190</td>
<td>0.021</td>
</tr>
<tr>
<td>Merus length</td>
<td>-2.58</td>
<td>1.14</td>
<td>-1.79</td>
<td>0.047</td>
</tr>
<tr>
<td>Dactyl heel width</td>
<td>-34.80</td>
<td>18.77</td>
<td>-1.85</td>
<td>0.050</td>
</tr>
<tr>
<td>Saddle length</td>
<td>21.07</td>
<td>10.40</td>
<td>2.02</td>
<td>0.059</td>
</tr>
<tr>
<td>Body length : Body mass</td>
<td>-0.82</td>
<td>0.54</td>
<td>-1.50</td>
<td>0.033</td>
</tr>
<tr>
<td>Body mass : Merus length</td>
<td>-1.30</td>
<td>0.68</td>
<td>-1.91</td>
<td>0.053</td>
</tr>
<tr>
<td>Body mass: Dactyl heel width</td>
<td>19.06</td>
<td>12.45</td>
<td>1.53</td>
<td>0.035</td>
</tr>
<tr>
<td>Merus length: Dactyl heel width</td>
<td>3.27</td>
<td>1.70</td>
<td>1.92</td>
<td>0.054</td>
</tr>
<tr>
<td>Body length: Saddle length</td>
<td>-0.63</td>
<td>0.52</td>
<td>-1.20</td>
<td>0.022</td>
</tr>
<tr>
<td>Body mass: Saddle length</td>
<td>0.81</td>
<td>7.39</td>
<td>0.11</td>
<td>0.0002</td>
</tr>
<tr>
<td>Dactyl heel width: Saddle length</td>
<td>-1.03</td>
<td>14.93</td>
<td>-0.07</td>
<td>0.00007</td>
</tr>
<tr>
<td>Body length: Body mass: Saddle length</td>
<td>0.43</td>
<td>0.29</td>
<td>1.47</td>
<td>0.032</td>
</tr>
<tr>
<td>Body mass: Dactyl heel width: Saddle length</td>
<td>-10.18</td>
<td>6.78</td>
<td>-1.50</td>
<td>0.034</td>
</tr>
</tbody>
</table>
Figure 9: Residuals of maximum strike force on body length show a positive correlation with residuals of saddle length (B), but not with residuals of merus length (A) or dactyl heel width (C). However, high variation (low $r^2$) in the residual results suggests saddle length explains a low amount of variation in strike force (see Results). Raptorial appendage schematics (Insets, top left of each plot; distal to the left, dorsal to the top) depict where each measurement was taken on the appendage.

A bold font $p$-value and a regression line indicates a statistically significant correlation ($p < 0.05$).
Appendix B: Supplementary Methods, Tables, and Figures for Chapter 3

B.1 Methods

B.1.1 Animal Collection and Maintenance

*Neogonodactylus bredini* is widely distributed in seagrass bed habitats across the Caribbean. We collected *N. bredini* individuals from burrows in coral rubble in *Thalassia* spp. seagrass habitats on the Atlantic coast of Panama. All individuals were collected near the Galeta Marine Laboratory (GML) of the Smithsonian Tropical Research Institute (STRI, 09° 24.279’ N, 079° 52.245’ W) between January and April 2014 (body length-matched contests; ANAM Collection Permit #SE/A-115-13) and in October 2015 (randomly-matched contests; ANAM Collection Permit #SE/A-92-15).

Individuals were maintained in a flow-through seawater system at GML in small, perforated plastic bottles. Some individuals were later transported to the STRI Naos Island Laboratory (body length-matched contests) or to Duke University (Durham, NC, USA; 37 body length-matched individuals, ANAM Export Permit #SEX/A-23-14; 57 randomly-matched individuals, ANAM Export Permit SEX/A-106-15). All individuals housed at Naos Island or Duke were maintained in plastic cups filled with natural seawater (Naos Island) or artificial seawater (Duke, mean water temperature 27.2 deg C, 12h/12h light/dark schedule), or in dedicated tanks filled with circulating artificial
seawater (Duke, mean water temperature 27.2 deg C, 12h/12h light/dark schedule). All cups and tanks contained sandy substrate and a piece of PVC or vinyl tubing as an artificial burrow. Individuals were fed frozen krill or frozen seafood twice weekly and had at least 10% of their water changed daily.

**B.1.2 Morphological Measurements**

We measured each individual’s body length – defined as the distance from the tip of the rostrum to the left medial apex of the telson (Ahyong, 2001) – within one week of collection using digital callipers (Mitutoyo Digimatic Caliper, range: 0-150mm, resolution: 0.01mm, Mitutoyo Corp., Kawasaki, Japan).

We measured body mass either on the morning of each individual’s use in a contest (body length-matched contests; Mettler AE163 balance; range: 0-31g, readability: 0.00001g, Mettler-Toledo, LLC, OH, USA. Denver Instruments APX-3202 balance; range: 0-3100g, readability: 0.01g, Sartorius AG, Goettingen, Germany) or in the evening after each individual’s use in a contest (randomly-matched contests; Denver Instruments APX-3202). For both body length and body mass, we used the mean of three replicate measurements in all analyses.

**B.1.3 Contest Protocol**

All contests were conducted in clear plastic arenas (32.5cm length x 20cm width x 11.5cm depth) with 3 cm of sandy substrate on the bottom and a laminated piece of centimetre graph paper on one of the broad sides.
B.1.3.1 Randomly-matched Contest Protocol

In randomly-matched contests, the night before a contest we randomly selected one individual (using the sample function in R version 3.0.1; Core Team, 2008) as a “resident”. We gave this individual 8-18 hours to establish residency in a single-entrance artificial burrow made of clear vinyl tubing. Except for a clear area at the top (to facilitate observations inside the burrow), the burrow was wrapped in black vinyl tape. Burrow volume was matched for resident body length according to *N. bredini*-specific equations established by (Steger, 1987). The burrow was placed approximately 5cm from one of the short ends of the arena. If the resident did not establish residency in the burrow before the contest began, it was removed from the arena and was used as an intruder in another, randomly-selected contest. We then randomly selected another individual and allowed it to establish residency in the burrow over the following night.

Once a resident had established itself in a burrow, we placed a grey laminated barrier approximately 9cm from the burrow entrance and introduced a second randomly-selected individual (the “intruder”) behind the barrier. We allowed the intruder to acclimate for 10 minutes, then raised the barrier and began videotaping contest behaviours.

Contests were videotaped for 20 minutes using two video cameras (Hero 3+ Silver Edition, GoPro, Inc., USA, 60fps, 1920 x 1080px resolution) placed orthogonal to each other; one facing down over the arena and one facing the broad side of the arena. If
competitors did not interact with each other for 20 minutes, we removed both to be randomly 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 contests (~20% of initiated trials) and 7 body length-matched contests (~17% of initiated trials) due to a lack of interactions or the presence of one-sided interactions.

**B.1.3.2 Body length-matched contest protocol**

Our dataset from body length-matched contests is a more detailed analysis of contests previously studied in (P. A. Green & Patek, 2015). In (P. A. Green & Patek, 2015), we matched competitors for body length, because we had not yet established a metric of RHP. In (P. A. Green & Patek, 2015), we also analysed the first contest behaviour, the number of strikes delivered during contests, each individual’s body and appendage morphology, and each individual’s strike performance (performance sensu Arnold, 1983; maximum peak strike force; Irschick, Meyers, Husak, & Le Galliard, 2008). In the present study, we used body length and body mass data from the same individuals, but expanded our behavioural analysis beyond the first contest behaviour and the total number of strikes to include contest duration and all contest behaviours.
B.1.3.3 Differences between randomly-matched and body length-matched contest protocols

The protocol for body length-matched contests is similar to that of randomly-matched contests (details in P. A. Green & Patek, 2015). However, there are differences between the protocols and data of randomly-matched and body length-matched contests. For example, body length-matched residents were given 3-6 hours to establish residency, whereas randomly-matched residents were given 8-18 hours. Additionally, randomly-matched and body length-matched contests differed in the average body mass of competitors (randomly-matched mean ± SD: 1.03 ± 0.49; body length-matched mean ± SD: 1.65 ± 0.54).

Despite the differences in protocol, we found no evidence that these differences significantly affected our results. Our previous work (P. A. Green & Patek, 2015) found no effect of resident acclimation time between 3-6 hours on the likelihood of a resident winning a body length-matched contest. Therefore, it is unlikely that differences in acclimation time significantly affect our conclusions.

In a post-hoc analysis, we found six data points from randomly-matched competitors that were paired within a similar degree of body length as our body length-matched competitors. We plotted these points within the RHP-cost plots of our body length-matched dataset, and found that the randomly-matched points fit well within the range of our body length-matched points (Figure 11). While there is insufficient power
to test for statistical differences between the groups, we believe this evidence shows that our conclusions should not be affected by differences in protocol.

**B.1.4 Contest Data**

We divided contests into “bouts”: a bout began when individuals first made eye contact (a “track” behaviour) or when one individual approached the other (an “approach” behaviour). A bout ended when one individual made a clear, directed movement away from its competitor and toward the edge of the contest arena (i.e., after a “retreat” behaviour). The winner was the individual that resided in the burrow after its competitor’s retreat. We only analysed data from the first contest bout.

From the first bout, we analysed total bout duration in seconds (also termed “total contest duration”). We also coded all contest behaviours using JWatcher (Blumstein et al., 2007), following an ethogram similar to that of Dingle & Caldwell (Dingle & Caldwell, 1969). Our ethogram consisted of 14 mutually exclusive contest behaviours (Table 7).

**B.1.5 Analysis Techniques**

We conducted separate analyses for the body length-matched and randomly-matched datasets. We removed four total outliers from the datasets – two from the randomly-matched and two from the body length-matched datasets. We statistically identified outliers based on the criterion that they had a contest duration greater than three standard deviations from the mean (uv.outliers function, BIOSTATS package, R...
Version 3.0.1; McGarigal, 2013). The resulting trends were the same with and without inclusion of the outliers (Tables 13 & 14). All data were analysed using R Version 3.0.1 (Core Team, 2008).

B.1.5.1 RHP Analysis

Body size is the most common metric of RHP across the assessment literature (Briffa et al., 2013). We compared two measures of body size as potential RHP metrics, body mass (g) and body length (mm), following the method of (2013). We randomly selected one competitor from each contest as a “focal” individual. We then created a metric of focal body length or mass relative to opponent body length or mass:

\[
\text{relative measure} = 1 - \frac{\text{opponent measure}}{\text{focal measure}}
\]

We tested the effect of relative mass and relative body length, as well as their interaction, on focal contest outcome (win or lose) using a binomial generalized linear model (GLM) with a log link function (Crawley, 2007). The full model was:

\[
focal \ outcome \sim relative \ mass + relative \ body \ length + relative \ mass : relative \ body \ length
\]

Because we were interested in which metric(s) best predicted outcome, not necessarily the significance of the models, we compared the support of all possible models using an Akaike Information Criterion (AIC) score. We looked at which models
had the lowest AIC score (within 2 delta AIC; Burnham et al., 2010) and chose which variable appeared most frequently in those best-supported models as our metric of RHP.

**B.1.5.2 Effects of RHP and Residency on Contest Outcome**

Given that both RHP and resource ownership can affect contest outcome (Kokko et al., 2006), once we established a metric of RHP we tested for the effects of RHP and burrow residency on contest outcome in two ways. First, we ran a binomial GLM with focal outcome (win/lose) as the dependent variable and relative RHP, focal role (resident or intruder), and their interaction as explanatory variables:

\[
\text{focal outcome} \sim \text{relative RHP} + \text{focal role} + \text{relative RHP : focal role}
\]

Since we were interested in how these combined predictors affected contest outcome and we were not focusing on which variable best predicted outcome, we did not use AIC to compare the support of individual models. Instead, we assessed the relative contributions of RHP, residency, and their interaction to the determination of contest outcome by significance values and effect sizes (z-values) in the full model. As recommended by (2009) and (2010), to avoid increasing the possibility of Type I error we did not simplify models and test significance.

We also tested for the “effective RHP increase” experienced by residents, which is defined as the benefit, in RHP units, of residency (Kasumovic et al., 2010; Kokko et al., 2006). In this case, we ran a binomial GLM with the (binary, win/lose) probability of an
intruder winning the contest as the dependent variable and the intruder minus resident RHP as the independent variable:

\[ P(\text{intruder win}) \sim (\text{intruder RHP} - \text{resident RHP}) \]

We used the slope and intercept estimates of this binomial equation to calculate the RHP value at which the intruder had a 50% chance of winning the contest (the inflection point of the resulting binomial plot). If there were no resident effect, the intruder would have a 50% chance of winning at an intruder minus resident RHP value of 0. If there were a residency effect, we would expect this relationship to be right-shifted; that is, an intruder would have a 50% chance of winning the contest only if its RHP were greater than that of the resident (Kokko et al., 2006).

**B.1.5.3 Correlational tests of assessment models**

We used multiple linear regression to test the effects of RHP and residency on two measures of contest costs: total contest duration (in seconds) and the total number of strikes exchanged during a contest. Both contest duration and number of strikes were log-transformed [\( \log_{10}(\text{duration}) \) and \( \log_{10}(1 + \text{number of strikes}) \)] to meet assumptions of normality.

For randomly-matched contests, we followed methods recommended by (2003) and (2008) to test correlations between winner and loser RHP and contest costs. We ran two multiple regressions, each with one cost variable (duration or number of strikes) as
the dependent variable. 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. We included these interaction terms because contests are inherently an interaction between two individuals; including interaction terms makes the statistical model more realistic, given the biology of the system (Elias et al., 2008). The full model was:

$$\log_{10}(\text{cost variable}) \sim (\text{winner RHP} \times \text{winner role}) + (\text{loser RHP} \times \text{loser role}) + (\text{winner RHP} : \text{loser RHP interaction}) + (\text{winner role} : \text{loser role interaction})$$

We did not simplify the multiple regression models to avoid statistical issues associated with testing significance on simplified models (Mundry & Nunn, 2009). Instead, we assessed the direction, strength, and statistical significance of winner and loser effects for the full models (P. W. Taylor & Elwood, 2003).

For body length-matched contests, we tested for the presence of correlations across averaged competitor RHP, competitor residency, and contest costs (Arnott & Elwood, 2009; P. W. Taylor & Elwood, 2003). Our multiple regression approach was similar to analyses performed on the randomly-matched contests; however, our independent variables included the averaged RHP of both competitors and the interaction term between winner and loser role:
\[ \log_{10}(\text{cost variable}) \sim \text{averaged competitor } RHP \]

\[ + \text{(winner role : loser role interaction)} \]

As in randomly-matched contests, we did not simplify the multiple regression models for body length-matched contests; instead, we assessed the directionality and statistical strength of each effect for the full model (P. W. Taylor & Elwood, 2003).

**B.1.5.6 Sequential Behavioural Analysis**

Sequential analysis is a technique developed by social scientists to analyse trends in inter- and intra-individual behaviour. We used sequential analysis to visualise how competing \textit{N. bredini} individuals progressed through behaviours during contests, and we then assessed whether these trends were congruent with predictions made by competitive assessment models (Briffa & Elwood, 2009; Egge et al., 2010; Enquist et al., 1990).

One purpose of a sequential analysis is to test if the observed set of behavioural transitions are different from what one would expect by chance. Additionally, one might ask which transitions contributed to this difference. In these cases, it is useful to simplify the overall dataset to identify individual transitions that occurred more frequently than expected. We used common methods in sequential analysis to simplify our behavioural sequence data.
After coding all contest behaviours from the first contest bout, we combined these behaviours into sequences for all body length-matched competitors and, separately, all randomly-matched competitors.

In these sequential datasets, each behavioural transition is represented in one row of data. For each row, column 1 represents the first behaviour in the transition, and column 2 the second behaviour. In each subsequent row, column 1 shows the second behaviour (i.e., column 2) from the preceding row and column 2 shows the behaviour that follows from it. In [row, column] notation, [r1, c1] is the first behaviour in a sequence, and [r1, c2] is the behaviour that follows from it. In the next row, [r2, c1] = [r1, c2], and [r2, c2] is the behaviour that follows from [r2, c1]. The table below shows a representative sample of coding from the dataset of randomly-matched competitors:

<table>
<thead>
<tr>
<th>Transition 1</th>
<th>Preceding behaviour</th>
<th>Subsequent behaviour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transition 2</td>
<td>t</td>
<td>a</td>
</tr>
<tr>
<td>Transition 3</td>
<td>a</td>
<td>m</td>
</tr>
<tr>
<td>Transition 4</td>
<td>m</td>
<td>r</td>
</tr>
<tr>
<td>Transition 5</td>
<td>r</td>
<td>b</td>
</tr>
</tbody>
</table>

This dataset represents the sequence of transitions “x $\rightarrow$ t $\rightarrow$ a $\rightarrow$ m $\rightarrow$ r $\rightarrow$ b”.

The behavioural transitions are organized into a contingency table (Bakeman & Quera, 1995) or, in the network analysis notation used by the igraph package in R (Csardi & Nepusz, 2006), an “adjacency matrix”. The adjacency matrix simply tallies the
number of times one behaviour transitioned to another across the total dataset. Each row in the adjacency matrix represents the preceding behaviour in a sequence, and each column represents a subsequent behaviour. The table below shows a portion of the adjacency matrix from our full dataset on randomly-matched competitors:

<table>
<thead>
<tr>
<th>Preceding behaviours</th>
<th>Subsequent behaviours</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bout start</td>
</tr>
<tr>
<td>Bout start</td>
<td>0</td>
</tr>
<tr>
<td>Track</td>
<td>0</td>
</tr>
<tr>
<td>Approach</td>
<td>0</td>
</tr>
<tr>
<td>Meral spread</td>
<td>0</td>
</tr>
</tbody>
</table>

In this adjacency matrix, “bout start” was followed by “track” 44 times, by “approach” 11 times, and by “meral spread” once. Also, “track” was followed by “approach” 28 times, and by “meral spread” 8 times.

At this point, the dataset may be simplified such that we can identify which transitions occurred more frequently than would be expected by chance. For example, in our data above, we might ask if the eight transitions from “track” to “meral spread” are more common that we would expect, given the frequencies of each individual behaviour.

Permutation tests can be used to simplify the sequential data to only transitions that occurred more frequently than expected by chance (Bakeman et al., 1996). We
developed a permutation procedure for sequential data in R. The code for executing this procedure is available upon request.

The permutation procedure establishes the 95% quantile of 10,000 randomized datasets holding constant only the number of times each behaviour was observed in the original dataset. This creates a null distribution of expected transitions where any behaviour can transition to any other behaviour, only constrained by the frequency of each behaviour’s occurrence.

The procedure first resamples the second column of the two-column raw behavioural transitions dataset. This technique keeps constant the frequency of individual behaviours, but reshuffles the sequence of these behaviours.

The dataset is resampled 10,000 times. After each resampling, the raw transition data is organized into and saved as an adjacency matrix. Therefore, at the end of the permutation procedure, 10,000 randomized adjacency matrices are saved.

Our goal is to identify which transitions in our observed dataset occurred more frequently than an upper quantile of our expected dataset. From our permuted adjacency matrices, we extracted the 95% quantile for each cell, over the entire adjacency matrix. We would expect transitions that significantly contribute to the structure of our observed dataset to have a higher adjacency matrix value than the 95% quantile. The table below shows the observed transition values and (in parentheses) the 95% quantile
values for a subset of our randomly-matched dataset. Cells in bold and italics represent transitions that occurred more frequently than expected:

<table>
<thead>
<tr>
<th>Preceding behaviours</th>
<th>Subsequent behaviours</th>
<th>Bout start</th>
<th>Track</th>
<th>Approach</th>
<th>Meral spread</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bout start</td>
<td></td>
<td>0 (0)</td>
<td></td>
<td>44 (17)</td>
<td>11 (11)</td>
</tr>
<tr>
<td>Track</td>
<td></td>
<td>0 (0)</td>
<td>0 (20)</td>
<td></td>
<td>28 (13)</td>
</tr>
<tr>
<td>Approach</td>
<td></td>
<td>0 (0)</td>
<td>15 (13)</td>
<td>0 (8)</td>
<td></td>
</tr>
<tr>
<td>Meral spread</td>
<td></td>
<td>0 (0)</td>
<td>0 (9)</td>
<td>2 (6)</td>
<td>0 (4)</td>
</tr>
</tbody>
</table>

From this example, the transitions “bout start $\rightarrow$ track”, “bout start $\rightarrow$ approach”, “track $\rightarrow$ approach”, “approach $\rightarrow$ track” and “approach $\rightarrow$ meral spread” occurred more frequently than expected. While “track $\rightarrow$ meral spread” was observed 8 times, it was not observed more frequently than expected (9 times).

Once the observed dataset has been simplified to those behaviours that occurred more frequently than expected, the transition data can be plotted using the igraph package or any other data visualisation technique.

We converted the matrices of significant transitions into graph objects (igraph package), then plotted these graphs as networks (one network for each of randomly-matched and body length-matched behaviours). Individual behaviours are represented as network vertices (circles), and statistically significant transitions between behaviours are represented as directed network edges (arrows). We grouped the size of the vertices to be proportional to five categories of scaled degree – the percentage of total contest...
behaviours made up by one behavioural state. Similarly, we grouped the size of the edges to be proportional to five categories of transitional probability – the number of times a transition from one behaviour to another occurred divided by the sum of transitions from that behaviour to all others (higher values are more likely transitions).

**B.1.6 Ethics Note**

All methods complied with legal requirements of the Panamanian Minesterio 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 #SEX/A-23-14 and #SEX/A-106-15).
### Table 7: Ethogram of contest behaviours. All behaviours are mutually exclusive. Further definitions are provided in Methods.

<table>
<thead>
<tr>
<th>Behaviour name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bout start</td>
<td>Start of contest bout, coding begins</td>
</tr>
<tr>
<td>Still</td>
<td>No movement</td>
</tr>
<tr>
<td>Out of sight</td>
<td>Individual cannot be seen by human observer</td>
</tr>
<tr>
<td>Track</td>
<td>Track competitor; often clear eyestalk movement or “eye contact”</td>
</tr>
<tr>
<td>Approach</td>
<td>Clear, directed movement toward competitor</td>
</tr>
<tr>
<td>Antennular flick</td>
<td>Antennae/antennules are rapidly moved medially and laterally</td>
</tr>
<tr>
<td>Meral spread</td>
<td>Raptorial appendages spread laterally</td>
</tr>
<tr>
<td>Lunge</td>
<td>Rapid movement toward competitor, much faster than approach</td>
</tr>
<tr>
<td>Strike</td>
<td>Strike competitor by using raptorial appendages</td>
</tr>
<tr>
<td>Telson coil</td>
<td>Coil telson in front of body</td>
</tr>
<tr>
<td>Telson push</td>
<td>While in coiled position, rapidly push telson toward competitor</td>
</tr>
<tr>
<td>Retreat</td>
<td>Clear, directed movement away from competitor, toward edge of arena</td>
</tr>
<tr>
<td>Chase</td>
<td>Rapidly follow retreating competitor</td>
</tr>
<tr>
<td>Bout end</td>
<td>End of contest bout, coding stops</td>
</tr>
</tbody>
</table>
Table 8: In randomly-matched contests, body mass was a better predictor of contest outcome than body length (BL); body length and body mass had similar support in body length-matched contests. For randomly-matched and body lengthmatched contests, models are ordered by AIC score (lowest score indicates best support). Models with ΔAIC of < 2 from best-supported model are italicized, representing approximately equivalent support. Terms separated by a colon represent an interaction term; e.g., “mass : BL” is the interaction between mass and body length.

<table>
<thead>
<tr>
<th>Model terms</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomly-matched</td>
<td></td>
</tr>
<tr>
<td>BL+ mass + mass : BL</td>
<td>39.4</td>
</tr>
<tr>
<td>Mass + mass : BL</td>
<td>39.7</td>
</tr>
<tr>
<td>Mass</td>
<td>44.1</td>
</tr>
<tr>
<td>BL + mass : BL</td>
<td>44.2</td>
</tr>
<tr>
<td>BL</td>
<td>44.8</td>
</tr>
<tr>
<td>BL + mass</td>
<td>46.0</td>
</tr>
<tr>
<td>Mass : BL</td>
<td>47.5</td>
</tr>
<tr>
<td>Body length-matched</td>
<td></td>
</tr>
<tr>
<td>BL</td>
<td>39.4</td>
</tr>
<tr>
<td>Mass</td>
<td>39.8</td>
</tr>
<tr>
<td>BL + mass</td>
<td>40.9</td>
</tr>
<tr>
<td>BL + mass : BL</td>
<td>40.9</td>
</tr>
<tr>
<td>Mass + mass : BL</td>
<td>41.8</td>
</tr>
<tr>
<td>Mass : BL</td>
<td>42.5</td>
</tr>
<tr>
<td>BL+ mass + mass : BL</td>
<td>42.7</td>
</tr>
</tbody>
</table>
Table 9: Pure self-assessment was ruled out in randomly-matched contests because contest duration and the total number of contest strikes correlated negatively with winner mass. Using multiple regressions, winner mass, loser mass, winner residency and loser residency were predictor variables, while log-corrected contest duration and log-corrected total number of strikes exchanged were cost (dependent) variables. Significant terms are italicized. *NA results reflect factors with repetitive and exactly opposite effects; e.g., winner role and loser role are repetitive and exactly opposite. NA results do not affect the directionality or strength of any other relationships, even when factors are re-ordered in the full model equation.

<table>
<thead>
<tr>
<th>Cost variable</th>
<th>Predictor</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log(_10)(contest duration)</td>
<td>Intercept</td>
<td>1.23</td>
<td>0.44</td>
<td>2.78</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Winner mass</td>
<td>-0.60</td>
<td>0.28</td>
<td>-2.10</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>Winner residency</td>
<td>-0.24</td>
<td>0.46</td>
<td>-0.05</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>Loser mass</td>
<td>-0.81</td>
<td>0.47</td>
<td>-0.17</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>Loser residency</td>
<td>NA*</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Winner mass : winner residency</td>
<td>0.51</td>
<td>0.27</td>
<td>1.84</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>Loser mass : loser residency</td>
<td>1.01</td>
<td>0.41</td>
<td>2.49</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>Winner mass : loser mass</td>
<td>0.08</td>
<td>0.32</td>
<td>0.25</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>Winner residency : loser residency</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Log(_10)(1 + total number of contest strikes)</td>
<td>Intercept</td>
<td>0.35</td>
<td>0.31</td>
<td>1.12</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>Winner mass</td>
<td>-0.48</td>
<td>0.18</td>
<td>-2.74</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Winner residency</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Loser mass</td>
<td>-0.12</td>
<td>0.29</td>
<td>-0.42</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>Loser residency</td>
<td>0.30</td>
<td>0.29</td>
<td>1.06</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>Winner mass : winner residency</td>
<td>0.23</td>
<td>0.17</td>
<td>1.34</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>Loser mass : loser residency</td>
<td>0.32</td>
<td>0.25</td>
<td>1.27</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>Winner mass : loser mass</td>
<td>0.21</td>
<td>0.20</td>
<td>1.09</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>Winner residency : loser residency</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>
Table 10: Cumulative assessment was ruled out according to the criteria of (Arnott & Elwood, 2009) but not (Fawcett & Mowles, 2013), in body length-matched contests because averaged competitor body mass did not correlate with contest duration or the total number of contest strikes. Using multiple regressions, average competitor mass, winner residency and loser residency are predictor variables, while log-corrected contest duration and log-corrected total number of strikes exchanged during contests are cost (dependent) variables. Significant terms are italicized. *NA results reflect factors with repetitive and exactly opposite effects; e.g., winner role and loser role are repetitive and exactly opposite. NA results do not affect the directionality or strength of any other relationships, even when factors are moved in the full model equation.

<table>
<thead>
<tr>
<th>Cost variable</th>
<th>Predictor variable</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log10(contest duration)</td>
<td>Intercept</td>
<td>1.61</td>
<td>0.32</td>
<td>4.95</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Average mass</td>
<td>0.07</td>
<td>0.17</td>
<td>0.38</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>Winner intruder : loser intruder</td>
<td>NA*</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Winner resident : loser intruder</td>
<td>-0.40</td>
<td>0.20</td>
<td>-2.02</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>Winner intruder : loser resident</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Winner resident : loser resident</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Log10(1 + total number of contest strikes)</td>
<td>Intercept</td>
<td>0.97</td>
<td>0.18</td>
<td>5.22</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Average mass</td>
<td>-0.01</td>
<td>0.10</td>
<td>-0.08</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>Winner intruder : loser intruder</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Winner resident : loser intruder</td>
<td>-0.39</td>
<td>0.11</td>
<td>-3.48</td>
<td>&lt;0.02</td>
</tr>
<tr>
<td></td>
<td>Winner intruder : loser resident</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Winner resident : loser resident</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>
Table 11: In randomly-matched contests, correlations of contest costs as a function of body length and residency were similar to correlations of costs as a function of body mass and residency (Table 9). Using multiple regressions, winner body length, loser body length, winner residency, and loser residency are predictor variables, while log-corrected contest duration and log-corrected total number of strikes exchanged are cost (dependent) variables. Significant terms are italicized.

Duration model: $F_{6,28}=2.86$, $R^2 = 0.25$, $p=0.03$

Number of strikes model: $F_{6,28}=3.06$, $R^2=0.27$, $p=0.02$

<table>
<thead>
<tr>
<th>Cost variable</th>
<th>Predictor</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\log_{10}(\text{contest duration})$</td>
<td>Intercept</td>
<td>4.60</td>
<td>3.46</td>
<td>1.33</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>Winner body length (BL)</td>
<td>-0.15</td>
<td>0.09</td>
<td>-1.71</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>Winner residency</td>
<td>1.49</td>
<td>1.49</td>
<td>1.00</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>Loser BL</td>
<td>-0.13</td>
<td>0.11</td>
<td>-1.14</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>Loser residency</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Winner BL : winner residency</td>
<td>0.02</td>
<td>0.04</td>
<td>0.59</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>$\text{Loser BL : loser residency}$</td>
<td>0.07</td>
<td>0.03</td>
<td>2.69</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Winner BL : loser BL</td>
<td>0.00</td>
<td>0.00</td>
<td>1.16</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>Winner residency : loser residency</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

| $\log_{10}(1 + \text{total number of contest strikes})$ | Intercept                     | 3.33     | 2.30       | 1.45    | 0.16    |
|                                                       | Winner BL                     | -0.11    | 0.06       | -1.81   | 0.08    |
|                                                       | Winner residency             | 0.63     | 0.99       | 0.63    | 0.53    |
|                                                       | Loser BL                     | -0.10    | 0.08       | 1.29    | 0.21    |
|                                                       | Loser residency              | NA       | NA         | NA      | NA      |
|                                                       | Winner BL : winner residency | 0.00     | 0.23       | 0.13    | 0.89    |
|                                                       | Winner BL : loser residency  | 0.30     | 0.02       | 1.66    | 0.11    |
|                                                       | Winner BL : loser BL         | 0.00     | 0.00       | 1.36    | 0.18    |
|                                                       | Winner residency : loser residency | NA   | NA         | NA      | NA      |
Table 12: In body length-matched contests, correlations of contest costs as a function of body length and residency were similar to correlations of costs as a function of body mass and residency (Table 10). Using multiple regressions, average competitor body length, winner residency, and loser residency are predictor variables, while contest duration and the total number of strikes exchanged are cost (dependent) variables. Significant terms are italicized.

**Duration model:** $F_{2,26}=2.12$, $R^2 =0.07$, $p=0.14$

**Strikes model:** $F_{2,26}=6.05$, $R^2 =0.27$, $p<0.01$

<table>
<thead>
<tr>
<th>Cost variable</th>
<th>Predictor variable</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log_{10}(contest duration)</td>
<td>Intercept</td>
<td>1.42</td>
<td>0.79</td>
<td>1.80</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>Average BL</td>
<td>0.00</td>
<td>0.02</td>
<td>0.37</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>Winner intruder : loser intruder</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Winner resident : loser intruder</td>
<td>-0.39</td>
<td>0.20</td>
<td>-2.01</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>Winner intruder : loser resident</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Winner resident : loser resident</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Log_{10}(1+ total number of contest strikes)</td>
<td>Intercept</td>
<td>0.90</td>
<td>0.45</td>
<td>2.00</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>Average BL</td>
<td>0.00</td>
<td>0.01</td>
<td>0.12</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>Winner intruder : loser intruder</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Winner resident : loser intruder</td>
<td>-0.39</td>
<td>0.11</td>
<td>-3.47</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Winner intruder : loser resident</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Winner resident : loser resident</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>
Table 13: Previously-removed outliers were added to the dataset and the effects of RHP and residency on randomly-matched contest costs were re-analyzed here. The overall trends ruling out cumulative and mutual assessment remain consistent across analyses. Compare this table to Table 9. NA* codes are explained in Table 9.

<table>
<thead>
<tr>
<th>Cost variable</th>
<th>Predictor</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contest duration</td>
<td>Intercept</td>
<td>1.68</td>
<td>0.58</td>
<td>2.9</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Winner mass</td>
<td>-0.65</td>
<td>0.38</td>
<td>-1.72</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>Winner residency</td>
<td>-0.40</td>
<td>0.61</td>
<td>-0.67</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>Loser mass</td>
<td>-0.15</td>
<td>0.64</td>
<td>-0.24</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>Loser residency</td>
<td>NA*</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Winner mass : winner residency</td>
<td>0.50</td>
<td>0.36</td>
<td>1.38</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>Loser mass : loser residency</td>
<td>0.78</td>
<td>0.56</td>
<td>1.40</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>Winner mass : loser mass</td>
<td>0.13</td>
<td>0.44</td>
<td>0.31</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>Winner residency : loser residency</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Total number of contest strikes</td>
<td>Intercept</td>
<td>0.87</td>
<td>0.33</td>
<td>2.63</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Winner mass</td>
<td>-0.49</td>
<td>0.22</td>
<td>-2.23</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Winner residency</td>
<td>-0.48</td>
<td>0.35</td>
<td>-1.38</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>Loser mass</td>
<td>-0.16</td>
<td>0.37</td>
<td>-0.43</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>Loser residency</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Winner mass : winner residency</td>
<td>0.21</td>
<td>0.21</td>
<td>1.01</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>Loser mass : loser residency</td>
<td>0.20</td>
<td>0.32</td>
<td>0.62</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>Winner mass : loser mass</td>
<td>0.24</td>
<td>0.25</td>
<td>0.98</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>Winner residency : loser residency</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>
Table 14: Previously-removed outliers were added to the dataset and the effects of RHP and residency on body length-matched contest costs were re-analyzed. The overall trends ruling out cumulative assessment according to (Arnott & Elwood, 2009) remain consistent across analyses. Compare this table to Table 10. NA* codes are explained in Table 10.

<table>
<thead>
<tr>
<th>Cost variable</th>
<th>Predictor variable</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contest duration</td>
<td>Intercept</td>
<td>1.63</td>
<td>0.35</td>
<td>4.68</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Average mass</td>
<td>0.16</td>
<td>0.18</td>
<td>0.89</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>Winner intruder : loser intruder</td>
<td>NA*</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Winner resident : loser intruder</td>
<td>-0.57</td>
<td>0.20</td>
<td>-2.85</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Winner intruder : loser resident</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Winner resident : loser resident</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Total number of contest strikes</td>
<td>Intercept</td>
<td>0.98</td>
<td>0.19</td>
<td>5.17</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Average mass</td>
<td>0.03</td>
<td>0.10</td>
<td>0.31</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>Winner intruder : loser intruder</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Winner resident : loser intruder</td>
<td>-0.47</td>
<td>0.11</td>
<td>-4.3</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Winner intruder : loser resident</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Winner resident : loser resident</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>
Table 15: Summary statistics of body length, body mass, the difference in body length and body mass, contest duration, total number of strikes exchanged, and total number of contest behaviours between competitors in randomly-matched and body length-matched contests. Each cell shows the mean ± standard deviation and (in parentheses) the minimum and maximum values. Body length difference and body mass differences are in absolute value.

<table>
<thead>
<tr>
<th></th>
<th>Body length (mm)</th>
<th>Body mass (g)</th>
<th>Absolute value of body length difference (mm)</th>
<th>Absolute value of body mass difference (g)</th>
<th>Total contest duration (s)</th>
<th>Total number of strikes exchanged</th>
<th>Total number of contest behaviours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomly-matched contests (n=35)</td>
<td>35.68 ± 5.63 (25.35, 50.62)</td>
<td>1.03 ± 0.49 (0.40, 2.61)</td>
<td>6.46 ± 4.55 (0.25, 18.72)</td>
<td>0.59 ± 0.45 (0.00, 1.82)</td>
<td>20 ± 19 (2, 78)</td>
<td>1 ± 1 (0, 5)</td>
<td>14 ± 8 (6, 36)</td>
</tr>
<tr>
<td>Body length-matched contests (n=29)</td>
<td>43.69 ± 5.12 (34.23, 52.30)</td>
<td>1.65 ± 0.54 (0.72, 2.56)</td>
<td>0.80 ± 0.47 (0.01, 1.83)</td>
<td>0.15 ± 0.13 (0.00, 0.59)</td>
<td>51 ± 58 (4, 188)</td>
<td>5 ± 5 (0, 20)</td>
<td>30 ± 21 (9, 89)</td>
</tr>
</tbody>
</table>
Table 16: Full transitional data matrix for randomly-matched contests. Each cell indicates the number of transitions from one behaviour (rows) to a subsequent behaviour (columns). Values outside parentheses and brackets represent the observed number of transitions, values in parentheses represent the 95% quantile of the expected number of transitions based on our permutation procedure (see Methods), and values in brackets represent transitional probability. In our sequential behavioural analysis, we simplified the matrix to only those transitions that were more frequent than expected (bolded, italicized cells; see Methods).

<table>
<thead>
<tr>
<th>Bout start</th>
<th>Still</th>
<th>Out of sight</th>
<th>Track</th>
<th>Approach</th>
<th>Antennular flick</th>
<th>Meral spread</th>
<th>Lunge</th>
<th>Strike</th>
<th>Telson coil</th>
<th>Telson push</th>
<th>Retreat</th>
<th>Chase</th>
<th>Bout end</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bout start</td>
<td>0 (0)</td>
<td>1 (2)</td>
<td>5 (3)</td>
<td>44 (17)</td>
<td>11 (11)</td>
<td>0 (7)</td>
<td>1 (7)</td>
<td>0 (12)</td>
<td>0 (9)</td>
<td>8 (18)</td>
<td>0 (1)</td>
<td>0 (8)</td>
<td>0 (14)</td>
</tr>
<tr>
<td>Still</td>
<td>[0.00]</td>
<td>[0.01]</td>
<td>[0.07]</td>
<td>[0.63]</td>
<td>[0.16]</td>
<td>[0.00]</td>
<td>[0.01]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.11]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
</tr>
<tr>
<td>Out of sight</td>
<td>0 (0)</td>
<td>0 (1)</td>
<td>2 (3)</td>
<td>2 (2)</td>
<td>[0.33]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.17]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.17]</td>
</tr>
<tr>
<td>Track</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.14]</td>
<td>[0.14]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.43]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.29]</td>
</tr>
<tr>
<td>Approach</td>
<td>0 (0)</td>
<td>1 (2)</td>
<td>15 (13)</td>
<td>8 (8)</td>
<td>13 (6)</td>
<td>0 (9)</td>
<td>1 (7)</td>
<td>4 (13)</td>
<td>0 (1)</td>
<td>10 (10)</td>
<td>0 (1)</td>
<td>6 (17)</td>
<td>0 (19)</td>
</tr>
<tr>
<td>Antennular flick</td>
<td>0 (0)</td>
<td>1 (1)</td>
<td>0 (2)</td>
<td>1 (8)</td>
<td>3 (6)</td>
<td>0 (4)</td>
<td>3 (4)</td>
<td>11 (6)</td>
<td>1 (5)</td>
<td>6 (9)</td>
<td>0 (1)</td>
<td>3 (4)</td>
<td>0 (1)</td>
</tr>
<tr>
<td>Meral spread</td>
<td>[0.00]</td>
<td>[0.03]</td>
<td>[0.00]</td>
<td>[0.03]</td>
<td>[0.10]</td>
<td>[0.00]</td>
<td>[0.10]</td>
<td>[0.38]</td>
<td>[0.03]</td>
<td>[0.21]</td>
<td>[0.00]</td>
<td>[0.10]</td>
<td>[0.00]</td>
</tr>
<tr>
<td>Lunge</td>
<td>[0.00]</td>
<td>[0.02]</td>
<td>[0.00]</td>
<td>[0.13]</td>
<td>[0.00]</td>
<td>[0.35]</td>
<td>[0.00]</td>
<td>[0.23]</td>
<td>[0.00]</td>
<td>[0.19]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.03]</td>
</tr>
<tr>
<td></td>
<td>[0.00]</td>
<td>[0.02]</td>
<td>[0.05]</td>
<td>[0.00]</td>
<td>[0.05]</td>
<td>[0.59]</td>
<td>[0.20]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.08]</td>
</tr>
<tr>
<td>Bout start</td>
<td>Still Out of sight</td>
<td>Track Approach</td>
<td>Antennular flick</td>
<td>Meral spread</td>
<td>Lunge Strike</td>
<td>Telson coil</td>
<td>Telson push</td>
<td>Retreat</td>
<td>Chase</td>
<td>Bout end</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>--------------------</td>
<td>----------------</td>
<td>------------------</td>
<td>--------------</td>
<td>--------------</td>
<td>-------------</td>
<td>-------------</td>
<td>---------</td>
<td>-------</td>
<td>----------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strike</td>
<td>0 (0)</td>
<td>0 (2)</td>
<td>2 (11)</td>
<td>0 (7)</td>
<td>0 (5)</td>
<td>0 (5)</td>
<td>0 (8)</td>
<td>0 (6)</td>
<td><strong>36 (11)</strong></td>
<td>0 (1)</td>
<td>1 (6)</td>
<td>0 (1)</td>
<td>1 (9)</td>
</tr>
<tr>
<td>Telson coil</td>
<td>0 (0)</td>
<td>1 (3)</td>
<td>1 (3)</td>
<td>16 (21)</td>
<td>2 (13)</td>
<td>5 (9)</td>
<td>3 (9)</td>
<td><strong>27 (15)</strong></td>
<td>2 (11)</td>
<td>0 (22)</td>
<td><strong>1 (1)</strong></td>
<td><strong>15 (10)</strong></td>
<td><strong>1 (1)</strong></td>
</tr>
<tr>
<td>Telson push</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (1)</td>
<td>0 (1)</td>
<td>0 (1)</td>
<td>0 (1)</td>
<td>0 (1)</td>
<td>1 (1)</td>
<td>0 (0)</td>
<td>0 (1)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (1)</td>
</tr>
<tr>
<td>Retreat</td>
<td><strong>0.00</strong></td>
<td><strong>0.00</strong></td>
<td><strong>0.00</strong></td>
<td><strong>0.05</strong></td>
<td><strong>0.00</strong></td>
<td><strong>0.00</strong></td>
<td><strong>0.00</strong></td>
<td><strong>0.00</strong></td>
<td><strong>0.00</strong></td>
<td><strong>0.00</strong></td>
<td><strong>0.00</strong></td>
<td><strong>0.00</strong></td>
<td><strong>0.00</strong></td>
</tr>
<tr>
<td>Chase</td>
<td><strong>0.00</strong></td>
<td><strong>0.00</strong></td>
<td><strong>0.00</strong></td>
<td><strong>0.00</strong></td>
<td><strong>0.00</strong></td>
<td><strong>0.00</strong></td>
<td><strong>0.00</strong></td>
<td><strong>0.00</strong></td>
<td><strong>0.00</strong></td>
<td><strong>0.00</strong></td>
<td><strong>0.00</strong></td>
<td><strong>0.00</strong></td>
<td><strong>0.00</strong></td>
</tr>
<tr>
<td>Bout end</td>
<td><strong>0.00</strong></td>
<td><strong>0.00</strong></td>
<td><strong>0.00</strong></td>
<td><strong>0.00</strong></td>
<td><strong>0.00</strong></td>
<td><strong>0.00</strong></td>
<td><strong>0.00</strong></td>
<td><strong>0.00</strong></td>
<td><strong>0.00</strong></td>
<td><strong>0.00</strong></td>
<td><strong>0.00</strong></td>
<td><strong>0.00</strong></td>
<td><strong>0.00</strong></td>
</tr>
</tbody>
</table>
Table 17: Full transitional data matrix for body length-matched contests. Each cell indicates the number of transitions from one behaviour (rows) to a subsequent behaviour (columns). Values outside parentheses and brackets represent the observed number of transitions, values in parentheses represent the 95% quantile of the expected number of transitions based on our permutation procedure (see Methods), and values in brackets represent transitional probability. In our sequential behavioural analysis, we simplified the matrix to only those transitions that were more frequent than expected (bolded, italicized cells; see Methods).
<table>
<thead>
<tr>
<th>Bout start</th>
<th>Still</th>
<th>Out of sight</th>
<th>Track</th>
<th>Approach</th>
<th>Antennular flick</th>
<th>Meral spread</th>
<th>Lunge</th>
<th>Strike</th>
<th>Telson coil</th>
<th>Telson push</th>
<th>Retreat</th>
<th>Chase</th>
<th>Bout end</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bout start</td>
<td>0 (0)</td>
<td>2 (3)</td>
<td>5 (2)</td>
<td>22 (7)</td>
<td>17 (5)</td>
<td>0 (5)</td>
<td>0 (5)</td>
<td>1 (16)</td>
<td>0 (14)</td>
<td>9 (21)</td>
<td>0 (3)</td>
<td>0 (4)</td>
<td>0 (1)</td>
</tr>
<tr>
<td>Still</td>
<td>0 (0)</td>
<td>0 (1)</td>
<td>2 (1)</td>
<td>3 (3)</td>
<td>1 (2)</td>
<td>2 (2)</td>
<td>1 (2)</td>
<td>0 (6)</td>
<td>0 (5)</td>
<td>0 (3)</td>
<td>0 (1)</td>
<td>2 (2)</td>
<td>0 (1)</td>
</tr>
<tr>
<td>Out of sight</td>
<td>0 (0)</td>
<td>0 (1)</td>
<td>0 (1)</td>
<td>4 (2)</td>
<td>0 (2)</td>
<td>0 (2)</td>
<td>1 (2)</td>
<td>3 (5)</td>
<td>0 (4)</td>
<td>2 (6)</td>
<td>0 (1)</td>
<td>1 (2)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Track</td>
<td>0 (0)</td>
<td>3 (3)</td>
<td>0 (2)</td>
<td>0 (6)</td>
<td>14 (5)</td>
<td>9 (5)</td>
<td>8 (5)</td>
<td>15 (5)</td>
<td>0 (13)</td>
<td>9 (20)</td>
<td>0 (3)</td>
<td>1 (4)</td>
<td>0 (1)</td>
</tr>
<tr>
<td>Approach</td>
<td>0 (0)</td>
<td>2 (2)</td>
<td>0 (2)</td>
<td>6 (5)</td>
<td>0 (4)</td>
<td>10 (4)</td>
<td>6 (4)</td>
<td>6 (12)</td>
<td>0 (11)</td>
<td>10 (16)</td>
<td>0 (3)</td>
<td>0 (3)</td>
<td>0 (1)</td>
</tr>
</tbody>
</table>

Antennular flick | 0 (0) | 1 (2) | 0 (2) | 0 (5) | 1 (4) | 0 (4) | 9 (4) | 18 (11) | 1 (10) | 4 (14) | 0 (2) | 1 (3) | 0 (1) | 1 (5) |

Meral spread | 0 (0) | 0 (2) | 3 (5) | 3 (4) | 9 (4) | 3 (4) | 13 (11) | 0 (10) | 6 (15) | 0 (3) | 1 (3) | 0 (1) | 0 (5) |

Lunge | 0 (0) | 0 (6) | 0 (5) | 1 (15) | 0 (12) | 0 (11) | 3 (11) | 2 (39) | 121 (35) | 35 (54) | 0 (7) | 1 (9) | 0 (1) | 0 (16) |

Strike | 0 (0) | 1 (5) | 1 (4) | 0 (14) | 0 (11) | 0 (10) | 0 (10) | 0 (34) | 2 (31) | 136 (47) | 0 (6) | 0 (8) | 0 (1) | 1 (14) |

Telson coil | 0 (0) | 6 (7) | 4 (6) | 15 (20) | 4 (16) | 6 (14) | 6 (16) | 111 (53) | 17 (47) | 0 (74) | 18 (9) | 21 (12) | 2 (2) | 21 (21) |

Telson push | 0 (0) | 0 (1) | 0 (1) | 0 (3) | 0 (3) | 0 (2) | 0 (2) | 0 (7) | 0 (6) | 18 (8) | 1 (2) | 0 (2) | 0 (0) | 0 (3) |
<table>
<thead>
<tr>
<th>Bout start</th>
<th>Out of sight</th>
<th>Track</th>
<th>Approach</th>
<th>Antennular flick</th>
<th>Meral spread</th>
<th>Lunge</th>
<th>Strike</th>
<th>Telson coil</th>
<th>Telson push</th>
<th>Retreat</th>
<th>Chase</th>
<th>Bout end</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (0)</td>
<td>0 (2)</td>
<td>0 (4)</td>
<td>0 (3)</td>
<td>0 (3)</td>
<td>0 (9)</td>
<td>0 (8)</td>
<td>0 (12)</td>
<td>0 (2)</td>
<td>0 (3)</td>
<td>0 (1)</td>
<td>28 (4)</td>
<td></td>
</tr>
<tr>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[1.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[1.00]</td>
<td></td>
</tr>
<tr>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (1)</td>
<td>0 (1)</td>
<td>0 (1)</td>
<td>0 (1)</td>
<td>0 (1)</td>
<td>0 (2)</td>
<td>0 (0)</td>
<td>0 (1)</td>
<td>0 (0)</td>
<td>2 (1)</td>
<td></td>
</tr>
<tr>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[1.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[1.00]</td>
<td></td>
</tr>
<tr>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td>[0.00]</td>
<td></td>
</tr>
</tbody>
</table>
Figure 10: To have an equal chance of winning a contest, intruders needed (a) 0.51g more mass than residents in randomly-matched contests and (b) 0.16g more mass than residents in body length-matched contests. In (a) and (b), the black solid lines represent the observed GLM fit and the grey solid lines represent a null scenario of no residency effect. In (a) and (b), the black dashed line represents the inflection point of the observed GLM fit, where the probability of an intruder win is 50%; the grey dashed line represents the inflection point of a null scenario of no residency effect. The y-axes are equivalent in (a) and (b).
Figure 11: The distribution of contest costs for randomly-matched competitors within 5% of total body length matches closely to that of body length-matched competitors, suggesting that differences between randomly-matched and body length-matched protocols did not cause substantial differences in contest dynamics. Plots of average competitor body mass against log-corrected (a) contest duration and (b) total contest strikes. Solid circles represent residents in body length-matched contests, open diamonds represent intruders in body length-matched contests.
Asterisks represent randomly-matched competitors that were within 5% of total body length, a range like that of our body length-matched competitors.

References


Biography

Patrick Andrew Green was born in Sacramento, CA on November 13, 1986; he grew up in Cameron Park, CA. Patrick graduated with his B.S. in Ecology, Behavior, and Evolution from the University of California, Los Angeles in 2009. He started his PhD in 2011 in the Organismal and Evolutionary Biology Program at the University of Massachusetts, Amherst, before transferring to the Biology Program at Duke University in 2013. Patrick received an Honorable Mention for the National Science Foundation Graduate Research Fellowship Program (2012) and has received fellowships from the Smithsonian Tropical Research Association (2013) and Duke University Biology (2017). He has also received the UMass, Amherst OEB Outstanding TA Award (2012), Duke Data Expeditions Teaching Awards (2017, 2018), and the Duke Dean’s Award for Excellence in Mentoring (2018).

Publications (reverse chronological order):


