doi:10.1111/j.1558-5646.2011.01477.x

SURVIVAL OF THE WEAKEST IN *N*-PERSON DUELS AND THE MAINTENANCE OF VARIATION UNDER CONSTANT SELECTION

Marco Archetti[1,2,3]

[1]*Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, Massachusetts 02138*

[3]*E-mail: marco.archetti@unibas.ch*

Received May 9, 2011
Accepted September 27, 2011

The persistence of extensive variation in nature seems to stand against the most general principle of evolution by natural selection: in antagonistic interactions, the stronger type is expected to replace the weaker. Game theory shows that, however, in contrast to this intuitive expectation for interactions between two players, strategic considerations on fitness maximization in repeated pairwise interactions between three players (truels) or more (*N*-person duels) lead to what can be dubbed "survival of the weakest": the weakest individual can have the highest fitness. A paradox arises: competitive skills cannot be improved by natural selection, unless we assume mutations with strong effects or unless we assume that interactions are exclusively between two individuals. The paradox disappears, however, with more realistic assumptions (a mixture of duels and truels; the attacked individual backfires; the contest can end without a winner; defensive and offensive skills are correlated; players not directly involved in the contest suffer collateral damage). An unexpected new result emerges: the weaker types can persist in a population in the absence of recurrent mutations, migration, and fluctuating selection. Game theory and the analysis of *N*-person duels, therefore, help understand one of the most enduring puzzles in evolutionary biology: the maintenance of variation under constant selection.

KEY WORDS: **Conflict, duel, game theory, *N*-person duel, strategy, truel, variation.**

THE DUEL: SURVIVAL OF THE FITTEST

The most general principle of evolution by natural selection is what Darwin (1869) dubbed "survival of the fittest" (Spencer 1864): the stronger types in the struggle for survival and reproduction increase in frequency, thereby leading to the design we see in nature. Although in certain cases selection can improve robustness (e.g., Wilke et al. 2001) or antirobustness (e.g., Archetti 2009) of genotypes at the expenses of fitness, it is understood that if a stronger type has an advantage over a weaker type in an antagonistic interaction it also has the highest probability of surviving.

Consider the duel as a model for antagonistic interactions. Two individuals, A and B, shoot at each other, with accuracies (probabilities to hit the opponent) a and b, respectively. If they shoot at the same time and $a > b$, clearly A has a higher probability of winning the contest. Consider then a sequential, repeated duel: at the beginning, and after each shot, who shoots next is chosen at random: again clearly A has a higher probability of winning the contest. There is nothing surprising here, but a problem arises: if selection is constant (a is always greater than b) the stronger type A will eventually replace the weaker type B in the population; selection erodes variation. What maintains variation under constant selection is one of the most enduring puzzles of population genetics (Barton and Turelli 1989; Charlesworth and Hughes 1999; Barton and Keightley 2002).

[2]Present address: Department of Business and Economics, University of Basel, Peter Merian-Weg 6, 4002 Basel, Switzerland.

Game theory shows, however, that in contrast to this intuitive expectation for interactions between two players, in duels between three or more players another apparent paradox arises, which can be dubbed "survival of the weakest."

THE TRUEL: SURVIVAL OF THE WEAKEST

Consider a three-person version of the duel (a "truel"; Shubik 1954, 1964, 1982). Three individuals, A, B, and C, shoot at each other with accuracies a, b, and c, respectively. At the beginning, and after each shot, who shoots next is chosen at random among the players still in the contest (being hit means quitting the game with payoff 0, not necessarily being killed). Who will be the most likely to win? The answer here is not so simple as in the two-person duel; one must take a strategic decision: whom to shoot at?

Suppose that C has been eliminated. With the same probability (1/2) A or B will shoot next. The payoff for A can be calculated as follows: if A shoots and hits the target B (which happens with probability a) A's payoff is 1; if B shoots and hits the target A (which happens with probability b) A's payoff is 0; if both miss the target (which happens with probability $1 - a/2 - b/2$), the status quo is repeated. The probability that A will ultimately win against B alone, therefore, is

$$P_{\mathrm{A(B)}} = a/2 + (1 - a/2 - b/2)P_{\mathrm{A(B)}},$$

which gives

$$P_{\mathrm{A(B)}} = a/(a+b).$$

Similarly, for each type $X,Y \in \{\mathrm{A,B,C}\}$ and probability $x,y \in \{a,b,c\}$ the probability that X will ultimately win against Y in a two-person duel is (Shubik 1954, 1964, 1982)

$$P_{X(Y)} = x/(x+y).$$

Suppose the players are ranked in skill by $a > b > c$ and that these values are common knowledge. Then, because the probability of hitting the target does not depend on the target, whom to shoot at first depends only on $P_{X(Y)}$; that is, on whom one prefers to face in the two-person duel.

Which type (T_X) should type X shoot at in the three-person duel? It is easy to see that A would prefer to fight against C than against B in a two-person duel, because $P_{\mathrm{A(C)}} > P_{\mathrm{A(B)}}$; therefore A should shoot at B first ($T_{\mathrm{A}} = \mathrm{B}$). B would prefer to fight against C than against A in a two-person duel, because $P_{\mathrm{B(C)}} > P_{\mathrm{B(A)}}$; therefore B should shoot at A first ($T_{\mathrm{B}} = \mathrm{A}$). C would prefer to fight against B than against A in a two-person duel, because $P_{\mathrm{C(B)}} > P_{\mathrm{C(A)}}$; therefore C should shoot at A first ($T_{\mathrm{C}} = \mathrm{A}$). In summary, in a three-person duel, the best strategy is to shoot at the strongest opponent: if the three players are still in the game, both B and C will shoot at A; A will shoot at B; nobody will shoot at C (Shubik 1954, 1964, 1982).

Note that the model does not allow or predict the possibility of "shooting in the air"; that is, of passing one's turn. This could be a sensible strategy in a different type of duel in which players shoot in a predefined order (if C starts, e.g., and he hits A, then B will shoot at C; if C hits B, then A will shoot at C; if C shoots in the air, A will then shoot at B, and B will shoot at A, therefore it is better for C to shoot in the air). In the model discussed here, however (in which who shoots next is chosen at random), if C hits A, C will have the payoff of a duel with B, which is always greater than what C gets by shooting in the air (in which case the status quo is repeated). Throughout this article, I will ignore the possibility of "shooting in the air."

We can write the payoff $P_{X(Y,Z)}$ of type X as the probability of winning against Y and Z. Let us work out the example of $P_{\mathrm{A(B,C)}}$. With probability 1/3, A is the one who starts and with probability a he hits the target (which, as we have seen, is B); if he hits the target (which happens with probability a) he will eventually have to face C, in which case his chance of winning is $P_{\mathrm{A(C)}}$. B and C also start with probability 1/3 each, and they hit the target (which, as we have seen, is A) with probabilities b and c, respectively, in which case the payoff for A is 0. With probability $(1 - a/3 - b/3 - c/3)$, instead, A will miss the target and so will B and C, in which case the process is repeated. Therefore, the probability that A will ultimately win against B and C in a three-person duel is

$$P_{\mathrm{A(B,C)}} = (a/3)P_{\mathrm{A(C)}} + (b/3)0 + (c/3)0 + (1 - a/3 - b/3 - c/3)P_{\mathrm{A(B,C)}}$$

and, in a similar way, for B against A and C, and for C against A and B:

$$P_{\mathrm{B(A,C)}} = (a/3)0 + (b/3)P_{\mathrm{B(C)}} + (c/3)P_{\mathrm{B(C)}} + (1 - a/3 - b/3 - c/3)P_{\mathrm{B(A,C)}}$$

$$P_{\mathrm{C(A,B)}} = (a/3)P_{\mathrm{C(A)}} + (b/3)P_{\mathrm{C(B)}} + (c/3)P_{\mathrm{C(B)}} + (1 - a/3 - b/3 - c/3)P_{\mathrm{C(A,B)}}$$

Some algebra gives (Shubik 1954, 1964, 1982):

$$P_{\mathrm{A(B,C)}} = [a^2]/[(a+b+c)(a+c)]$$

$$P_{\mathrm{B(A,C)}} = [b(a+c)]/[(a+b+c)(a+c)]$$

$$P_{\mathrm{C(A,B)}} = [c(2a+c)]/[(a+b+c)(a+c)].$$

An apparent paradox emerges: the weakest player can have the highest probability of winning and the strongest player can

have the lowest, unless the differences in skills are extremely large; more specifically, the weakest player C has the highest probability of winning if

$$c > a(\sqrt{2} - 1)$$

$$c > b/2 - a + \sqrt{(a^2 + b^2/4)}.$$

For example, with $a = 0.8$, $b = 0.6$, and $c = 0.4$, we get $P_{A(B,C)} = 0.296$, $P_{B(A,C)} = 0.333$, $P_{C(A,B)} = 0.370$. Note that what matters is not the absolute value of the skills but their relative values. What seems paradoxical (the weakest type can have the highest fitness) is actually the result of rational, strategic considerations.

The three-person duel has been known since the beginnings of game theory. Shubik (1954, 1964, 1982) used the model described above to show that strategic interactions can lead to seemingly paradoxical results. A similar game has been known for a long time as a mathematical curiosity (Kinnaird 1946; Larsen 1948) and it has been discussed in political economy (Kilgour 1972, 1975, 1978; Kilgour and Brams 1997), where it has implications for strategic voting in multiple-party, winner-take-all elections.

Here I extend Shubik's model to interactions in evolving populations with multiple players, I generalize it to *N*-person duels, and to situations with more realistic assumptions (the attacked individual backfires; the contest ends without a winner; defensive and offensive skills are correlated; players not directly involved in the fight suffer collateral damage). This allows to apply the logic of the truel not only to antagonistic interactions in shooting contests, but also to more general fighting contests and to cases in which the players compete indirectly.

Methods and Results

TRUELS IN EVOLVING POPULATIONS

Shubik's truel assumes repeated pairwise interactions between three different types. In a population, however, the players will be chosen at random, with probabilities proportional to their frequencies, which change over time. For each type $X,Y,Z \in \{A,B,C\}$ and skill $x,y,z \in \{a,b,c\}$, the probabilities $P_{X(Y,Z)}$ that X will win against Y and Z in all possible truels are shown in Table 1.

The fitness of type X in a population, where X has frequency f_X and the other two types Y and Z have frequencies f_Y and f_Z, respectively, is

$$W_X = f_X^2 P_{X(X,X)} + f_Y^2 P_{X(Y,Y)} + f_Z^2 P_{X(Z,Z)} + 2 f_X f_Y P_{X(X,Y)} + 2 f_X f_Z P_{X(X,Z)} + 2 f_Y f_Z P_{X(Y,Z)}$$

Stability of X requires that

$$P_{X(X,X)} > P_{Y(X,X)}$$

and

$$P_{X(X,X)} > P_{Z(X,X)}$$

It is easy to prove the following conditions for the stability of each type:

$$\text{A}: \quad a > 2b$$

$$\text{B}: \quad a < 2b \text{ and } b > 2c$$

$$\text{C}: \quad a < 2c \text{ and } b < 2c$$

Table 1. Payoffs in a population version of Shubik's truel.

$P_{A(B,C)}$	$(a/3)P_{A(C)}+(b/3)0+(c/3)0+(1-a/3-b/3-c/3)P_{A(B,C)}$
$P_{B(A,C)}$	$(a/3)0+(b/3)P_{B(C)}+(c/3)P_{B(C)}+(1-a/3-b/3-c/3)P_{B(A,C)}$
$P_{C(A,B)}$	$(a/3)P_{C(A)}+(b/3)P_{C(B)}+(c/3)P_{C(B)}+(1-a/3-b/3-c/3)P_{C(A,B)}$
$P_{A(A,B)}$	$(a/3)P_{A(B)}+(a/3)0+(b/3)(0+P_{A(B)})/2+(1-2a/3-b/3)P_{A(A,B)}$
$P_{A(A,C)}$	$(a/3)P_{A(C)}+(a/3)0+(c/3)(0+P_{A(C)})/2+(1-2a/3-c/3)P_{A(A,C)}$
$P_{A(B,B)}$	$(a/3)P_{A(B)}+2(b/3)0+(1-a/3-2b/3)P_{A(B,B)}$
$P_{A(C,C)}$	$(a/3)P_{A(C)}+2(c/3)0+(1-a/3-2c/3)P_{A(C,C)}$
$P_{A(A,A)}$	$(a/3)(1/2)+2(a/3)(0+1/2)/2+(1-a)P_{A(A,A)}$
$P_{B(B,A)}$	$2(b/3)(1/2)+(a/3)(0+P_{B(A)})/2+(1-a/3-2b/3)P_{B(B,A)}$
$P_{B(B,C)}$	$(b/3)P_{B(C)}+(b/3)0+(c/3)(0+P_{B(C)})/2+(1-2b/3-c/3)P_{B(B,C)}$
$P_{B(A,A)}$	$(b/3)P_{B(A)}+2(a/3)P_{B(A)}+(1-2a/3-b/3)P_{B(A,A)}$
$P_{B(C,C)}$	$(b/3)P_{B(C)}+2(c/3)0+(1-b/3-2c/3)P_{B(C,C)}$
$P_{B(B,B)}$	$(b/3)(1/2)+2(b/3)(0+1/2)/2+(1-b)P_{B(B,B)}$
$P_{C(B,C)}$	$2(c/3)(1/2)+(b/3)(0+P_{C(B)})/2+(1-b/3-2c/3)P_{C(B,C)}$
$P_{C(A,C)}$	$2(c/3)(1/2)+(a/3)(0+P_{C(A)})/2+(1-a/3-2c/3)P_{C(A,C)}$
$P_{C(A,A)}$	$(c/3)P_{C(A)}+2(a/3)P_{C(A)}+(1-2a/3-c/3)P_{C(A,A)}$
$P_{C(B,B)}$	$(c/3)P_{C(B)}+2(b/3)P_{C(B)}+(1-2b/3-c/3)P_{C(B,B)}$
$P_{C(C,C)}$	$(c/3)(1/2)+2(c/3)(0+1/2)/2+(1-c)P_{C(C,C)}$

Otherwise there is no equilibrium of pure types. Therefore, if differences in skill are extremely large, either A (if both b and c are much smaller than a) or B (if c is much smaller than a but b is not) are stable. At intermediate values of b and c none of the pure types is stable and the frequencies of the three types change cyclically. If differences in skill are not extreme, however, only type C, the weakest, is stable.

Thus, another related, apparent paradox arises in three-person duels played in evolving populations: only mutants with extremely higher competitive abilities can invade a population fixed on a weaker type. This means that selection cannot lead to a gradual improvement of competitive abilities in interactions between more than two players. The logic of the theory is indisputable, but the result is disturbing: either we assume that competition occurs exclusively in two-person interactions (which seems unlikely), or we accept that only mutations with strong effects have an actual impact on evolutionary change (which contrasts with the standard view of gradual evolutionary change).

Is there a solution to the paradox? In the next sections I will show how more realistic assumptions can lead not only to a solution, but also to understand how variation persists in nature under constant selection.

A MIXTURE OF DUELS AND TRUELS

If the frequency of truels is τ (<1) and all other interactions are two-person duels, the payoff of type X in a population where it has frequency f_X and the other two types Y and Z have frequencies f_Y and f_Z, respectively, is

$$(1-\tau)(f_X P_{X(X)} + f_Y P_{X(Y)} + f_Z P_{X(Z)} +)\tau W_X.$$

Stability of X requires that

$$(1-\tau)P_{X(X)} + \tau P_{X(XX)} > (1-\tau)P_{Y(X)} + \tau P_{Y(XX)}$$

and

$$(1-\tau)P_{X(X)} + \tau P_{X(XX)} > (1-\tau)P_{Z(X)} + \tau P_{Z(XX)}$$

The conditions for the stability of each type when $\tau < 1$ are

A: $a > 2b$ or [$a < 2b$ and $\tau < 6a/(a+b) - 3$]

B: $a < 2b$ and $(3+\tau)c < b(3-\tau)$ and $\tau > \dfrac{3(2b^2 - a^2 - ab)}{a^2 + 2b^2 - 9ab}$

C: $b < 2c$ and $a < 2c$ and $\tau > \dfrac{3(2c^2 - a^2 - ac)}{a^2 + 2c^2 - 9ac}$,

Again, unless the differences in skill are extreme, the weakest type C is the only stable type irrespective of skills or τ, and will go to fixation; otherwise B or C will go to fixation, or the three types will coexist in a cyclical polymorphism (Fig. 1). Variation can persist at intermediate values of τ and skills without any kind of fluctuating selection, recurrent mutations, or migration: fluctuations in the frequencies of the three types depend entirely on the strategic nature of the interactions, under constant selection. Note that the period of the oscillations can be very long, in the order of thousands of generations (Fig. 1), therefore this could look like a stable polymorphism in short-term data from natural populations.

LIMITED TRUELS: CONTESTS CAN END WITHOUT A WINNER

In Shubik's truel, the three players fight repeatedly until only one player is left. In a *limited* truel, instead, I assume that at any time there is a probability that the interaction ends and there is no winner at all (it is also possible to assume that if an interaction ends without a winner, the players still in the contest share the reward of the contest, but results are not very different in this case and will not be shown here). The assumptions of the model are the same as in Shubik's truel, except that after each shot, the interaction goes on with probability $\omega < 1$. In a duel, therefore, for each type $X,Y \in \{A,B,C\}$ and probability $x,y \in \{a,b,c\}$:

$$P_{X(Y)} = x/2 + (1 - x/2 - y/2)\omega P_{X(Y)},$$

that is,

$$P_{X(Y)} = x/[2(1-\omega) + \omega(x+y)].$$

The payoffs of the truel can be obtained simply by multiplying by ω the expressions for $P_{X(Y,Z)}$ in Table 1.

When interactions have a probability ($\omega < 1$) of continuing after each shot, C goes to fixation if all interactions are truels ($\tau = 1$). With a combination of duels and truels ($\tau < 1$), three further types of dynamics are possible, besides the four (stability of A, B, or C, cyclical polymorphisms) we have already seen in the previous section. Both A and B, or both B and C can be stable, or even all three types depending on the parameters. The conditions for the stability of each type are too cumbersome to be reported here; a graphical representation of the results is given in Figure 2. Which equilibrium will be achieved depends on the initial frequencies of the three types.

DEFENSIVE TRUELS: DEFENSIVE AND OFFENSIVE SKILLS ARE CORRELATED

Another unrealistic assumption in Shubik's truel is the fact that the probability of hitting the target depends only on the skill of the shooter and not on the skill of the target. In a *defensive* truel

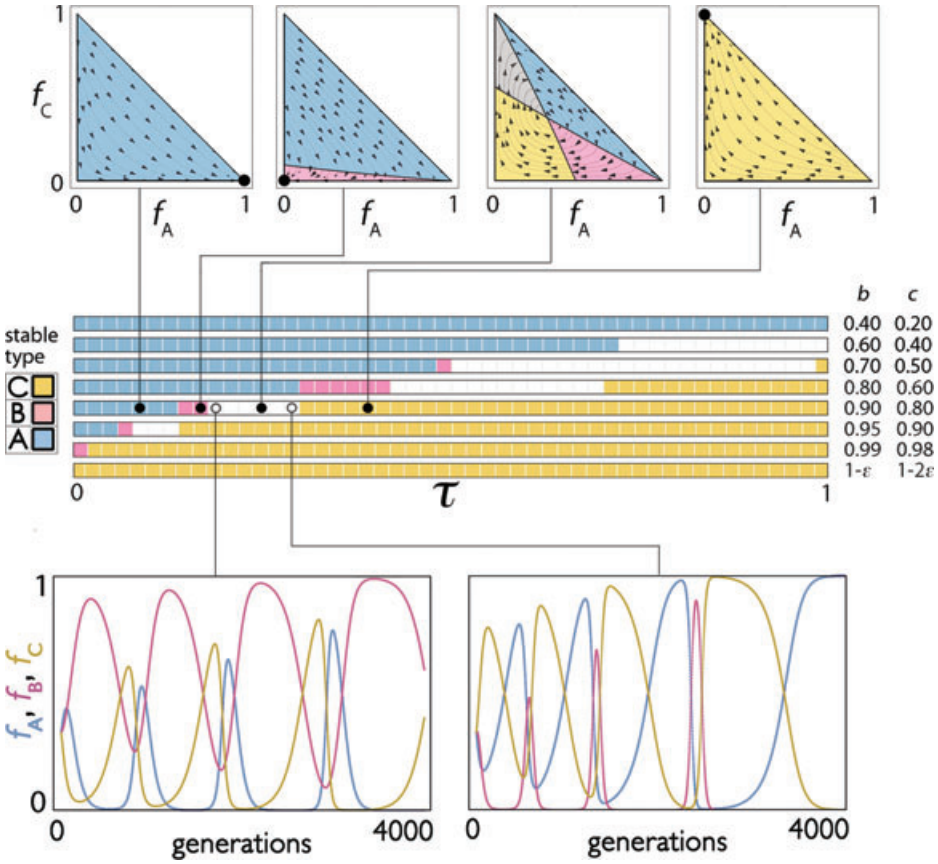

Figure 1. Equilibria and evolutionary dynamics of 2/3-person duels. The top plots, drawn for $a = 1$, $b = 0.9$, $c = 0.8$ (the skills of the three players A, B, and C) and four different values of τ (the proportion of three-person duels: 0.1, 0.17, 0.25, and 0.4), show the direction of change of the frequency of A (the strongest player) and C (the weakest player) (f_A and f_C, respectively; $f_B = 1 - f_A - f_C$); in each region either f_A or f_C or both increase; the black circles show the stable equilibria. The bars show which type goes to fixation, for different values of skills b and c ($a = 1$; $\varepsilon = 0.0001$) as a function of τ; when no type is stable the population fluctuates between the three types. The bottom plots are also drawn for $a = 1$, $b = 0.9$, $c = 0.8$ and for two different values of τ for which the dynamics is cyclical ($\tau = 0.2$ and 0.3), and show how f_A, f_B, and f_C change over time.

instead I assume that a player with strong offensive skills may also have strong defensive skills. The assumptions of the model are the same as in Shubik's truel, except that the probability that type X hits target Y is $x_Y = x - \delta y$, where δy is a measure of the defensive skills of Y. I assume that the correlation between offensive and defensive skills $\delta$is the same for all types, and that δ is small enough to keep $x_Y > 0$. In a duel, therefore, for each type $X,Y \in \{A,B,C\}$ and probability $x,y \in \{a,b,c\}$

$$P_{X(Y)} = x_Y/(x_Y + y_X)$$

and

$$P_{A(B,C)} = (a_B/3)P_{A(C)} + (b_A/3)0 + (c_A/3)0 + (1 - a_B/3 - b_A/3 - c_A/3)P_{A(B,C)}$$

$$P_{B(A,C)} = (a_B/3)0 + (b_A/3)P_{B(C)} + (c_A/3)P_{B(C)} + (1 - a_B/3 - b_A/3 - c_A/3)P_{B(A,C)}$$

$$P_{C(A,B)} = (a_B/3)P_{C(A)} + (b_A/3)P_{C(B)} + (c_A/3)P_{C(B)} + (1 - a_B/3 - b_A/3 - c_A/3)P_{C(A,B)}$$

The payoffs for all other possible truels in a population are shown in Table 2.

When defensive skills are correlated with offensive skills, most cases have either one equilibrium or a cyclical dynamics; only for a very limited parameter range bistability is possible (both A and B are stable; Fig. 2).

INTERFERENCE TRUELS: COLLATERAL DAMAGE

Shubik's truel assumes a physical contest or any antagonistic interaction in which it is possible to direct one's competitive skills exclusively toward one other opponent. In this section, I extend the model to antagonistic interactions in which competition is indirect, for example, competition for shared resources. Consider three individuals competing for the same resource: they do not

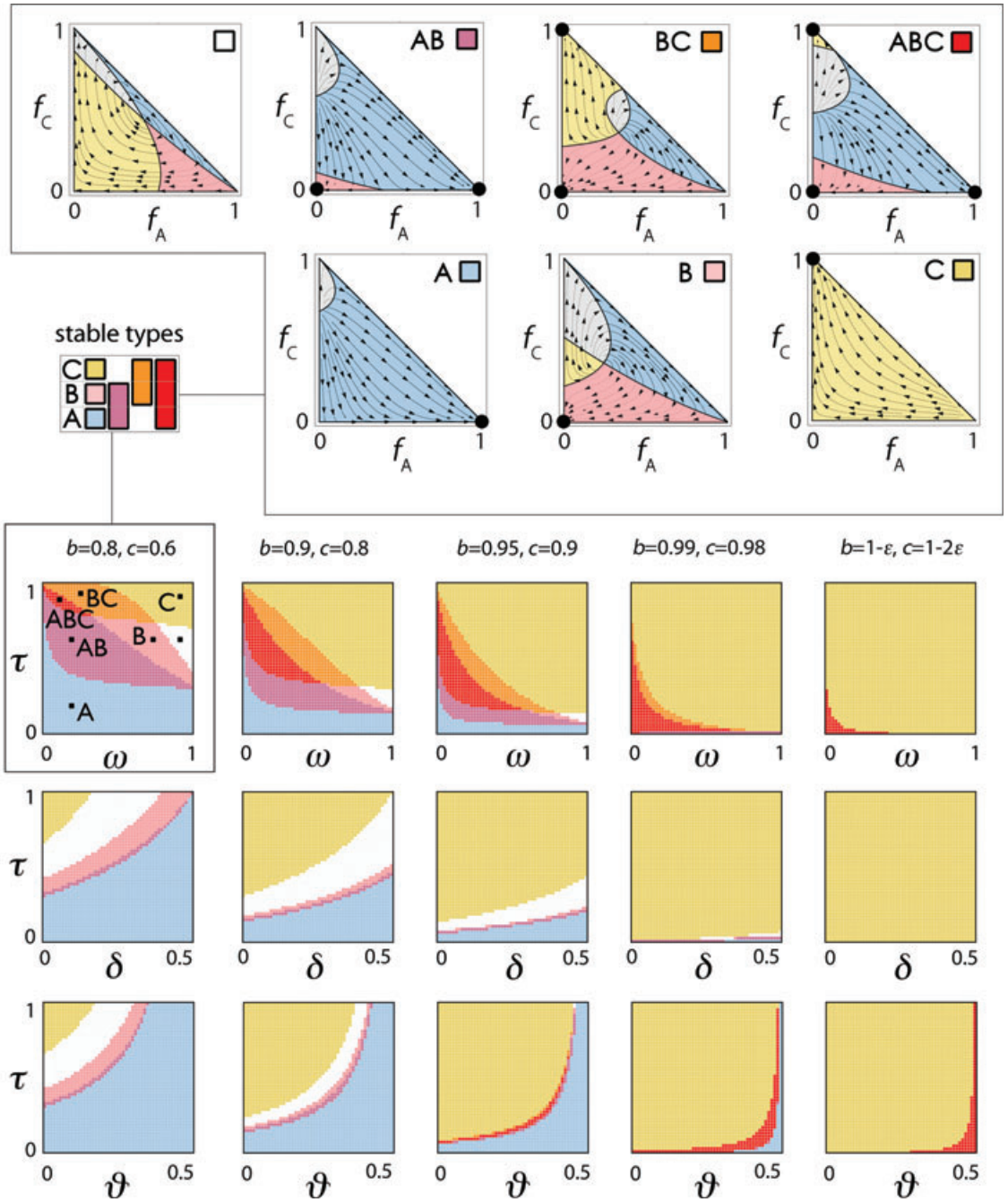

Figure 2. Equilibria and evolutionary dynamics of limited truels, defensive truels, and interference truels. The top plots show, for different values of τ (the proportion of three-person duels) and ω (the probability of continuing the interaction after each shot) corresponding to different types of equilibrium (indicated by the one to three letter code on the top right of each panel, each corresponding to one of the dots indicated by the same letters in the first of the bottom plots), the direction of change of the frequencies A (the strongest player) and C (the weakest player) (f_A and f_C, respectively; $f_B = 1 - f_A - f_C$); in each region either f_A or f_C or both increase; the black circles show the stable equilibria. The skills of the three players A, B, and C are $a = 1$, $b = 0.8$, $c = 0.6$, respectively. The bottom plots show, for different values of b and c ($a = 1$), the type of equilibria as a function of τ and one other parameter: ω (the probability of continuing the interaction after each shot), δ (the correlation between defensive and offensive ability), or ϑ (amount of collateral damage). Each black dot in the upper left plot shows the parameters used in the top panel associated with each equilibrium type.

Table 2. Payoffs of the defensive truels.

$P_{A(B,C)}$	$(a_B/3)P_{A(C)}+(b_A/3)0+(c_A/3)0+(1-a_B/3-b_A/3-c_A/3)P_{A(B,C)}$
$P_{B(A,C)}$	$(a_B/3)0+(b_A/3)P_{B(C)}+(c_A/3)P_{B(C)}+(1-a_B/3-b_A/3-c_A/3)P_{B(A,C)}$
$P_{C(A,B)}$	$(a_B/3)P_{C(A)}+(b_A/3)P_{C(B)}+(c_A/3)P_{C(B)}+(1-a_B/3-b_A/3-c_A/3)P_{C(A,B)}$
$P_{A(A,B)}$	$(a_A/3)P_{A(B)}+(a_A/3)0+(b_A/3)(0+P_{A(B)})/2+(1-2a_A/3-b_A/3)P_{A(A,B)}$
$P_{A(A,C)}$	$(a_A/3)P_{A(C)}+(a_A/3)0+(c_A/3)(0+P_{A(C)})/2+(1-2a_A/3-c_A/3)P_{A(A,C)}$
$P_{A(B,B)}$	$(a_B/3)P_{A(B)}+2(b_A/3)0+(1-a_B/3-2b_A/3)P_{A(B,B)}$
$P_{A(C,C)}$	$(a_C/3)P_{A(C)}+2(c_A/3)0+(1-a_C/3-2c_A/3)P_{A(C,C)}$
$P_{A(A,A)}$	$(a_A/3)(1/2)+2(a_A/3)(0+1/2)/2+(1-a_A)P_{A(A,A)}$
$P_{B(B,A)}$	$2(b_A/3)(1/2)+(a_B/3)(0+P_{B(A)})/2+(1-a_B/3-2b_A/3)P_{B(B,A)}$
$P_{B(B,C)}$	$(b_B/3)P_{B(C)}+(b_B/3)0+(c_B/3)(0+P_{B(C)})/2+(1-2b_B/3-c_B/3)P_{B(B,C)}$
$P_{B(A,A)}$	$(b_A/3)P_{B(A)}+2(a_A/3)P_{B(A)}+(1-2a_A/3-b_A/3)P_{B(A,A)}$
$P_{B(C,C)}$	$(b_C/3)P_{B(C)}+2(c_B/3)0+(1-b_C/3-2c_B/3)P_{B(C,C)}$
$P_{B(B,B)}$	$(b_B/3)(1/2)+2(b_B/3)(0+1/2)/2+(1-b_B)P_{B(B,B)}$
$P_{C(B,C)}$	$2(c_B/3)(1/2)+(b_C/3)(0+P_{C(B)})/2+(1-b_C/3-2c_B/3)P_{C(B,C)}$
$P_{C(A,C)}$	$2(c_A/3)(1/2)+(a_C/3)(0+P_{C(A)})/2+(1-a_C/3-2c_A/3)P_{C(A,C)}$
$P_{C(A,A)}$	$(c_A/3)P_{C(A)}+2(a_A/3)P_{C(A)}+(1-2a_A/3-c_A/3)P_{C(A,A)}$
$P_{C(B,B)}$	$(c_B/3)P_{C(B)}+2(b_B/3)P_{C(B)}+(1-2b_B/3-c_B/3)P_{C(B,B)}$
$P_{C(C,C)}$	$(c_C/3)(1/2)+2(c_C/3)(0+1/2)/2+(1-c_C)P_{C(C,C)}$

shoot at each other, or fight physically; instead, they take actions that interfere with the actions of both other players. Although one cannot decide to target exclusively one of the other players, one can bias this action to interfere preferentially with one of the two. We can think of this as a duel with weapons that produce *collateral damage*, measured by the parameter ϑ ($0 \le \vartheta \le 0.5$; if $\vartheta = 0$ the effects of shooting are entirely paid by the target, as in Shubik's truel; if $\vartheta = 0.5$ the effects of shooting are paid equally by the target and the nontarget).

Suppose that C has been eliminated. With the same probability (1/2) A or B will shoot next. Consider the payoff for A: if A shoots and hits the target B, which happens with probability $(1-\vartheta)a$, A's payoff is 1; if B shoots and hits the target A, which happens with probability $(1-\vartheta)b$, A's payoff is 0; if both miss the target, which happens with probability $1-(1-\vartheta)a/2-(1-\vartheta)b/2$, the status quo is repeated. The probability that A will ultimately win against B alone, therefore, is $P_{A(B)} = (1-\vartheta)a/2 + [1-(1-\vartheta)a/2-(1-\vartheta)b/2]P_{A(B)}$, which gives $P_{A(B)} = a/(a+b)$, as in Shubik's truel. In general, therefore, $P_{X(Y)} = x/(x+y)$, irrespective of ϑ; this is clear, because collateral damage has no consequences in a two-person duel, it only reduces accuracies for both players.

In the truel, however, ϑ does affect payoffs. Consider the payoff for A. With probability 1/3, A shoots: he hits his favorite target (which is B) with probability $a(1-\vartheta)$, in which case A's payoff is $P_{A(C)}$ (because he faces C in the eventual duel); but because $\vartheta > 0$, A hits C (not his favorite target) with probability $a\vartheta$, in which case A's payoff is $P_{A(B)}$ (because he faces B in the eventual duel). With probability 1/3, B shoots: he hits his favorite target (which is A) with probability $b(1-\vartheta)$, in which case A's payoff is 0; but because $\vartheta > 0$, B hits C (not his favorite target) with probability $b\,\vartheta$, in which case A's payoff is $P_{A(B)}$ (because he faces B in the eventual duel). With probability 1/3, it is C that shoots; he hits his favorite target (which is A) with probability $c\,(1-\vartheta)$, in which case A's payoff is 0; but because $\vartheta > 0$, C hits B (not his favorite target) with probability $c\,\vartheta$, in which case A's payoff is $P_{A(C)}$ (because he faces C in the eventual duel). With probability $(1 - a/3 - b/3 - c/3)$ nobody is hit and the process is repeated. This yields

$$\begin{aligned}P_{A(B,C)} = {} & (a/3)[(1-\vartheta)P_{A(C)} + \vartheta P_{A(B)}] + (b/3)[(1-\vartheta)0 + \vartheta P_{A(B)}] \\ & + (c/3)[(1-\vartheta)0 + \vartheta P_{A(C)}] \\ & + (1 - a/3 - b/3 - c/3)P_{A(B,C)}\end{aligned}$$

Similarly,

$$\begin{aligned}P_{B(A,C)} = {} & (a/3)[(1-\vartheta)0 + \vartheta P_{B(A)}] + (b/3)[(1-\vartheta)P_{B(C)} + \vartheta P_{B(A)}] \\ & + (c/3)[(1-\vartheta)P_{B(C)} + \vartheta 0] \\ & + (1 - a/3 - b/3 - c/3)P_{B(A,C)}\end{aligned}$$

$$\begin{aligned}P_{C(A,B)} = {} & (a/3)[(1-\vartheta)P_{C(A)} + \vartheta 0] + (b/3)[(1-\vartheta)P_{C(B)} + \vartheta 0] \\ & + (c/3)[(1-\vartheta)P_{C(B)} + \vartheta P_{C(A)}] \\ & + (1 - a/3 - b/3 - c/3)P_{C(A,B)}\end{aligned}$$

That is,

$$P_{A(B,C)} = [a(a+2c\vartheta)]/[(a+b+c)(a+c)]$$

$$P_{B(A,C)} = b/(a+b+c)$$

Table 3. Payoffs of the interference truels.

$P_{A(B,C)}$	$(a/3)[(1-\vartheta)P_{A(C)}+\vartheta P_{A(B)}]+(b/3)[(1-\vartheta)0+\vartheta P_{A(B)}]+(c/3)[(1-\vartheta)0+\vartheta P_{A(C)}]+(1-a/3-b/3-c/3)P_{A(B,C)}$
$P_{B(A,C)}$	$(a/3)[(1-\vartheta)0+\vartheta P_{B(A)}]+(b/3)[(1-\vartheta)P_{B(C)}+\vartheta P_{B(A)}]+(c/3)[(1-\vartheta)P_{B(C)}+\vartheta 0]+(1-a/3-b/3-c/3)P_{B(A,C)}$
$P_{C(A,B)}$	$(a/3)[(1-\vartheta)P_{C(A)}+\vartheta 0]+(b/3)[(1-\vartheta)P_{C(B)}+\vartheta 0]+(c/3)[(1-\vartheta)P_{C(B)}+\vartheta P_{C(A)}]+(1-a/3-b/3-c/3)P_{C(A,B)}$
$P_{A(A,B)}$	$(a/3)[(1-\vartheta)P_{A(B)}+\vartheta/2]+(a/3)[(1-\vartheta)0+\vartheta/2]+(b/3)(0+P_{A(B)})/2+(1-2a/3-b/3)P_{A(A,B)}$
$P_{A(A,C)}$	$(a/3)[(1-\vartheta)P_{A(C)}+\vartheta/2]+(a/3)[(1-\vartheta)0+\vartheta/2]+(c/3)(0+P_{A(C)})/2+(1-2a/3-c/3)P_{A(A,C)}$
$P_{A(B,B)}$	$(a/3)P_{A(B)}+2(b/3)[(1-\vartheta)0+\vartheta P_{A(B)}]+(1-a/3-2b/3)P_{A(B,B)}$
$P_{A(C,C)}$	$(a/3)P_{A(C)}+2(c/3)[(1-\vartheta)0+\vartheta P_{A(C)}]+(1-a/3-2c/3)P_{A(C,C)}$
$P_{A(A,A)}$	1/3
$P_{B(B,A)}$	$(b/3)[(1-\vartheta)/2+\vartheta P_{B(A)}]+(b/3)[(1-\vartheta)/2+\vartheta 0]+(a/3)(0+P_{B(A)})/2+(1-a/3-2b/3)P_{B(B,A)}$
$P_{B(B,C)}$	$(b/3)[(1-\vartheta)P_{B(C)}+\vartheta/2]+(b/3)[(1-\vartheta)0+\vartheta/2]+(c/3)(0+P_{B(C)})/2+(1-2b/3-c/3)P_{B(B,C)}$
$P_{B(A,A)}$	$(b/3)P_{B(A)}+2(a/3)[(1-\vartheta)P_{B(A)}+\vartheta 0]+(1-2a/3-b/3)P_{B(A,A)}$
$P_{B(C,C)}$	$(b/3)P_{B(C)}+2(c/3)[(1-\vartheta)0+\vartheta P_{B(C)}]+(1-b/3-2c/3)P_{B(C,C)}$
$P_{B(B,B)}$	1/3
$P_{C(B,C)}$	$(c/3)[(1-\vartheta)/2+\vartheta P_{C(B)}]+(c/3)[(1-\vartheta)/2+\vartheta 0]+(b/3)(0+P_{C(B)})/2+(1-b/3-2c/3)P_{C(B,C)}$
$P_{C(A,C)}$	$(c/3)[(1-\vartheta)/2+\vartheta P_{C(A)}]+(c/3)[(1-\vartheta)/2+\vartheta 0]+(a/3)(0+P_{C(A)})/2+(1-a/3-2c/3)P_{C(A,C)}$
$P_{C(A,A)}$	$(c/3)P_{C(A)}+2(a/3)[(1-\vartheta)P_{C(A)}+\vartheta 0]+(1-2a/3-c/3)P_{C(A,A)}$
$P_{C(B,B)}$	$(c/3)P_{C(B)}+2(b/3)[(1-\vartheta)P_{C(B)}+\vartheta 0]+(1-2b/3-c/3)P_{C(B,B)}$
$P_{C(C,C)}$	1/3

$$P_{C(A,B)}=c[c-2a(\vartheta-1)]/[(a+b+c)(a+c)].$$

The payoffs for all possible truels in a population are shown in Table 3.

With collateral damage, besides bistability and cyclical dynamics, tristability is also possible, in particular at low values of τ and high values of ϑ (nonnegligible collateral damage for the third type; Fig. 2).

TRUELS WITH BACKFIRE: IMMEDIATE RETALIATION

The most unrealistic feature of Shubik's truel applied to animal behavior is that the probability of winning a fight, in the model, does not depend on the skill of the attacked individual. In real contests, instead, attacking a strong individual is likely to be more costly than attacking a weak individual if the attacked individual can immediately backfire (as he is attacked). Here I take into account this backfire effect. In a backfire truel, "shooting" at one player simply means choosing whom to fight with: after X has chosen his opponent Y, they attack each other, and they can both hit each other in the same turn.

If σ is the degree of immediate backfire by the target Y when attacked by X (if $\sigma = 1$, Y backfires efficiently, hitting X with the same probability y with which Y would hit X had Y shot first; if $\sigma = 0$, Y does not backfire at all and the model is equivalent to Shubik's truel), for each type $X,Y \in \{A,B,C\}$ and probability $x,y \in \{a,b,c\}$, in a duel between X and Y

$$P_{X(Y)} = \sigma[x(1-y)+(1-x)(1-y)P_{X(Y)}] + (1-\sigma)[x/2+(1-x/2-y/2)P_{X(Y)}]$$

because if the target Y backfires, the shooter X has payoff 1 if he hits the target (Y) but only if Y misses the target (X), whereas the status quo is repeated if both miss, and X has payoff 0 if Y hits the target; if the target does not react, as in Shubik's model, X has a payoff 1 when he shoots first and hits the target, 0 when Y shoots first and hits the target, and the status quo is repeated if both miss.

Whom to shoot at first still depends on $P_{X(Y)}$ but now this must be conditional on not being hit by the backfire of the target. The payoff A gets for shooting at B first is

$$P_A|[T_A=B] = (1-\sigma b)\{aP_{A(C)}+(1-a)P_A|[T_A=B]\}$$

whereas the payoff for shooting at C first is

$$P_A|[T_A=C] = (1-\sigma c)\{aP_{A(B)}+(1-a)P_A|[T_A=C]\}$$

because with probability σy the target Y backfires successfully and the shooter's payoff is 0; with probability $(1-\sigma y)$, the target does not backfire (or does but unsuccessfully) and in this case the shooter will go on to face the other player if he hits the target; if he does not, the status quo is repeated. It is easy to verify that $P_A|[T_A=B] > P_A|[T_A=C]$ if $\sigma > 0$; that is, it is still the best choice for A to shoot at B first (and, with the same logic, for B and C to shoot at A first); as σ increases the advantage of shooting at one's preferred target decreases, and when $\sigma = 1$ (the target backfires perfectly) a player is indifferent about shooting at either opponent first. Even a very slight advantage for shooting first, due, for example, to the fact that the target is not ready to backfire, therefore, still leads to the same shooting preferences as in Shubik's model (A will prefer to shoot at B first, and both B and C will prefer to shoot at A first).

Table 4. Payoffs of the backfire truels.

$P_{A(B,C)}$	$(2/3)\{a[(1-b)P_{A(C)}]+(1-a)[(1-b)P_{A(B,C)}]\}+(1/3)\{a[(1-c)P_{A(B)}]+(1-a)[(1-c)P_{A(B,C)}]\}$
$P_{B(A,C)}$	$(2/3)(1-a)[bP_{B(C)}+(1-b)P_{B(A,C)}]+(1/3)\{a[c+(1-c)P_{B(A)}]+(1-a)[cP_{B(C)}+(1-c)P_{B(A,C)}]\}$
$P_{C(A,B)}$	$(2/3)\{a[b+(1-b)P_{C(A)}]+(1-a)[bP_{C(B)}+(1-b)P_{C(A,B)}]\}+(1/3)(1-a)[cP_{C(B)}+(1-c)P_{C(A,B)}]$
$P_{A(A,B)}$	$(2/3)\{a[(1-a)P_{A(B)}]+(1-a)[(1-a)P_{A(A,B)}]\}+(1/6)\{a[(1-b)P_{A(A)}]+(1-a)[(1-b)P_{A(A,B)}]\}$ $+(1/6)\{a[b+(1-b)P_{A(A)}]+(1-a)[bP_{A(B)}+(1-b)P_{A(A,B)}]\}$
$P_{A(A,C)}$	$(2/3)\{a[(1-a)P_{A(C)}]+(1-a)[(1-a)P_{A(A,C)}]\}+(1/6)\{a[(1-c)P_{A(A)}]+(1-a)[(1-c)P_{A(A,C)}]\}$ $+(1/6)\{a[c+(1-c)P_{A(A)}]+(1-a)[cP_{A(C)}+(1-c)P_{A(A,C)}]\}$
$P_{B(B,C)}$	$(2/3)\{b[(1-b)P_{B(C)}]+(1-b)[(1-b)P_{B(B,C)}]\}+(1/6)\{b[(1-c)P_{B(B)}]+(1-b)[(1-c)P_{B(B,C)}]\}$ $+(1/6)\{b[c+(1-c)P_{B(B)}]+(1-b)[cP_{B(C)}+(1-c)P_{B(B,C)}]\}$
$P_{B(B,A)}$	$(1/2)\{b[(1-a)P_{B(B)}]+(1-b)[(1-a)P_{B(B,A)}]\}+(1/2)\{b[a+(1-a)P_{B(B)}]+(1-b)[aP_{B(A)}+(1-a)P_{B(B,A)}]\}$
$P_{C(C,A)}$	$(1/2)\{c[(1-a)P_{C(C)}]+(1-c)[(1-a)P_{C(C,A)}]\}+(1/2)\{c[a+(1-a)P_{C(C)}]+(1-c)[aP_{C(A)}+(1-a)P_{C(C,A)}]\}$
$P_{C(C,B)}$	$(1/2)\{c[(1-b)P_{C(C)}]+(1-c)[(1-b)P_{C(C,B)}]\}+(1/2)\{c[b+(1-b)P_{C(C)}]+(1-c)[bP_{C(B)}+(1-b)P_{C(C,B)}]\}$
$P_{A(B,B)}$	$a(1-b)P_{A(B)}+(1-a)(1-b)P_{A(B,B)}$
$P_{A(C,C)}$	$a(1-c)P_{A(C)}+(1-a)(1-c)P_{A(C,C)}$
$P_{B(C,C)}$	$b(1-c)P_{B(C)}+(1-b)(1-c)P_{B(C,C)}$
$P_{B(A,A)}$	$(2/3)\{a[a+(1-a)P_{B(A)}]+(1-a)[aP_{B(A)}+(1-a)P_{B(A,A)}]\}+(1/3)[b(1-a)P_{B(A)}+(1-b)(1-a)P_{B(A,A)}]$
$P_{C(A,A)}$	$(2/3)\{a[a+(1-a)P_{C(A)}]+(1-a)[aP_{C(A)}+(1-a)P_{C(A,A)}]\}+(1/3)[c(1-a)P_{C(A)}+(1-c)(1-a)P_{C(A,A)}]$
$P_{C(B,B)}$	$(2/3)\{b[b+(1-b)P_{C(B)}]+(1-b)[bP_{C(B)}+(1-b)P_{C(B,B)}]\}+(1/3)[c(1-b)P_{C(B)}+(1-c)(1-b)P_{C(B,B)}]$
$P_{A(A,A)}$	$(2/3)\{a(1-a)P_{A(A)}+(1-a)(1-a)P_{A(A,A)}\}+(1/3)\{a[a+(1-a)P_{A(A)}]+(1-a)[aP_{A(A)}+(1-a)P_{B(A,A)}]\}$
$P_{B(B,B)}$	$(2/3)\{b(1-b)P_{B(B)}+(1-b)(1-b)P_{B(B,B)}\}+(1/3)\{b[b+(1-b)P_{B(B)}]+(1-b)[bP_{B(B)}+(1-b)P_{B(B,B)}]\}$
$P_{C(C,C)}$	$(2/3)\{c(1-c)P_{C(C)}+(1-c)(1-c)P_{C(C,C)}\}+(1/3)\{c[c+(1-c)P_{C(C)}]+(1-c)[cP_{C(C)}+(1-c)P_{C(C,C)}]\}$

Now, however, because one suffers higher immediate damage from shooting at a stronger opponent, it is not obvious that the result will be the same as in Shubik's model. A strong opponent will backfire more than a weaker opponent, which clearly is detrimental for weak types—an effect that is absent in Shubik's model; on the other hand, now one's opponents also backfire at each other, which may give an advantage to the weakest player. Consider C; the immediate result of shooting at A now will be worse than in Shubik's model, because A backfires more strongly than B; however, now A and B also backfire at each other, which gives a higher payoff to C.

In what follows, I assume $\sigma \approx 1$ (efficient backfire). We can write the payoff $P_{X(Y,Z)}$ as in Shubik's truel. Let us work out the example of $P_{A(B,C)}$. With probability 1/3, A starts and picks up a fight with B; with probability 1/3, B starts and picks up a fight with A; with probability 1/3, C starts and picks up a fight with A. In the first case, A's payoff is 0 with probability b; A hits B and B does not hit A with probability $(1-b)a$, in which case A's payoff is $P_{A(C)}$; with probability $(1-a)(1-b)$ the status quo is repeated. The second case is equivalent. In the third case, A's payoff is 0 with probability c; A hits C and C does not hit A with probability $(1-c)a$, in which case A's payoff is $P_{A(B)}$; with probability $(1-a)(1-c)$ the status quo is repeated. Therefore,

$$\begin{aligned} P_{A(B,C)} = {} & (2/3)\{a[(1-b)P_{A(C)}]+(1-a)[(1-b)P_{A(B,C)}]\} \\ & +(1/3)\{a[(1-c)P_{A(B)}] \\ & +(1-a)[(1-c)P_{A(B,C)}]\} \end{aligned}$$

$P_{B(A,C)}$ can be derived in a similar way. In a fight between A and B (which happens 2/3 times) B's payoff is 0 with probability a; B hits A and A does not hit B with probability $(1-a)b$, in which case B's payoff is $P_{B(C)}$; with probability $(1-a)(1-b)$ the status quo is repeated. In a fight between C and A (which happens with probability 1/3), B's payoff is 1 if A and C hit each other, which happens with probability ac; it is $P_{B(C)}$ if only A hits C and $P_{B(A)}$ if only C hits A; the status quo is repeated if both A and C miss. Therefore,

$$\begin{aligned} P_{B(A,C)} = {} & (2/3)(1-a)[bP_{B(C)}+(1-b)P_{B(A,C)}] \\ & +(1/3)\{a[c+(1-c)P_{B(A)}] \\ & +(1-a)[cP_{B(C)}+(1-c)P_{B(A,C)}]\} \end{aligned}$$

With a similar logic it is easy to derive

$$\begin{aligned} P_{C(A,B)} = {} & (2/3)\{a[b+(1-b)P_{C(A)}] \\ & +(1-a)[bP_{C(B)}+(1-b)P_{C(A,B)}]\} \\ & +(1/3)(1-a)[cP_{C(B)}+(1-c)P_{C(A,B)}] \end{aligned}$$

and the payoffs for all possible truels (Table 4).

The apparent paradox observed in Shubik's truel persists even in backfire truels: the weakest player can have the highest probability of winning and the strongest player can have the lowest, unless the differences in skills are extreme. In fact, if skills are high, the "survival of the weakest" effect occurs for an even larger parameter range; for example, with $a = 0.9$ and $b = 0.5$, if c is as low as 0.25 we still observe the "survival of the weakest"; with $a = 0.8$, $b = 0.6$, and $c = 0.4$, we get $P_{A(B,C)} = 0.190$,

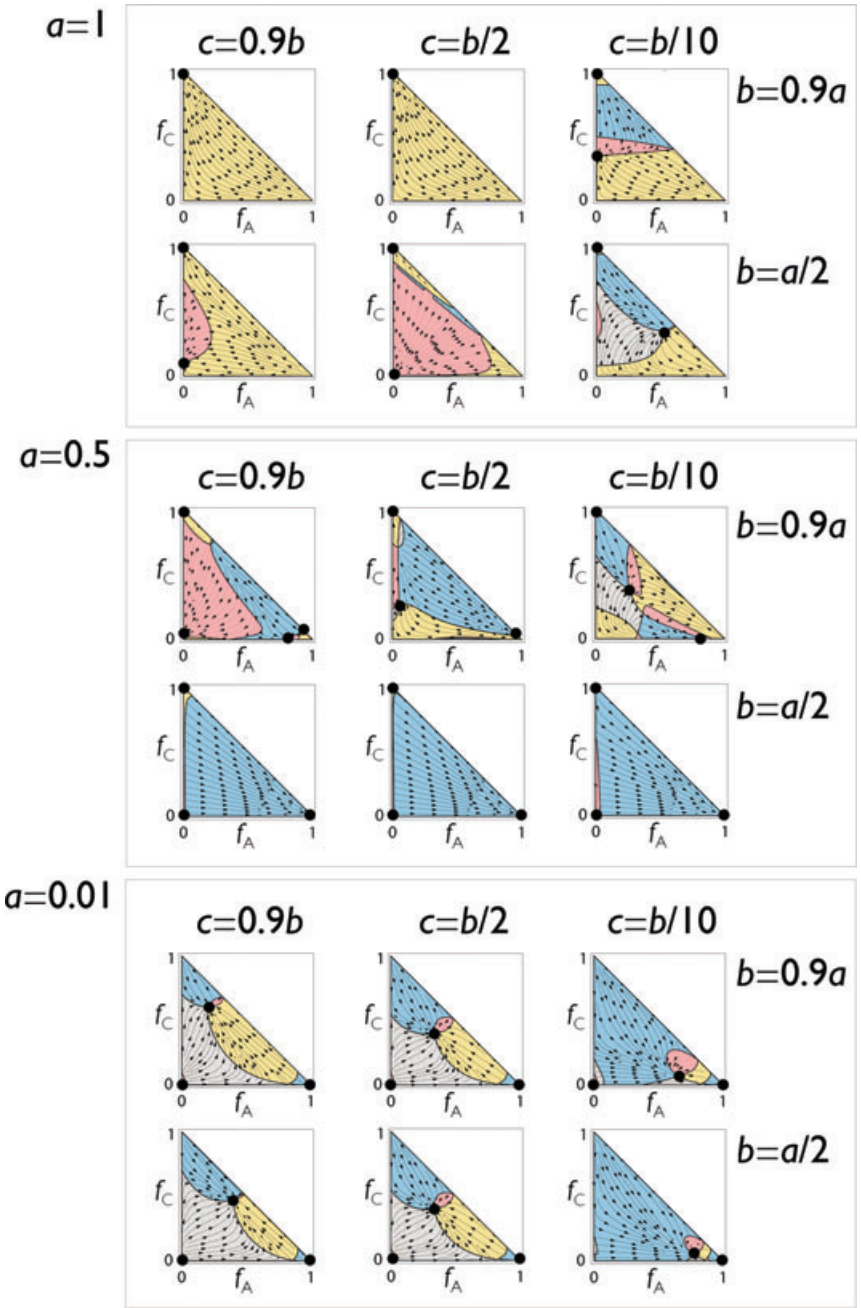

Figure 3. Equilibria and evolutionary dynamics of backfire truels. For different values of *a*, *b*, and *c* (the skills of the three players A, B, and C), each plot shows the direction of change of the frequencies of A (the strongest player) and C (the weakest player) (f_A and f_C, respectively; $f_B = 1 - f_A - f_C$); in each region either f_A or f_C or both increase; the black circles show the stable equilibria; τ (the proportion of three-person duels) = 1.

$P_{B(A,C)} = 0.196$, $P_{C(A,B)} = 0.399$; with $a = 0.1$, $b = 0.08$, $c = 0.065$, we get $P_{A(B,C)} = 0.315$, $P_{B(A,C)} = 0.306$, $P_{C(A,B)} = 0.339$. Note that here absolute skills matter, rather than just relative skills as in Shubik's truel. The "survival of the weakest" in backfire truels might seems even less intuitive than in Shubik's truel, because the weakest individual has a higher *direct* disadvantage in fights with any other opponent. Remember, however, that the "survival of the weakest" effect is due to the fact that the other two types (the average and the strongest) preferentially fight against each other and that this gives an *indirect* advantage to the weakest type; in backfire truels this indirect advantage to the weakest type can increase because the other two types backfire at each other.

The possible combinations of skills are so many that it is difficult to classify the results for all of them. Some relevant examples are shown in Figure 3. The most important result to

point out, which is absent in the models that we have seen so far, is the possibility of a *stable* polymorphism of two or even three types.

N-PERSON DUELS

An obvious extension of the theory of three-person duels is the study of *N*-person duels. This section extends Shubik's truel to four and more players.

In a four-person duel, player X should shoot at the opponent whose disappearance would confer X the highest payoff in the eventual three-person duel. In general, this depends on the combinations of the skills of the players. The possible combinations are too many to lead to any practical result unless we make some simplifying assumption. Let us assume, therefore, we are in the parameter region in which the most likely winners in the three-person duel are, in descending order, the weakest, the average, and finally the strongest (as we have seen, this requires only that the differences in skills are not extreme).

Therefore, in the four-person duel player X should chose to shoot at the player whose elimination would make X the weakest player in the three-person duel. If this is not possible, X should shoot at the type whose elimination would make X the average player in the three-person duel. So whom should each type shoot at in the four-person duel?

A can never become the average nor the weakest player in the three-person duel, because

$$\text{if } T_{\mathrm{A}} = \mathrm{B}: \quad P_{\mathrm{A},(\mathrm{C},\mathrm{D})} = [a^2]/[(a+c+d)(a+d)];$$

$$\text{if } T_{\mathrm{A}} = \mathrm{C}: \quad P_{\mathrm{A},(\mathrm{B},\mathrm{D})} = [a^2]/[(a+b+d)(a+d)];$$

$$\text{if } T_{\mathrm{A}} = \mathrm{D}: \quad P_{\mathrm{A},(\mathrm{B},\mathrm{C})} = [a^2]/[(a+b+c)(a+c)].$$

Because $P_{\mathrm{A},(\mathrm{C},\mathrm{D})}$ is the highest possible payoff, A will shoot at B.

B can never become the weakest player in the three-person duel, because

$$\text{if } T_{\mathrm{B}} = \mathrm{A}: \quad P_{\mathrm{B},(\mathrm{C},\mathrm{D})} = [b^2]/[(b+c+d)(b+d)];$$

$$\text{if } T_{\mathrm{B}} = \mathrm{C}: \quad P_{\mathrm{B},(\mathrm{A},\mathrm{D})} = [b(a+d)]/[(a+b+d)(a+d)];$$

$$\text{if } T_{\mathrm{B}} = \mathrm{D}: \quad P_{\mathrm{B},(\mathrm{A},\mathrm{C})} = [b(a+c)]/[(a+b+c)(a+c)].$$

B can become the strongest player (which is the worst option) in the three-person duel if he hits A; if he hits D or C, he becomes the average player in the three-person duel; but if he hits C, the weakest player in the three-person duel is weaker (D instead of C), which is better for the average type in the three-person duel (because $P_{\mathrm{B},(\mathrm{A},\mathrm{D})} > P_{\mathrm{B},(\mathrm{A},\mathrm{C})}$). Therefore, B will shoot at C.

C remains the weakest in the three-person duel (which is the best option) if he hits D:

$$\text{if } T_{\mathrm{C}} = \mathrm{A}: \quad P_{\mathrm{C},(\mathrm{B},\mathrm{D})} = [c(b+d)]/[(b+c+d)(b+d)];$$

$$\text{if } T_{\mathrm{C}} = \mathrm{B}: \quad P_{\mathrm{C},(\mathrm{A},\mathrm{D})} = [c(a+d)]/[(a+c+d)(a+d)];$$

$$\text{if } T_{\mathrm{C}} = \mathrm{D}: \quad P_{\mathrm{C},(\mathrm{A},\mathrm{B})} = [c(2a+c)]/[(a+b+c)(a+c)].$$

Therefore, C will shoot at D.

D would prefer to shoot at B than at C; and would prefer to shoot at A than at B (both would remain anyway the strongest players, so it is better to eliminate the stronger of the two):

$$\text{if } T_{\mathrm{D}} = \mathrm{A}: \quad P_{\mathrm{D},(\mathrm{B},\mathrm{C})} = [d(2b+d)]/[(b+c+d)(b+d)];$$

$$\text{if } T_{\mathrm{D}} = \mathrm{B}: \quad P_{\mathrm{D},(\mathrm{A},\mathrm{C})} = [d(2a+d)]/[(a+c+d)(a+d)];$$

$$\text{if } T_{\mathrm{D}} = \mathrm{C}: \quad P_{\mathrm{D},(\mathrm{A},\mathrm{B})} = [d(2a+d)]/[(a+b+d)(a+d)].$$

Therefore, D will shoot at A.

The rule that emerges is to *shoot at the type immediately weaker than oneself*, unless one is the weakest type, in which case he should shoot at the strongest type. Note that the rule that applies to the three-person duel is a degenerate version of this: A shoots at B, and C shoots at A; B, however, does not shoot at C (as he should according to the general rule) in a three-person duel because we are one step away from a two-person duel, in which he prefers to face C than A.

The probabilities of winning the four-person duel, therefore, are

$$\begin{aligned}[P_{\mathrm{A},(\mathrm{B},\mathrm{C},\mathrm{D})} &= (a/4)P_{\mathrm{A}(\mathrm{C},\mathrm{D})} + (b/4)P_{\mathrm{A}(\mathrm{B},\mathrm{D})} + (c/4)P_{\mathrm{A}(\mathrm{B},\mathrm{C})} \\ &\quad + (d/4)0 + (1 - a/4 - b/4 - c/4 - d/4)P_{\mathrm{A},(\mathrm{B},\mathrm{C},\mathrm{D})}\end{aligned}$$

$$\begin{aligned}P_{\mathrm{B},(\mathrm{A},\mathrm{C},\mathrm{D})} &= (a/4)0 + (b/4)P_{\mathrm{B}(\mathrm{A},\mathrm{D})} + (c/4)P_{\mathrm{B}(\mathrm{A},\mathrm{C})} \\ &\quad + (d/4)P_{\mathrm{B}(\mathrm{C},\mathrm{D})} \\ &\quad + (1 - a/4 - b/4 - c/4 - d/4)P_{\mathrm{B},(\mathrm{A},\mathrm{C},\mathrm{D})}\end{aligned}$$

$$\begin{aligned}P_{\mathrm{C},(\mathrm{A},\mathrm{B},\mathrm{D})} &= (a/4)P_{\mathrm{C},(\mathrm{A},\mathrm{D})} + (b/4)0 + (c/4)P_{\mathrm{C},(\mathrm{A},\mathrm{B})} \\ &\quad + (d/4)P_{\mathrm{C},(\mathrm{B},\mathrm{D})} \\ &\quad + (1 - a/4 - b/4 - c/4 - d/4)P_{\mathrm{C},(\mathrm{A},\mathrm{B},\mathrm{D})}\end{aligned}$$

$$\begin{aligned}P_{\mathrm{D},(\mathrm{A},\mathrm{B},\mathrm{C})} &= (a/4)P_{\mathrm{D},(\mathrm{A},\mathrm{C})} + (b/4)P_{\mathrm{D},(\mathrm{A},\mathrm{B})} + (c/4)0 \\ &\quad + (d/4)P_{\mathrm{D},(\mathrm{B},\mathrm{C})} \\ &\quad + (1 - a/4 - b/4 - c/4 - d/4)P_{\mathrm{D},(\mathrm{A},\mathrm{B},\mathrm{C})}\end{aligned}$$

More in general, in a duel between N individuals of type $X_i \in \mathbf{X} = \{X_1, X_2, \ldots, X_N\}$, with skills $x_i \in \{x_1, x_2, \ldots, x_N\}$, type X_i should target the individual of type $T_i \in \mathbf{T} = (\mathbf{X}: T_i \neq X_i)$ that, if eliminated, would give X_i the highest chance of winning in the eventual duel with the types left $Y_i \in (\mathbf{T}: Y_i \neq T_i)$. Therefore,

$$P_{X_i(\mathbf{T})} = \sum_{j=1}^{N} (x_j/N) \cdot p_j + \left[1 - \sum_{j=1}^{N} (x_j/N)\right] \cdot P_{X_i(\mathbf{T})}$$

where

$$p_j = \begin{cases} P_{X_i(\mathbf{T}: \mathbf{Y_i} \neq \mathbf{T_j})} & \text{if } T_j \neq X_i \\ 0 & \text{if } T_j = X_i \end{cases}$$

It is easy, if a bit tedious, to show that the same rule found for the four-person duel (shoot at the type immediately weaker than oneself, unless one is the weakest, in which case one should shoot at the strongest type) should be adopted in duels with more players.

A complete analysis of N-person duels for $N > 3$ is beyond the scope of this article. It is easy to verify, however, that the type with the lowest skill has again the highest probability of surviving and the type with the highest skill has the lowest probability, like in the three-person duel (unless differences in skills are too extreme). For example, in the four-person duel, if $a = 1$, $b = 0.9$, $c = 0.8$, $d = 0.7$, the result is that $P_{\mathrm{A},(\mathrm{B,C,D})} = 0.176$, $P_{\mathrm{B},(\mathrm{A,C,D})} = 0.213$, $P_{\mathrm{C},(\mathrm{A,B,D})} = 0.271$, and $P_{\mathrm{D},(\mathrm{A,B,C})} = 0.338$.

Discussion

THE TRUEL

An N-person duel is a series of sequential, repeated, pairwise interactions with opponents from a group of N players. This is different from both a two-person repeated game (pairwise interactions with the same opponent) and an N-person game (collective interactions) and can lead to surprising results. Simple strategic considerations on payoff maximization lead to what can be dubbed "survival of the weakest": the weakest player can have the highest fitness; the strongest player can have the lowest. This result arises from strategic considerations alone (Shubik 1954). The theory of three-person duels has been applied to the study of strategic voting in political economy, where it has implications for competition in multiple-party elections (Kilgour 1972, 1975, 1978; Kilgour and Brams 1997), but it can also be used in interactions in evolving populations. Although most of the analysis presented here, and all the literature to date, is limited to three-person duels, the logic of the truel (survival of the weakest) seems to apply to $N > 3$ as well, and future extensions to duels between more than three players seem possible. Most of the results reported here focus on three-person duels, as will the rest of this discussion.

TRUELS IN BIOLOGY

As in most games shared by economics and biology, the logic of rational choice (in economics) is replaced by the logic of natural selection (in biology): mutations that induce an individual to compete preferentially with the right opponent will have an advantage and increase in frequency; a rational, conscious ability to tell skills or ranks apart is not necessary. The main difference with simple static truels between three individuals (Shubik's truel) is that, in an evolving population, participants in the truel will be chosen with a probability proportional to their current frequency in the population—the approach adopted in this article—and therefore the outcome may change over time.

Although truels have been ignored so far in evolutionary biology, a certain amount of attention (e.g., Maynard-Smith 1983; Sinervo and Lively 1996; Alonzo and Sinervo 2001; Frean and Abraham 2001; Nahum et al. 2011) has been devoted to another game that might resemble the truel: the rock-paper-scissors (RPS) game. The RPS game, however, is a two-person game (with three strategies), whereas the truel is a three-person game; a cyclical dynamics arises in the RPS because of its peculiar ranking of payoffs (A defeats B, which defeats C, which defeats A), which is absent in the duels we have analyzed here (in which A defeats both B and C, and B defeats C) and is not necessary to produce cyclical dynamics in the truel. It is the fact that interactions are between more than two players, not the peculiar ranking of payoffs, which produces a diversity of equilibria and dynamics (including cyclical dynamics) in the truel. Although a few examples of RPS game have been described in nature (e.g., Sinervo and Lively 1996; Alonzo and Sinervo 2001) or created in the laboratory (Nahum et al. 2011), the simple ranking of payoffs assumed by the truel seems to have more general applicability. It is interesting that even in the RPS being the weakest competitor can be an advantage (Nahum et al. 2011). Note, however, that the "survival of the weakest" described by Frean and Abraham (2001) for the RPS game (in which it means that the weakest type does not disappear because of the cyclical dynamics) is different from Shubik's (1954) (in which it means that the weakest type can go to fixation).

The assumption of sequential, repeated interactions with multiple opponents is not very restrictive. Indeed, there is no reason to assume that antagonistic interactions are always between two individuals only and end after one shot. In some cases, individuals in a group have precise hierarchy dominance rankings and must decide whom to pick a fight with, as in a truel. Indeed, Shubik (1954), in his introduction to the truel, credits Konrad Lorenz for pointing out that this kind of interactions is common

in fights among animals. Examples of antagonistic interactions that can be modeled as simple Shubik's truels are physical contests between males for access to females, contests for dominance over territories and resources, fights for establishing dominance hierarchies.

Consider fights in red deer (*Cervus elaphus*). One of the first observations of fighting behavior (Clutton-Brock et al. 1979) was that weak stags attack strong stags more frequently than what would be expected by chance, and vice versa strong stags attack weak stags less frequently. This observation, however, may be biased by the fact that, because strong stags are more likely to be holders of a territory, weak stags have more to gain from attacking strong stags. More detailed analysis of fighting frequencies showed that fights usually occur between stags with similar fighting abilities; as one of the earliest accounts of social behavior in red deer (Darling 1937) noticed, "only stags of almost equal merit fight each other." More precisely, fights between stags that are more than two steps apart in the dominance hierarchy occur less frequently than one might expect by chance (Clutton-Brock et al. 1982). This result was confirmed by a more precise and recent analysis (Freeman et al. 1992) showing that stags pick up fights preferentially with individuals *immediately below* their own hierarchy level. This is a striking observation when compared to the logic of the truel.

Remember that, although in the three-person duel the logic is to shoot at the strongest opponent (A should shoot at B, and both B and C should shoot at A), this is a degenerate version of the more general rule that, in *N*-person duels, one should shoot at the type *immediately weaker than oneself* (unless one is the weakest, in which case he should shoot at the strongest type); in the three-person duel this rule degenerates into shooting at the strongest opponent because, as we have seen, it does no longer apply to the average player. The general rule, therefore, seems to match the observed preferences for picking up fights in *N*-person duels in red deer. It would be interesting to study actual three-person duels rather than *N*-person duels, to see whether the degenerate rule of shooting at the strongest opponent is observed; it would also be interesting to see whether the very weakest individuals in red deer actually prefers to pick up fights with the strongest (we lack data because weaker individuals usually fight less often).

MORE REALISTIC TRUELS: EFFECTS ON EQUILIBRIA AND DYNAMICS

Shubik's truel is a neat example of how strategic thinking can lead to counterintuitive results. In its basic form, however, it lacks the sophistication to describe interactions that go beyond actual shooting contests. Five possible extensions of the truel were discussed here: a mixture of two-person and *N*-person interactions (A Mixture of Duels and Truels), the possibility that the contest end without a winner (Limited Truels: Contests Can End without a Winner); a correlation between defensive and offensive skills (Defensive Truels: Defensive and Offensive Skills are Correlated); the possibility that players not directly involved in the contest suffer collateral damage (Interference Truels: Collateral Damage); most important: the possibility that the attacked individual backfires when attacked (Truels with Backfire: Immediate Retaliation). All these possibilities were studied in evolving populations, rather than in static games between three types.

The paradox arising from Shubik's truel (survival of the weakest) extends to interactions in evolving populations: because the weakest type has the highest fitness, selection cannot lead to a gradual increase of skills in antagonistic interactions. This means that design by natural selection in phenotypes related to competition can only be explained by mutations with very large effects (which seems contrary to the standard view of gradual evolutionary change) or if contests are always between two individuals (which seems unlikely). More realistic assumptions on the nature of the interactions, however, allow to solve the paradox and produce diverse equilibria and dynamics.

First, playing the truel in evolving populations can lead to different equilibria and even, for a narrow range of parameters, cyclical polymorphisms. If individuals engage in a mixture of duels and truels, the paradox generally disappears because the strongest type has an advantage in normal duels. At intermediate frequencies of truels, which type is stable depends on the differences in skill between the three players: the weakest player is more likely to be stable when the frequency of truels is high (Fig. 1).

More realistic assumptions on the nature of fights we have considered are the following: after each shot the interaction can end without a winner; defensive and offensive skills can be correlated so that strong players are also better defended against the other player's attack; an action against one individual can produce collateral damage for the third one (e.g., in indirect competition for a foraging territory). These extensions of Shubik's truel lead to different results: multiple stable equilibria are possible, and cyclical polymorphic equilibria become more likely; the period of the oscillations can be very long, in the order of thousands of generations, so that short-term population data might not show that the frequencies are changing over time (Fig. 2).

The most important modification of Shubik's truel, however, is the backfire model: here we no longer consider sneak attacks as in Shubik's truel (in which the target has no option to counterattack), but more general contests in which one player simply chooses whom he wants to fight with, and then the target opponent can immediately fight back (without waiting his turn to choose). Winning the fight, in this case, depends on the skill of the target too (whereas Shubik's truel assumes it is independent), and attacking a strong individual is more costly than attacking a weak individual. Taking into account this backfire effect, a new

unexpected result emerges: the three types can coexist in a *stable* polymorphism (Fig. 3). Note that this is not a trivial effect of negative frequency dependence, as in two-person duels under the same assumptions (backfire effect) there is no stable polymorphism and the strongest player always goes to fixation.

Note that we have assumed that individuals must be able to tell skills apart, and that there is no scope for cheating: in all models of truels analyzed so far skills are common knowledge and cannot be faked. This is a reasonable assumption for indices of quality such as body size or antler size; in other cases, however, one should take into account the possibility of mistakes in assessing ranks, or the possibility of cheating, and it would be interesting to analyze what happens in these cases.

IMPLICATIONS FOR THE MAINTENANCE OF VARIATION IN NATURAL POPULATIONS

Beside the precise characterization of the equilibria and of the dynamics, the general lesson arising from the analysis of three-person duels in evolving populations is that it is not simply the case that the strongest type survives and goes to fixation (survival of the fittest) as predicted by two-person games, or that the weakest type does (survival of the weakest) as in Shubik's truel, but that a variety of other possibilities exists instead. This may help understand the persistence of variation in natural populations.

What maintains variation in fitness-related traits is one of the major unresolved issues in evolutionary biology (Barton and Turelli 1989; Charlesworth and Hughes 1999; Barton and Keightley 2002). Two possible solutions have been proposed. The first is fluctuating selection: the optimal phenotype may vary in space or time, for instance, because parasites are continually evolving to overcome host defenses. The second is mutation–selection balance: recurrent mutations can generate new genetic variation as quickly as it is eroded by selection. As we have seen, if interactions are (even partially) between more than two players, there is no need to invoke recurrent mutations or fluctuating selection to explain the maintenance of variation. Constant selection can maintain existing variation simply because of the strategic nature of the interactions. This clearly does not question the importance of fluctuating selection or recurrent mutations for the maintenance of variation, but can help explain it under constant selection and low mutation rates.

CONCLUSION

The theory of *N*-person duels leads to counterintuitive results. By highlighting the strategic nature of competition, game theory sheds light on one of the most enduring problems in evolutionary theory: the persistence of variation under constant selection. By making more realistic assumptions on the nature of actual antagonistic interactions, evolutionary biology helps understand one of the oldest paradoxes of game theory.

ACKNOWLEDGMENTS

Thanks for comments to S. Stearns, D. Haig, I. Scheuring, D. Ebert, and G. Nöldeke.

LITERATURE CITED

Alonzo, S. H., and B. Sinervo. 2001. Mate choice games, context-dependent good genes, and genetic cycles in the side-blotched lizard, *Uta stansburiana*. Behav. Ecol. Sociobiol. 49:176–186.

Archetti, M. 2009. Survival of the steepest: hypersensitivity to mutations as an adaptation to soft selection. J. Evol. Biol. 22:740–750.

Barton, N. H., and P. D. Keightley. 2002. Understanding quantitative genetic variation. Nat. Rev. Genet. 3:11–21.

Barton, N. H., and M. Turelli. 1989. Evolutionary quantitative genetics: how little do we know? Annu. Rev. Genet. 23:337–370.

Charlesworth, B., and K. A. Hughes. 1999. The maintenance of genetic variation in life-history traits. Pp. 369–392 *in* R. S. Singh and C. B. Crimbas, eds. Evolutionary genetics: from molecules to morphology. Cambridge Univ. Press, Cambridge, UK.

Clutton-Brock, T. H., S. D. Albon, R. M. Gibson, and F. E. Guinness. 1979. The logical stag: adaptive aspects of fighting in red deer (*Cervus elaphus* L.). Anim. Behav. 27:211–225.

Clutton-Brock, T. H., F. E. Guinness, and S. D. Albon. 1982. Red deer. University of Chicago Press, Chicago, IL.

Darling F. F. 1937. A herd of red deer. Oxford Univ. Press, Oxford, UK.

Darwin, C. 1869. On the origin of species by means of natural selection. 5th ed. John Murray Ed. London, UK.

Frean, M., and E. R. Abraham. 2001. Rock-scissors-paper and the survival of the weakest. Proc. R. Soc. Lond. B 268:1323–1327.

Freeman L. C., S. C. Freeman, and K. Romney. 1992. The implications of social structure for dominance hierarchies in red deer, *Cervus elaphus* L. Anim. Behav. 44:239–245.

Kilgour, D. M. 1972. The simultaneous truel. Int. J. Game Theory 1:229–242.

———. 1975. The sequential truel. Int. J. Game Theory 4:151–174.

———. 1978. Equilibrium points of infinite sequential truels. Int. J. Game Theory 6:167–180.

Kilgour, D. M., and S. J. Brams. 1997. The truel Math. Mag. 70:315–326.

Kinnaird, C. 1946. Encyclopedia of puzzles and pastimes. Citodel Press, New York, NY.

Larsen, H. D. 1948. A dart game. Am. Math. Monthly Dec.:640–641.

Maynard-Smith, J. 1983. Evolution and the theory of games. Cambridge Univ. Press, Cambridge.

Nahum, J. R., B. N. Harding, and B. Kerr. 2011. Evolution of restraint in a structured rock-paper-scissors community. Proc. Natl. Acad. Sci. USA 108(Suppl 2):10831–10838.

Shubik, M. 1954. Does the fittest necessarily survive? Pp. 43–46 *in* M. Shubik, ed. Readings in game theory and political behavior. Doubleday, Garden City, NJ.

———. 1964. Game theory and the study of social behavior: an introductory exposition. *In* M. Shubik, ed. Game theory and related approaches to social behavior. John Wiley & Sons, New York, NY.

———. 1982. Game theory in the social sciences: concepts and solutions. MIT Press, Cambridge, MA.

Sinervo, B., and C. M. Lively. 1996. The rock-paper-scissors game and the evolution of alternative male strategies. Nature 380:240–243.

Spencer, H. 1864. Principles of biology. Williams & Norgate, London, UK.

Wilke, C. O., J. L. Wang, C. Ofria, R. E. Lenski, and C. Adami. 2001. Evolution of digital organisms at high mutation rates leads to survival of the flattest. Nature 412:331–333.

Associate Editor: A. Agrawal