# Evolution of warfare by resource raiding favors polymorphism in belligerence and bravery

Charles Mullon[1,*] and Laurent Lehmann[1]

[1]Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland
[*]Correspondence: `Charles.mullon@unil.ch`

**Abstract.** From protists to primates, intergroup aggression and warfare over resources has been observed in several taxa whose populations typically consist of groups connected by limited genetic mixing. Here, we model the co-evolution between four traits relevant to this setting: (i) investment into common-pool resource production within groups ("helping"); (ii) proclivity to raid other groups to appropriate their resources ("belligerence"); and investments into (iii) defense and (iv) offense of group contests ("defensive and offensive bravery"). We show that when traits co-evolve, the population often experiences disruptive selection favouring two morphs: "Hawks", who express high levels of both belligerence and offensive bravery; and "Doves", who express neither. This social polymorphism involves further among-traits associations when the fitness costs of helping and bravery interact. In particular if helping is antagonistic with both forms of bravery, co-evolution leads to the coexistence of individuals that either: (i) do not participate into common-pool resource production but only in its defense and appropriation ("Scrounger Hawks"); or (ii) only invest into common pool resource production ("Producer Doves"). Provided groups are not randomly mixed, these findings are robust to several modelling assumptions. This suggests that inter-group aggression is a potent mechanism in favoring within-group social diversity and behavioural syndromes.

**Keywords:** Warfare, polymorphism, trait associations, kin selection

## 1 Introduction

Warfare –the coalitionary aggression between groups of individuals – is one of the defining features of the human lineage. It is not only thought to have driven the advent of large-scale societies, but also that small scale hunter-gatherer societies regularly took part in coalitionary contests for material resources and reproductive opportunities [1–3]. Warfare is also well known to occur in chimpanzees [4], and has further been observed in several taxa outside of primates: in banded mongoose, where it is initiated by females who seek to mate with extragroup males [5]; in army ants, where colonies engage in contests for access to new territories [6]; and even in some strains of bacteria, who use a large assemblage of different offensive weapons to dislodge patches of rival cells [7]. As highlighted by these examples, warfare typically involves a resource over which conflict occurs between groups, and such conflict depends on individual contributions into offensive and defensive group-level activities [8]. This leads to an overall structure of interactions involving in-group solidarity and out-group hostility.

There is a large theoretical literature on the behavioral underpinnings of warfare that use different approaches to consider related questions about within-group solidarity and between-group hostility. In economics and game theory, models often focus on understanding the Nash equilibrium strategies for two types of behaviours under different scenarios of intergroup conflict: effort into material resource production and/or effort into fighting for appropriating these resources produced by others (e.g., [8–14]). This approach therefore assumes that individuals choose their behaviours freely with the goal of maximising their payoff, i.e., the rational actor model, while the material nature of the resources over which groups contest means that warfare can be regarded as a type of subsistence strategy in these formalizations. There is a parallel literature in evolutionary game theory, where instead of being freely chosen by individuals, strategies are genetically or culturally inherited (e.g., [15–18]). Both of these strands of the literature usually assume that groups are formed randomly, so that individuals mix freely between groups.

Social vertebrates and invertebrates, however, have in common that their evolution occurs in populations composed of groups of finite size with limited genetic mixing between groups [19, 20]. Such structure is relevant as it leads to interactions between genetically related individuals. Kin selection, which occurs whenever a trait expressed by an actor affects the fitness of others who are genetically related to the actor at the loci determining the trait [21–24], is therefore likely to affect the evolutionary dynamics of warfare. The literature that investigates the effects of limited genetic mixing on warfare evolution typically focuses on understanding how the demographic properties of groups, such as migration rate or sex-differences, influence the evolution of two types of behaviours: the propensity for between-groups contests and/or the fighting effort into contests [25–33]. Most of these models consider that the resources over which groups contest are reproductive in nature (e.g., mates or reproductive breeding spots). Warfare in this case should therefore be regarded as a reproductive rather than a subsistence strategy.

In this paper, we blend key elements from these different strands of the literature with the goal of contributing in two main ways. First, we decompose the individual behaviours that underpin in-group solidarity and out-group hostility by considering the genetic co-evolution between four individual traits in a group-structured population subject to limited genetic mixing: (i) the effort into a common-pool resource within groups; (ii) the proclivity to contest other groups to appropriate their common-pool resources; and efforts into (iii) defensive and (iv) offensive structures. Distinguishing between defense and offense allows to consider potential trade-offs between the two, something that has been been somewhat neglected by previous formalizations [34]. Hence, our model can be regarded as an extension of "Guns versus butter" models, which involve a trade-off between appropriating and producing resources ([14] for review), to a situation allowing for trade-offs between appropriating, defending, and producing resources. The second way we aim to contribute is by expanding current evolutionary analyses to understand the conditions that are conducive to the emergence of polymorphism in traits underlying warfare (instead of focusing on monomorphic evolutionary equilibria like all current models). In particular, we investigate whether there are situations where peaceful individuals, characterised by low levels of belligerence and bravery ("Doves"), can coexist with hostile ones, characterised by high levels of belligerence and bravery ("Hawks"). As it turns out, our analysis suggests that such a situation can readily arise.

# 2 The Model

## 2.1 Life-cycle and evolving traits

We consider a population of asexual haploid individuals that is subdivided among $N_\mathrm{g}$ groups (where $N_\mathrm{g} \gg 1$ is large), each containing $N$ adult individuals. All groups are subject to the same environmental conditions and are equally connected to one another by uniform random dispersal (i.e. Wright's island model [35]). We census this population at discrete demographic time points between which the following events occur in sequence: (1) Within groups, adult individuals produce a common-pool resource that yields a material payoff (e.g. calories). (2) Each group may raid another one to appropriate its common-pool resource, leading to contests (fights) among pairs of groups over resources. (3) Each adult individual produces many offspring in quantity that depends on outcome of the previous events, i.e. on the eventual amount of resources available for reproduction after raiding (we detail this in (2.2) below). (4) Independently from one another, each offspring then disperses with a probability $0 < m \leq 1$ to another randomly chosen group (or remains in its natal group otherwise). (5) Finally, a randomly chosen adult dies in each group and one offspring among those competing locally is randomly chosen to fill the open reproductive spot (i.e., a Moran reproductive process [36]). Generations are thus overlapping in our model (unlike Wright-Fisher models for e.g. [36]), with individuals living on average for $N-1$ demographic time periods. This assumption of a Moran life-cycle as well as of asexual reproduction improves mathematical tractability without loss of generality (Supplementary Material - SM for short - S5 for further discussion on this).

Against this demographic backdrop, we are interested in the co-evolution between four quantitative traits that influence interactions within- and between-groups: (1) the effort (or investment) $h \geq 0$ made by an individual into the production of a common-pool resource for its group (helping for short); (2) the proclivity or motivation $a \geq 0$ of an individual to raid another group, which we refer to as "belligerence"; (3) the effort $b \geq 0$ an individual makes into group contest when its group raids another one, which we refer to as "offensive bravery"; and finally (4) the effort $d \geq 0$ made by an individual into group contest when its group is attacked by others, which we refer to as "defensive bravery". We assume that each of these four traits ($a$, $b$, $d$ and $h$) is encoded by a separate genetic locus, at each of which we assume there is a continuum of possible alleles (i.e., the standard continuum of allele model of population genetics [37, 38]). Specifically, mutations occur during reproduction with a small probability $\mu$ at each locus, in which case the effect size of a mutation on the trait value is random, unbiased and weak; namely, a mutation causes a small trait deviation from the parental trait value and this deviation has mean 0 and small variance $\sigma^2$.

Under these assumptions, each individual expresses a potentially unique genetically-determined trait. Thankfully, we do not need to take into account the full breadth of this variation in order to evaluate the payoff, reproduction and survival of individuals that underlay our evolutionary analysis. To that end, it is sufficient to focus here on three levels of phenotypic specification. First, we consider a *focal individual* (i.e. a representative or randomly sampled individual in the population), whose phenotype we will denote by the vector $\boldsymbol{z}_\bullet = (a_\bullet, b_\bullet, d_\bullet, h_\bullet)$. Second, because we will assume that individuals within groups interact in a way that can be characterised by the average trait within groups (section 2.2 for details), we do not need to specify the traits of each individual in the focal group but rather summarise this by $\boldsymbol{z}_0 = (a_0, b_0, d_0, h_0)$, which collects the averages of each trait among all adults of the focal group (thus including the focal individual). Finally, since mutations are rare with small effects on the phenotype, variation among individuals in the rest of the population will typically be small. We can in fact ignore the variation among- and within- groups other than in the focal group

[39], and in effect consider that the rest of the population is monomorphic for the population average, which we denote by $\boldsymbol{z} = (a, b, d, h)$. In the next section, we specify under these assumptions how the evolving traits influence common-pool production within groups and raiding between groups (stages (1)-(2) stage of the life cycle in section 2.1), and in turn how this affects individual reproduction.

## 2.2 Helping, belligerence and bravery: costs and benefits

**Common-pool production.** We assume that the material payoff yielded by the common pool resource production in the focal group is given by a function $B(Nh_0)$, which increases in a decelerating manner with the total amount of investment, $Nh_0$, into helping within the group (i.e. exhibiting diminishing returns; $B(0) = 0$, $B'(x) > 0$, and $B''(x) < 0$ where throughout a prime ' denotes differentiation; and as with other relevant functions used in our model, we will later specify and explore different forms for $B(x)$, see section 3.2 and SM S5).

**Attacking.** We assume that the probability that the focal group raids another group is given by a function $0 \leq \alpha(a_0) \leq 1$, which increases with the average group belligerence ($\alpha(0) = 0$, $\alpha'(a_0) > 0$). Following the island model of warfare [29], we assume that when a group decides to attack another one, the group it raids is sampled randomly from the population. If two or more groups decide to raid the same group, one group is chosen at random from the attackers to perform the raid and engage into the contest for the appropriation of the common-pool resource of the raided group. As a result of these assumptions, the probability that a focal group engages into a fight over the resource of another group (i.e. decides to attack and is chosen among the attackers if there are more than one) is

$$\phi_0(a_0, a) = \alpha(a_0)\left(\frac{1 - \exp(-\alpha(a))}{\alpha(a)}\right), \tag{1}$$

where $a$ is the average level of belligerence in the rest of the population (Appendix S1 of [29] for derivation). Meanwhile, the probability that this same focal group is attacked by another group and engages locally in a contest over its own common-pool resources is

$$\phi_1(a) = 1 - \exp(-\alpha(a)). \tag{2}$$

**Winning a contest.** When a group raids another one, who wins the ensuing contest depends on how much the attacking group has invested into offense compared to how much the attacked group has invested into defense. More specifically, a raiding focal group with average level of offensive bravery $b_0$ is assumed to win the contest against a raided group with average level of defensive bravery $d$ with probability

$$v(b_0, d) = \frac{\omega g(Nb_0)}{\omega g(Nb_0) + (1 - \omega) g(Nd)}, \tag{3}$$

where $g$ is a positive, increasing function of its argument ($g(y) > 0$, $g'(y) > 0$). The parameter $0 < \omega < 1$ allows to tune the advantage of being offensive relative to being defensive (e.g., in the extreme case $\omega = 0$, attackers always lose whereas they always win when $\omega = 1$). Note that by symmetry of eq. (3), $v(b, d_0)$ gives the probability that a focal group with average level of defensive bravery $d_0$ loses a contest when it is attacked by another group with average level of offensive bravery $b$. In the context of conflict, eq. (3) is typically referred to as a "contest success function", in which the choice of the function $g$ allows to model different qualitative types of conflicts [8, 10, 11, 40].

**Resource distribution and fighting costs.** If the attacking group wins the contest, it appropriates all the collectively produced resources of the raided group (who thus loses all its collective resources). If the attacking group loses, then both the raiding and raided groups keep their own resources. Fighting, however, is costly (for instance due to lost opportunities). We assume that the payoff available to a group is reduced by a constant $-c_1 < 0$ when this group fought once (as an attacker or defender), and by $-c_2 < 0$ when this group fought twice (once as an attacker and once as a defender).

**Individual payoff benefits and costs.** After fighting is done, the resources that remain (if any) in a group are pooled and divided equally among its members. For example, consider the focal group with average level of helping $h_0$ that (i) raided (and won against) another group with average investment $h$ into helping, and (ii) was attacked by yet another group but also won this second fight. An individual from such a focal group will then obtain a payoff of $[B(Nh_0)+B(Nh)-c_2]/N$. Individuals also pay an individual cost due to the expression of their traits. For instance, an investment $\boldsymbol{h}_\bullet$ by a focal individual to the common-pool resource will incur a cost to that individual (that increases with the actual investment $\boldsymbol{h}_\bullet$). We additionally assume that both offensive ($\boldsymbol{b}_\bullet$) and defensive ($\boldsymbol{d}_\bullet$) bravery are costly to express, for instance due to individual resources being redirected towards offensive or defensive structures (e.g. bows, arrows or trenches). Belligerence ($\boldsymbol{a}_\bullet$), by contrast, is assumed to not be associated with a direct individual cost (there are however indirect costs due to fighting captured by $c_1$ and $c_2$, see above paragraph). To reflect these assumptions, we write $C(\boldsymbol{h}_\bullet, \boldsymbol{b}_\bullet, \boldsymbol{d}_\bullet)$ for the material costs that a focal individual pays when expressing trait values $\boldsymbol{h}_\bullet$, $\boldsymbol{b}_\bullet$, and $\boldsymbol{d}_\bullet$ (we assume these costs increase monotonically with each trait, at least linearly, i.e. $C(0,0,0) = 0$, $\partial C/\partial x > 0$ and $\partial^2 C/\partial x^2 \geq 0$ for $x \in \{\boldsymbol{h}_\bullet, \boldsymbol{b}_\bullet, \boldsymbol{d}_\bullet\}$). To continue with the above example, a focal individual with traits $\boldsymbol{h}_\bullet$, $\boldsymbol{b}_\bullet$ and $\boldsymbol{d}_\bullet$ that is a member of the focal group will then receive a payoff of $[B(Nh_0) + B(Nh) - c_2]/N - C(\boldsymbol{h}_\bullet, \boldsymbol{b}_\bullet, \boldsymbol{d}_\bullet)$.

Taking into account all possible outcomes, the expected material payoff to a focal individual can be written as

$$\begin{aligned}\pi(\boldsymbol{z}_\bullet, \boldsymbol{z}_0, \boldsymbol{z}) = v_{\mathrm{b}} &+ \left[\phi_0(a_0,a)\nu(b_0,d)\frac{B(Nh)}{N} + \big(1-\phi_1(a)[1-\nu(b,d_0)]\big)\frac{B(Nh_0)}{N}\right] \\ &- \left[\big(\phi_0(a_0,a)+\phi_1(a)\big)\frac{c_1}{N} + \phi_0(a_0,a)\phi_1(a)\frac{c_2-2c_1}{N}\right] - C(\boldsymbol{h}_\bullet, \boldsymbol{b}_\bullet, \boldsymbol{d}_\bullet),\end{aligned} \tag{4}$$

where the first term, $v_{\mathrm{b}} > 0$, is some baseline payoff; the term within square brackets on the first line is the average amount of resources (over offensive and defensive contests) an individual obtains; the term within square brackets on the second line is the decrease in the amount of such resources due to the costs of contests; and the final term is the individual costs of expressing helping and bravery (see SM S1 for a derivation). When fighting probabilities are equal to one ($\phi_0 = \phi_1 = 1$) and the cost of fighting one or two fights are the same ($c_2 = c_1 = c$), eq. (S1) reduces to

$$\pi(\boldsymbol{z}_\bullet, \boldsymbol{z}_0, \boldsymbol{z}) = v_{\mathrm{b}} + \nu(b_0,d)\frac{[B(Nh)-c]}{N} + (1-\nu(b,d_0))\frac{[B(Nh_0)-c]}{N} - C(\boldsymbol{h}_\bullet, \boldsymbol{b}_\bullet, \boldsymbol{d}_\bullet). \tag{5}$$

This has the same structure as the payoff function used in classical model of contests (e.g. eq. 7 of [8], eq. 1 of [14], second equation on p. 1017 of [18]), with the difference that here, a group experiences two contests: one in offense and another in defense.

**Evolutionary dynamics.** We assume that the fecundity of an individual increases with its payoff. From eq. (4), it is then only a matter of bookkeeping to calculate the individual fitness of a focal individual, which depends on the payoff of other individuals in the population (SM S1 for these calculations). This lays the basis of our method of analysis of evolutionary dynamics, which is detailed in our SM S2. One can then follow our mathematical analysis from SM S3 and via an accompanying Mathematica Notebook.

# 3 Results

## 3.1 Directional selection on helping and belligerence

From our assumptions, it is clear that belligerence and bravery can only evolve by selection if other groups have produced some amount of common resource that can be appropriated by raiding. It is therefore useful to first understand the evolution of the production of the common good within groups in the absence of belligerence and bravery ($a = b = d = 0$). We find that provided helping can increase when absent in the population, then there is a unique equilibrium $h^*$ for helping that satisfies the first-order condition

$$B'(Nh^*)\kappa^{\mathrm{R}} = \left.\frac{\partial C(h_\bullet, 0, 0)}{\partial h_\bullet}\right|_{h_\bullet = h^*}, \tag{6}$$

where

$$\kappa^{\mathrm{R}} = \frac{1}{N} + \frac{N-1}{N}\underbrace{\frac{1-m}{1-m+N}}_{=\kappa} \geq 0, \tag{7}$$

decreases with dispersal and group size (SM S3.1 for derivation). The equilibrium $h^*$ defined by eq. (6) is both *convergence stable* and *locally uninvadable* under our assumptions for the benefits and cost functions (SM S2.1 for a formal definition of these terms). The population mean of helping will therefore converge to $h^*$ and the phenotypic distribution will remain unimodally distributed around that mean (Suppl. Fig. 2 for e.g.). Eq. (6) says that the equilibrium investment into common pool resource production is such that the marginal cost of helping (right hand side of eq. 6) is offset by the marginal effect $B'(Nh)$ of helping by an individual on group productivity weighted by $\kappa^{\mathrm{R}}$. To understand this parameter $\kappa^{\mathrm{R}}$, let us note first that in eq. (7), $\kappa$ is the scaled relatedness between two individuals within groups taking kin competition into account [41, 42]. This $\kappa$ can be interpreted as the number of units of fecundity or payoff that a focal individual is willing to forgo to increase the fecundity or payoff of a randomly sampled neighbor by one unit (for details see [43]). The parameter $\kappa^{\mathrm{R}}$ can then be understood as the number of such sacrificed units when the one unit increase can be obtained by any individual in the group, including the focal individual (i.e. sampled with replacement from the group, hence the superscript R). This stems from the fact that public good production benefits all individuals equally within the group, including the focal actor. Since $\kappa$ and $\kappa^{\mathrm{R}}$ are both monotonically increasing functions of relatedness, they vary with demographic parameters in similar ways as relatedness (i.e. decrease with dispersal and group size, Suppl. Fig. 1). From this observation and eq. (6), we thus see that helping and common good production evolve to greater levels under directional selection when dispersal is weak and group size is small, which is a standard result across many incarnations of this problem in the literature ([41, 42] for reviews).

Let us now suppose that helping has evolved towards the equilibrium (eq. 6), but belligerence as well as bravery are absent ($a = b = d = 0$). By studying selection on belligerence in this situation (SM S3.1), we find that belligerence is favored by selection when the expected benefit from a raid exceeds its cost, i.e. when

$$\omega B(Nh^*) > c_1, \tag{8}$$

where $\omega$ is the probability of winning a fight when attacking when $b = d = 0$ in eq. (3). This shows that intergroup belligerence readily evolves, provided helping within groups leads to a sufficient amount of common goods that makes fighting for worth it. Once belligerence has evolved (so eq. 8 holds), this should set the stage for bravery to be favored by selection since it increases one's chances to win a contest, either in offense or defense. We investigate this in the next section.

## 3.2 The co-evolutionary equilibrium between belligerence, bravery and helping

In order to analyse the co-evolution of all traits, we assume the following relationships: (1) $B(Nh) = \beta\sqrt{Nh}$ for the benefits of the common good (where $\beta > 0$ is a constant); (2) $\alpha(a) = a$ for the probability of attacking in a group with belligerence $a$ (so that when $N = 1$, the trait $0 \leq a \leq 1$ is simply the probability of attacking); (3) $g(b) = b$ for the effect of bravery $b$ on the winning probability; and finally, (4) $C(h, b, d) = h + c_\mathrm{b} b + c_\mathrm{d} d$ for the individual cost of expressing helping and bravery, so that helping has a baseline cost of 1, and relative to this, offensive and defensive bravery have costs $c_\mathrm{b}$ and $c_\mathrm{d}$, respectively. These relationships capture the main properties of our model, while being simple enough to allow us to characterise analytically evolutionary equilibria (SM S5 for relaxation of these assumptions).

Under these assumptions, we show in SM S3.2 that there is a unique convergence stable strategy $\boldsymbol{z}^* = (a^*, b^*, d^*, h^*)$ that can be expressed as

$$a^* = \log\left(\frac{c_2 - 2c_1}{c_2 - c_1 - v^* B(Nh^*)}\right) \tag{9a}$$

$$b^* = \left[1 - e^{-a^*}\right] v^*(1 - v^*)\frac{B(Nh^*)}{Nc_\mathrm{b}}\kappa^\mathrm{R} \tag{9b}$$

$$d^* = \left[1 - e^{-a^*}\right] v^*(1 - v^*)\frac{B(Nh^*)}{Nc_\mathrm{d}}\kappa^\mathrm{R} \tag{9c}$$

$$h^* = \left[1 - \left(1 - e^{-a^*}\right) v^*\right]^2 \left(\frac{1}{2}\frac{\beta}{\sqrt{N}}\kappa^\mathrm{R}\right)^2, \tag{9d}$$

where we used

$$v^* = v(b^*, d^*) = \frac{c_\mathrm{d}\omega}{c_\mathrm{d}\omega + c_\mathrm{b}(1 - \omega)} \tag{10}$$

to denote the probability of winning a contest when attacking, or equivalently, of losing a contest when defending one's own resources at the equilibrium (found by substituting the equilibria for bravery, eqs. 9b-9c, into the contest function eq. 3 with $g(y) = y$).

**Belligerence.** To interpret eq. (9a), note first that $v^* B(Nh^*)$ corresponds to the equilibrium expected payoff that a group receives if it raids another one (since it wins the contest with probability $v^*$ and then obtains payoff $B(Nh^*)$). Thus, the equilibrium for belligerence $a^*$ increases with the expected group payoff of a raid and decreases with the cost of two contests relative to one contest (i.e. with $c_2/c_1$). Eq. (9a) also reveals that for $a^*$ to be an interior equilibrium (i.e. $0 < a^* < 1$ since here $a^*$ is a probability), the cost of two contests must be greater than twice the cost of a single contest ($c_2 > 2c_1$). If the cost of two contests is lower than this ($c_2 \leq 2c_1$), then the cost of contests for groups that both raid and are raided is relatively low. This favours the evolution of "total" belligerence where every group attempts to raid (i.e. $a \to 1$).

**Bravery.** To understand eqs. (9b)-(9c), we can decompose these as the product of four quantities with which both forms of bravery therefore increase: (1) the probability $[1 - e^{-a^*}]$ of raiding or being raided in a population where the average belligerence is $a^*$ (eqs. 1 and 2 with $a_0 = a = a^*$); (2) the variance or uncertainty, $v^*(1-v^*)$, in the outcome of a contest for a group either in offense or defense. Selection on both forms of bravery increases with this uncertainty because when one is certain to win or lose irrespective of the investment into offense or defense (e.g. when $\omega = 0$ or 1), then there is no incentive for such investment; (3) the ratio of the individual benefit in the case of the group winning a contest relative to the individual cost of the relevant bravery trait, which is $[B(Nh^*)/N]/c_\mathrm{b}$ when raiding and $[B(Nh^*)/N]/c_\mathrm{d}$ when raided; and finally (4) scaled relatedness with replacement $\kappa^\mathrm{R}$ (eq. 7). We further see that when the costs of investments into offense and defense are equal ($c_\mathrm{b} = c_\mathrm{d}$), individuals evolve to invest the same amount of resources into offense and defense (i.e. $b^* = d^*$) and that this amount is greatest when $\omega = 1/2$ (so that $v^* = 1/2$ and there is maximum uncertainty $v^*(1-v^*) = 1/4$ over outcome). The equilibrium given by eqs. (9b)-(9c) is consistent with previous models of bravery evolution for randomly mixed groups (i.e. when $m = 1$ so $\kappa^\mathrm{R} = 1/N$, e.g. eq. 10 of [8], first equation p. 1018 of [18]), but inconsistent with those in [17, 18] concerning evolution under limited genetic mixing. The equations presented in [17, 18], however, fail a number of sanity checks (SM S3.2.2 for details).

**Helping.** Finally, eq. (9d) can be understood by first recognising that the quantity $\psi(a^*) = [1-(1-e^{-a^*})v^*]^2$, is the probability that a group is left with just its own common good in a population at equilibrium. This can be seen by decomposing $\psi(a^*)$ as the product between the probabilities of two events: (1) that the group does not attack and win the ensuing contest (with probability $[1-(1-e^{-a^*})v^*]$); and (2) that the group does not get attacked and lose the ensuing contest (also with probability $[1-(1-e^{-a^*})v^*]$). In this light, eq. (9d) is intuitive. Selection for helping and participation to the common good increases with the certainty that this common good is the only source of payoff to oneself and to relatives (from $\kappa^\mathrm{R}$ in eq. 9d). Accordingly, belligerence between groups, $a^*$, reduces helping at equilibrium (as in the absence of belligerence, helping stabilises to $\left(\frac{1}{2}\frac{\beta}{\sqrt{N}}\kappa^\mathrm{R}\right)^2$, see also eq. 6).

**Payoff at equilibrium.** From eq. (9), it is straightforward to obtain an explicit solution for each trait (i.e. a solution that depends only on model parameters, eq. S21 in SM S3.2.3). As expected, one characteristic feature of these solutions is that all traits at equilibrium increase with relatedness and therefore decrease with dispersal between groups (Fig. 1 A-C). What is perhaps less intuitive is that the coevolution of all four traits leads to a non-monotonic relationship between payoff at the evolutionary equilibrium and relatedness (or dispersal, Fig. 1D-F). Specifically, depending on the individual cost of offensive bravery, $c_\mathrm{b}$, the payoff in an equilibrium population can: (1) decrease with relatedness (for small $c_\mathrm{b}$, Fig. 1D); (2) first increase and then decrease in a quadratic fashion with relatedness (for intermediate $c_\mathrm{b}$, Fig. 1E); or (3) increase with relatedness (for high $c_\mathrm{b}$, Fig. 1F). This reflects the fact that payoff increases with helping within groups and decreases with belligerence between groups. As dispersal goes down and relatedness increases, selection favours more helping (which increases payoff) but simultaneously also more fighting (which decreases payoff). These antagonistic effects on payoff can lead to a situation where payoff does not monotonically increase with relatedness (in contrast to most models of social evolution). In particular, when $c_\mathrm{b}$ is small, belligerence tends to increase compared to helping (Fig. 1 A) causing overall a decrease in payoff (Fig. 1 D). The coevolution of belligerence and helping can therefore lead to a somewhat counter-intuitive scenario where greater relatedness and greater prosociality within groups are associated with lower payoff at equilibrium.

### 3.3 The emergence and polymorphic coexistence of belligerent and pacifist individuals

**Additive individual costs and the emergence of Hawks and Doves.** Our results so far indicate that the mean phenotypes will gradually converge towards an interior evolutionary equilibrium point, provided the cost of two contests are greater than twice the cost of a single contest ($c_2 > 2c_1$, eq. (9)). However, we find that once such convergence has occurred, selection becomes disruptive and favors an increase in phenotypic variance, and in particular in the covariance between belligerence and offensive bravery (SM S3.2.5 for analysis). Specifically, selection favours belligerence and offensive bravery to become positively associated within individuals because genotypes that code for either more belligerence and greater investment into offense (i.e. with $a > a^*$ and $b > b^*$), or less belligerence and fewer resources in offense (i.e. with $a < a^*$ and $b < b^*$), have greater fitness than average when the population is at the evolutionary equilibrium. These two types gain fitness by employing two opposite strategies: to attack and win contests by investing more resources into offense, or not to attack and bypass the need to invest costly resources into offense. We respectively refer to these two types as "Hawks" and "Doves" owing to their overall phenotypic similarity to the strategies of the classical Hawk-Dove game [44, 45] as well as the nature of selection that acts upon them (see below).

We checked our mathematical analyses and investigated whether Hawks and Doves can coexist in the long run by running individual based simulations (SM S4 for procedure). Starting with a monomorphic population where each trait is absent (i.e. the value of each trait is zero), the population average of each trait rapidly converges to its predicted equilibrium (Fig. 2A). Concomitantly, the variance in belligerence also increases, with individuals gradually expressing either no or complete belligerence ($a = 0$ or 1, Fig. 2B-C). Offensive bravery $b$ meanwhile also becomes polymorphic, with individuals eventually investing either some resources into offense or none at all (Fig. 2D). In contrast, defensive bravery $d$ and helping $h$ remain unimodally distributed so that these traits do not become significantly differentiated in the population (Fig. 2E-F). The joint equilibrium distribution of belligerence and offensive bravery (Fig. 3A) confirms our analytical predictions that these two traits become positively associated, and further reveals that highly-differentiated Hawks and Doves coexist in the long run (respectively the top right and bottom left clusters in Fig. 3A). By contrast, there is no clearly discernible association among any other pair of traits, so that Hawks and Doves both express helping and defense bravery in about the same amount (Fig. 3B-F). These weak associations are confirmed by the weak equilibrium covariances among traits other than between belligerence and offensive bravery (Fig. 3G).

These analyses show that the evolution in our model leads to the gradual emergence and maintenance of highly differentiated-types owing to negative frequency-dependent interactions. The pattern of frequency-dependence can be understood by focusing on Hawks. These are particularly successful when they are rare as they tend to engage their group into a raid and to win contests against groups consisting mainly of Doves. But as Hawks become common, their groups suffer from attacks by other groups consisting mainly of Hawks, and as a result pay the high cost of two contests rather than one (i.e. pay $c_2$ rather than $c_1$, where $c_2 > 2c_1$). Passed a critical frequency, Hawks thus become less successful than Doves, allowing both types to be maintained in the population. This frequency-dependence is thus reminiscent to that of the Hawk-Dove game [44, 45], with interactions here mediated by group-structure rather than occurring strictly among individuals.

**Polymorphism within and between groups.** The group structure of the population raises the question whether coexistence occurs within or between groups, in other words whether groups tend to be composed of one type (only Hawks or only Doves) or a mix of both. The distribution of belligerence in a single focal group over time suggests the latter (Fig. 4A), with the group very often consisting of both types. Nevertheless, when we compare the average level of belligerence within a focal group to the population average (dots vs full lines in Fig. 4B), we see that the focal group experiences significant variation over time. In other words, even though both Hawks and Doves co-occur within a group, one morph will typically dominate at any given time. Groups will therefore tend to be differentiated according to whether they are composed of more or less of one type. We can quantify this at the level of the population by calculating the phenotypic differentiation among groups at the belligerence locus (Fig. 4C, red). We find that this differentiation is no different before and after the polymorphism emerged or to differentiation at the helping locus (which recall never becomes clearly polymorphic, Fig. 4C, blue). In fact, differentiation among groups at the belligerence locus is the same as expected for a neutral locus (Fig. 4C, black). This shows that negative frequency-dependent selection does not generate significant departures in morph distribution among groups compared to neutral expectation. Put differently, how the two different morphs are distributed among groups is primarily determined by dispersal and local genetic fluctuations (due to variance in reproductive success that generates identity-by-descent) and thus reflect the pattern of relatedness.

We additionally examined the effects of different levels of dispersal on polymorphism (i.e. different values of $m$). Although simulations generally show significant phenotypic variation at the belligerence and offensive bravery loci (Suppl. Fig. 3A-C), differentiation among the Hawk and Dove morph is clearest and most stable where relatedness within groups is high (Suppl. Fig. 3D). This reflects that much of the frequency-dependent selection in our model is due to interactions between groups. As a result, selection is most able to discriminate among morphs and therefore favour their differentiation when morphs are unequally distributed among groups, which as we saw in the preceding paragraph happens when relatedness is non-zero. In fact, polymorphism among Hawks and Doves is weak if non-existent under complete random group mixing (and relatedness is zero, Suppl. Fig. 3A & D).

**Non-additive costs and the emergence of Scrounging Hawks and Producing Doves.** One assumption we have made so far is that the individual costs of investing personal resources into bravery (offensive and defensive) and helping are additive (within and between traits, i.e. that $C(h_\bullet, b_\bullet, d_\bullet) = h_\bullet + c_\mathrm{b} b_\bullet + c_\mathrm{d} d_\bullet$). Complementarity or antagonistic effects between traits (i.e. when different traits respectively have positive or negative non-additive effects) on individual costs can significantly influence how selection shapes associations between social traits and therefore on the nature of adaptive polymorphism when fitness depends on multiple traits [e.g. 39, 46–48]. In the context of inter-group contests and intra-group helping, one relevant scenario to investigate is where bravery traits complement one another, for instance because weapons or behaviours that are useful in offense are also useful in defense, but antagonistic with helping, for e.g. because characteristics that are beneficial in situations of conflict are counter-productive in prosocial situations. One way to capture this scenario is to implement non-additive individual costs in eq. (4) of the form

$$C(h_\bullet, b_\bullet, d_\bullet) = c_\mathrm{h} h_\bullet^2 + c_\mathrm{b} b_\bullet^2 + c_\mathrm{d} d_\bullet^2 - \gamma_\mathrm{bd} b_\bullet d_\bullet + \gamma_\mathrm{bh} b_\bullet h_\bullet + \gamma_\mathrm{dh} d_\bullet h_\bullet, \tag{11}$$

where $\gamma_\mathrm{bd} > 0$ controls the complementarity among offensive and defensive bravery (so that individual costs are lower when one unit of resource is invested in both types of bravery compared to two units invested in a single type), while $\gamma_\mathrm{bh} > 0$ and $\gamma_\mathrm{dh} > 0$ tune the antagonism between each respective bravery trait and helping.

Numerical exploration of the mathematical model using eq. (11) suggests that polymorphism still emerges in this case and that it is still driven by the coevolution between belligerence and offensive bravery (Mathematica Notebook). However, phenotypic associations now involve more traits. In particular, selection now favours a negative association between helping with offensive and defensive bravery. Individual-based stochastic simulations confirm this and further highlight the maintenance of a negative association among belligerence and helping (Fig. 5). Specifically, the Hawk morph is now also characterised by little within-group helping (morph on the top left of Fig. 5A), and the Dove morph by high levels of helping (morph on the bottom right of Fig. 5A). In fact, all other traits are now negatively associated with helping (last column of Fig. 5B). This is due to the extra individual costs suffered by individuals that combine helping with any form of bravery. Since Hawks are characterised by high levels of bravery, they evolve lower helping because of these extra costs. Doves, meanwhile, can continue to invest resources in helping since they invest little or no resources into bravery.

These results show how complementarity and antagonistic effects among traits can lead the polymorphism in our model to become more complex and involve further traits (such as helping). The main characteristics of the Hawk and Dove morphs, however, have not changed due to such non-linear effects, with some individuals with a strong preference for raiding and investing resources into offense, and others favouring not to raid and investing no resources into offense. We test and discuss the robustness of this polymorphism further in our SM S5, where in particular we investigate the effects of various contest functions (via $g(y)$), group decision making (via $\alpha(a)$) and benefit functions (via $B(y)$). We find that in all examined cases, Hawks and Doves are still expected to emerge as in our baseline model.

## 4 Discussion

Our results indicate that selection can readily lead to the emergence of a social polymorphism where two highly-differentiated morphs relevant to warfare eventually co-exist: one that codes for belligerence and offensive bravery (Hawks) and the other for neither (Doves, Fig. 3). The frequency-dependent interactions that maintain this polymorphism are close to those characterising the classical Hawk-Dove game [44, 45], where Hawks are favored when rare but disfavored when common as they engage with other Hawks and suffer an extra (non-additive) cost due to fighting. This extra cost is typically captured in the classical Hawk-Dove game by the condition that the cost $C$ for a Hawk to lose against another Hawk is greater than the value $V$ of the resource obtained in case of a win (i.e. $C > V$; whereas in the additive case, $C = V$ which would disfavor Doves, always). Similarly, one prerequisite for frequency-dependent interactions to lead to co-existence in our model is that the cost of contests is greater than additive (with the cost of two contests $c_2$ greater than twice the cost of a single contest $c_1$, $c_2 > 2c_1$, eq. S21 in SM S3). These extra costs prevent groups of Hawks to dominate at all frequencies and allow Doves to thrive when rare.

There are nonetheless also significant differences between the classical Hawk-Dove game for inter-individual conflicts and our inter-group conflict interaction scenario. First, the two morphs that are held in stable polymorphism in our model are not preexisting, but rather emerge from disruptive selection and gradual evolution. In the same way that polymorphism does not emerge when belligerence evolves alone in our model (SM S3.2.5), the mixed strategy characterized by the probability of taking the action "Hawk", usually understood as "Attack", also converges towards an equilibrium point and does not experience disruptive selection in the classical Hawk-Dove game [49]. This brings us to a second difference: the strategies that the two coexisting types employ at equilibrium in our model are more complex than simply playing "Attack" or not. Rather those strategies consist of composite behaviours including the propensity to raid and investment into offensive abilities. Interestingly, a similar polymorphism has been found to emerge in well-mixed populations with inter-individual conflicts (as in the Hawk-Dove game), and where the probability of playing Hawk coevolves with a physiological trait that is costly to express but that increases the probability of a win against another Hawk [50]. Like in our model, it is the coevolution of these two traits that leads to polymorphism emergence. One further conceptual difference regarding strategies is that Hawks always beat Doves in the classical game, whereas here, individuals of low belligerence nevertheless also invest significant resources into defense. Accordingly, groups consisting mainly of Doves in our model have a non-zero probability of winning against raiding groups consisting of mostly Hawks.

But perhaps the most significant way that we depart from the classical Hawk-Dove (as well as [50]), is that frequency-dependent interactions are mediated through group-structure in our model. Rather than lone individuals, it is groups with a majority of Hawks that tend to partake in raids and groups with a majority of Doves that tend not to. So even though most groups are composed of a mix of both morphs due to dispersal (Fig. 4A), the maintenance of polymorphism through negative frequency-dependent interactions relies on variation in this mix among groups (Fig. 4B-C for e.g.). Accordingly, when groups are formed via complete random mixing (i.e. complete dispersal, $m = 1$ and relatedness is zero) or group size is infinitely large ($N \to \infty$), selection is unable to discriminate among morphs and cause their differentiation since groups show very little variation in their composition (Suppl. Fig. 3A). By contrast, where relatedness is positive, highly-divergent morphs can be observed (Suppl. Fig. 3B-D). In fact, limited dispersal tends to stabilise the polymorphism (Suppl. Fig. 3D), with selection remaining disruptive even where dispersal is severely limited (i.e. very close to $m = 0$, Mathematica Notebook). This differs also from previous models where frequency-dependent interactions happen only among individuals of the same group, in which limited dispersal and thus relatedness inhibits disruptive selection [as it reduces the amount of local genetic variation and thus differentiation within groups, 39, 51–53].

In spite of these group-effects, one should keep in mind that selection occurs at the level of the gene (or replicator) and that these are expressed by individuals. As a result, where the individual fitness costs of the different traits interact with one another, evolutionary dynamics can lead to more traits becoming associated to the social polymorphism. In particular, if helping is antagonistic with offensive and defensive bravery while both forms of bravery are complementary (eq. 11), then the two co-existing morphs we observe consist of individuals that either do not participate in common-pool resource production but only in its defense and appropriation ("Scrounger Hawks"); or only invest into common pool resource production ("Producer Doves", Fig. 5). The negative frequency-dependent interactions that maintain these two morphs are then reminiscent to those characterising the classical Scrounger-producer game [54]. Beyond this scenario, our results suggest that through fitness interactions with bravery traits, other relevant social traits (such as the tendency to lead or follow) may become associated to the social polymorphism we have described.

The apparent ease with which polymorphism emerges in our model raises the question why it has not been reported in previous papers that have investigated the (genetic) co-evolution of traits involved in warfare [26, 28–31, 33]. By comparing these models to ours, we find that this is likely due to divergent formulations for fitness (compare eqs. S5 and S6 in SM). This divergence comes from our perspective of warfare as a subsistence strategy, whereas previous papers allowing for the coevolution of belligerence and bravery considered it as a reproductive strategy (but see [30]). These models typically assume that (i) groups loosing contests are repopulated (partially or completely, or their females mated) by winning groups, and (ii) that belligerence has fixed or unconditional costs to the individual that expresses it (in contrast to our model where belligerence has conditional costs, which are incurred only if raids take place). The direct extra costs associated with greater belligerence under such assumptions make it more difficult for Hawks to differentiate from the population in these models compared to ours and therefore for polymorphism to emerge (SM S1.2.2 and S3.2.5 for a more formal explanation). But nothing in principle precludes from considering conditional costs in models of warfare as a reproductive strategy and thus allow for polymorphism. Meanwhile, models that consider warfare as a subsistence strategy typically ignore belligerence and focus on bravery evolution (e.g., [15–18]). But as we have established, this coevolution between belligerence and bravery is necessary to the emergence of polymorphism.

To sum up, we have proposed a model of warfare evolution in which an initially asocial and undifferentiated population evolves towards within-group solidarity and between-group hostility enacted by highly-differentiated individuals. This as long as group size is not too large and dispersal is limited. Indeed, with complete dispersal, no differentiation occurs in our model, and with large group size, no social trait evolves to begin with as the selection pressure on each trait is of the order of the inverse of group size ($1/N$). This effect of group size on the strength of selection is, provided a few exceptions, common to all models of evolution of prosocial traits affecting group members indiscriminately [55], and thus applies to those aforementioned on warfare evolution as recently illustrated in simulations [33]. These models and ours are therefore most relevant to small-scale societies (or populations with small local *effective* size and *effective* migration rate, e.g., eq. 9.59 in [24]). For such societies, our model suggests that between-group aggression can be a potent mechanism for the maintenance of within-group trait diversity and behavioral syndromes, in particular favouring a positive association between belligerence and bravery.

## Acknowledgements

We thank Carsten De Dreu and Zegni Triki for their invitation to contribute to this special issue and for their comments on a previous version of this paper, as well as three anonymous reviewers.

## References

[1] Gat A. 2006 *War in Human Civilization.* Oxford: Oxford University Press.

[2] Gat A. 2017 *The Causes of War and the Spread of Peace: But Will War Rebound?* Oxford: Oxford University Press.

[3] Allen MW, Jones TH, editors. 2014 *Violence and Warfare among Hunter-Gatherers.* Walnut Creek, CA: Left Coast Press.

[4] Crofoot M, Wrangham RW. 2010 Intergroup aggression in primates and humans: the case for a unified theory. In Kappeler PM, Silk JB, editors, *Mind the Gap: Tracing the Origins of Human Universals* pp. 171–195. Berlin-Heidelberg: Springer-Verlag.

[5] Johnstone RA, Cant M, Cram D, Thompson FJ. 2020 Exploitative leaders incite intergroup warfare in a social mammal. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 29759–29766.

[6] Wilson EO, Hölldobler B. 1990 *The Ants.* Harvard: The Belkap Press of Harvard University Press.

[7] Granato ET, Meiller-Legrand TA, Foster KR. 2019 The Evolution and Ecology of Bacterial Warfare. *Current Biology* **29**, R521–R537.

[8] Garfinkel MR, Skaperdas S. 2007 Economics of conflict: an overview. In T. S, Hartley K, editors, *Handbook of Defense Economics, Volume 2* pp. 649–709. Amsterdam: North-Holland.

[9] Hirshleifer J. 1987 Conflict and rent-seeking success functions: ratio vs. difference models of relative success. *Public Choice* **63**, 101–112.

[10] Hirshleifer J. 1991 The technology of conflict as an economic activity. *The American Economic Review* **81**, 130–134.

[11] Hirshleifer J. 2001 *The Dark Side of the Force: Economic Foundations of Conflict Theory.* Cambridge: Cambridge University Press.

[12] Hausken K. 2004 Mutual raiding of production and the emergence of exchange. *Economic Inquiry* **42**, 572–586.

[13] Hausken K. 2008 Exchange, raiding, and the shadow of the future. *Defence and Peace Economics* **19**, 89–106.

[14] Kimbrough EO, Laughren K, Sheremeta R. 2020 War and conflict in economics: Theories, applications,and recent trends. *Journal of Economic Behavior and Organization* **178**, 998–1013.

[15] Hausken K. 1995 The dynamics of within-group and between-group interaction. *Journal of Matematical Economics* **24**, 655–657.

[16] Konrad K, Morath F. 2012 Evolutionarily stable in-group favoritism and out-group spite in intergroup conflict. *Journal of Theoretical Biology* **306**, 61–67.

[17] Gavrilets S, Fortunato L. 2014 A solution to the collective action problem in between-group conflict with within-group inequality. *Nature Communications.*

[18] Rusch H, Gavrilets S. 2020 The logic of animal intergroup conflict: A review. *Journal of Economic Behavior and Organization* **178**, 1014–1030.

[19] Cavalli-Sforza LL, Bodmer WF. 1999 *The genetics of human populations.* New York: Dover Publications.

[20] Hartl D, Clark AG. 2007 *Principles of Population Genetics.* Massachusetts: Sinauer 4th edition.

[21] Hamilton WD. 1964 The genetical evolution of social behaviour, 1. *Journal of Theoretical Biology* **7**, 1–16.

[22] Michod RE. 1982 The theory of kin selection. *Annual Review of Ecology and Systematics* **13**, 23–55.

[23] Frank SA. 1998 *Foundations of Social Evolution.* Princeton, NJ: Princeton University Press.

[24] Rousset F. 2004 *Genetic Structure and Selection in Subdivided Populations.* Princeton, NJ: Princeton University Press.

[25] Bowles S. 2006 Group competition, reproductive leveling, and the evolution of human altruism.. *Science* **314**, 1569–1572.

[26] Choi JK, Bowles S. 2007 The coevolution of parochial altruism and war. *Science* **318**, 636–640.

[27] Smirnov O, Arrow H, Kennett D, Orbell J. 2007 Ancestral war and the evolutionary origins of "heroism". *The Journal of Politics* **69**, 927–940.

[28] Lehmann L, Feldman MW. 2008 War and the evolution of belligerence and bravery. *Proceedings of the Royal Society of London Series B-Biological Sciences* **275**, 2877–2885.

[29] Lehmann L. 2011 The demographic benefits of belligerence and bravery in the island model of warfare: defeated group repopulation or victorious group size expansion?. *Plos One* **6**, 1–13.

[30] Rusch H. 2014 The two sides of warfare: an extended model of altruistic behavior in ancestral human intergroup conflict. *Human Nature* **25**, 359–377.

[31] Micheletti AJC, Ruxton GD, Gardner A. 2017 Intrafamily and intragenomic conflicts in human warfare. *Proceedings of the Royal Society B-Biological Sciences* **284**.

[32] Michelleti A, Ruxton G, Gardner A. 2020 The demography of human warfare can drive sex differences in altruism. *Evolutionary Human Sciences* **2**.

[33] Dyble M. 2021 The evolution of altruism through war is highly sensitive to population structure and to civilian and fighter mortality. *Proceedings of the National Academy of Sciences of the United States of America* p. e2011142118.

[34] De Dreu CKW, Gross J. 2019 Revisiting the form and function of conflict: Neurobiological, psychological, and cultural mechanisms for attack and defense within and between groups. *Behavioral and Brain Sciences* pp. 1–66.

[35] Wright S. 1931 Evolution in Mendelian populations. *Genetics* **16**, 97–159.

[36] Ewens WJ. 2004 *Mathematical Population Genetics*. New York: Springer-Verlag.

[37] Kimura M. 1965 A stochastic model concerning the maintenance of genetic variability in quantitative character. *Proceedings of the National Academy of Sciences of the United States of America* **54**, 731–736.

[38] Bürger R. 2000 *The Mathematical Theory of Selection, Recombination, and Mutation*. New York: John Wiley and Sons.

[39] Mullon C, Lehmann L. 2019 An evolutionary quantitative genetics model for phenotypic (co)variances under limited dispersal, with an application to socially synergistic traits. *Evolution* **73**, 1695–1728.

[40] Skaperdas S. 1996 Contest success functions. *Economic Theory* **7**, 283–290.

[41] Lehmann L, Rousset F. 2010 How life-history and demography promote or inhibit the evolution of helping behaviors. *Philosophical Transactions of the Royal Society B* **365**, 2599–2617.

[42] Van Cleve J. 2015 Social evolution and genetic interactions in the short and long term. *Theoretical Population Biology* **103**, 2–26.

[43] Alger I, Weibull JW, Lehmann L. 2020 Evolution of preferences in structured populations: Genes, guns, and culture. *Journal of Economic Theory* **185**, 1–45.

[44] Maynard Smith J, Price GR. 1973 The logic of animal conflict. *Nature* **246**, 15–18.

[45] Maynard Smith J. 1982 *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.

[46] Débarre F, Nuismer SL, Doebeli M. 2014 Multidimensional (Co)Evolutionary Stability. *The American Naturalist* **184**, 158–171.

[47] Mullon C, Keller L, Lehmann L. 2016 Evolutionary stability of jointly evolving traits in subdivided populations. *American Naturalist* **188**, 175–195.

[48] Geritz SAH, Metz JAJ, Rueffler C. 2016 Mutual invadability near evolutionarily singular strategies for multivariate traits, with special reference to the strongly convergence stable case. *Journal of Mathematical Biology* **72**, 1081–1099.

[49] Meszéna G, Kisdi E, Dieckmann U, a. H. Geritz S, a. J. Metz J. 2002 Evolutionary Optimisation Models and Matrix Games in the Unified Perspective of Adaptive Dynamics. *Selection* **2**, 193–220.

[50] Wolf M, McNamara JM. 2012 On the Evolution of Personalities via Frequency-Dependent Selection. *The American Naturalist* **179**, 679–692.

[51] Day T. 2001 Population structure inhibits evolutionary diversification under competition for resources. *Genetica* **112**, 71–86.

[52] Ajar E. 2003 Analysis of disruptive selection in subdivided populations. *BMC Evolutionary Biology* **3**, 22.

[53] Wakano JY, Lehmann L. 2014 Evolutionary branching in deme-structured populations. *Journal of Theoretical Biology* **351**, 83–95.

[54] Barnard CJ, Sibly RM. 1981 Producers and scroungers: a general model and its application to captive flocks of house sparrows. *Animal Behaviour* **29**, 543–550.

[55] Powers S, Lehmann L. 2017 When is bigger better? The effects of group size on the evolution of helping behaviours. *Biological Reviews* **92**, 902–920.

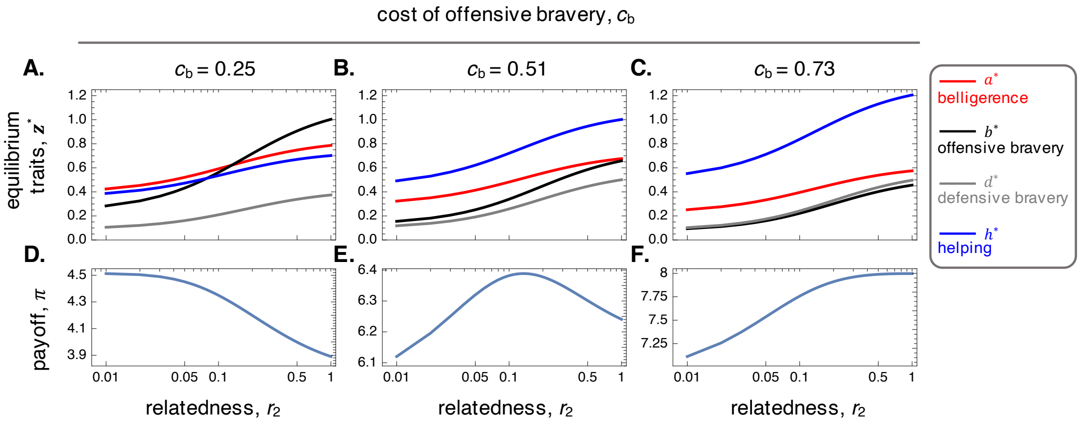


**Figure 1: Traits $\boldsymbol{z}^* = (a^*, b^*, d^*, h^*)$ and payoff $\pi(\boldsymbol{z}^*, \boldsymbol{z}^*, \boldsymbol{z}^*)$ at evolutionary equilibrium.** **A**-**C**: Evolutionary equilibria of belligerence ($a^*$ in red), offensive bravery ($b^*$ in black), defensive bravery ($d^*$ in gray) and helping ($h^*$ in blue) as a function of pairwise relatedness (found by fixing $N = 8$ and solving $r_2$ eq. (S12) for $m$, which is then substituted into equilibrium eqs. 9 with eqs. 7 and (S21); other parameters: $c_1 = 18$, $c_2 = 115$, $c_\text{d} = 0.67$, $\omega = 0.5$, $\beta\sqrt{N} = 100$, $v_\text{b} = 0$, see legend for $c_\text{b}$). All traits increase with relatedness. **D**-**F**: Payoff as a function of pairwise relatedness (using eqs. (S1)- (S3) and traits in top row). When the cost of offensive bravery is low ($c_\text{b} = 0.25$), payoff decreases with relatedness (in D). This is because belligerence (red in A) increases more than helping (blue in A) with relatedness. By contrast, when $c_\text{b} = 0.73$, helping increases more than attack (in C) with relatedness causing a simultaneous increase in fecundity (in F).

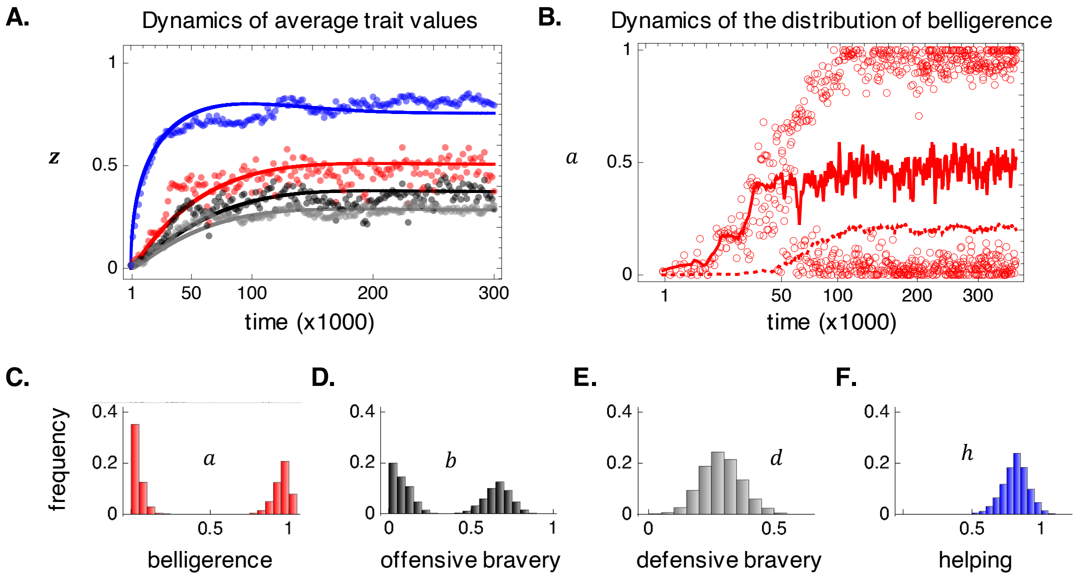
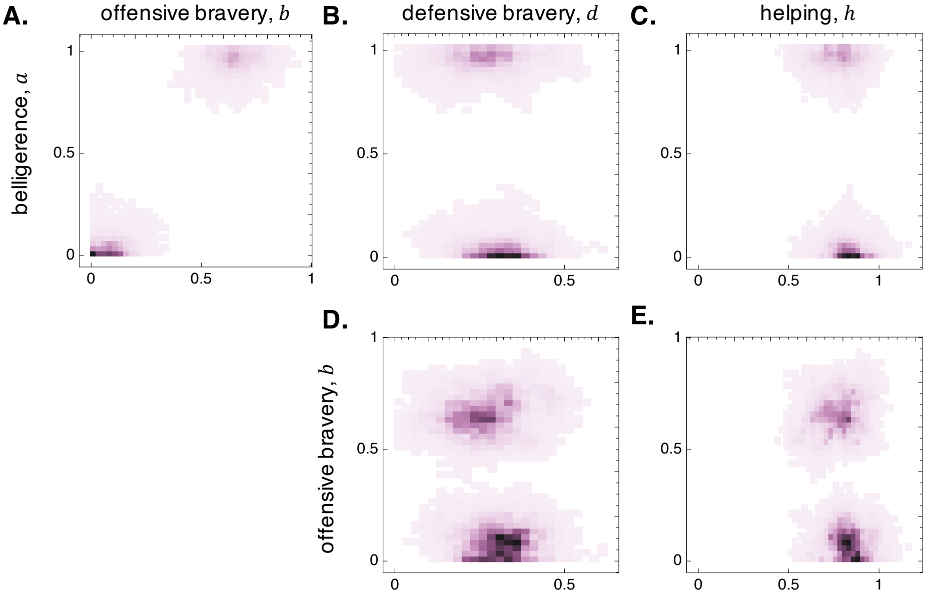


**Figure 2: The emergence of polymorphism.** **A**: Average trait values in the population as a function of time (with a unit of time corresponding to an iteration of the life cycle) in a population where all traits are initially absent ($a_t$ in red; $b_t$ in black; $d_t$ in gray; $h_t$ in blue; observed in a simulation in dots, SM S4 for simulation details; analytically predicted in full lines, from eq. (S9) with variance-covariance matrix $\mathbf{G} = \delta^2\mathbf{I}$ where $\mathbf{I}$ is the identity matrix and $\delta = 0.043$ was chosen heuristically; $N = 8$ and $m = 0.467$ so that $r_2 = 0.125$; otherwise same parameters as Fig. 1 middle). **B**: Individual values of belligerence observed in a simulation (empty circles, shown for two individuals randomly sampled every 800 time points, same parameters as A), as well as the observed trait population average (full line) and variance (dotted line, time on a log scale). Note that polymorphism occurs before the population average has completely converged to its equilibrium in simulations due to mutations being relatively large (to speed up computation time). But as predicted, the trait variance starts to increase significantly only once the population average has stabilised. **C**-**F**: Distribution of each trait in a simulation (calculated from time 250'000 for 100'000 time steps, same parameters as A).



**G.** **Variance-covariance matrix**

| | $a$ | $b$ | $d$ | $h$ |
|---|---|---|---|---|
| $a$ | 0.21<br>± 0.010 | 0.13<br>± 0.011 | −0.014<br>± 0.0093 | −0.015<br>± 0.014 |
| $b$ | - | 0.089<br>± 0.012 | −0.0091<br>± 0.0063 | −0.0092<br>± 0.0082 |
| $d$ | - | - | 0.0062<br>± 0.0022 | 0.0012<br>± 0.0027 |
| $h$ | - | - | - | 0.0081<br>± 0.0042 |

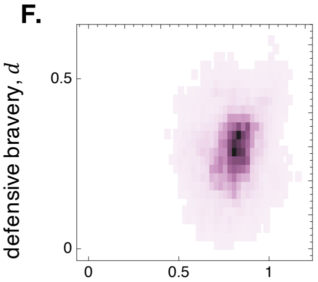


**Figure 3: Correlations among traits. A.-F.** Joint distribution of each pair of traits in a simulated population at equilibrium (calculated from time 250'000 for 100'000 time steps, same simulation as Fig. 2, darker shade means greater density). **G.** Mean ± standard deviation of the variance-covariance matrix evaluated in the same simulated population and time points as A-F; variances are on the diagonal while covariances among each pair traits are on the off-diagonal (see SM eq. (S8) for definition).

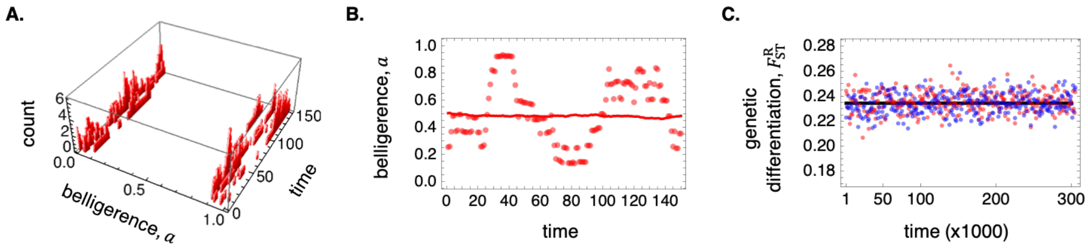
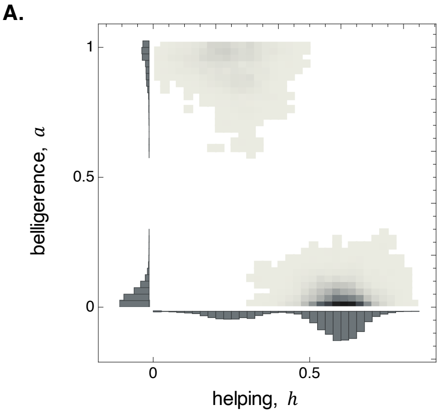


**Figure 4: Polymorphism within and between groups. A.** Distribution of belligerence (number of individuals) in a focal group over time (150 steps, same simulation as in figure 2 from time 299'000 onwards). **B.** Focal group- (dots) and population- (full line) average belligerence (same group as in A). **C.** Genetic differentiation among groups $F_{\mathrm{ST}}^{\mathrm{R}}$ ($F_{\mathrm{ST}}$ with replacement, i.e. the ratio of the variance among groups of group-averages to the total trait variance in the population) over time at the belligerence (in red) and helping (in blue) loci in a simulated population, against neutral expectation (in black full line, calculated from $F_{\mathrm{ST}}^{\mathrm{R}} = 1/N + (N-1)r_2/N$ where $r_2$ is the neutral relatedness coefficient, eq. S12, same simulation as in figure 2).



**B.**

**Variance-covariance matrix**

| | $a$ | $b$ | $d$ | $h$ |
|---|---|---|---|---|
| $a$ | 0.15<br>± 0.022 | 0.051<br>± 0.007 | 0.018<br>± 0.0056 | −0.059<br>± 0.011 |
| $b$ | - | 0.017<br>± 0.003 | 0.0062<br>± 0.0018 | −0.020<br>± 0.0044 |
| $d$ | - | - | 0.0040<br>± 0.0012 | −0.0074<br>± 0.0029 |
| $h$ | - | - | - | 0.026<br>± 0.0075 |

**Figure 5: Negative association between helping and contest traits. A.** Joint distribution of belligerence and helping in a simulated population at equilibrium, where offensive and defensive bravery have complementarity effects among each other but antagonistic effect with helping (at the individual level), i.e. where $C(h_{\bullet}, b_{\bullet}, d_{\bullet})$ is given by eq. (11) (with $c_{\text{b}} = 0.7$, $c_{\text{d}} = 1.3$,$c_{\text{h}} = 1$, $\gamma_{\text{bd}} = 0.5$, $\gamma_{\text{bh}} = 2.2$, $\gamma_{\text{dh}} = 2.2$; other parameters: same as Fig. 2; joint distribution calculated over 100'000 time points after 150'000 of evolution). **B.** Mean ± standard deviation of the variance-covariance matrix evaluated for the same simulated population as A. In contrast to Fig. 3G, the covariance among helping ($h$) and belligerence ($a$) is now significantly different to zero and negative. And while this covariance may seem small, the correlation among belligerence ($a$) and helping ($h$) is large (as shown in **A**).

# Supplementary Material for "Evolution of warfare by resource raiding favors polymorphism in belligerence and bravery"

Charles Mullon[1,*] and Laurent Lehmann[1]

[1]Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland
[*]Correspondence: `Charles.mullon@unil.ch`

## S1 Payoff and fitness

In this first supplement, we derive the expected payoff that a focal individual obtains (in S1.1) from which we then characterise its fitness (in S1.2).

### S1.1 Expected payoff

From the assumptions and notations of the main text (section 2.2), it follows that the expected material payoff to a focal individual with traits $\boldsymbol{z}_\bullet$ in a group with average trait values $\boldsymbol{z}_0$ when the population average trait values are $\boldsymbol{z}$ can be written as

$$\pi(\boldsymbol{z}_\bullet, \boldsymbol{z}_0, \boldsymbol{z}) = v_\mathrm{b} + \phi_0(a_0, a)\pi_\mathrm{a}(\boldsymbol{z}_0, \boldsymbol{z}) + (1 - \phi_0(a_0, a))\pi_\mathrm{b}(\boldsymbol{z}_0, \boldsymbol{z}) - C(h_\bullet, b_\bullet, d_\bullet), \tag{S1}$$

where $v_\mathrm{b} > 0$ is some baseline payoff, $\pi_\mathrm{a}(\boldsymbol{z}_0, \boldsymbol{z})$ is the expected payoff to the focal individual, conditional on its group engaging into a raid [which occurs with probability $\phi_0(a_0, a)$], and $\pi_\mathrm{b}(\boldsymbol{z}_0, \boldsymbol{z})$ is the expected payoff to the focal, conditional on the focal group not raiding another group [which occurs with probability $1 - \phi_0(a_0, a)$]. Let us first consider $\pi_\mathrm{b}(\boldsymbol{z}_0, \boldsymbol{z})$ as it is simpler to obtain. Given that the focal group has not raided another one, the expected payoff to the focal individual is

$$\pi_\mathrm{b}(\boldsymbol{z}_0, \boldsymbol{z}) = \phi_1(a)\left[-\frac{c_1}{N} + (1 - v(b, d_0))\frac{B(Nh_0)}{N}\right] + (1 - \phi_1(a))\frac{B(Nh_0)}{N}, \tag{S2}$$

which can be understood as follows. With probability $\phi_1(a)$, the focal group is attacked. In this case, the focal individual necessarily pays a cost $-c_1/N$ for one fight but only gets to retain its share $B(Nh_0)/N$ of the common resource if its group wins the fight against its raiders, which occurs with probability $1 - v(b, d_0)$. With probability $(1-\phi_1(a))$, the focal group is not attacked and therefore the focal individual always gets its share $B(Nh_0)/N$. Using similar arguments, we find that the expected payoff of the focal individual given that its group has participated in a raid is

$$\begin{aligned}\pi_{\mathrm{a}}(\boldsymbol{z}_0,\boldsymbol{z}) = {} & v(b_0,d)\left\{\frac{B(Nh)}{N}+\phi_1(a)\left[-\frac{c_2}{N}+(1-v(b,d_0))\frac{B(Nh_0)}{N}\right]+(1-\phi_1(a))\left[\frac{-c_1+B(Nh_0)}{N}\right]\right\} \\ & +(1-v(b_0,d))\left\{\phi_1(a)\left[-\frac{c_2}{N}+(1-v(b,d_0))\frac{B(Nh_0)}{N}\right]+(1-\phi_1(a))\left[\frac{-c_1+B(Nh_0)}{N}\right]\right\},\end{aligned} \tag{S3}$$

where the first line consists of the probability $v(b_0,d)$ that the focal group wins the raid it engaged in, multiplied to the expected payoff in this case (between curly brackets). Conversely, the second line is the probability $1-v(b_0,d)$ that the focal group loses the raid it engaged in multiplied by the relevant payoff obtained in that case (between curly brackets also). Substituting eqs. (S2)–(S3) into eq. (S1) and re-arrangements yield eq. (4).

## S1.2 Fitness

We assume that an individual's fecundity, i.e. the number of offspring produced during stage (3) of the life cycle, increases with its expected payoff according to a function $F$, so that the fecundity of a focal individual with payoff $\pi$ is written as $F(\pi)$. We assume that this function is positive and decelerating with expected payoff (i.e. $F'' \leq 0$). Formally, the fecundity of an individual should really be a function of its *realised* payoffs, which depend on a specific sequence of events (such as whether fighting took place, whether the fight was won and so on), rather than *expected* payoffs, which are averaged over all possible outcomes (as in eqs. S1-S3). In writing fecundity directly as a function of expected payoffs only, we are essentially assuming that the deviation between realised and expected payoffs is small (specifically, with $\Pi$ denoting the random variable for the payoff to the focal individual, we ignore terms of order $\mathrm{E}[(\Pi-\pi)^2]$ and higher, where $\mathrm{E}[\cdot]$ stands for the expectation operator over all relevant stochastic effects). This is a standard albeit often left unspecified assumption in evolutionary game theory. One alternative to such an assumption is to define payoff directly in units in fecundity, in which case payoff is simply fecundity (i.e. $\pi = F(\pi)$). In any event, once we have specified its fecundity we can determine the *fitness* of an individual, which is its expected number of surviving offspring produced over one full iteration of the life-cycle (e.g., [1, 2]) and lays the foundation of our evolutionary analysis. We detail such fitness function for our model, which considers warfare as a subsistence strategy, below (in S1.2.1), as well as the typical fitness function that characterises models where warfare should rather be considered as a reproductive strategy for contrast (in S1.2.2).

### S1.2.1 Warfare as a subsistence strategy

In order to obtain the expression for the fitness of a focal individual under the assumptions of our model (section 2.1), let us first label the other individuals in the focal group (i.e., the focal's neighbours) as individual "2", "3" until "$N$", and denote their respective traits as $\boldsymbol{z}_2$, $\boldsymbol{z}_3$ and so on, whereby we can write the average trait in the focal group as

$$\boldsymbol{z}_0=\frac{\boldsymbol{z}_\bullet+\boldsymbol{z}_2+\boldsymbol{z}_3+\sum_{j=4}^{N}\boldsymbol{z}_j}{N}. \tag{S4}$$

It will also be useful to collect all the trait values of the neighbours of the focal into the vector $\boldsymbol{z}_{-\bullet}=(\boldsymbol{z}_2,\boldsymbol{z}_3,\ldots)$. Then, according to the life-cycle detailed in section 2.1, the fitness of the focal individual is

$$w(\boldsymbol{z}_\bullet,\boldsymbol{z}_{-\bullet},\boldsymbol{z})=\frac{N-1}{N}+\frac{(1-m)F(\pi(\boldsymbol{z}_\bullet,\boldsymbol{z}_0,\boldsymbol{z}))}{(1-m)[F(\pi(\boldsymbol{z}_\bullet,\boldsymbol{z}_0,\boldsymbol{z}))+\sum_{j=2}^{N}F(\pi(\boldsymbol{z}_j,\boldsymbol{z}_0,\boldsymbol{z}))]+mNF(\pi(\boldsymbol{z},\boldsymbol{z},\boldsymbol{z}))}+\frac{mF(\pi(\boldsymbol{z}_\bullet,\boldsymbol{z}_0,\boldsymbol{z}))}{NF(\pi(\boldsymbol{z},\boldsymbol{z},\boldsymbol{z}))}, \tag{S5}$$

where recall $\pi$ is the payoff function (eq. S1) and $F(\pi)$ is the fecundity of an individual with payoff $\pi$. Equation (S5) can be read as the sum of three fitness components. (1) The first summand is the probability that the focal survives. (2) The second summand is the probability that the open philopatric spot is filled by one of its offspring. This consists of the ratio of the number of the focal's offspring that remain philopatric to the total number of offspring that enter competition in the focal patch (comprised of all those that remain philopatric – factored by $(1-m)$ – and all those that disperse from other patches – factored by $m$). (3) The last summand is the expected number of spots filled in other patches, given by the ratio of the number of the focal's offspring that disperse to the expected total number of offspring that compete in another patch.

Equation (S5) has the standard form of individual fitness under the island model of dispersal coupled with a Moran process (Box 1 in [3]). It could be straightforwardly expanded to consider the case where payoffs affect survival rather than fecundity [3]. Note also that to obtain the expression for fitness under a Wright-Fisher process (i.e. where all individuals are replaced per life-cycle iteration), one simply removes the first summand of eq. (S5) and multiplies the rest by $N$. More generally, assuming that the spoils of warfare only influence fecundity or survival, i.e. where warfare is conceived as a subsistence strategy, eq. (S5) can easily be amended to consider other common variations, such as where regulation occurs before dispersal, or where individuals survive with fixed probability.

### S1.2.2 Warfare as a reproductive strategy

In order to highlight the key difference in fitness between co-evolutionary models where warfare is a subsistence (as in eq. S5) vs. a reproductive strategy (as in [4–9]), let us consider a simple representative case of the latter under the following assumptions: that the winning group takes over all the breeding spots of the losing group; that density-dependent regulation of offspring occurs before dispersal (i.e. soft selection); and that there is no difference between offensive and defensive bravery ($b = d$). In this case, the fitness of a focal individual can be written as

$$w(\boldsymbol{z}_{\bullet}, \boldsymbol{z}_0, \boldsymbol{z}) = \left[1 + \underbrace{\phi_0(a_0, a)v(b_0, b)}_{\text{gaining a patch}} - \underbrace{\phi_1(a)v(b, b_0)}_{\text{loosing own patch}}\right] \times \frac{1 - C(a_{\bullet}, b_{\bullet})}{1 - C(a_0, b_0)} \tag{S6}$$

(obtained from eq. 11 of [6] by setting $m = 0$ and $h = 0$). The main differences between eq. (S5) and eq. (S6) is that for the latter: (i) there is no common pool resource production; (ii) the cost of belligerence is fixed, i.e., it now appears along the other traits in the cost function $C(a_{\bullet}, b_{\bullet})$ (in contrast to eq. S1 where belligerence has conditional costs, $c_1$ and $c_2$); (iii) the benefits of warfare and the costs of trait expression are multiplicative (i.e., the benefits of warfare, which are in the square brackets, multiply the costs, while in eq. S1 the benefits of warfare and these costs of trait expression add up); and finally, (iv) the gains of warfare will not be partly destroyed by competition, i.e., the gains of warfare affect the denominator in eq. (S5) but not in eq. (S6). Having these differences in mind is useful to contrast our results with those of previous studies (see section S3.2.5).

## S2 Mathematical evolutionary approach

In this supplement, we outline our mathematical analysis for the joint evolution of the four traits of interest ($a$, $b$, $d$ and $h$). It is based on an evolutionary quantitative genetics and adaptive dynamics model for group-structured populations that is tightly connected to invasion analysis [10]. This model tracks the dynamics of a multi-trait phenotypic distribution, assuming that the processes of selection and mutation are such that this distribution is approximately multi-variate Gaussian across the whole population (i.e. over all individuals and all groups) with small (co)variance (specifically, assuming that the largest absolute value among all traits' (co)variances can be written as $\delta^2$ where $0 < \delta \ll 1$ is a small parameter). Note that this assumption does not require that the realized distribution of phenotypes within a focal group at any given demographic time period is Gaussian, but rather that its time average is. Such assumption of normality has been shown to give accurate predictions for the evolution of traits' means and (co)variances, even where selection generates significant deviations from normality (refs. [11] for well-mixed, and [10] for dispersal-limited populations).

With the assumption of multivariate normality, the phenotypic distribution at any time $t$ for our warfare model can be described by the vector of means,

$$\boldsymbol{z}_t = \begin{pmatrix} a_t \\ b_t \\ d_t \\ h_t \end{pmatrix}, \tag{S7}$$

and the (symmetric) variance/covariance matrix

$$\mathbf{G}_t = \begin{pmatrix} G_{aa,t} & G_{ab,t} & G_{ad,t} & G_{ah,t} \\ G_{ba,t} & G_{bb,t} & G_{bd,t} & G_{bh,t} \\ G_{da,t} & G_{db,t} & G_{dd,t} & G_{dh,t} \\ G_{ha,t} & G_{hb,t} & G_{hd,t} & G_{hh,t}, \end{pmatrix} \tag{S8}$$

where $G_{uv,t} \sim \mathcal{O}\left(\delta^2\right)$ is the genetic covariance among traits $u$ and $v$ at time $t$ (so that $G_{uv,t} = G_{vu,t}$). Hence the dynamics of the phenotypic distribution are given by the joint dynamics of the vector $\boldsymbol{z}_t$ and the matrix $\mathbf{G}_t$. However, as we detail below, we do not need to consider these dynamics jointly when $\delta$ is small.

## S2.1 Evolution in two time scales

In brief, the upshot of the analysis we follow is that when mutations are rare and have small effects on phenotypes (so that $\delta$ is small), the evolutionary dynamic can be decomposed into two time scales. First, the population evolves under directional selection whereby the average trait values $\boldsymbol{z}_t$ in the population change gradually, but the variance in each trait and covariance among each pair of traits remains small and approximately constant (i.e., $\boldsymbol{z}_t$ changes while $\mathbf{G}_t$ can be held constant). Once the population average has converged to an equilibrium for directional selection, a so-called "convergence stable strategy" (that is thus an attractor of the evolutionary dynamics), selection shapes the traits' (co)variances (i.e. $\mathbf{G}_t$ changes while $\boldsymbol{z}_t$ remains fixed for its equilibrium). An analysis of selection close to convergence stable strategies then allows to establish whether selection is (a) stabilising, keeping traits' (co)variances small so that the phenotypic distribution remains unimodal and centred around the equilibrium which is thus "locally evolutionary stable" or "uninvadable" (so that $\mathbf{G}_t$ converges); or (b) disruptive, favoring an increase in the variance (and possibly covariance) of some traits (so that $\mathbf{G}_t$ diverges). Due to disruptive selection, the population may eventually undergo "evolutionary branching" [12], whereby the phenotypic distribution becomes multi-modal so that two or more clearly differentiated morphs emerge. These morphs may differ in multiple traits owing to correlational selection, which favours specific associations between traits within individuals. We detail mathematically the two time scales of evolutionary dynamics and corresponding effects of selection in the next two sections (directional selection in S2.2 and stabilising/disruptive selection in S2.3).

### S2.2 Directional selection

**The selection gradient.** First, the population evolves under directional selection. To the leading order in $\delta$, the change $\Delta \boldsymbol{z}_t = \boldsymbol{z}_{t+1} - \boldsymbol{z}_t$ in average trait is given by

$$\Delta \boldsymbol{z}_t = \mathbf{G} \cdot \boldsymbol{s}(\boldsymbol{z}_t), \tag{S9}$$

(eq. 3 of [10]) where $\mathbf{G}$ is the matrix of traits' genetic (co)variance (which is assumed to be constant during the evolution of the means so we can drop its time index for this section S2.2; in fact since we assume that each trait is encoded by a separate locus and all loci mutate in a similar way, it is reasonable to assume that $\mathbf{G} = \delta^2 \mathbf{I}$ here with $\mathbf{I}$ being the identity matrix), and

$$\boldsymbol{s}(\boldsymbol{z}_t) = \begin{pmatrix} s_a(\boldsymbol{z}_t) \\ s_b(\boldsymbol{z}_t) \\ s_d(\boldsymbol{z}_t) \\ s_h(\boldsymbol{z}_t) \end{pmatrix}, \tag{S10}$$

is the so-called selection gradient, which is a vector where each entry tells us whether selection favours an increase (when $s_u(\boldsymbol{z}) > 0$) or decrease (when $s_u(\boldsymbol{z}) < 0$) in the corresponding trait ($u \in \{a, b, d, h\}$) when the population average is $\boldsymbol{z}$. Such selection coefficient is given by

$$s_u(\boldsymbol{z}) = \frac{\partial w(\boldsymbol{z}_\bullet, \boldsymbol{z}_{-\bullet}, \boldsymbol{z})}{\partial u_\bullet} + (N-1) r_2 \frac{\partial w(\boldsymbol{z}_\bullet, \boldsymbol{z}_{-\bullet}, \boldsymbol{z})}{\partial u_2}, \tag{S11}$$

where here and hereafter, derivatives are evaluated where all individuals have the average phenotype, $\boldsymbol{z}_\bullet = \boldsymbol{z} = \boldsymbol{z}_2 = \ldots = \boldsymbol{z}$. The first term of eq. (S11) corresponds to the direct fitness effect of trait $u$: the marginal effect of a change in trait $u$ in the focal on its own fitness. The second term of eq. (S11), meanwhile, is the relatedness-weighted indirect fitness effect: the effect of a change in trait $u$ in a neighbour on the fitness of the focal individual, weighted by the coefficient $r_2$ of pairwise of relatedness, which is the probability that two individuals randomly sampled in a group are identical-by-descent, IBD, under neutrality. For the specific life-cycle described in the main text, this coefficient is given by

$$r_2 = \frac{1-m}{1+m(N-1)}, \tag{S12}$$

(e.g., [3] for the Moran model). As such, eq. (S11) can be seen as the marginal form of Hamilton's rule, $-c + r_2 b$, with direct effect as cost, $-c$, and indirect effect as a benefit, $b$.

**Singular strategy and convergence stability.** The dynamics given by eq. (S9) may eventually converge to an equilibrium, $\boldsymbol{z}^*$, so that the means no longer change (i.e. $\Delta \boldsymbol{z}_t = \mathbf{0}$). Such convergence first requires that

$$\boldsymbol{s}(\boldsymbol{z}^*) = \mathbf{0}. \tag{S13}$$

Indeed, because $\mathbf{G}$ is a positive-definite matrix (since it is a covariance matrix), condition eq. (S13) is the only way for $\boldsymbol{z}_{t+1} = \boldsymbol{z}_t = \boldsymbol{z}^*$. A strategy satisfying eq. (S13) is typically referred to as a singular strategy [13]. Whether a singular strategy is an attractor for directional selection (i.e. whether means will converge to $\boldsymbol{z}^*$ defined by eq. S13) can be investigated from the Jacobian matrix

$$\mathbf{J}(\boldsymbol{z}^*) = \begin{pmatrix} J_{aa}(\boldsymbol{z}^*) & J_{ab}(\boldsymbol{z}^*) & J_{ad}(\boldsymbol{z}^*) & J_{ah}(\boldsymbol{z}^*) \\ J_{ba}(\boldsymbol{z}^*) & J_{bb}(\boldsymbol{z}^*) & J_{bd}(\boldsymbol{z}^*) & J_{bh}(\boldsymbol{z}^*) \\ J_{da}(\boldsymbol{z}^*) & J_{db}(\boldsymbol{z}^*) & J_{dd}(\boldsymbol{z}^*) & J_{dh}(\boldsymbol{z}^*) \\ J_{ha}(\boldsymbol{z}^*) & J_{hb}(\boldsymbol{z}^*) & J_{hd}(\boldsymbol{z}^*) & J_{hh}(\boldsymbol{z}^*) \end{pmatrix} \tag{S14}$$

with $(u, v)$-entry

$$J_{uv}(\boldsymbol{z}^*) = \left.\frac{\partial s_u(\boldsymbol{z})}{\partial v}\right|_{\boldsymbol{z}=\boldsymbol{z}^*}. \tag{S15}$$

A *necessary* condition for a singular strategy to be an attractor is that the real parts of the eigenvalues of $\mathbf{J}(\boldsymbol{z}^*)$ are all negative. If so, we say that $\boldsymbol{z}^*$ is weakly convergence stable, "weakly" because it is still possible that the evolutionary dynamics do not converge towards $\boldsymbol{z}^*$ in the presence of genetic correlations among traits (i.e. there may exist non-diagonal matrix $\mathbf{G}$ so that iteration of eq. (S9) do not converge to $\boldsymbol{z}^*$). A *sufficient* condition for a singular strategy to be an attractor is that $\mathbf{J}(\boldsymbol{z}^*)$ is negative-definite [14, 15], i.e. that its symmetric part, $[\mathbf{J}(\boldsymbol{z}^*) + \mathbf{J}(\boldsymbol{z}^*)^{\mathrm{T}}]/2$ has only negative eigenvalues (where T denotes transpose; note that since $\mathbf{J}(\boldsymbol{z}^*) + \mathbf{J}(\boldsymbol{z}^*)^{\mathrm{T}}$ is symmetric with real entries, all its eigenvalues are real). A singular strategy $\boldsymbol{z}^*$ satisfying this sufficiency condition is said to be strongly convergence stable [15] in reference to the fact that as a result of mutation and selection, a population close to $\boldsymbol{z}^*$ will always gradually converge to $\boldsymbol{z}^*$, whatever the genetic correlations among traits (i.e, whatever the positive-definite $\mathbf{G}$ matrix). But since we assume that each trait is encoded by an independently mutating locus, there should be no genetic correlation among traits (under weak selection at least) . So the necessary condition for convergence stability should also be sufficient for trait averages to converge towards an equilibrium in our model.

### S2.3 Stabilising, disruptive and correlational selection

**Hessian matrix.** Once the average traits in the population have converged to an equilibrium $\boldsymbol{z}^*$ under directional selection, the traits' (co)variances given by $\mathbf{G}$ around this population mean then start changing under the actions of mutations and selection. To the leading order of $\delta$, this change $\Delta\mathbf{G}_t = \mathbf{G}_{t+1} - \mathbf{G}_t$ over one demographic time step when the population mean is at a convergence stable phenotype is captured by

$$\Delta\mathbf{G}_t = \mathbf{M} + \mathbf{G}_t \cdot \mathbf{H}(\boldsymbol{z}^*) \cdot \mathbf{G}_t \tag{S16}$$

(eq. 3b of [10] with vanishing selection gradient), where the constant positive-definite matrix $\mathbf{M}$ captures the input of mutations (in the absence of pleiotropy and each trait mutating with the same probability and effects, as in our model, we can write this equation as $\mathbf{M} = \delta^2\mathbf{I}$), and the symmetric Hessian matrix

$$\mathbf{H}(\boldsymbol{z}^*) = \begin{pmatrix} H_{aa}(\boldsymbol{z}^*) & H_{ab}(\boldsymbol{z}^*) & H_{ad}(\boldsymbol{z}^*) & H_{ah}(\boldsymbol{z}^*) \\ H_{ba}(\boldsymbol{z}^*) & H_{bb}(\boldsymbol{z}^*) & H_{bd}(\boldsymbol{z}^*) & H_{bh}(\boldsymbol{z}^*) \\ H_{da}(\boldsymbol{z}^*) & H_{db}(\boldsymbol{z}^*) & H_{dd}(\boldsymbol{z}^*) & H_{dh}(\boldsymbol{z}^*) \\ H_{ha}(\boldsymbol{z}^*) & H_{hb}(\boldsymbol{z}^*) & H_{hd}(\boldsymbol{z}^*) & H_{hh}(\boldsymbol{z}^*), \end{pmatrix} \tag{S17}$$

captures the effects of selection. Each entry of this matrix informs on the nature of selection on traits' (co)variances at a singular strategy. Specifically, the sign of each diagonal entry indicates whether selection favours a decrease (when $H_{uu}(\boldsymbol{z}^*) < 0$) or increase (when $H_{uu}(\boldsymbol{z}^*) > 0$) in the variance of the corresponding trait (here $u$) when this trait evolves in isolation of the other traits [16]. In other words, $H_{uu}(\boldsymbol{z}^*)$ tells us whether selection on trait $u$ alone is stabilising (when $H_{uu}(\boldsymbol{z}^*) < 0$) or disruptive (when $H_{uu}(\boldsymbol{z}^*) > 0$). Further, the sign of each off-diagonal entry indicates whether selection favours a positive (when $H_{uv}(\boldsymbol{z}^*) > 0$) or negative (when $H_{uv}(\boldsymbol{z}^*) > 0$) correlation among the two corresponding traits (here $u$ and $v$) when these traits evolve in isolation from the others. The quantity $H_{uv}(\boldsymbol{z}^*)$ (with $u \neq v$) has accordingly been coined as the coefficient of "correlational selection" [16].

Under our life-cycle assumptions (i.e. Moran reproductive process), the $(u, v)$-entry of the Hessian matrix is given by

$$\begin{aligned} H_{uv}(\boldsymbol{z}^*) = \frac{\partial^2 w(\boldsymbol{z}_\bullet, \boldsymbol{z}_{-\bullet}, \boldsymbol{z})}{\partial u_\bullet \partial v_\bullet} &+ (N-1) r_2 \left[ \frac{\partial^2 w(\boldsymbol{z}_\bullet, \boldsymbol{z}_{-\bullet}, \boldsymbol{z})}{\partial u_2 \partial v_2} + \frac{\partial^2 w(\boldsymbol{z}_\bullet, \boldsymbol{z}_{-\bullet}, \boldsymbol{z})}{\partial u_\bullet \partial v_2} + \frac{\partial^2 w(\boldsymbol{z}_\bullet, \boldsymbol{z}_{-\bullet}, \boldsymbol{z})}{\partial u_2 \partial v_\bullet} \right] \\ &+ (N-1)(N-2) r_3 \frac{\partial^2 w(\boldsymbol{z}_\bullet, \boldsymbol{z}_{-\bullet}, \boldsymbol{z})}{\partial u_2 \partial v_3} \end{aligned} \tag{S18}$$

(eq. 7.b of [10]), where derivatives are all evaluated at the singular phenotype, $\boldsymbol{z}_\bullet = \boldsymbol{z} = \boldsymbol{z}_2 = \ldots = \boldsymbol{z}^*$, and

$$r_3 = \frac{2(1-m)}{2+m(N-2)} r_2 = \frac{2(1-m)^2}{[1+m(N-1)][2+m(N-2)]} \tag{S19}$$

is the three way relatedness coefficient for the Moran model, i.e., the probability that three individuals randomly sampled from the same group are IBD under neutrality (more generally, the Hessian matrix eq. S17 is comprised of extra terms that capture the effects of traits on relatedness, eq. 7.a & c of [10], but these vanish under fecundity effects under a Moran life-cycle at a singular phenotype, eq. 16 of [3] for details, and so we can ignore them here).

**Stabilising and disruptive selection.** With all traits coevolving, whether selection is: (1) stabilising, keeping traits' (co)variances small so that the phenotypic distribution remains unimodal around the equilibrium; or (2) disruptive, favoring an increase in the variance of some traits (and possibly some covariances), depends on the leading eigenvalue of the Hessian matrix $\rho(\mathbf{H}(\boldsymbol{z}^*))$ (where $\rho(\mathbf{A})$ denotes the leading eigenvalue of a matrix $\mathbf{A}$). Selection is stabilising when $\rho(\mathbf{H}(\boldsymbol{z}^*)) < 0$. In this case, selection purges genetic variation that deviates from the singular strategy. Such a strategy $\boldsymbol{z}^*$ is said to be uninvadable. As a result of stabilising selection combined with mutation, traits in the population reach an equilibrium that is characterised by a distribution concentrated around the singular strategy. By contrast, selection is disruptive when $\rho(\mathbf{H}(\boldsymbol{z}^*)) > 0$. In this case, genetic variation increases along the eigenvector associated with the leading eigenvalue. This may lead to evolutionary branching whereby the phenotypic distribution goes from being unimodal to bimodal so that two highly differentiated morphs or types coexist in the population (for further considerations on this when multiple traits coevolve, [17]).

The analysis of the eigenvalue $\rho(\mathbf{H}(\boldsymbol{z}^*))$ can be prohibitively complicated. Fortunately there exists simpler conditions that are sufficient for disruptive selection to occur (i.e. for $\rho(\mathbf{H}(\boldsymbol{z}^*)) > 0$), which use the fact that the Hessian is a symmetric matrix [18]. In particular, if any diagonal entry is positive ($H_{uu}(\boldsymbol{z}^*) > 0$), then $\rho(\mathbf{H}(\boldsymbol{z}^*)) > 0$, i.e. if selection is disruptive on any trait when it evolves in isolation from the others, then selection is disruptive when they all co-evolve. Alternatively when $H_{uu}(\boldsymbol{z}^*) < 0$ for all $u$, selection is disruptive ($\rho(\mathbf{H}(\boldsymbol{z}^*)) > 0$) if the off-diagonal entry of any $2 \times 2$ submatrix of $\mathbf{H}(\boldsymbol{z}^*)$ is large relative to the diagonal entries of this submatrix so that $H_{uv}(\boldsymbol{z}^*)^2 > H_{uu}(\boldsymbol{z}^*) H_{vv}(\boldsymbol{z}^*)$ (for some $u \neq v$). Put differently, selection is disruptive if correlational selection among two traits is large relative to stabilising selection on both isolated traits.

# S3 Analyses

In this supplement, we detail the mathematical analyses underlying the results summarized in the main text. In particular, we derive the singular strategies (eq. 9 of the main text), show that we expect these strategies to be convergence stable but not uninvadable, i.e., that we expect evolutionary branching to happen in our model. The basis of all our results is obtained by first computing payoff (substitute eqs. (1)–(3) into eq. (S1)) that is substituted into eq. (S5) to calculate fitness, which is in turn substituted into the selection gradient vector eq. (S11) and Hessian matrix eq. (S18). All the relevant quantities for our evolutionary analysis unfold from these operations. We provide a Mathematica notebook to follow and check all computations reported below ([19], see attached M-file).

## S3.1 Helping and belligerence

To begin with, we derive the conditions under which belligerence emerge (eq. 6 of main text). First, we set the selection gradient on helping to zero when belligerence and both forms of bravery are absent in the population, i.e. we set $s_h(\boldsymbol{z}) = 0$ with $\boldsymbol{z} = (0, 0, 0, h)$. After rearrangements, we obtain eq. (6) of the main text which gives the first order condition to the equilibrium of helping when the other traits are absent. Condition eq. (6) shows that helping equilibrium increases with the parameter $\kappa^{\mathrm{R}}$ which increases with relatedness (eq. 7, Fig. 4). This parameter $\kappa^{\mathrm{R}}$ in fact incorporates two antagonistic effects of limited dispersal on the evolution of helping or other pro-social traits. On one hand, high relatedness due to limited dispersal favors prosocial behavior within groups because in this case, the recipients of the actions of an individual tend to bear the same genes underlying those actions (i.e. kin selection operates). On the other hand, group members typically also compete more strongly for the same local resources than two randomly sampled individuals in the populations (which is the case in our model, section 2.1). As a result, the positive effects of relatedness on the evolution of behaviour tend to be mitigated by competition between relatives, referred to as "kin competition". This is typically reflected in models of social evolution under limited dispersal where evolutionary stable trait values depend on relatedness scaled by local competition, which is captured by the parameter $\kappa < r_2$ (eq. 7, Fig. 4; e.g., [2, 20–24]). As mentioned in the main text, our condition for helping evolution depends on $\kappa^{\mathrm{R}}$ rather than $\kappa$ because the benefits of helping are shared equally within the group. An individual therefore always recoup a share $1/N$ of its own investment, which increases selection on helping (i.e. $\kappa^{\mathrm{R}} > \kappa$, eq. 7, Fig. 4). Since all traits we study are in effect pro-social and benefit the whole group equally, this quantity $\kappa^{\mathrm{R}}$ will also emerge in the selection gradients of the other traits other than helping ($a$, $b$ and $d$).

Second, we look at where the selection gradient on belligerence is positive when helping is present in the population but bravery is not, i.e. look at where $s_a(\boldsymbol{z}) > 0$ with $\boldsymbol{z} = (0, 0, 0, h)$. This gives us eq. (8) of the main text.

## S3.2 All traits co-evolving

### S3.2.1 Singular strategies

Recall that in order to go further in our analysis and analyse the case where all traits are co-evolving, we make the assumptions: (1) $B(Nh) = \beta\sqrt{Nh}$ (where $\beta > 0$ is a constant); (2) $\alpha(a) = a$; (3) $g(b) = b$; (4) $C(h, b, d) = h + c_{\mathrm{b}} b + c_{\mathrm{d}} d$; and (5) $F(\pi) = \pi$.

First, the singular values for offensive and defensive bravery (eqs. 9b-9c) are found by setting the selection gradients for offensive and defensive bravery to zero and solving these equations for $b^*$ and $d^*$ (i.e. find $b^*$ and $d^*$ in $\boldsymbol{z}^* = (a^*, b^*, d^*, h^*)$ such that $s_b(\boldsymbol{z}^*) = s_d(\boldsymbol{z}^*) = 0$). The singular value for helping (eq. 9d) is in turn found by solving $s_h(\boldsymbol{z}^*) = 0$ for $h^*$ where $b^*$ and $d^*$ are given by eqs. (9b)-(9c). Similarly, the (implicit) singular value for belligerence (eq. 9a) is found by solving $s_a(\boldsymbol{z}^*) = 0$ for $a^*$ where $b^*$ and $d^*$ are given by eqs. (9b)-(9c).

### S3.2.2 Connections with previous results on bravery evolution

The equilibrium for bravery in our model (eq. 9b-9c) is consistent with previous models. In particular, when fighting is certain (setting $\phi = 1 - e^{-a^*} = 1$), offensive and defensive bravery costs are equal ($c_\mathrm{b} = c_\mathrm{b} = c$), and there is no inherent advantage to being in an offensive or defensive position ($\omega = 1/2$), eqs. (9b)-(9c) reduce to

$$b^* = d^* = \frac{1}{4} \times \frac{B}{Nc} \times \kappa^\mathrm{R}. \tag{S20}$$

Under complete dispersal (i.e. random group formation, $m = 1$), we have $\kappa^\mathrm{R} = 1/N$ (eq. 7) so that eq. (S20) reads as, $b^* = d^* = (1/4)B/(cN^2)$. This is equal to the equilibrium found in classical models of investment into contest (e.g. eq. 10 of [25], first equation p. 1018 of [26]). Such congruence follows from the connection between our payoff function and that used in classical model (eq. 5).

Rusch and Gavrilets [26] also present an expression for investment into contest where it is claimed groups consists of relatives due limited dispersal (their first equation p. 1023 of ref. [26], referred to as eq. (RG) hereafter). This eq. (RG) is inconsistent with our eq. (S20). While the exact source of this inconsistency is not fully clear to us, there are several problems with eq. (RG). First, it is in conflict with [26]'s own equation without relatedness (i.e. in the same paper, on p. 1018), as eq. (RG) with $r_2 = 0$ does not reduce to the latter. Second, it disagrees with the notion that when $r_2 = 1$, the equilibrium strategy should maximizes group payoff (as there is no conflict within groups of clones). Third, eq. (RG) was taken from [27] in which there are several discrepancies between biological assumptions and fitness accounting[1]. These issues lead us to believe that eq. (RG) is erroneous.

### S3.2.3 Explicit solutions

Our implicit expressions for the equilibria (eq. 9) highlight the inter-dependence between the four co-evolving traits. To obtain explicit expressions in terms of model parameters only, we substitute for $h^*$ (eq. 9d) into eq. (9a) and solve the resulting equation yielding

$$a^* = \log\left(\frac{c_2 - 2c_1 + {v^*}^2\beta^2\kappa^\mathrm{R}/2}{c_2 - c_1 - v^*(1-v^*)\beta^2\kappa^\mathrm{R}/2}\right). \tag{S21}$$

Substituting eq. (S21) into eq. (9d) and in turn these into eqs. (9b)-(9c) gives explicit solutions for the equilibria of the other traits.

---

[1] For e.g.: fitness in the model of [27] goes to zero when the number of groups in the population becomes large, see their eqs. (2) and (4); the "correction" factor in their eq. 6, i.e. their eq. 13, is just stated, it is neither derived nor supported by reference to previous literature; the equation for relatedness (above their eq. 14) is for a model with isolation by distance but nowhere in the manuscript is such isolation by distance evoked and the equations are more consistent with uniform dispersal.

### S3.2.4 Convergence stability

From our general assumption that benefits of the common good decelerate ($B''(h) < 0$) but that costs associated with obtaining this common good (either through production or attacking) do not, it seems reasonable to expect that traits will not grow indefinitely. In other words, we expect that provided helping and belligerence emerge (section 3.1), the joint singular trait value given by eqs. (9) is an evolutionary attractor (i.e. convergence stable). Although the Jacobian matrix (eq. S14) for our model is too complicated to check this expectation analytically, a numerical approach supports it (Mathematica Notebook for results). Indeed, when we sampled $10^6$ random combinations of model parameters such that helping and belligerence emerged, we found that in 99.9% of cases the real part of the dominant eigenvalue of the Jacobian matrix at the singular value was negative (i.e. that $\boldsymbol{z}^*$ is – weakly – convergence stable), and that for 55% of those combinations, the Jacobian matrix was further negative-definite (i.e. that $\boldsymbol{z}^*$ is – strongly – convergence stable). This tells us that in the majority of cases, the population will gradually converge to a state where its phenotypic mean is given by the singular strategy $\boldsymbol{z}^*$, whatever the genetic correlations among traits. If traits are not genetically correlated, which should be the case under our assumption that each trait mutates independently (provided selection is not too strong), such convergence should happen in essentially all cases.

### S3.2.5 Local evolutionary stability

The above analysis suggests that first, the mean phenotype in the population will converge to the singular value $\boldsymbol{z}^*$ while the traits' (co)variances remain small. Selection on these (co)variances then depend on the Hessian matrix $\mathbf{H}(\boldsymbol{z}^*)$ (eqs. S17–S18). Although the Hessian matrix have complicated entries for our model, it turns out to have a simple sign structure,

$$\mathbf{H}(\boldsymbol{z}^*) = \begin{pmatrix} 0 & >0 & 0 & 0 \\ >0 & <0 & 0 & 0 \\ 0 & 0 & <0 & >0 \\ 0 & 0 & >0 & <0 \end{pmatrix} \tag{S22}$$

(Mathematica Notebook). This sign structure tells us a few things. The first is that since none of the diagonal element is positive, none of the traits are under disruptive selection when they evolve in isolation from one another. Further, since $H_{aa}(\boldsymbol{z}^*) = 0$, selection on bravery alone is neither stabilizing nor disruptive at the singular strategy, and this holds for all scenarios investigated in this paper (section S5). This entails that in the $2\times 2$ upper left submatrix, $H_{ab}(\boldsymbol{z}^*)^2 > H_{aa}(\boldsymbol{z}^*)H_{bb}(\boldsymbol{z}^*) = 0$, which means that whenever belligerence coevolves with offensive bravery, selection is disruptive favouring polymorphism. In addition, since $H_{ab}(\boldsymbol{z}^*) > 0$ and $H_{dh}(\boldsymbol{z}^*) > 0$, we expect this polymorphism to be characterised by a positive correlation between belligerence and offensive bravery, and between helping and defensive bravery. Note that because the Hessian matrix provides information on the nature of selection locally, i.e. based on the assumption that the phenotypic distribution is peaked around the singular strategy, our conclusions on correlations hold at least for when the polymorphism initially emerges. As a result, we cannot say anything about the long term nature of the polymorphism [17]. Our simulations provide nonetheless insights into this (Figs. 2-3).

The key feature of eq. (S22) that promotes the emergence of polymorphism is the fact that $H_{aa}(\boldsymbol{z}^*) = 0$. This property will typically not hold when warfare is modelled as a reproductive strategy. This can be seen by considering that when eq. (S6) is substituted into eq. (S18), we are likely to have $H_{aa}(\boldsymbol{z}^*) < 0$. In other words, selection on belligerence alone will tend to be stabilising and thus inhibit disruptive selection.

## S4 Individual based simulations

To confirm our mathematical analysis and investigate trait associations in the longer term, we used individual based stochastic simulations (with finite number of groups $N_\mathrm{g} < \infty$). Such simulations have been carried out extensively across several papers, in which they been shown to generally be in excellent agreement with results from local analyses in group-structured populations, irrespective of group-size and dispersal as long as $m > 0$ and $N_\mathrm{g}$ is sufficiently large for genetic drift to be negligible ([3, 10, 28, 29]; for discussions on the effects of drift on polymorphism in well-mixed populations, see [30, 31]). With this in mind and given the time taken for such simulations to run, we focused simulations here on a specific set of parameter values representative of our model.

Our individual based simulations follow a population composed of $N_\mathrm{g} = 1250$ groups, each populated by $N = 8$ individuals, using Mathematica 10.2.0.0 (see attached M-file, [19]). Starting with a monomorphic population, we track the evolution of the phenotypic distribution for a fixed number of generations. Each individual $i \in \{1, \ldots, N_\mathrm{g}N\}$ at each generation is characterised by a vector of traits $(a_i, b_i, d_i, h_i)$. At the beginning of a generation, we first calculate the payoff $\pi_i$ of each individual according to its traits, those of its neighbours and the average traits in the population (using eqs. S1-S3). Fecundity is taken as payoff (i.e. $F(\pi_i) = \pi_i$). We also ran simulations where we explicitly modelled individual battles following the (finite) island model of warfare [6] so that individual fecundity depended on a specific sequence of events (e.g. whether a raid took place, whether it was won, how many units of resources were present in the specific group raided). As expected from the considerations of section S1.2, these simulations were consistent with those where fecundity was given by expected payoff (using eqs. S1-S3) provided baseline fecundity $v_\mathrm{b}$ was high enough. Since the former were more stochastic and significantly more time consuming, we focused on the latter (i.e. using eqs. S1-S3 to calculate fecundity).

After fecundity is calculated, an individual is randomly sampled in each group to be replaced by an offspring. This offspring is then chosen independently in each group by sampling among the population an individual according to its group and the group in which an breeding spot is being filled. Specifically, if an individual belongs to the same group in which the breeding spot is filled, then its weight is $\pi_i(1-m)$, where $m$ is the dispersal probability. If it belongs to another group, then its weight is $\pi_i m/(N_\mathrm{g}-1)$. Once an individual is chosen to fill the breeding spot, each of its traits mutate independently with probability $\mu = 0.01$. If a trait does not mutate, then it has the same value as in the parent. If a trait does mutate, then we add to parental values a small perturbation that is sampled from a normal distribution with mean 0 and variance $\sigma^2 = 0.02^2$. The resulting phenotypic values are truncated to remain positive (and less than 1 for belligerence $a$ as it is a probability in our examples). We repeat the procedure for a fixed number of generations (Figure legends for parameter values used).

## S5 Robustness of results

In addition to exploring different individual cost function (eq. 11 in main text), we relaxed our baseline model (detailed at the beginning of section 3.2) in three other directions, and (1) considered two further contest functions that have been suggested in the literature: $g(y) = y^\lambda$ and $g(y) = \exp(\lambda y)$ (e.g., [32, 33]); (2) explored the effect of group decision by modeling it as majority "voting" $\alpha(a_0) = a_0^\lambda / [a_0^\lambda + (1-a_0)^\lambda]$, so that as $\lambda$ increases, the decision to raid increasingly becomes binary according to whether the group-average of belligerence $a_0$

is below or above 1/2; and finally (3) allowed for a sigmoidal relationship between total investment into helping in the group and the benefits of this common-pool resource, with $B(Nh) = \beta(Nh)^{\lambda}/((Nh)^{\lambda} + \chi)$. We find that in all examined cases, these different functions influence the value of the equilibrium for each trait in a quantitative way (Suppl. Fig. 4A-D) but not the qualitative nature of these equilibria. Those are still internal evolutionary attractors under directional selection when the costs of fighting are non-additive ($c_2 > 2c_1$), and once the population has converged to the joint equilibrium, selection becomes disruptive favouring the emergence of polymorphism (Suppl. Fig. 5A-D). Our analysis additionally indicates that this polymorphism should again be characterised by a positive association between belligerence and offensive bravery (as indicated by a positive correlational selection coefficient among these two traits, Suppl. Fig. 5A-D).

These extensions suggest that the polymorphism we observed under our baseline model assumptions (section 3.2) is robust to changing the behavioral rules of within- and between-groups interactions. Variations of the demographic assumptions are also unlikely to change these results. We chose the Moran process for simplicity but of course many different alternative life-cycle assumptions are possible (e.g., all individuals die per time step, each individual survives with a fixed probability; density-dependent regulation occurs before dispersal "soft-selection"; dispersal occurs through propagules of individuals). Yet we know from previous social evolution models that these alternatives do not qualitatively affect equilibrium conditions (eq. 9), as all these life-cycle variations can be accounted by varying the scaled-relatedness coefficient [23, 24]. Such variations are also unlikely to qualitatively alter the analysis of disruptive selection (and therefore polymorphism) as disruptive selection can also be expressed in terms of summary demographic variables [3].

One particularly strong assumption when applying our model to animals is that individuals are haploid and reproduce asexually. Thankfully, neither the condition for equilibrium nor for disruptive selection will be qualitatively influenced by diploidy and sexual reproduction when genes have additive effects within individuals [2, 34]. The emergence of polymorphism due to correlational selection may however depend on the genetic architecture of traits [29]. Nonetheless, if the genetic architecture of belligerence and bravery are such that their associations are heritable (e.g. tightly linked or encoded by the same pleiotropic locus), or alternatively if such architecture is allowed to evolve, then the emergence of polymorphism will unfold as in our model [29, 34–36].

## References

[1] Hamilton WD. 1964 The genetical evolution of social behaviour, 1. *Journal of Theoretical Biology* **7**, 1–16.

[2] Rousset F. 2004 *Genetic Structure and Selection in Subdivided Populations.* Princeton, NJ: Princeton University Press.

[3] Mullon C, Keller L, Lehmann L. 2016 Evolutionary stability of jointly evolving traits in subdivided populations. *American Naturalist* **188**, 175–195.

[4] Choi JK, Bowles S. 2007 The coevolution of parochial altruism and war. *Science* **318**, 636–640.

[5] Lehmann L, Feldman MW. 2008 War and the evolution of belligerence and bravery. *Proceedings of the Royal Society of London Series B-Biological Sciences* **275**, 2877–2885.

[6] Lehmann L. 2011 The demographic benefits of belligerence and bravery in the island model of warfare: defeated group repopulation or victorious group size expansion?. *Plos One* **6**, 1–13.

[7] Rusch H. 2014 The two sides of warfare: an extended model of altruistic behavior in ancestral human intergroup conflict. *Human Nature* **25**, 359–377.

[8] Micheletti AJC, Ruxton GD, Gardner A. 2017 Intrafamily and intragenomic conflicts in human warfare. *Proceedings of the Royal Society B-Biological Sciences* **284**.

[9] Dyble M. 2021 The evolution of altruism through war is highly sensitive to population structure and to civilian and fighter mortality. *Proceedings of the National Academy of Sciences of the United States of America* p. e2011142118.

[10] Mullon C, Lehmann L. 2019 An evolutionary quantitative genetics model for phenotypic (co)variances under limited dispersal, with an application to socially synergistic traits. *Evolution* **73**, 1695–1728.

[11] Turelli M. 1994 Evolution of incompatibility-inducing microbes and their hosts. *Evolution* **48**, 1500–1513.

[12] Geritz SAH, Kisdi E, Meszéna G, Metz JAJ. 1998 Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evolutionary Ecology* **12**, 35–57.

[13] Dercole F, Rinaldi S. 2008 *Analysis of Evolutionary Processes: The Adaptive Dynamics Approach and Its Applications.* Princeton, NJ: Princeton University Press.

[14] Leimar O. 2005 The evolution of phenotypic polymorphism: randomized strategies versus evolutionary branching. *American Naturalist* **165**, 669–681.

[15] Leimar O. 2009 Multidimensional convergence stability. *Evolutionary Ecology Research* **11**, 191–208.

[16] Phillips PC, Arnold SJ. 1989 Visualizing multivariate selection. *Evolution* **43**, 1209–1222.

[17] Geritz SAH, Metz JAJ, Rueffler C. 2016 Mutual invadability near evolutionarily singular strategies for multivariate traits, with special reference to the strongly convergence stable case. *Journal of Mathematical Biology* **72**, 1081–1099.

[18] Bhatia R. 2015 *Positive Definite Matrices.* Princeton University Press.

[19] Wolfram Research I. 2016 *Mathematica.* Champaign, Illinois: Wolfram Research, Inc.

[20] Taylor PD. 1992 Altruism in viscous populations - an inclusive fitness model. *Evolutionary Ecology* **6**, 352–356.

[21] Queller DC. 1994 Genetic relatedness in viscous populations. *Evolutionary Ecology* **8**, 70–73.

[22] Frank SA. 1998 *Foundations of Social Evolution.* Princeton, NJ: Princeton University Press.

[23] Lehmann L, Rousset F. 2010 How life-history and demography promote or inhibit the evolution of helping behaviors. *Philosophical Transactions of the Royal Society B* **365**, 2599–2617.

[24] Van Cleve J. 2015 Social evolution and genetic interactions in the short and long term. *Theoretical Population Biology* **103**, 2–26.

[25] Garfinkel MR, Skaperdas S. 2007 Economics of conflict: an overview. In T. S, Hartley K, editors, *Handbook of Defense Economics, Volume 2* pp. 649–709. Amsterdam: North-Holland.

[26] Rusch H, Gavrilets S. 2020 The logic of animal intergroup conflict: A review. *Journal of Economic Behavior and Organization* **178**, 1014–1030.

[27] Gavrilets S, Fortunato L. 2014 A solution to the collective action problem in between-group conflict with within-group inequality. *Nature Communications.*

[28] Wakano JY, Lehmann L. 2014 Evolutionary branching in deme-structured populations. *Journal of Theoretical Biology* **351**, 83–95.

[29] Mullon C, Keller L, Lehmann L. 2018 Social polymorphism is favoured by the co-evolution of dispersal with social behaviour. *Nat Ecol Evol* **2**, 132–140.

[30] Wakano JY, Iwasa Y. 2013 Evolutionary branching in a finite population: deterministic branching vs. stochastic branching. *Genetics* **193**, 229–241.

[31] Débarre F, Otto SP. 2016 Evolutionary dynamics of a quantitative trait in a finite asexual population. *Theoretical Population Biology* **108**, 75–88.

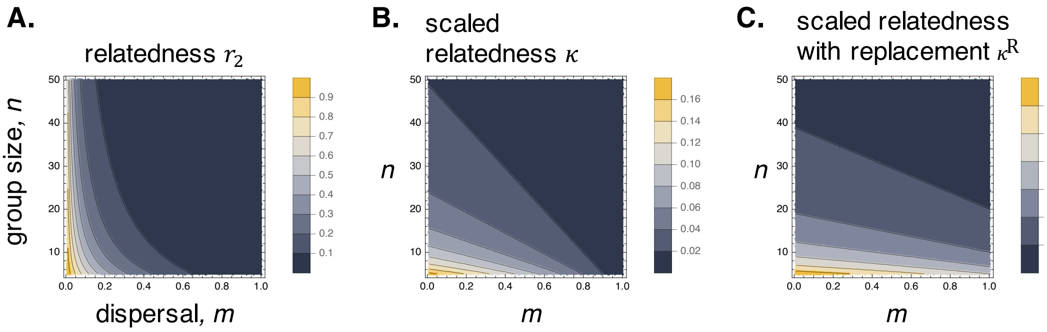
[32] Hirshleifer J. 1991 The technology of conflict as an economic activity. *The American Economic Review* **81**, 130–134.

[33] Skaperdas S. 1996 Contest success functions. *Economic Theory* **7**, 283–290.

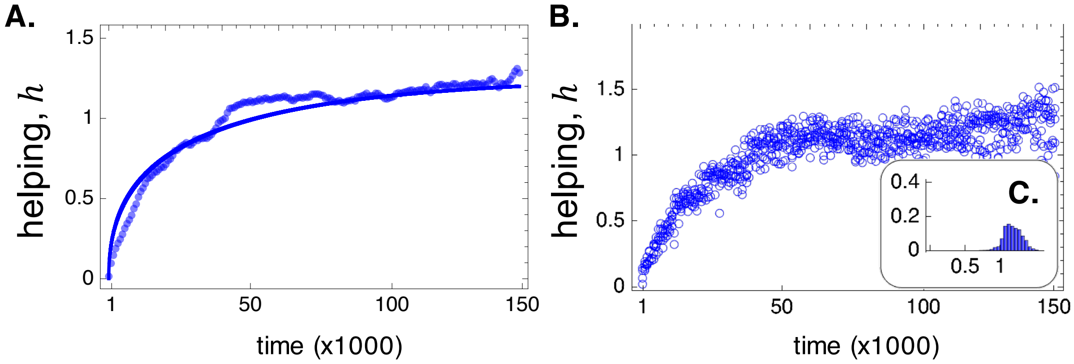
[34] Geritz SAH, Kisdi E. 2000 Adaptive dynamics in diploid, sexual populations and the evolution of reproductive isolation. *Proceedings of the Royal Society B: Biological Sciences* **267**, 1671–1678.

[35] Van Dooren T. 1999 The evolutionary ecology of dominance-recessivity. *Journal of Theoretical Biology* **198**, 519–532.

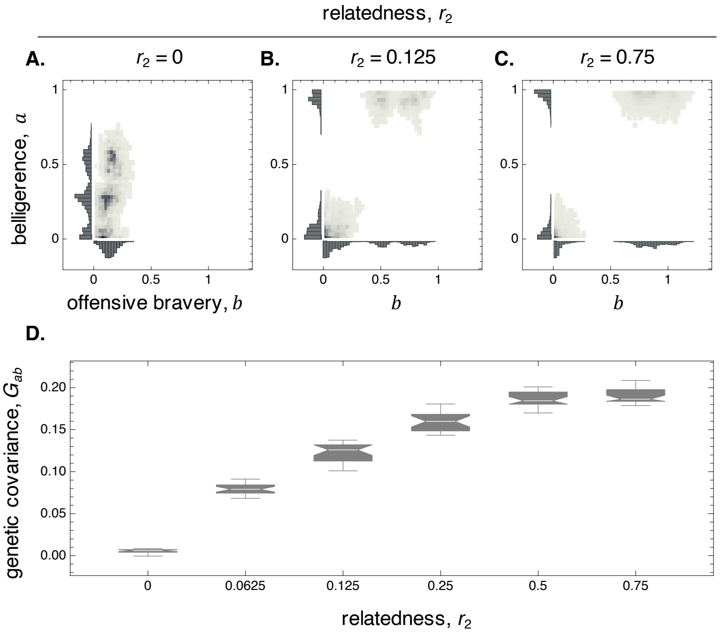
[36] Durinx M, Van Dooren T. 2009 Assortative mate choice and dominance modification: Alternative ways of removing heterozygote disadvantage. *Evolution* **63**, 334–352.



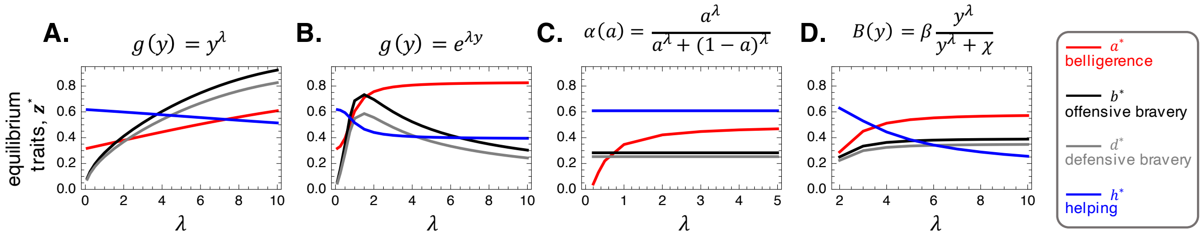
**Supplementary Figure 1: (Scaled) Relatedness, with and without replacement.** A. Relatedness $r_2$ (eq. S12), in the island model when a single reproductive spot is replaced in each generation; B. Scaled relatedness $\kappa$ (eq. 7); C. Scaled relatedness with replacement $\kappa^{\mathrm{R}}$ (eq. 7) as a function of dispersal $m$ and group size $n$ (legend for values).



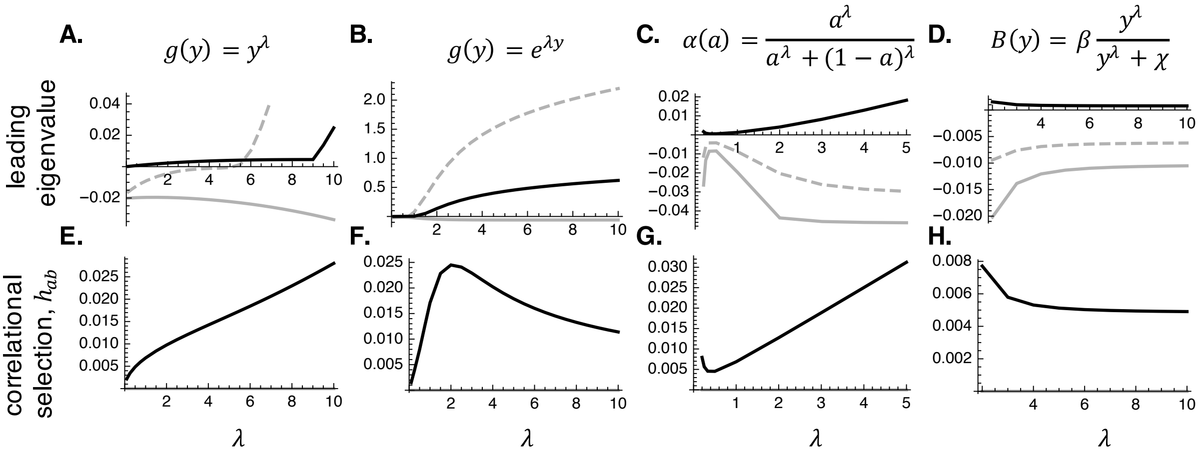
**Supplementary Figure 2: The evolution of helping.** **A**: The average level of helping as a function of demographic time $t$ (dots: observed in simulations, section S4 for details; full line: analytical predictions from eq. (S9) with variance-covariance matrix $\mathbf{G}$ composed of all zeroes except $G_{hh} = 0.00375$, chosen heuristically) when the other traits are absent in the population, i.e., $a = d = b = 0$ for all individuals throughout; with $B(Nh) = \beta\sqrt{Nh}$, $C(h, b, d) = h$, $\beta\sqrt{N} = 100$, $\upsilon_{\mathrm{b}} = 0$, $N = 8$, $m = 0.476$. **B**: Individual values of helping observed in a simulation (shown for 5 individuals randomly sampled every 800 time points). **C**: Distribution of helping in a simulated population (calculated from time 100’000 for 50’000 time steps).



**Supplementary Figure 3: The effect of limited dispersal and relatedness on polymorphism.** **A-C** Joint distribution of belligerence and offensive bravery in simulated populations at equilibrium for different levels of relatedness (found by fixing $N = 8$ and varying $m$ in $r_2$ eq. (S12)) with: A. $r_2 = 0$ (so $m = 1$); B. $r_2 = 0.125$ (so $m = 0.467$); $r_2 = 0.75$ (so $m = 0.04$) (other parameters, same as Fig. 1 middle; joint distribution calculated over 100'000 time points after 150'000 of evolution). **D.** Distribution of covariance between belligerence and offensive bravery over time at equilibrium according to relatedness within groups (same as A other than $m$).



**Supplementary Figure 4: The effect of changing functional relationships on equilibrium.** Equilibrium value of each trait ($a$ in red, $b$ in black, $d$ in gray, $h$ in blue) against $\lambda$ that is used as the parameter in the different following function: **A.** $g(y) = y^\lambda$; **B.** $g(y) = \exp(\lambda y)$; **C.** $\alpha(a_0) = a_0^\lambda/[a_0^\lambda + (1-a_0)^\lambda]$; **D.** $B(Nh) = \beta(Nh)^\lambda/((Nh)^\lambda + \chi)$ (with $\beta = 100$ and $\chi = 10$, varying $\lambda$ for the steepness of the sigmoid; unless otherwise stated: $\alpha(a) = a$, $g(y) = y$, and $B(Nh) = \beta\sqrt{Nh}$; Other parameters: $C(h,b,d) = h^2 + c_\text{b} b^2 + c_\text{d} d^2$, $c_1 = 18$, $c_2 = 115$, $c_\text{d} = 1$, $c_\text{b} = 0.8$, $\omega = 0.5$, $\beta\sqrt{N} = 100$, $v_\text{b} = 0$, $N = 8$, $m = 0.467$ so that $r_2 = 0.125$). All computed from solving the selection gradients numerically for singular values (see Mathematica Notebook). All these equilibria strategies are at least weakly convergence stable (see text below eq. S14) but not locally stable due to correlational selection among belligerence and offensive bravery (Suppl. Fig. 5). This indicates that they are attractors under directional selection but that once the population expresses these traits on average, selection becomes disruptive, favoring an association between belligerence and offensive bravery.



**Supplementary Figure 5: Summary analysis of the effect of contest types and group decisions. A-D** Leading eigenvalues of the Jacobian (in dashed gray), symmetric part of the Jacobian (in full gray), and Hessian (in black), showing that the singular values plotted in Fig. 4 are all (at least weakly) convergence stable but none is locally stable. This suggests that polymorphism also emerges under these different assumptions (Fig. 4 for details on parameters values). **E-H** Correlational selection among belligerence and offensive bravery at the singular strategy. It is always positive, highlighting that selection still favours a positive selection.