Evolutionary Ecology **17:** 509–522, 2003.

Research article

The Napoleon Complex: why smaller males pick fights

WINFRIED JUST [1,*] and MOLLY R. MORRIS[2]
[1]*Department of Mathematics, Ohio University, Athens, OH, USA;* [2]*Department of Biological Sciences, Ohio University, Athens, OH, USA*
*(*author for correspondence, e-mail: just@math.ohiou.edu)*

Received 25 June 2002; accepted 20 August 2003

Co-ordinating editor: O. Leimar

Abstract. Does it ever pay for smaller animals to initiate fights even when they are likely to lose? Asymmetry in payoffs between opponents or a suboptimal strategy resulting from likely losers misperceiving themselves as likely winners have both been proposed as possible explanations for the aggressive behavior of smaller males. The model presented here suggests that in some cases, even without a payoff asymmetry and allowing for only a small error in perception, likely losers are expected to attack first. If the value of the resource exceeds the cost of losing a fight, the cost of displaying is sufficiently small, and assessment of resource holding power is reasonably accurate but not perfect, the evolutionarily stable strategy (ESS) prompts those contestants who perceive themselves as the likely losers to initiate fights, while it prompts those contestants who perceive themselves as the likely winners to wait for the adversary to attack or retreat.

Key words: aggression, assessment of fighting ability, escalation, evolutionarily stable strategy, resource holding power

Introduction

Ritualized displays are often used to resolve animal conflicts, but contests over limited resources can escalate to costly fights (Enquist and Leimar, 1990). Differences in fighting ability, or resource holding power (abbreviated RHP; Parker, 1974), should influence the decisions animals make in these contests (Hammerstein, 1981; Maynard Smith, 1982). Therefore, Parker (1974) predicted that the probability of an escalated contest should increase as asymmetry in RHP between the contestants decreases. In many organisms, body size is a good indicator of an individual's ability to acquire or hold a resource (Archer, 1988). If individuals are able to assess their relative RHP in a contest, it would be expected that in general, larger individuals should benefit from escalation since smaller individuals are not likely to gain resources by fighting but will often pay a cost of fighting (Parker, 1974). However, it is less clear which of the contestants should escalate first if both contestants would prefer escalation to giving up. For example, a standard model of escalated fighting, the sequential assessment game of Enquist and Leimar (1983) postulates repeated *simultaneous* bouts of potentially costly interaction and does not address the issue of who escalates first.

Evidence from numerous empirical studies supports the prediction that the probability of escalation increases as asymmetry in RHP decreases (for a review see Archer, 1988). Asymmetry in RHP, therefore, can be used to predict whether or not escalation will occur. It is observed that when asymmetry in RHP is large, the individuals with the relatively lower RHP, the probable losers, retreat without escalation. However, when asymmetry in RHP is small and contests do escalate to the use of costly fighting behavior, which individual attacks first (i.e. *initiates* escalation)? Individuals with the relatively higher RHP (likely winners) initiate escalated contests in several different organisms (e.g. mollusks, Zack, 1975; sea anemones, Brace and Pavey, 1978; fishes, Figler and Einhorn, 1983; Barlow *et al.*, 1986; Turner and Huntingford, 1986; hermit crabs, Dowds and Elwood, 1983; Keeley and Grant, 1993), although it should be noted that the confounding interactions between RHP and intruder/resident effects may be important in some of these cases. Recently, however, empirical studies have detected cases where likely losers are more aggressive than expected (Dow *et al.*, 1976; Enquist and Jackobsson, 1986; Morris *et al.*, 1995). In particular, Morris *et al.* (1995) studied the aggressive behavior of initiating contests between males in two species of swordtail fishes (*Xiphophorus nigrensis* and *X. multilineatus*) in the laboratory. When the difference in size between fish was very large, contests were settled without fights with the smaller animal retreating. However, 78% of the observed fights were initiated by the smaller animal, and in 70% of the fights, the fish that delivered the first bite lost the contest. Interest in formulating models that can explain similar counterintuitive observations is growing (Payne, 1998).

Dugatkin and Ohlsen (1990) examined fighting behavior in the pumpkinseed sunfish *Lepomis gibbosus* and reported that holding all else equal, individuals with higher RHP escalated first. However, when they manipulated the value of the contested resource by providing the probable loser with more food, the males that lost more often became more likely to attack first. Ribowski and Frank (1993) detected aggressive losers also in behavioral studies of the swordtail *Xiphophorus helleri*. They demonstrated that omega males, who were losers in previous contests, were more likely to bite first as compared to alpha males. Several mechanisms have been suggested that could explain aggressive behavior in these cases. Losers might be expected to be more aggressive if the value of the resource is greater for the likely loser than the likely winner, in other words, if there is an asymmetry in the value of the resource to the two contestants (Parker, 1974; Dugatkin and Ohlsen, 1990; Dugatkin and Biederman, 1991). Another possible mechanism for the aggressive behavior of lower

RHP individuals is that they might misperceive themselves as having higher RHP and initiate escalation by mistake (Bradbury and Vehrencamp, 1998). This mechanism, however, would be an unlikely explanation of a situation where more than half of all contests are escalated by their eventual losers, because this would entail that role is misperceived in more than half of all fights. Another possible explanation could be the Desperado Effect (Grafen, 1987) where omega males attack because they have no alternative opportunities to obtain resources. Note that this effect in essence postulates an asymmetry in value of the resource. However, the Desperado Effect seems an unlikely explanation for systems in which no escalation is observed when the difference in body size is very large.

In contrast, we wanted to determine if a mechanism that directly involves difference in RHP could also explain the phenomenon of aggressive small males. Note that for example in the experimental setup of Morris *et al.* (1995), none of the asymmetries studied by Parker (1974) seem to be present. In the model we propose, the only asymmetry is a moderate difference in RHP itself, and assessment of RHP is only slightly less than perfect. The purpose of our model is to elucidate a possible scenario under which a unique evolutionarily stable strategy (ESS) prompts contestants to initiate fights precisely when they perceive the odds as slightly (but not overwhelmingly) against them. The model we present demonstrates that this scenario is possible even when there is no payoff asymmetry. While we do not claim that the mechanism discussed in this paper is actually responsible for the observed behavior in any concrete species (e.g. Morris *et al.*, 1995), we do formulate an empirically verifiable prediction of the model. Aggressiveness of potential losers is a phenomenon that could potentially occur in a number of species with a wide range of behavior patterns and mechanisms for assessing fighting ability. Therefore, we have attempted to keep the assumptions and number of parameters to a minimum, which will keep the model as versatile as possible.

Description of the model

We assume that each contest has up to two stages: a 'display' stage, during which no physical contact occurs, followed in some cases by a 'fight' stage during which physical contact occurs. Note that this structure implies that passage from the display stage to the fight stage requires escalation by only one of the contestants.

We assume that the objective of a contest is to gain access to a resource that has a value V for each of the contestants (thus there is no payoff asymmetry). The payoff for each individual is determined by the outcomes of contests, and an individual may be involved in one or more contests during his lifetime.

Displaying carries a cost to both contestants. The costs involved could include losing the opportunity to search for an alternative, undefended resource, as well as energy expended displaying and increased risk of predation. We assume that displaying cost is roughly proportional to the time the display lasts. This assumption is similar to the one made in the classical War of Attrition model of Maynard Smith (1974), but it may not apply to all species (Payne and Pagel, 1996; Payne, 1998). Thus, in our model, the total cost of displaying will be equal to d times t, where d is a parameter of the model and t is the total duration of the display stage. The cost of a fight will normally be considerably higher than the cost of displaying. Although the cost of a fight may depend to some extent on its duration, we will not model the duration of a fight but will instead model costs of a fight by fixed constants that represent average costs and subsume, among other things, the costs associated with fight duration. We allow for a component of the cost of the fight born only by the loser (denoted by K) and a component that is incurred by both the winner and the loser (denoted by L).

We assume that during the display stage contestants try to assess the probability of winning a fight, and that this probability is entirely determined, via a one-to-one correspondence, by the difference in the resource holding power (ΔRHP) of the contestants. This conceptualization of RHP follows Parker (1974) and Archer (1988). Thus we sometimes write that the two contestants assess ΔRHP instead of writing that they assess the probability of winning a fight. Note that the probability of winning the fight is assumed not to depend on who initiates. An easy calculation shows that if x is the probability of winning the fight, then the expected payoff from the fight is $xV - (1 - x)K - L$, which is positive if and only if $x > (K + L)/(V + K)$. Thus a player should escalate only if his probability of winning a fight is above the threshold of $(K + L)/(V + K)$, which we call the escalation threshold.

If the cost of losing a fight is small relative to the resource value V, then $(K + L)/(V + K)$ will be less than 0.5, and we may have a situation where the probability of winning the fight is above the threshold both for the player with the larger RHP and the player with the smaller RHP. Clearly, in this situation each player should prefer to engage in a fight rather than retreat unilaterally. But which player should initiate escalation? Intuitively, one could argue as follows. During the display stage, both players will try to determine whether their probability of winning is above or below this threshold. If information about relative fighting ability is perfect, then whichever contestant first reaches the conclusion that the probability of winning is above the threshold for escalation should initiate escalation. Thus fights should be initiated 50% of the time by the lower RHP individual and 50% of time by the higher RHP individual. However, if we assume that assessment of relative fighting ability is reasonably accurate but not perfect, the fact that the player with the higher

RHP has a probability of winning that is further above the threshold than for the player with the lower RHP will lead to a situation in which the lower RHP individual attacks first. While the player with the higher RHP will still decide to fight rather than retreat, even with a possible small error in estimation of $|\Delta \mathrm{RHP}|$, the player with the lower RHP is much more likely to misperceive his fighting ability as being below the threshold for escalation and, as a consequence, mistakenly give up without a fight. The player with the larger RHP can benefit from such mistakes and avoid the cost of fighting in some cases by delaying escalation. Thus, if the cost of displaying is small relatively to the cost of losing a fight, the player with the larger RHP should leave the initiative to the player with the smaller RHP, while the player with the smaller RHP should escalate to the fighting stage as soon as he perceives that his probability of winning is below 0.5 (i.e. that he is the likely loser) but above the threshold for escalation. This is the basic idea behind our model.

To give mathematically precise meaning to this idea, assume that the parameters K, L, V are such that the escalation threshold $(K + L)/(V + K) < 0.5$. Let x denote the probability of winning a fight for the focal player. Note that the corresponding probability for the adversary will be $1-x$. We will say that the probability x of winning is *very small* if $x < (K + L)/(V + K)$, that it is *small* if $(K + L)/(V + K) < x < 0.5$, that it is *large* if $0.5 < x < 1 - (K + L)/(V + K)$, and that it is *very large* if $1 - (K + L)/(V + K) < x$. These four possibilities will be called the *probability classes*. For simplicity, we will represent the whole distribution of probabilities within each of the four classes by a single representative probability that is treated as a parameter of our model. Thus if the probability of winning is *very small* it will be treated as equal to a, if it is *small* it will be treated as b, if it is *large* it will be treated as $1 - b$, if it is *very large* it will be treated as $1 - a$, where the parameters a and b obey the inequalities $a < (K + L)/(V + K) < b < 0.5$.

The basic assumption of our model is that at any given stage in the displaying process, a player has a choice between three actions: to continue displaying, to retreat, or to escalate to the fighting stage. Moreover, the action of a player should depend on whether he perceives the probability of winning as *very small*, *small*, *large*, or *very large*. Our model allows for partial information about this probability. One can argue that the correct way of modeling partial information would be a distribution of perceived winning probabilities, but this approach is mathematically more complicated and might not be as useful in elucidating our basic ideas. Therefore, we assume for simplicity that at any stage of the displaying process a player perceives the probability of winning as falling into a subset of adjacent probability classes (e.g. {*small*, *large*} would be a subset of adjacent classes, while {*very small*, *very large*} would not). We also allow for perception error, which is a crucial parameter of our model. Thus if the probability of winning a fight is really *small*, a player may eventually misperceive it as *large* (one adjacent probability class) with probability $q[0]$ and as *very small* (the other adjacent probability class) with probability $q[1]$.

A model that takes into account most of the possibilities for the acquisition of partial information was explored by Just and Sun (2003) with the help of computer simulations. In that model, there are altogether 6561 possible strategies, which precludes an analytical treatment of the model. In the model we present here, we investigate one particular scenario of information acquisition that we found plausible as well as analytically tractable. We assume that the encounter starts in a situation where one of the players correctly perceives the probability of winning as either *very small* or *small*, while the other player correctly perceives it as either *very large* or *large*. This scenario seems likely in encounters where the actual probability of winning for the lower RHP contestant is close to the escalation threshold. Note that under these simplifying assumptions we have $q[0] = 0$, since the lower RHP contestant will not misperceive himself as having the higher probability of winning or *vice versa*. The player with the lower RHP will be denoted by P^l; the player with the higher RHP will be denoted by P^H.

We assume that P^l determines at time $t(l)$ whether the difference in RHP is above or below the escalation threshold, and P^H reaches a similar decision at time $t(H)$. We assume that $t(l)$ and $t(H)$ are independent, normally distributed random variables with mean value μ and standard deviation σ. The parameters μ and σ should be such that the probability of $t(l)$ or $t(H)$ assuming negative values is negligible.

It will be convenient to treat a strategy of a player in this restricted version of our model as a pair of *tactics*, i.e. prescriptions of how a player should proceed when he finds himself in a given role (P^l or P^H). If a contestant is in the role of P^l, he may at any given time in the contest perceive the probability of winning as either *very small*, *small*, or he may not yet have reached a decision whether his probability of winning is above the escalation threshold. In each of these three perception states he has a choice between three actions: to continue displaying, to initiate a fight, or to retreat. Thus, there are as many as $3^3 = 27$ possible tactics for P^l. Similarly, P^H has 27 possible tactics, resulting in altogether 729 possible strategies in this game. In order to keep the model analytically tractable, we concentrate on a set of several strategies that appear most plausible.

Initiating a fight at time $t = 0$ by the lower RHP player will be called tactic e ('escalate'). Tactic c ('conditional wait') for the lower RHP individual is to continue to display until a decision about the magnitude of $|\Delta\text{RHP}|$ has been reached, at which time ($t(l)$) he retreats if the probability of winning is perceived as *very small* or initiates a fight if the probability of winning is perceived as *small*. Player P^H is in a somewhat different situation, since the displaying process will never convince him that retreat is preferable to a fight. He has four plausible tactics available to him: tactic E ('escalate') is initiating a fight at $t = 0$; tactic B ('bully') is to display until time $t(H)$, when he determines whether the difference in RHP is large or small, and then to attack if the probability of winning is perceived as *very large* or to continue to display until P^l attacks or retreats if he perceives the probability of winning as *large*; tactic C ('conditional wait') is to display until time $t(H)$, and then to attack if the probability of winning is perceived as *large* or to continue to display until P^l attacks or retreats if he perceives the probability of winning as *very large*; and tactic W ('wait'), which is simply to leave the initiative to the opponent and display until P^l attacks or retreats. Note that in the tactics c, B, and C above, actions are contingent upon whether a player perceives the probability of winning as above or below the escalation threshold.

Our model allows us to construct payoff matrices for the eight strategies described above and to predict ESS's for any given parameter setting. The payoffs were calculated under the assumption that a player would on average find himself in the role of P^l in half of his encounters and in the role of P^H in the other half, and that the probability of winning for P^l would be below the escalation threshold in half of all encounters and above this threshold in the other half. We chose to make these assumptions rather than introducing additional parameters that would make the model even more complicated. One can argue that the assumption of playing the role of P^l in half of the encounters of an individual may be quite realistic in species with a characteristic lifetime history of change in RHP. Alternatively, one could adopt an evolutionary point of view and argue that if there is no linkage between genes that determine strategy and genes that influence RHP, than over evolutionary time, genes that determine a given strategy will half of the time reside in the lower RHP player and half of the time in the higher RHP player. The assumption that the winning probability for P^l is above the threshold for escalation exactly half of the time is admittedly arbitrary. Table 1 shows the payoff matrix for one typical parameter setting. If there is no error in perception (i.e. if $q[1] = 0$), then strategy cC is predicted to be an ESS. However, if a positive probability of misperception of $|\Delta \mathrm{RHP}|$ is assumed ($q[1] = 0.033$ in this example), then cW is predicted to be an ESS.

The fact that a given strategy is an ESS only means that it is stable against invasion by mutant strategies once it is established in a population. It does not necessarily imply that it will evolve in a population that starts out with a mix of different strategies (Hofbauer and Sigmund, 1998). In order to test the evolvability of our predicted ESS's, we wrote a simulation program that monitors the evolution of the percentages of the eight strategies (eE, eB, eC, eW, cE, cB, cC, and cW) given that fitness is proportional to expected payoff. We ran this program for 100 randomly generated parameter settings. In 50 of these parameter settings, the probability of misperception $q[1]$ was set to zero; in the

Table 1. Payoffs for tactics and ESS for specific cases within the parameter space where escalation is expected

P^H/P^l	E	W	B	C
No error in perception of $\|\Delta RHP$ ($q[1] = 0$)				
e	*(22.02) (−12.02)	(22.02) (−12.02)	(22.02) (−12.02)	(22.02) (−12.02)
c	(22.02) (−12.02)	(43.48) (0.02)	(32.33) (−6.42)	(43.56) (0.10) **
Error in perception of $\|\Delta RHP$ ($q[1] = 0.033$)				
e	*(22.02) (−12.02)	(22.02) (−12.02)	(22.02) (−12.02)	(22.02) (−12.02)
c	(22.02) (−12.02)	(43.94) (−0.44) **	(33.51) (−6.46)	(43.09) (−0.55)

Equilibria* are nonstrict and do not give rise to ESS (Selten, 1980). Equilibria ** give rise to ESS. P^H is the player with the higher RHP and P^l is the player with the lower RHP. The payoffs for the higher RHP player are on top. Arrows point to the higher payoff for each player and assist in identification of ESS (Gardner, 1995). Values for parameters other than $q[1]$ in both cases were:

Table 2. Winning strategy and average percentage (standard deviation) of population in which the loser and the lower RHP player attacks first from simulations in which we examined 50 parameter settings where information was not perfect ($0 < q[1] < 0.2$) as well as 50 parameter settings with perfect ($q[1] = 0$) information about $|\Delta \text{RHP}|$.

Value of $q[1]$	Winning strategies	Average (SD) % attacks first	
		Loser	P^l
$0 < q[1] < 0.2$	78% *cW*	62.3 (6.43)	100.0 (0.00)
	22% not *cW*	67.8 (12.01)	89.5 (19.55)
$q[1] = 0$	82% *cC*	50.0 (0.00)	50.0 (0.00)
	18% not *cC*	67.8 (7.70)	98.7 (2.43)

Contests were between all strategies (*eE*, *eC*, *eW*, *eB*, *cE*, *cC*, *cW*, *cB*). Starting frequencies for all strategies were equal (12.5%). The values for the parameter settings were randomly chosen within the following ranges: $V = 100$, $0 < K < 50$, $0 < L < 50$, $0 < a < 0.5$, $0 < b < 0.5$, $0 < d < 5$, $\mu = 1$, $\sigma = 0.3$. The inequalities $0 < a < (K + L)/(V + K) < b < 0.5$ were always satisfied.

other 50 simulations $q[1]$ was randomly chosen from the interval (0, 2). All simulations were started with an even mix of the eight strategies mentioned above. Table 2 summarizes the results of these simulations. In 39 of the 50 simulations with a positive error of perception, *cW* emerged as the winning strategy, with 100% of all fights being initiated by P^l and an average of 62.3% of all fights being initiated by their eventual losers. In nine of the remaining simulations with these parameter settings we saw mixtures of *eB*, *eC*, *eW* and sometimes *eE* evolving; and again a majority of the fights being initiated by P^l and the eventual loser. The latter effect could not be predicted from the payoff matrix alone; it seems to be due at least partially to our choice of the initial mix of strategies. In two of these simulations (with $q[1] = 0.00775542$ and $q[1] = 0.0046444$), *cC* emerged as the winning strategy. This confirms that *cC* may be an ESS not only when the perception error is exactly zero, but also when it is very small. It is interesting to note that if $q[1] > 0$, in a population of *cC* players, a slight majority of fights will still be initiated by P^l. The reason for this is that if the winning probability for the player with the smaller RHP is in reality small and is misperceived by P^H (but not P^l) as very small, we always will see a fight that is initiated by P^l. In contrast, if P^l (but not P^H) misperceives his winning probability as very small, than we will see fights (initiated by P^H) only half of the time, and unilateral retreat by P^l in the other half of such encounters.

In 41 out of the 50 simulations with $q[1] = 0$, strategy *cC* evolved as the ESS. Note that when the perception error is zero, the model dictates that P^l will initiate exactly half of all fights, and the eventual loser will initiate exactly half. In the remaining nine simulations with $q[1] = 0$, we again saw mixtures of *eB*, *eC*, *eW* and *eE* evolving.

Discussion

Counter-intuitive aggressiveness by likely losers has sometimes been attributed to payoff asymmetries (Dugatkin and Ohlsen, 1990), likely losers misperceiving themselves as likely winners (Bradbury and Vehrencamp, 1998), or differences in personal history (Ribowski and Franck, 1993). In contrast, our model stipulates that under certain parameter constellations with equal payoffs to both players, even if assessment of RHP is only slightly less than perfect, it is precisely the perception of having somewhat lower RHP than the opponent (being the likely loser) that drives the lower RHP contestant to pick a fight. For the player with the higher RHP, the probability of winning is far enough above the escalation threshold $(K + L)/(V + K)$ that engaging in a fight is always preferable to a retreat. On the other hand, the optimal course of action for the player with the lower RHP will depend more on any error in perception of the difference in RHP; the closer his probability of winning is to the threshold, the more likely he is to make a mistake and retreat when above the threshold or attack when below. Therefore, if perception of this probability is somewhat less than perfect and the cost of displaying is sufficiently small, then the player with the higher RHP can benefit by leaving the initiative to the opponent and hoping for him to retreat, either because of realistic perception of his chances to win the fight or by mistake. In some senses then, the mechanism we have described for the aggressive behavior of probable losers is not based on losers being more likely to attack first but on probable winners being less likely to attack first. In this sense, our proposed explanation for what is commonly considered the 'Napoleon Complex' might be more appropriately identified as the 'Gentle Giant Syndrome.' This possible explanation for the aggressive behavior of probable losers involving the initiation of contests has not been previously identified in the literature.

The model presented here indicates that the mechanism described above is possible. The simulations we have presented show that for a wide range of parameter settings with a positive perception error, the strategy in which fights are initiated by likely losers (*cW*) will evolve as the ESS. If perception error is zero, or very close to zero, most of the time the strategy that calls for initiation of a fight by whomever first perceives the winning probability of P^l as above the escalation threshold (*cC*) is the ESS. If the whole population follows *cC* and the perception error is exactly zero, then exactly half of all fights will be initiated by P^l and exactly half of the fights will be initiated by their eventual losers. This shows that the perception error $q[I]$ is a crucial parameter of our model.

In parameter settings with a relatively high cost of displaying and very low costs of fight, strategies tended to become dominant that favor immediate initiation of fights. In our simulations, this led to populations with much higher proportions of strategies *eW*, *eC*, and *eB* than *eE* or *cE*, so that the great majority of fights were still initiated by P^l. The latter, however, is most likely an artifact due to the initial mix of strategies in our simulations.

Several of our simplifying assumptions, especially linearity of displaying costs, normality and independence of the times $t(l)$ and $t(H)$, and picking representative winning probabilities for each of the four probability classes, were necessary to make the model analytically tractable and because we do not have data on these functions and probability distributions in real animals. Even with all these simplifying assumptions, the formulas for our payoffs are rather complicated, which makes it impractical to precisely characterize the regions of the parameter space where strategy *cW* is the only ESS. Therefore, the statistical approach taken in this paper seems justified. We discuss the possible impact of some of the assumptions of our model below.

Some of our simplifying assumptions on acquisition of partial information during the display stage have been relaxed in related work. First of all, one can ask what happens if one does not concentrate just on the eight most plausible strategies but takes into account all 729 possible strategies in our game. This number of strategies precludes an analytical treatment, but the model still can be explored with simulated evolution for selected parameter settings. Such an approach was taken in (Just *et al.*, 2000), and the results of the simulations largely confirm the predictions of the present paper, with cW often evolving when $q[1] > 0$ and cC evolving when $q[1] = 0$. One can go one step further and allow for more possibilities of partial information than we do here. Such a model has been explored for selected parameter settings by Just and Sun (2003) with the help of computer simulations. Their results partially confirm the predictions of the model presented here. A major difference is that for parameter settings with both $q[0] > 0$ and $q[1] > 1$, ESS's tend to evolve that are mixtures of cW-like strategies and cC-like strategies, rather than pure cW-like strategies. A similar prediction was made by Just *et al.* (in review), where the process of information acquisition was not explicitly modeled. Note that a mix of cW-like strategies and cC-like strategies still results in a situation where most fights are initiated by P^l and thus most fights are initiated by their eventual loser.

We assume in our model that the cost of displaying is the same for the lower and higher RHP individual, and if information concerning relative RHP is gained during the display stage, it is not based on differences in cost of displaying. However, if one way to honestly signal higher RHP is to display longer, then the lower RHP individual may choose to end the display stage sooner than the higher RHP individual in some regions of the parameter space. This would suggest yet another mechanism, in addition to the one we present here, by which lower RHP individuals might be driven to initiate contests more often than higher RHP individuals. However, if to display more vigorously is an honest signal of higher RHP, then the cost of displaying would be higher for the higher RHP individual. In our model, differences in RHP are going to be small if both contestants are above the escalation threshold, and therefore any difference in the costs of displaying is likely to be small as well. We predict that a slightly greater display cost for the higher RHP individual would, to some extent, reduce the parameter space in which the lower RHP individual would initiate contests. It is possible that cases where the cost of displaying is higher for the higher RHP individual could help explain why current empirical data suggests that higher RHP individuals initiate contests more often than our model predicts.

The most likely reason why we do not see pure cW-like ESS's evolving if we allow for misperception of role as well as absolute value of $|\Delta\text{RHP}|$ is that the presence of cC-like players in the population acts as an insurance policy against too many encounters where both players persist indefinitely. The fact that pure cW appears to be an ESS if no perception error of role is allowed, while we see a cW/cC-like mix evolving in the presence of such errors leads us to the conjecture that in a more realistic setting, the ratio of fights that are escalated by the player with the lower RHP grows larger as the actual difference in RHP gets closer to the threshold, since in these cases misperception of role should become much less likely than misperception of absolute value of the difference. This conjecture will be explored elsewhere.

Our results open several other new directions of research. First, the phenomenon of aggressive losers predicted by our model may be more common in nature than previously recognized, but may have gone largely unnoticed due to the paucity of published data on how often which contestant initiates fights. In order to settle this question, more experimental data are needed. We suggest that experimentalists record more data on which contestant initiates which stage of escalation.

Second, recall that in the experiments on swordtail fishes by Morris *et al.* (1995) 70% of all fights were initiated by the eventual loser and 78% of the fights were initiated by the smaller animal. This could be consistent with a cW-like strategy in the population, if, as the authors of the study (Morris *et al.*, 1995) conjectured, RHP for swordtail fishes is determined largely but not exclusively by body size. Alternatively, the observed behavior could be the result of a cW/cC-like mix of strategies. More data allowing for more sophisticated statistical analyses of RHP are needed to determine whether almost all fights are initiated by the contestant who perceives himself as the likely loser (i.e. cW adopted by the whole population), or whether likely losers just pick fights more often than likely winners (as in a cW/cC mix). One of our conjectures is that with increasing asymmetry in size (within the range where fights are predicted), an increasing ratio of the observed fights will be initiated by the smaller contestant. This effect should become observable if sufficiently many contests are staged between animals. Experiments on swordtail fish are under way to test these hypotheses.

Third, while the model presented here explains how counterintuitive aggressiveness of potential losers could in principle evolve, it would be interesting to determine the robustness of our findings given increasingly realistic models of contests. For example, we want to study the effect of a finer gradation of perception of $|\Delta\text{RHP}|$ than the simple large/small dichotomy. The development of a new kind of simulation algorithm that would allow testing of the above and similar scenarios is under way.

Acknowledgements

We thank Jeff Holt for running some of the computer simulations for us, Xiaolu Sun for help with the simulation program, and Shala Hankison, and Harold R. Swardson for valuable comments on earlier versions of the manuscript. Special thanks are due to the referees of earlier versions for asking pointed questions that led us to clarify our assumptions and to considerably improve our presentation. This research was supported by a grant to WJ from the National Science Foundation (DBI-9904799).

References

Archer, J. (1988) *The Behavioural Biology of Aggression*. Cambridge University Press, Cambridge.

Barlow, G.W., Rogers, W. and Fraley, N. (1986) Do Midas Cichlids win through prowess or daring? It depends. *Behav. Ecol. Sociobiol.* **19**, 1–8.

Brace, R.C. and Pavey, J. (1978) Size-dependent dominance hierarchy in the anemone *Actinia equina*. *Nature* **273**, 752–753.

Bradbury, J.W. and Vehrencamp, S.L. (1998) *Principles of Animal Communication*. Sinauer Associates, Sunderland, MA, pp. 667–710.

Dow, M., Ewing, A.W. and Sutherland, I. (1976) Studies on the behavior of cyprinodont fish III. The temporal patterning of aggresssion in *Aphyosemion striatum* (Boulenger). *Behaviour* **59**, 252–268.

Dowds, B.M. and Elwood, R.W. (1983) Shell wars: assessment strategies and the timing of decisions in hermit crab shell fights. *Behaviour* **85**, 1–24.

Dugatkin, L.A. and Biederman, L. (1991) Balancing asymmetries in resource holding power and resource value in the pumpkinseed sunfish. *Anim. Behav.* **42**, 691–692.

Dugatkin, L.A. and Ohlsen, S.R. (1990) Contrasting asymmetries in value expectation and resource holding power: effects on attack behavior and dominance in the pumpkinseed sunfish, *Lepomis gibbosus*. *Anim. Behav.* **39**, 802–804.

Enquist, M. and Jackobsson, S. (1986) Decision making and assessment in the fighting behavior of *Nannacara anomala* (Cichlidae, Pisces). *Ethology* **72**, 143–153.

Enquist, M. and Leimar, O. (1983) Evolution of fighting behaviour: decision rules and assessment of relative strength. *J. Theor. Biol.* **102**, 387–410.

Enquist, M. and Leimar, O. (1990) The evolution of fatal fighting. *Anim. Behav.* **39**, 1–9.

Figler, M.H. and Einhorn, D.M. (1983) The territorial prior residence effect in convict cichlids (*Chichlasoma nigrofasciatum* Gunther): temporal aspects of establishment and retention and proximate mechanisms. *Behaviour* **85**, 157–181.

Gardner, R. (1995) *Games for Business and Economics*. John Wiley and Sons, Inc., New York.

Grafen, A. (1987) The logic of diversively asymmetric contests: respect for ownership and the desperado effect. *Anim. Behav.* **35**, 462–467.

Hammerstein, P. (1981) The role of asymmetries in animal contests. *Anim. Behav.* **19**, 193–205.

Hofbauer, J. and Sigmund, K. (1998) *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge, UK.

Just, W. and Sun, X. (2003) Simulating the evolution of contest escalation. In A. Barry (ed.) *Workshop Program for GECCO 2003 (Genetic and Evolutionary Computation Conference)*, 75–77.

Just, W., Wu, M. and Holt, J. (2000) How to evolve a Napoleon Complex. In A. Zalzala (ed.) *Proceedings of 2000 Congress on Evolutionary Computation*. IEEE Press, pp. 851–856.

Keeley, E.R. and Grant, J.W.A. (1993) Asymmetries in the expected value of food do not predict the outcome of contests between convict cichlids. *Anim. Behav.* **45**, 1035–1037.

Maynard Smith, J. (1974) The theory of games and the evolution of animal conflicts. *J. Theor. Biol.* **47**, 209–221.

Maynard Smith, J. (1982) *Evolution and the Theory of Games*. Cambridge University Press, Cambridge.

Morris, M.R., Gass, L. and Ryan, M.J. (1995) Assessment and individual recognition of opponents in the swordtails *Xiphophorus nigrensis* and *X. multilineatus*. *Behav. Ecol. Sociobiol.* **37**, 303–310.

Parker, G.A. (1974) Assessment strategy and evolution of fighting behaviour. *J. Theor. Biol.* **47**, 223–243.
Payne, R.J.H. (1998) Gradually escalating fights and displays: the cumulative assessment model. *Anim. Behav.* **56**, 651–662.
Payne, R.J.H. and Pagel, M. (1996) Escalation and time costs in displays of endurance. *J. Theor. Biol.* **183**, 185–193.
Ribowski, A. and Franck, D. (1993) Demonstration of strength and concealment of weakness in escalating fights of male swordtails (*Xiphophorus helleri*). *Ethology* **93**, 265–271.
Selten, R.J. (1980) A note on evolutionarily stable strategies in asymmetric animal conflicts. *J. Theor. Biol.* **84**, 93–101.
Turner, G.F. and Huntingford, F. (1986) A problem for game theory analysis: assessment and intention in male mouthbrooder contests. *Anim. Behav.* **34**, 961–970.
Zack, S. (1975) A description and analysis of agonistic behaviour patterns in an opisthobranch mollusk, *Hermissenda crassicornis*. *Behaviour* **5**, 238–267.