PROCEEDINGS OF THE ROYAL SOCIETY B

Proc. R. Soc. B (2006) **273**, 1555–1561
doi:10.1098/rspb.2006.3468
Published online 28 February 2006

Testing game theory models: fighting ability and decision rules in chameleon contests

Devi Stuart-Fox*

School of Animal, Plant and Environmental Sciences, University of the Witwatersrand, Private Bag 3, Wits 2050, South Africa

Game theory models of animal contests make many non-mutually exclusive predictions, complicating empirical tests. These predictions regard the relationship between contest parameters and fighting ability, for which body size is usually used as a proxy. However, in many systems, body size may be a limited proxy since multiple traits and contextual factors such as experience influence fighting ability. Using contests between male Cape dwarf chameleons, *Bradypodion pumilum*, I test alternative game theory models of extended contests. I show how the most likely candidate model can be identified through a process of elimination, based on tests of key predictions. In addition, I present a measure of fighting ability based on multiple traits that allows ability to change as experience changes. In dwarf chameleons, persistence is based on loser thresholds rather than assessment of relative ability, ruling out the sequential assessment model. Winners and losers do not match behaviours in early parts of the contest, arguing against all types of war of attrition models. Although the cumulative assessment model remained as the most likely candidate model, not all specific predictions of this model were upheld.

Keywords: cumulative assessment; sequential assessment; war of attrition; resource-holding potential; contest competition; signalling

1. INTRODUCTION

Animal contests often involve a series of complex behaviours varying from energetically demanding but non-dangerous displays to physical fighting with risk of serious injury. During contests, animals must make decisions about which behaviours to use and when to give up. Game theory models are useful for understanding the evolution of fighting behaviour and decision rules because they make explicit predictions that can be tested empirically (Maynard Smith 1982; Riechert 1998). Currently available game theory models, however, do not make mutually exclusive predictions and support for those of a particular model may not rule out alternative models (Nuyts 1994; Payne 1998; Taylor & Elwood 2003). For instance, a negative relationship between contest duration and asymmetry among contestants has traditionally been interpreted as a support for the sequential assessment model (SAM; Enquist & Leimar 1983; Enquist *et al.* 1990). However, Taylor & Elwood (2003) recently showed that such a relationship can result from fundamentally different decision rules. Consequently, it is important to consider multiple models simultaneously.

Game theory models of extended contests make predictions regarding the relationship between contestant resource-holding potential (RHP, defined as 'absolute fighting ability'; Parker 1974) and various measures of contest duration, intensity and structure. To test these predictions, most studies have used body size as a proxy for RHP, since size is often a good predictor of contest outcome. However, when multiple traits influence fighting ability, body size may be a poor proxy (e.g. Zamudio *et al.* 1995; Barki *et al.* 1997; Hernandez & Benson 1998). Furthermore, as RHP is manifested as fighting effort, it may be greatly influenced by contest experience (previous wins or losses), which can override effects of size or other traits (Zucker & Murray 1996; Daws *et al.* 2002; Hoefler 2002; Stuart-Fox & Johnston 2005). As a result, an animal's fighting ability can change between contests as experience changes (Payne 1998). A measure of RHP that can incorporate the influence of multiple traits as well as context-specific effects such as experience, would represent a significant advance in attempts to test the game theory models.

Here, I examine the fit of game theory models that are applicable to the extended contests using Cape dwarf chameleons, *Bradypodion pumilum*. I present a means by which to identify the most likely candidate model through a process of elimination (falsification of key predictions) because lack of support for critical model predictions can unequivocally rule out candidate models, whereas confirmation can be inconclusive when predictions are not mutually exclusive (Payne 1998). I then test specific predictions of the best candidate model to assess the fit of the model to the data. Dwarf chameleon contests vary greatly in duration and degree of escalation from simple lateral displays to prolonged jaw-locks (Stuart-Fox *et al.* in press). Contest outcome is influenced by the height of the ornamental casque, the relative size of the pink patch in the centre of the flank and by previous experience (Stuart-Fox *et al.* in press). Body size does not predict male fighting ability in this species (in which males are smaller than females, Stuart-Fox & Whiting 2005), except in the case of extreme size differences. Thus, size cannot be used as a proxy for fighting ability. Instead, I derive a 'multivariate' estimate of relative fighting ability using

*devi@gecko.biol.wits.ac.za

Received 14 October 2005
Accepted 28 December 2005

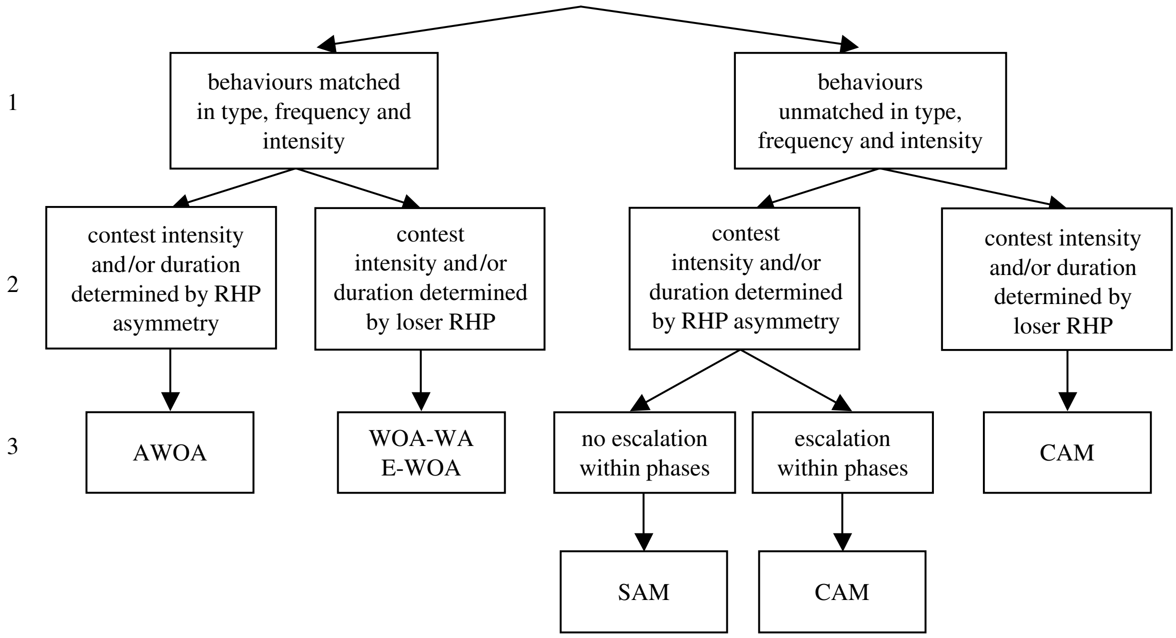

Figure 1. Flow chart showing means to distinguish between main types of game theoretic models applicable to extended contests (see text) in three steps: (1) test for behavioural matching; (2) test for a relationship between contest duration and/or intensity and both loser RHP and RHP asymmetry (see Taylor & Elwood 2003); (3) test for escalation within phases. Based on these three tests, any of the four main types of model of extended contests may be distinguished. The order of tests was chosen for ease of schematic representation and makes no difference to conclusions drawn.

a form of generalized linear model (the structured Bradley–Terry (B–T) model; Firth 2005; Stuart-Fox *et al.* in press) that incorporates the effects of casque height, flank patch and contest experience on contest outcome, allowing fighting ability to change between contests as experience changes. I use these 'multivariate' estimates of fighting ability and a detailed analysis of behavioural dynamics to assess decision rules in chameleon contests.

(a) *Model comparison*

There are four main types of models which are applicable to extended contests (figure 1): (i) the traditional asymmetric war of attrition (AWOA; Parker & Rubenstein 1981; Hammerstein & Parker 1982); (ii) wars of attrition (WOA) based on individual energetic thresholds such as the WOA without assessment (WOA-WA; Mesterton-Gibbons *et al.* 1996) and the energetic WOA (E-WOA; Payne & Pagel 1996*a*; Payne & Pagel 1997)—the WOA-WA and E-WOA make qualitatively similar predictions and are therefore treated together; (iii) the SAM (Enquist & Leimar 1983; Enquist *et al.* 1990) and (iv) the cumulative assessment model (CAM; Payne 1998). Although there are other game theory models of animal contests such as the 'best so far rule' (Payne & Pagel 1996*b*; Payne & Pagel 1997), these are either not applicable to extended or escalated contests or tend to be context-specific variations of the more general models considered here.

The four types of models differ in three key predictions that can be used to identify the most likely candidate model because, in each case, predictions differ for at least two of the models. First, crucial to all WOA models is that contestants must match each other in the behaviours used and their frequencies (i.e. matching of intensity or energy expenditure) because otherwise a cheat may delay display until the opponent is exhausted (Payne 1998). Thus, both types of WOA model predict that behaviours should be matched between contestants until near the end of the contest. Both SAM and CAM allow behaviours of opponents to be unmatched in rate or intensity, even at the beginning of the contest (Payne 1998).

A second fundamental difference between models is whether individuals make decisions based on assessment of their opponent's relative qualities or based only on their own thresholds. The AWOA and SAM both involve opponent assessment whereas the WOA-WA, E-WOA and CAM do not. Both the AWOA and SAM predict a negative relationship between a measure of asymmetry in RHP between contestants and the duration and intensity of a contest since small differences result in greater potential error in assessment. In contrast, in the WOA-WA (Mesterton-Gibbons *et al.* 1996) and E-WOA (Payne & Pagel 1996*a*; Payne & Pagel 1997) decisions are based solely on individual energetic thresholds while the CAM involves decisions based on the threshold of costs inflicted by the opponent that an individual is willing to bear (individual cost threshold; Payne 1998). In both cases, contests end when the loser reaches its threshold, so contest duration and intensity are expected to be a function of loser RHP rather than RHP asymmetry. However, as noted by Taylor & Elwood (2003), the CAM could also yield a true relationship between contestant asymmetry and contest duration or intensity. This is because the decision to withdraw could be influenced by both an individual's own RHP (poor quality individuals can bear fewer costs) and the opponent's RHP (higher quality individuals can inflict costs at a higher rate), even in the absence of mutual assessment (Payne 1998). Consequently, the CAM cannot be unequivocally ruled out when contest duration or intensity are a function of contestant asymmetry (as predicted by SAM and AWOA).

Based on the above factors, under both the SAM and CAM, behaviours may be unmatched and contest duration and intensity may be predicted by RHP asymmetry. To unequivocally distinguish CAM from SAM, rates of escalation must be examined. The SAM cannot accommodate escalation within phases (periods of the contest characterized by behaviours of similar intensity) and phases must proceed in order of increasing intensity (Enquist & Leimar 1983; Enquist *et al.* 1990). In contrast, the CAM predicts that winners and losers should differ in rates of escalation within phases for long and intermediate duration contests (where duration is defined relative to the time-scale of escalation; Payne 1998). By assessing (i) behavioural matching, (ii) whether decisions are based on individual RHP or RHP asymmetries and (iii) rates of escalation, three of the four types of model can be eliminated (or considered unlikely) as potential candidates (see figure 1 for schematic representation of model comparison). Once the most likely candidate model has been identified, model-specific predictions can then be tested to assess the fit of the model to the data.

2. MATERIAL AND METHODS

(a) *Study species and contests*

The Cape dwarf chameleon is a small (up to 90 mm snout–vent length) lizard that occurs in a variety of mesic habitats. Males are highly intolerant of other males. Contests can result in surface wounds and scarring and, more rarely, serious injury, but death from contests is highly unlikely (Burrage 1973; Stuart-Fox *et al.* in press).

I used data from 107 staged contests between wild-caught Cape dwarf chameleons from Stellenbosch in the Western Cape Province, South Africa (33°56′ S, 18°52′ E). These contests are the same as those in Stuart-Fox *et al.* (in press), in which full details of capture, husbandry and contests are provided. Briefly, trials were conducted in an arena (60×40×50 cm) with four horizontal, intersecting dowel sticks 30 cm above the floor and a vertical dowel to the floor at each intersection. The arena was in a constant temperature (CT) room at 28 °C and lighting was with Osram L36W/72-965 Biolux fluorescent lights, which approximates natural sunlight. Contests were videotaped and behaviours scored from digital footage at 25 frames per second.

(b) *Fighting ability*

Contests were organized as a tournament where each male ($n=36$) competed with an average of approximately six others (Stuart-Fox *et al.* in press). From this, estimates of fighting ability (from hereon referred to as ‘ability’ for brevity) were derived using a structured B–T model (Firth 2005; Stuart-Fox *et al.* in press). Full details of the model and its application to the chameleon contests are provided in Stuart-Fox *et al.* (in press) and details of the more general application of this model to contest data can be found in Firth (2005) and Stuart-Fox *et al.* (in press). Briefly, the structured B–T model is a form of generalized linear model that estimates the probability of individual i beating individual j as the difference in their estimated abilities (Firth 2005). Estimates of abilities are derived from a matrix of contest outcomes (win or lose) between participants of the tournament, related to a series of male traits through a linear predictor. As each male contested several different opponents, the model employed by Stuart-Fox *et al.* (in press) incorporated an additional term that allows for the potential effect of experience; where the extra predictor z_{ik} summarizes the contest history of individual i at the time of contest k. Thus, the full model can be expressed as

$$\text{logit}[\text{probability}(i \text{ beats } j \text{ in contest } k)] = \lambda_i - \lambda_j + \delta(z_{ik} - z_{jk}),$$

where λ_i and λ_j represent the abilities of the two individuals.

Ability estimates were derived from the model presented in Stuart-Fox *et al.* (in press), which showed that fighting ability was a function of the height of the casque, the relative area occupied by the flank patch and experience (number of previous wins in the two prior contests). As contestant abilities can change as their experience changes, estimates of ability of each male were derived for each contest.

(c) *Behavioural elements, intensity and escalation*

I recorded contest duration (from initial aggressive behaviour to the end of the contest). Contests were considered to have ended when the loser dropped to the ground or fled/retreated more than once. Winners were easily identified by their colour: winners remained bright while losers became dark and fled repeatedly.

Chameleon contests involve a series of complex behaviours (table 1). Males begin by displaying bright coloration with a lateral aggressive display or head-shakes. Often, one male retreats immediately after the opponent initiates with one of these behaviours. Contests that involve mutual display often escalate further to aggressive displays with open mouth threat and chasing followed by biting and/or mouth wrestling and/or jaw-locking (table 1). I recorded the frequency and duration of these six behavioural elements: (i) aggressive display, (ii) head-shake, (iii) aggressive display with open mouth threat or chasing, (iv) biting, (v) mouth wrestling and (vi) jaw-locking (table 1). I classified behaviours into three levels of intensity, consistent with likely associated costs. Aggressive displays and head-shaking were given an intensity score of one; aggressive display with open mouth threat and/or chasing was given a score of two; and biting, mouth-wrestling and jaw-locking (behaviours that involved physical contact) were given a score of three.

Under the CAM, temporal trends of escalation between winners and losers differ according to the duration of the contests, which is relative to the time-scale of escalation (Payne 1998). Short contests are generally one sided (only one contestant displays), precluding analysis of differences in rates of escalation between winners and losers. Long contests are those in which the initial rates of escalation become insignificant in the calculation of costs (Payne 1998). Thus, I defined long contests as two-sided contests (both contestants display) lasting more than 10 times as long as the period prior to physical contact. In other words, at least nine-tenths of the contest consisted of physical combat during which the great majority of costs will be accrued. Two-sided contests that lasted for less than 10 times the period prior to contact were classed as ‘intermediate duration’.

(d) *Analysis*

The analysis proceeded according to the three steps described in figure 1. First, to assess behavioural matching (as predicted by all WOA models), I tested for differences in the frequency and duration of behaviours of each contestant in the first half of contests involving mutual display (paired t-tests). I chose to

Table 1. Description of male aggressive behaviours. Mean, standard deviation (s.d.) and range of each behaviour performed by the winner per trial. In each case, descriptive statistics for frequencies of behaviours are presented on the top row and duration (seconds) underneath.

behaviour	description	mean ± s.d.	range
aggressive display	body strongly laterally compressed, gular pouch expanded, casque 'raised', tail usually coiled, bright coloration. Laterally oriented towards opponent	1.8±1.5 32.5±34.1	0–9 0–218.4
aggressive display with open mouth threat	as above but with mouth open, often combined with advance towards opponent or chase	1.3±1.4 5.3±7.4	0–8 0–57.6
head-shake	short, rapid, discrete side-to-side shakes of the head, usually combined with bright coloration and body laterally compressed	14.5±24.7 10.6±17.1	0–135 0–85.8
bite	biting any part of the body (except jaw-lock)	1.0±2.1 132.8±532.2	0–14 0–3655.3
mouth-wrestle	males head-to-head on perch, nipping each other. Usually precedes jaw-lock and may be an attempt to obtain an optimum position for jaw-locking	0.4±0.7 1.6±3.4	0–4 0–16.8
jaw-lock	each male biting the upper or lower jaw of his opponent, often pushing each other along perch in an apparent show of strength	0.7±1.5 87.7±335.9	0–8 0–2819.1

examine the first half of contests because lack of behavioural matching, even in the early parts of the contest, would constitute strong evidence against WOA models, which predict behavioural matching until the near end of contests. Sample sizes for these tests vary from $n=21$ to 54 depending on the behaviour, because not all behaviours were performed in every contest. Mouth-wrestling and jaw-locking are necessarily matched since they involve both animals and so were not included in this analysis. Second, whether decisions are based on mutual assessment or individual thresholds, I used a combination of simple and multiple regressions with winner ability, loser ability and measures of asymmetry between opponents as predictors of contest duration, intensity and the number of different behavioural elements used by one or both individuals (ranging from 1 to 6; Taylor & Elwood 2003). Mutual assessment predicts a negative relationship with winner ability and a positive relationship with loser ability of approximately equal magnitude in both simple regressions and multiple regressions (Taylor & Elwood 2003). On the other hand, assessment of own thresholds predicts a strong negative relationship with loser RHP but only a weak positive relationship with winner RHP in simple regressions. In multiple regressions, the relationship with winner RHP is expected to become even weaker or non-significant (Taylor & Elwood 2003). Third, I assessed the differences between winners and losers in rate of escalation. The CAM predicts that, in long contests, winners should begin at a similar intensity but escalate at a greater rate than losers while, for intermediate duration contests, winners should begin at higher intensities but escalate slower than losers. Thus, I tested for an interaction between outcome (winner or loser) and duration (long or intermediate) on initiating behaviours, time to physical contact and rate of escalation (rate of display up to time of physical contact) using two-way analyses of variance. Estimates of fighting ability were derived using the statistical package 'R' (Ihaka & Gentleman 1996). All other statistical tests were done in SAS v. 9.01 (SAS Institute 2003).

3. RESULTS

There was great variation in the duration and intensity of contests, as well as in the frequency and duration of particular behavioural elements used (table 1). Among the 107 fights, 56 (52%) involved displays by both participants. In the other 48% of contests, only one male displayed and the other retreated immediately. Among the 56 contests with mutual display, 42 (75%) escalated to biting whereas the attacker bit its opponent in only 10 of the contests (21%) where only one male displayed. Winners initiated 87 of 112 contests (binomial test: $z=5.86$, two-tailed $p<0.001$).

(a) *Behavioural matching*

In contests involving mutual display, winners performed significantly more aggressive displays than losers in the first half of the contest (paired *t*-test: $t_{\text{d.f.}}=2.55_{53}$, $p=0.014$) and performed aggressive displays for longer ($t_{\text{d.f.}}=2.66_{53}$, $p=0.01$). Winners also bit their opponents more frequently ($t_{\text{d.f.}}=2.13_{20}$, $p=0.046$) and there was a trend for winners to bite their opponents for longer ($t_{\text{d.f.}}=1.73_{20}$, $p=0.098$) and to head-shake more frequently ($t_{\text{d.f.}}=1.73_{39}$, $p=0.09$). There was no consistent difference in the frequency or duration of aggressive display with open mouth threat (frequency, $t_{\text{d.f.}}=1.14_{36}$, $p=0.26$; duration, $t_{\text{d.f.}}=-1.2_{36}$, $p=0.24$), or the duration of head-shaking; ($t_{\text{d.f.}}=1.22_{39}$, $p=0.23$) between contestants. The lack of behavioural matching for all other behaviours, however, even within the first half of the contests, argues against all types of WOA models.

(b) *Individual RHP versus RHP asymmetry*

Contest duration, intensity and the number of behavioural elements used were all significantly associated with winner ability, loser ability and all three measures of asymmetry between contestants (see table 2) in simple regressions, although loser ability was consistently the strongest predictor (table 2). When winner ability, loser ability and

Table 2. Simple regressions of contest characteristics (duration, intensity, number of behavioural elements) against measures of individual RHP (winner or loser ability) and three different measures of RHP asymmetry. SCE refers to standardized coefficient estimate. All regressions are significant after table-wide sequential Bonferroni adjustment for multiple tests (Rice 1989).

	duration			intensity			number of behavioural elements		
	SCE	p	r^2	SCE	p	r^2	SCE	p	r^2
winner ability	0.14	0.02	0.05	0.22	<0.01	0.08	0.35	0.01	0.06
loser ability	0.33	<0.0001	0.26	0.40	<0.0001	0.27	0.66	<0.0001	0.26
winner−loser	−0.25	<0.01	0.08	−0.15	0.01	0.06	−0.41	<0.01	0.07
winner/loser	−0.44	<0.01	0.06	−0.53	<0.001	0.12	−0.97	<0.001	0.13
(winner−loser)/ mean ability	−0.29	<0.01	0.09	−0.32	<0.001	0.13	−0.59	<0.001	0.12

a measure of contestant asymmetry were included as predictors in multiple regressions, in each case loser ability was the only variable retained in the final model after stepwise selection (the final models are thus identical to the simple regressions with loser ability in table 2). The weak positive relationship with winner ability and the strong negative relationship with loser ability in simple regressions, as well as the retention of loser ability as the only significant variable in multiple regressions, are consistent with assessment based on own RHP only (Taylor & Elwood 2003).

To confirm that contest duration and intensity are a function of absolute individual abilities, I examined the subset of contests in which contestants were broadly matched ($n=77$), as judged by an overlap in the standard errors of the ability estimates. Contest duration, intensity and number of behavioural elements were all significantly associated with the mean ability of the two contestants (contest duration, standardized coefficient estimate (SCE)=0.3, $p<0.0001$, partial $r^2=0.25$; intensity, SCE=0.4, $p<0.0001$, partial $r^2=0.24$; number of behavioural elements, SCE=0.62, $p<0.0001$, partial $r^2=0.19$). In terms of specific behaviours, male ability was significantly associated with the frequency of aggressive displays with open mouth threat and jaw-locks (multiple regression: aggressive posture with open mouth threat, SCE=0.72, $p=0.006$, partial $r^2=0.11$; jaw-lock, SCE=2.08, $p<0.0001$, partial $r^2=0.46$). Thus, absolute individual abilities (especially of the loser), rather than relative ability, determines contest duration, intensity and number and type of behavioural elements used, ruling out the SAM.

(c) *Rates of escalation*

Based on preceding results, the CAM is the most likely remaining candidate model. For contests that are shorter than the time-scale of escalation, the CAM predicts that winners will perform at a higher level than losers. Winners necessarily perform at higher levels than losers for contests that are one-sided because the loser does not display at all. Not surprisingly, one-sided contests are much shorter than two-sided contests (mean±s.e. of one-sided log duration=1.72±0.04; two-sided=2.29±0.06; t-test, $t=8.45$, $p<0.0001$). Among the 36 two-sided contests that involved physical contact, 17 were of intermediate duration and 19 were long. However, contrary to the predictions of the CAM, there was no significant interaction between outcome (winner or loser) and duration of contest (intermediate or long) in determining choice of initial behaviour, time to physical contact or rate of escalation (table 3).

Table 3. Tests of predictions of the CAM. $n=36$ Contests with both mutual display and physical contact. Initiating behaviour was treated as a categorical variable and consisted of either aggressive display, aggressive display with approach or head-shake.

dependent variable	factor	$F_{\text{d.f.}}$	p
initiating behaviour	outcome	0.31_1	0.58
	duration	0.13_1	0.72
	outcome × duration	0.01_1	0.92
time to intensify to physical contact	outcome	0.05_1	0.83
	duration	7.3_1	0.009
	outcome × duration	1.19_1	0.28
rate of display before physical contact	outcome	1.84_1	0.18
	duration	0.62_1	0.43
	outcome × duration	0.12_1	0.73

4. DISCUSSION

The structure of chameleon contests contradicts the key predictions of WOA models and SAM. Eventual winners and losers did not match the frequency and/or duration of aggressive displays or bites even in the first half of the contest, suggesting that contestants are not matching energy expenditure as required by WOA models. This view is supported by the finding that winners performed at consistently higher levels. In addition, the duration, intensity and complexity (number of behavioural elements) of contests was determined by the loser's fighting ability rather than asymmetries between contestants, ruling out the SAM. In dwarf chameleons, males with low fighting ability avoid fighting, rarely initiate contests and are less likely to perform aggressive displays with open mouth threat or enter into jaw-locks, regardless of their opponent. Males with high-fighting ability, on the other hand, are very aggressive. They initiate contests, persist for longer and are more likely to use costly and dangerous behaviours such as jaw-locks. This indicates that behavioural dynamics and contest structure are influenced by inherent individual abilities rather than assessment of relative abilities.

Results also confirm the simulation study of Taylor & Elwood (2003), showing that a relationship between contest duration and asymmetry in RHP can be the incidental result of an underlying relationship with the RHP of the weaker contestant. There was a significant relationship between all measures of RHP asymmetry and contest duration and intensity in simple regressions, but this relationship disappeared once loser RHP was taken into account in multiple regressions. A negative relationship between contest duration or intensity and a measure of RHP asymmetry, which is usually interpreted as support for the SAM, has been found for numerous systems (reviewed in Taylor & Elwood 2003). The majority of these studies, however, did not test for a relationship of contest duration with loser or winner RHP. Among the studies that have attempted to distinguish between decisions based on individual (loser) RHP and RHP asymmetries, results have been mixed with evidence for opponent assessment (Pratt *et al.* 2003), duration based on individual RHP only (Bridge *et al.* 2000; Morrell *et al.* 2005) or neither (Jennings *et al.* 2004).

Other contextual clues also make WOA models and the SAM unlikely to apply to dwarf chameleon contests. For instance, WOA models are generally applicable only to the contests that do not carry risk of serious injury. A fundamental assumption of WOA models is that accumulation of injuries in physical contests does not affect basic ability to keep fighting (Hammerstein & Parker 1982; Nuyts 1994). In other words, costs accrue only as a function of time or energy expenditure and not as a function of injury inflicted by the opponent. This assumption is unlikely to hold for prolonged and escalated chameleon contests. Similarly, the SAM predicts that behaviours should be of constant intensity within phases, but there should be an escalation between phases since more costly behaviours carry more reliable information on RHP (Enquist & Leimar 1983; Enquist *et al.* 1990). However, there was clear temporal overlap of behaviours within contests. Although there was a distinct pattern of escalation from lateral aggressive displays and head-shaking at the beginning to jaw-locking later on, low-intensity behaviours were also performed at later stages of the contest. This is consistent with CAM, which allows a gradual change in the proportion of behaviours used (Payne 1998). Such contextual clues, however, are not sufficient to exclude a particular model from consideration, reinforcing the need to systematically assess the key predictions of competing models.

Although based on tests of three key predictions CAM remained as the most likely candidate model, additional-specific predictions of this model were not upheld. Winners and losers did not differ in their initial behaviours, the time they took to escalate to physical contact or their rate of display. The specific predictions of the CAM, however, can be difficult to test, especially in stages of the contest that involve physical contact. The majority of the costs are likely to be accrued in the period of physical contact, yet differences in intensity during contact (e.g. in bite force) could not be determined. Assessing such costs is crucial to a proper test of the CAM since it relies on externally derived costs—either directly from the damage inflicted by the opponent or indirectly from opportunity costs or increased predation risk when performing conspicuous displays (Payne 1998). Indeed, the source and nature of costs represent a crucial difference between game theory models (see Briffa & Elwood 2001, 2002, 2004). Measuring costs directly, however, is challenging, especially in complex systems with multiple behavioural elements. The most promising way in which to test specific predictions of models is to experimentally manipulate specific costs (e.g. Brick 1999). Whether or how well chameleon contests fit the CAM, therefore, requires further testing. Another possibility is that different decision rules may be applied during different stages of the contest (Jennings *et al.* 2005). For instance, individuals may assess an opponent's initiating display when deciding whether or not to respond, then base decisions on individual cost thresholds once the contest escalates. Attempts to model such conditional strategies could prove fruitful.

To assess decision rules, I employed a 'multivariate' measure of fighting ability or RHP. Even in systems where size is a predictor of contest outcome, body size alone may explain only a moderate proportion of the variation and a 'multivariate' measure of ability will be very likely a more accurate predictor of outcome. This is especially true when contextual factors, such as experience or energy reserves strongly influence ability (Payne 1998). A measure of fighting ability that is allowed to change with variation in contextual factors is a useful advance on the previous reliance on a single proxy such as body size for RHP. Relative 'ability' estimated using a structured B–T model represents a realistic measure of fighting ability that takes into account multiple male traits as well as experience (Stuart-Fox *et al.* in press), both of which will affect perceptions of own ability or relative ability and thus decision rules. This study illustrates how model comparison and assessment based on core predictions, combined with a 'multivariate' measure of fighting ability can provide a robust and comprehensive approach to empirically testing game theory models of animal contests.

I am grateful to David Firth for 'R' syntax to output contest-specific ability estimates. Constructive comments from Adnan Moussalli, Terry Ord, Jess Stapley and Martin Whiting improved the manuscript. Funding was from Claude Leon Foundation and National Research Foundation Fellowships and a UNESCO-L'Oreal Grant for Women in Science. University Animal Ethics Committee permit, 2003262A; Cape Nature permit, 234/2003.

REFERENCES

Barki, A., Harpaz, S. & Karplus, I. 1997 Contradictory asymmetries in body and weapon size and assessment in fighting male prawns, *Macrobrachium rosenbergii*. *Aggressive Behav.* **23**, 81–91. (doi:10.1002/(SICI)1098-2337(1997)23:2<81::AID-AB1>3.0.CO;2-W)

Brick, O. 1999 A test of the sequential assessment game: the effect of increased cost of sampling. *Behav. Ecol.* **10**, 726–732. (doi:10.1093/beheco/10.6.726)

Bridge, A. P., Elwood, R. W. & Dick, J. T. A. 2000 Imperfect assessment and limited information preclude optimal strategies in male–male fights in the orb-weaving spider *Metellina mengei*. *Proc. R. Soc. B* **267**, 273–279. (doi:10.1098/rspb.2000.0997)

Briffa, M. & Elwood, R. W. 2001 Decision rules, energy metabolism and vigour of hermit crab fights. *Proc. R. Soc. B* **268**, 1841–1848. (doi:10.1098/rspb.2001.1752)

Briffa, M. & Elwood, R. W. 2002 Power of shell-rapping signals influences physiological costs and subsequent decisions during hermit crab fights. *Proc. R. Soc. B* **269**, 2331–2336. (doi:10.1098/rspb.2002.2158)

Briffa, M. & Elwood, R. W. 2004 Use of energy reserves in fighting hermit crabs. *Proc. R. Soc. B* **271**, 373–379. (doi:10.1098/rspb.2003.2633)

Burrage, B. R. 1973 Comparative ecology and behaviour of *Chamaeleo pumilis pumilis* (Gmelin) and *C. namaquensis* A. Smith (Sauria: Chamaeleonidae). *Ann. S. Afr. Mus.* **61**, 1–158.

Daws, A. G., Grills, J., Konzen, K. & Moore, P. A. 2002 Previous experiences alter the outcome of aggressive interactions between males in the crayfish, *Procambarus clarkii*. *Mar. Freshw. Behav. Physiol.* **35**, 139–148. (doi:10.1080/1023624021000014725)

Enquist, M. & Leimar, O. 1983 Evolution of fighting behaviour: decision rules and assessment of relative strength. *J. Theor. Biol.* **102**, 387–410. (doi:10.1016/0022-5193(83)90376-4)

Enquist, M., Leimar, O., Ljungberg, T., Mallner, Y. & Segerdahl, N. 1990 A test of the sequential assessment game: fighting in the cichlid fish *Nannacara anomala*. *Anim. Behav.* **40**, 1–14.

Firth, D. 2005 Bradley–Terry models in R. *J. Stat. Softw.* **12**, 1–12.

Hammerstein, P. & Parker, G. A. 1982 The asymmetric war of attrition. *J. Theor. Biol.* **96**, 647–682. (doi:10.1016/0022-5193(82)90235-1)

Hernandez, M. I. M. & Benson, W. W. 1998 Small-male advantage in the territorial tropical butterfly *Heliconius sara* (Nymphalidae)—a paradoxical strategy. *Anim. Behav.* **56**, 533–540. (doi:10.1006/anbe.1998.0840)

Hoefler, C. D. 2002 Is contest experience a trump card? The interaction of residency status, experience, and body size on fighting success in *Misumenoides formosipes* (Araneae: Thomisidae). *J. Insect Behav.* **15**, 779–790. (doi:10.1023/A:1021123523307)

Ihaka, R. & Gentleman, R. 1996 R: a language for data analysis and graphics. *J. Comput. Graph. Stat.* **5**, 299–314.

Jennings, D. J., Gammell, M. P., Carlin, C. M. & Hayden, T. J. 2004 Effect of body weight, antler length, resource value and experience on fight duration and intensity in fallow deer. *Anim. Behav.* **68**, 213–221. (doi:10.1016/j.anbehav.2003.11.005)

Jennings, D. J., Gammell, M. P., Payne, R. J. H. & Hayden, T. J. 2005 An investigation of assessment games during fallow deer fights. *Ethology* **111**, 511–525. (doi:10.1111/j.1439-0310.2005.01068.x)

Maynard Smith, J. 1982 *Evolution and the theory of games*. Cambridge, UK: Cambridge University Press.

Mesterton-Gibbons, M., Marden, J. H. & Dugatkin, L. A. 1996 On wars of attrition without assessment. *J. Theor. Biol.* **181**, 65–83. (doi:10.1006/jtbi.1996.0115)

Morrell, L. J., Backwell, P. R. Y. & Metcalfe, N. B. 2005 Fighting in fiddler crabs *Uca mjoebergi*: what determines duration? *Anim. Behav.* **73**, 653–662. (doi:10.1016/j.anbehav.2004.11.014)

Nuyts, E. 1994 Testing for the asymmetric war of attrition when only roles and fight duration are known. *J. Theor. Biol.* **169**, 1–13. (doi:10.1006/jtbi.1994.1125)

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

Parker, G. A. & Rubenstein, D. I. 1981 Role assessment, reserve strategy and acquisition of information in asymmetric animal contests. *Anim. Behav.* **26**, 221–240.

Payne, R. J. H. 1998 Gradually escalating fights and displays: the cumulative assessment model. *Anim. Behav.* **56**, 651–662. (doi:10.1006/anbe.1998.0835)

Payne, R. J. H. & Pagel, M. 1996*a* Escalation and time costs in displays of endurance. *J. Theor. Biol.* **183**, 185–193. (doi:10.1006/jtbi.1996.0212)

Payne, R. J. H. & Pagel, M. 1996*b* When is false modesty a false economy? An optimality model of escalating signals. *Proc. R. Soc. B* **263**, 1545–1550.

Payne, R. J. H. & Pagel, M. 1997 Why do animals repeat displays? *Anim. Behav.* **54**, 109–119. (doi:10.1006/anbe.1996.0391)

Pratt, A. E., McLain, D. K. & Lathrop, G. R. 2003 The assessment game in sand fiddler crab contests for breeding burrows. *Anim. Behav.* **65**, 945–955. (doi:10.1006/anbe.2003.2152)

Rice, W. R. 1989 Analyzing tables of statistical tests. *Evolution* **43**, 223–225.

Riechert, S. E. 1998 Game theory and animal contests. In *Game theory and animal behaviour* (ed. L. A. Dugatkin & H. K. Reeve), pp. 64–93. Oxford, UK: Oxford University Press.

SAS Institute 2003 *The SAS system for Windows*. Cary, NC: SAS Institute Inc.

Stuart-Fox, D. M. & Johnston, G. R. 2005 Experience overrides colour in lizard contests. *Behaviour* **142**, 329–350. (doi:10.1163/1568539053778265)

Stuart-Fox, D. M. & Whiting, M. J. 2005 Male dwarf chameleons assess risk of courting large, aggressive females. *Biol. Lett.* **1**, 231–234. (doi:10.1098/rsbl.2005.0299)

Stuart-Fox, D. M., Firth, D., Moussalli, A., Whiting, M. In press. Multiple traits in chameleon contests: designing and analysing animal contests as a tournament. *Anim. Behav.*

Taylor, P. W. & Elwood, R. W. 2003 The mismeasure of animal contests. *Anim. Behav.* **65**, 1195–1202. (doi:10.1006/anbe.2003.2169)

Zamudio, K. R., Huey, R. B. & Crill, W. D. 1995 Bigger isn't always better—body size, developmental and parental temperature and male territorial success in *Drosophila melanogaster*. *Anim. Behav.* **49**, 671–677.

Zucker, N. & Murray, L. 1996 Determinants of dominance in the tree lizard *Urosaurus ornatus*: the relative importance of mass, previous experience and coloration. *Ethology* **102**, 812–825.