

# The Immuno-Dynamics of Conflict Intervention in Social Systems

David C. Krakauer<sup>1\*</sup>, Karen Page<sup>2</sup>, Jessica Flack<sup>1,3</sup>

**1** Santa Fe Institute, Santa Fe, New Mexico, United States of America, **2** Department of Mathematics, University College London, London, United Kingdom, **3** Yerkes National Primate Research Center, Atlanta, Georgia, United States of America

## Abstract

We present statistical evidence and dynamical models for the management of conflict and a division of labor (task specialization) in a primate society. Two broad intervention strategy classes are observed— a dyadic strategy – pacifying interventions, and a triadic strategy –policing interventions. These strategies, their respective degrees of specialization, and their consequences for conflict dynamics can be captured through empirically-grounded mathematical models inspired by immuno-dynamics. The spread of aggression, analogous to the proliferation of pathogens, is an epidemiological problem. We show analytically and computationally that policing is an *efficient* strategy as it requires only a small proportion of a population to police to reduce conflict contagion. Policing, but not pacifying, is capable of effectively eliminating conflict. These results suggest that despite implementation differences there might be universal features of conflict management mechanisms for reducing contagion-like dynamics that apply across biological and social levels. Our analyses further suggest that it can be profitable to conceive of conflict management strategies at the behavioral level as mechanisms of social immunity.

**Citation:** Krakauer DC, Page K, Flack J (2011) The Immuno-Dynamics of Conflict Intervention in Social Systems. PLoS ONE 6(8): e22709. doi:10.1371/journal.pone.0022709

**Editor:** Matjaz Perc, University of Maribor, Slovenia

**Received:** April 3, 2011; **Accepted:** June 28, 2011; **Published:** August 24, 2011

**Copyright:** © 2011 Krakauer et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** DK and JF acknowledge support from National Science Foundation grant 0904863 on network conflict theory. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. KP thanks the National Institute for Mathematical and Biological Synthesis for a sabbatical fellowship during which part of the work was completed.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: krakauer@santafe.edu

## Introduction

In large societies of individuals or cells, sophisticated regulatory mechanisms are required to control conflict and promote coordination [1–7]. These conflict management mechanisms can involve specialization through to a full division of labor [8,9], or the creation of social norms that reinforce roles and behavioral patterns [10]. The organism has a dedicated suite of conflict management mechanisms, including an immune system, that regulates cell-cell and cell-pathogen interactions [11–13]. Beyond research on policing in social insects [14,15] and punishment in human societies (e.g. [15–17]), relatively little is known about conflict management dynamics in animal societies. The common presumption that organismal conflicts of interest promote a tragedy of the commons has generated low expectations for group-level regulation.

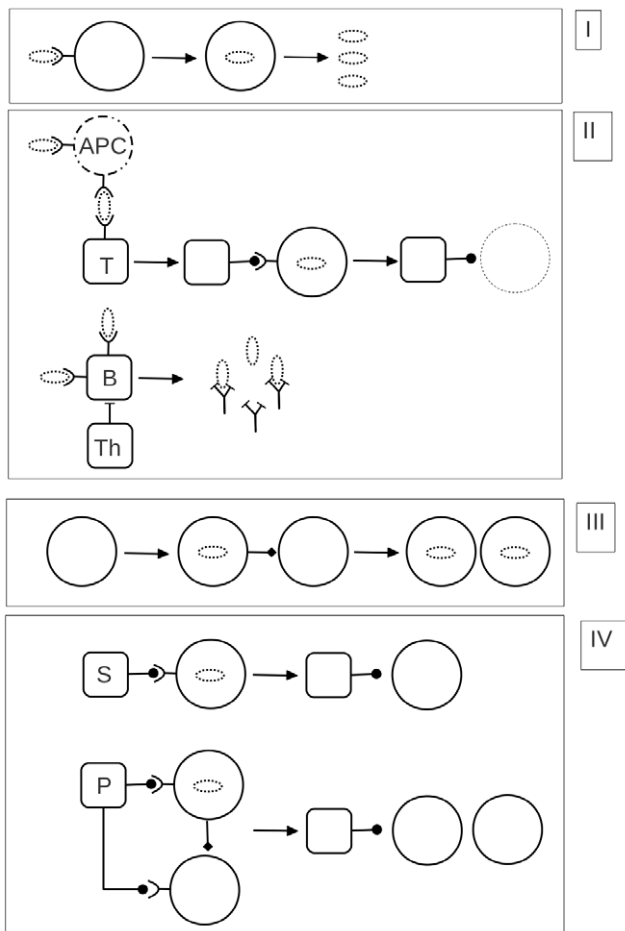
Nonetheless, conflicts in animal societies lead to fights, and these do not typically overwhelm the group or result in severe injury or death as there are a variety of mechanisms animals use for resolving disputes or mitigating the effects of aggression [2]. In some societies conflicts are managed by other individuals in the group [2,6]. One management strategy third-parties use is to calm agitated individuals (see box 13.1 in [18]), forestalling aggression. These interventions are called *pacifying* interventions. A second strategy is an impartial intervention –a *policing* intervention– in which all conflict participants are targeted indiscriminately by a third-party through aggression, or through the implicit threat of aggression inherent in an approach by the third-party [6]. Both

pacifying and policing interventions can cause the fight to terminate and/or aggression to dissipate [6].

In this paper we use an immuno-dynamics modeling approach [19,20] to explore the consequences of third-party conflict management for the contagion of aggression. Our primary goal is to determine whether policing and pacifying interventions have different effects on aggression dynamics and hence different mechanistic benefits. A second, more ambitious goal is to ask why, as with cellular immunity, we observe in social systems triadic as well as dyadic strategies for managing conflict. We use the macaque genus, a model system for social evolution [21,22], to develop the theory, working with a data set describing conflict dynamics collected from a large pigtailed macaque (*Macaca nemestrina*) society housed at the Yerkes National Primate Research Center (Methods 1).

## A multi-scale immuno-dynamics approach

In the case of policing in animal societies, the aggressive dyad or fight-complex, is targeted, eliminated, and resolved into peaceful or “passive” individuals. As illustrated in Fig. 1., the triadic character of policing interventions (P) is roughly comparable to the action of T cells that identify infected cells (a complex of pathogen and host cell) and eliminate them from a population. During a pacifying intervention an aggressive or agitated individual is targeted and resolved into a peaceful state: Pacifiers (S) identify aggressive individuals on route to fight, or while the fight momentarily abates, and through direct interaction, induce them to transform to a passive state (Fig. 1 IV upper schematic).



**Figure 1. Comparison of the structure of cellular and social immunity.** The top two panels illustrate cellular infection and immunity. The bottom two panels illustrate social infectivity and social immunity. (I) Pathogens infect cells and thereby proliferate. (II) Naive T cells are presented with antigen epitopes by antigen presenting cells (APC), inducing them to target infected cells and eliminate them. B cells sequester antigen and are induced to generate antibodies to these by T helper cells. (III) Individuals become aggressive and by direct contact with others redirect aggression infectiously. (IV) Pacifiers (S) engage with aggressive individuals, calming them down thereby preventing redirection. Police (P) intervene directly into conflicts, resolving disputes and returning combatants to the passive state. In animal conflict it is a state of behavior that is transmitted that arises spontaneously from within the population.

doi:10.1371/journal.pone.0022709.g001

Pacifying interventions are comparable to the behavior of antibodies engaging pairwise with pathogens to prevent infection of susceptible cells.

Without an immune response infection threatens to exponentially encompass a population of susceptible cells (Fig. 1 I). The T cell response requires activation of naive T cells by antigen presenting cells. The T cells are then able to recognize “foreign” antigens on the infected cell surface and remove these cells clearing the infection (Fig. 1 II upper schematic). Homeostatic regulation of the cell population restores depleted cells. B cells bind antigen directly, and through interaction with helper T cells, generate antibody to antigen, leading to antigen clearance prior to cellular infection (Fig. 1 II lower schematic). Similarly in the absence of social immunity aggression in an animal society threatens to exponentially encompass a group through behavioral redirection of aggression (Fig. 1 III) [23,24].

Unlike cellular immunity, however, neither policing nor pacifying interventions kills the target. Instead these strategies terminate deleterious behavioral patterns, transforming the target and keeping the population size constant. Neither behavioral strategy requires “fourth” party activation of the conflict management mechanism—hence there is no apparent social immune analog to cellular immune priming. In contrast to cellular immunity, the effectiveness of policing and pacifying interventions is not assured, as effectiveness depends in part on characteristics of the individual performing the intervention, on the character of the conflict itself, and on properties of the conflict time-series [6,24].

Adaptive immune systems possess the property of antigen mediated clonal selection and expansion. In animal societies, in contrast, passive individuals can spontaneously adopt a policing strategy and remain in that state until the number of aggressive individuals decreases to some threshold value. Another important difference is that within the biological immune system, pathogens form a population independent from immune cells, whereas behavioral conflict management strategies are implemented by individuals belonging to the same group as the conflict participants. Hence functional constraints on conflict management due to partially aligned interests among group members are likely more significant in social systems than in the case of pathogen control.

However, in mechanical terms the fundamental pathogenic property is the ability to transform the state of a cell, increase cellular rates of mortality and proliferate through a population. These are also properties of aggression that can be thought of as a transmissible state of behavior. Proliferation of aggression leads to larger fights and an increased probability of individual mortality [23,24]. The contagion property of both pathogens and aggression, coupled to similar management mechanics instantiated in dyadic and triadic interactions, suggests that comparison are warranted despite critical differences in implementation.

In the empirical section of the Results, we further describe the mechanics of policing and pacifying interventions in our pigtailed macaque study system. In the theory section of the Results we develop empirically-grounded immuno-dynamics models to explore the implications of these alternative strategies for containing aggression.

## Results

### Empirical description of pacifying and policing in pigtailed macaques

In previous work [6], it was shown empirically that pigtailed macaques use both policing and pacifying interventions. It was also shown that the frequency distribution of policing interventions is heavy-tailed. The frequency distribution of pacifying interventions on the other hand is normal. This difference suggests that there might be a proto-division of labor for policing interventions but not for pacifying interventions. Here we determine whether this is the case.

We define a proto-division of labor as specialization on a group beneficial task by a subset of components in the absence of complementary specialization by a second subset. A full division of labor minimally involves two subsets, each specializing on complementary tasks [25]. Role specialization, defined here as either individual, age-sex class, or other subgroup specific strategy sets, is a foundational assumption of game dynamics yet is rarely empirically evaluated outside the study of cellular immunity or social insect societies. Role specialization on strategy sets can be operationalized statistically.

An individual is said to specialize on *conflict management generally* if its policing (POL) frequency is  $> \mu_{POL} + \sigma_{POL}$  AND pacifying

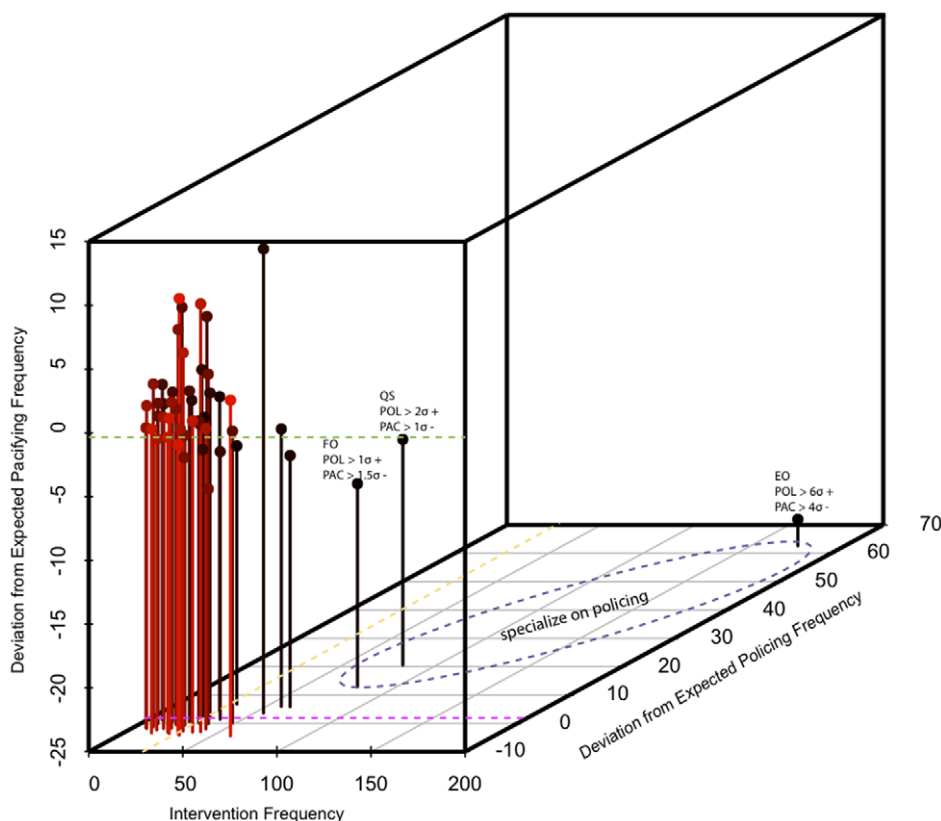
(PAC) frequency is  $> \mu_{PAC} + \sigma_{PAC}$ , where  $\mu$  is the sample mean and  $\sigma$  denotes one standard deviation about the mean. *Policing specialization* is  $POL > \mu_{POL} + \sigma_{POL}$  and  $PAC < \mu_{PAC} - \sigma_{PAC}$ . *Pacifying specialization* is  $POL < \mu_{POL} - \sigma_{POL}$  and  $PAC > \mu_{PAC} + \sigma_{PAC}$  (unless otherwise noted, all empirical analyses in this paper use corrected frequency data, in which an individual's frequency is equal to a measured deviation from an expected score, see Methods 1). Note that whereas the distribution of pacifying interventions is normal, the distribution of policing interventions is roughly heavy-tailed. Given that we are interested in evidence for concerted deviations from the average behavior of *all* individuals in the group, including those in the tail, we operationalize specialization with respect to the standard deviation of the policing and pacifying distributions regardless of the form of those distributions (e.g. normal, heavy tailed, etc).

Considering the 48 socially mature animals in the group, we find no evidence for general conflict management specialization and no evidence for pacifying specialization. Only three adult males, EO, QS, FO, exhibit policing specialization Fig. 2. These three individuals account for 39% of the 477 policing interventions but only 10% of the 304 pacifying interventions. The remaining 45 socially-mature group members show no preference for either policing or pacifying, performing as many policing as pacifying interventions (deviation from expected frequencies (DEF), policing data are non-normal,  $n = 45$ , Wilcoxon signed rank test,  $V = 460$ ,

$p = .52$ ). Elsewhere we have shown that the distribution of social power (degree of consensus among group members that an individual can use force successfully during fights), by modulating the cost of social interaction, influences individual strategy choice [6,26,27]. In our study group the power distribution is not significantly different than lognormal (log-transformed  $\ln(x+1)$ ) data, Lilliefors KS test,  $n = 48$ ,  $D = 0.123$ ,  $p = .08$ ). The three individuals specializing on policing (henceforth, "the policers") occupy the tail of this power distribution and respectively have 24.29, 8.85 and 7.82 times more power than the average individual among the remaining 45.

Approaching conflicts, required for both pacifying and policing interventions, is costly because conflict participants frequently redirect aggression to interveners [6,28]. The cost of pacifying and policing interventions can be measured as aggression or threat received in response to intervention (Methods 1). We find that cost decreases with increasing power for policing but not pacifying interventions (data include policers and the 42 of 45 non-policers who perform both pacifying and policing interventions, DEF POL: data nonnormal,  $n = 45$ , Kendall's Rank Correlation:  $Tau = -0.26$ ,  $p = 0.01$ , DEF PAC: data nonnormal,  $n = 45$ ,  $Tau = -0.08$ ,  $p = .48$ ).

We find no significant difference between cost of pacifying and cost of policing for the 42 non-policers (deviation from expected cost (DEC), data are non-normal,  $n = 42$ , Wilcoxon signed rank test,



**Figure 2. Specialization on policing produces a proto-division of labor in a macaque society.** Only three individuals, EO, QS, and FO, enclosed by the dashed purple ellipse, perform policing at a minimum of one standard deviation greater than the mean deviation from expected frequencies for the population (DEF; see text for definition) and pacifying at one standard deviation less than the mean DEF. EO's policing frequency is six standard deviations greater than the mean DEF. The dashed green line indicates the mean deviation from expected pacifying frequency ( $n = 48$ ). The dashed fuchsia line indicates the mean deviation from expected policing frequency ( $n = 48$ ). The dashed yellow line indicates the mean frequency of intervention performed by the 48 socially-mature individuals. The bar color is graded from red to black to make individual differences easier to see. The patchy distribution of the state space supports the interpretation of behavior in terms of statistically-defined strategy classes. doi:10.1371/journal.pone.0022709.g002

$V = 375, p = 0.35$ ). The cost paid by policers for policing is two orders of magnitude less than the average deviation from expected cost paid by the other 42 animals (DEC:  $n = 45, m = -1.36, sd = 5.11, EO = -30.30, QS = -6.59, FO = -6.691$ , remaining 42 individuals:  $m = -0.32, sd = 2.25$ ). EO was not observed to receive aggression in response to any of his 104 policing interventions.

The policers pay a slightly higher cost than expected for pacifying (DEC:  $n = 45, m = 0.097, sd = 1.63, EO = -1.34, QS = 0.04, FO = 2.32$ , remaining 42 individuals,  $m = 0.06, sd = 1.66$ ), but this cost is still negligible as they receive threats in response to fewer than 5% of their pacifying interventions and never receive contact aggression.

We find that the effectiveness—the ability to terminate a fight or reduce the severity of aggression (Methods 1)—of policing, but not pacifying, increases with increasing power (data include the three policers and the 42 nonpolicers who perform both intervention types, DEF POL: data nonnormal,  $n = 45$ , Kendall’s Rank Correlation:  $Tau = 0.34, p = 0.001$ , DEF PAC: data nonnormal,  $n = 45, Tau = 0.19, p = .07$ ). The policers are more than four orders of magnitude more effective at policing than the remaining 42 individuals (DEF: all 45 individuals,  $m = 1.9, sd = 9.61, EO = 61.64, QS = 15.07, FO = 8.90$ , remaining 42 individuals,  $m = .0005, sd = 1.67$ ).

The policers are 11 times more effective than expected at breaking up or reducing the intensity of fights using policing interventions than they are when using pacifying interventions (DEF POL:  $EO = 61.64, QS = 15.07, FO = 8.90$ , DEF PAC:  $EO = 4.30, QS = 1.50, FO = 2.15$ ). The remaining 42 socially-mature group members perform as many effective policing as effective pacifying interventions (DEF, data are non-normal,  $n = 42$ , Wilcoxon signed rank test,  $D = 474, p = .78$ ).

To summarize, the data indicate that a small subset of the group performs policing, everyone engages in pacifying, and policing is better than pacifying at controlling the escalation of aggression when the policers are powerful. The effectiveness and cost of policing appear to depend on relative power in a heavy-tailed power distribution. Hence a power-based *state dependence* supports specialization on policing but does not influence pacifying. We find no evidence suggesting that pacifying is state dependent.

### Social immuno-dynamics modeling results

We have explored elsewhere why a high variance distribution of power is required to support policing [6,26]. We and others have also considered how conflict management mechanisms such as policing evolve [3,5,15,26]. We seek an ontogenetic explanation for how policing, when performed by few individuals, can effectively control conflict or reduce its frequency in social groups. And, why pacifying strategies, performed by many individuals, are less effective. If we are able to reproduce these conflict management patterns, we shall have succeeded in accounting for the division of labor in policing, and the widespread, undifferentiated, adoption of pacifying. This will provide the beginnings of an account for diverse forms of conflict management at the social level, and the grounds for a more informed comparison with the control of contagion among populations of cells.

We develop two classes of models—one for pacifying and one for policing. We explore how the degree of specialization on a management strategy influences conflict dynamics (for a review of this approach [19,20]). Degree of specialization in these models is operationalized as the proportion of individuals in the group performing conflict management.

In the pacification model, we assume a population of passive individuals  $x$  that spontaneously become aggressive  $y$  at a rate  $f_1 x$  (Methods 2). Once in an aggressive state, these are capable of

“infecting” further individuals through social contagion inducing them to become aggressive. From the resulting aggressive dyad or complex  $D$ , emerge two aggressive individuals. Monitoring is performed by a population of individuals – conflict managers  $z$  – who identify aggressive states prior to the formation of the complex and form an intervention dyad  $D_z$ . This resolves into a single manager and pacified individual. In this model, conflict managers have no influence over conflicts that have already begun. The initial conditions corresponding to the start of observations of behavior are,  $x(0) = x_0, y(0) = y_0, z(0) = z_0$  and  $D(0) = D_z = 0$ .

In Figure 3B, we illustrate the steady state frequencies of the each of the state variables as a function of the proportion of individuals in the population assuming a pacifying, conflict management role. We find that the total number of fights declines monotonically with increased pacifying as do the number of aggressive individuals. For a large decrease in aggression, there needs to be a concomitant large increase in the proportion of individuals in the population assuming the pacifier role.

Given that pacifiers exist in only two states, and their total number is assumed to be conserved, we make the observation that,

$$\frac{dz}{dt} + \frac{d(D_z)}{dt} = 0 \Rightarrow z(t) + D_z(t) = z_0 \tag{1}$$

This allows us to express the state variable  $D_z$  in terms of  $z$  and the initial number of pacifiers at the start of observations,  $z_0$ . Furthermore, we assume that the total population size remains constant,

$$x(t) + y(t) + z(t) + 2D(t) + 2D_z = K. \tag{2}$$

This allows us to express the fight complex  $D$  in terms of  $x, y$  and  $z$  and two constants:

$$D(t) = \frac{1}{2}(K - 2z_0 - x(t) - y(t) + z(t)) \tag{3}$$

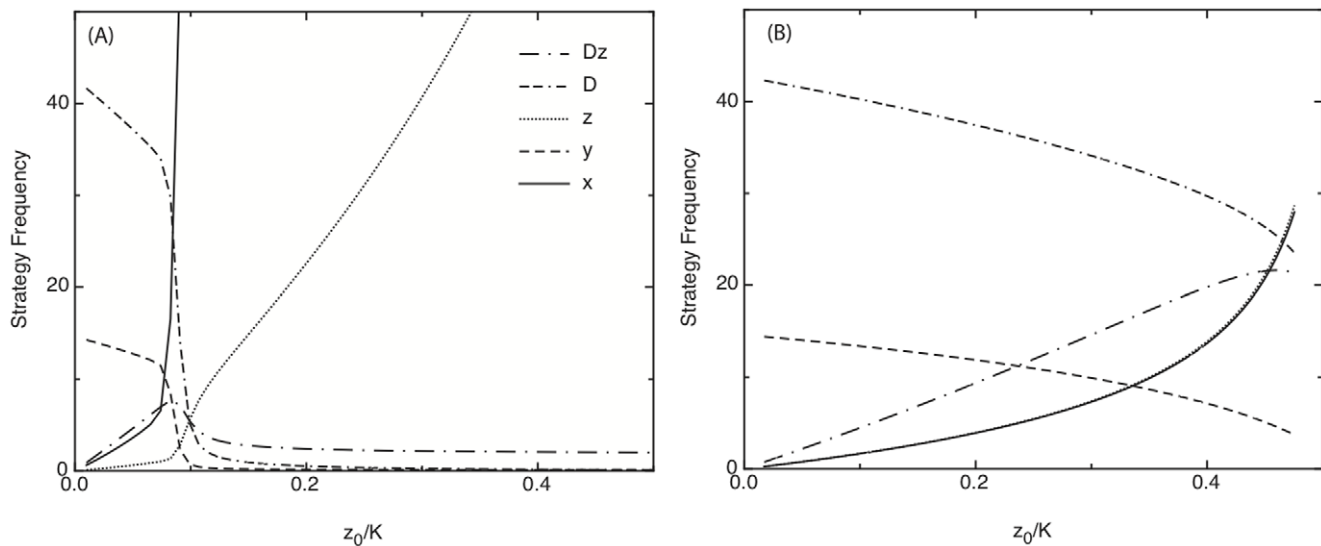
We minimize the number of parameters by assuming that there are three time scales in the dynamics. A time scale at which fights are initiated and monitored, a second time scale at which fights are resolved, and a very slow time scale at which aggression emerges. Thus  $f_1 = c, f_3 = f_5 = f$  and  $f_2 = f_4 = s$ . These assumptions allow us to write down a 3-dimensional dynamical system (rather than 5-dimensional), that describes the conflict management dynamics.

$$\dot{x} = -cx - sxy + f(z_0 - z) \tag{4}$$

$$\dot{y} = cx - sy(x + 2y + z) + f(K - 2z_0 - x - y + z) \tag{5}$$

$$\dot{z} = -szy + f(z_0 - z) \tag{6}$$

As per our assumptions:  $c \ll s, f$ , and furthermore, that the population size is significantly greater than the number of police,  $K > 2z_0$ . We find that there are two steady states. The first steady state is at  $z \approx z_0, y \approx 0, x \approx K - z_0, D_z \approx 0 \approx D$ . This state is always unstable. The other steady state is at  $y + 2D \approx K - 2z_0, D_z = z_0 - z$  and  $x \approx z \approx \frac{fz_0}{f + sy}$ , where  $y$  is approximately the unique positive solution to  $2s^2y^3 + 3fsy^2 + f(f + 4sz_0 - sK)y - f^2(K - 2z_0) = 0$ .



**Figure 3. Steady state frequencies of aggressive individuals  $y$ , passive individuals  $x$ , fighting pairs,  $D$ , and police  $z$  as a function of the proportion of policing individuals  $z(0)/K$ .** Assuming policing interventions (A) we observe a threshold value of police above which conflict is effectively eradicated. Under pacifying interventions (B), conflict declines monotonically in the proportion of pacifiers. For these figures we have used parameter values:  $f = 0.5$ ;  $s = 0.01$ ;  $k = 110$ . The parametric sensitivity of these results are indicated analytically in the text. doi:10.1371/journal.pone.0022709.g003

Thus if there are few police,  $z_0 \ll K$ , the population is made up largely of aggressors ( $y$ ) and “fighters” ( $2D$ ). For fixed values of the remaining parameters, the steady state values of  $y$  and  $D$  are monotonically decreasing in  $z_0$ .

In the policing model, the conflict managers, called pacifiers  $z$ , do not target aggressives  $y$  before they engage in conflict as in the antibody model, but resolve disputes directly, by intervening and eliminating fights between pairs forming in this case a triadic variable  $D_z$  composed of both the complex  $D$  and the pacifier  $z$ . As before, the conflict managers exist in two states, and their total number is conserved:

$$\frac{dz}{dt} + \frac{d(D_z)}{dt} = 0 \Rightarrow z(t) + D_z(t) = z_0 \quad (7)$$

The total population size is constant,

$$x(t) + y(t) + z(t) + 2D(t) + 3D_z(t) = K. \quad (8)$$

Thus

$$D = \frac{1}{2}(K - 3z_0 - x - y + 2z). \quad (9)$$

We assume three timescales, with  $c \ll s, f$  and  $K > 2z_0$ . In addition, we assume  $cK \ll f^2/s$ .

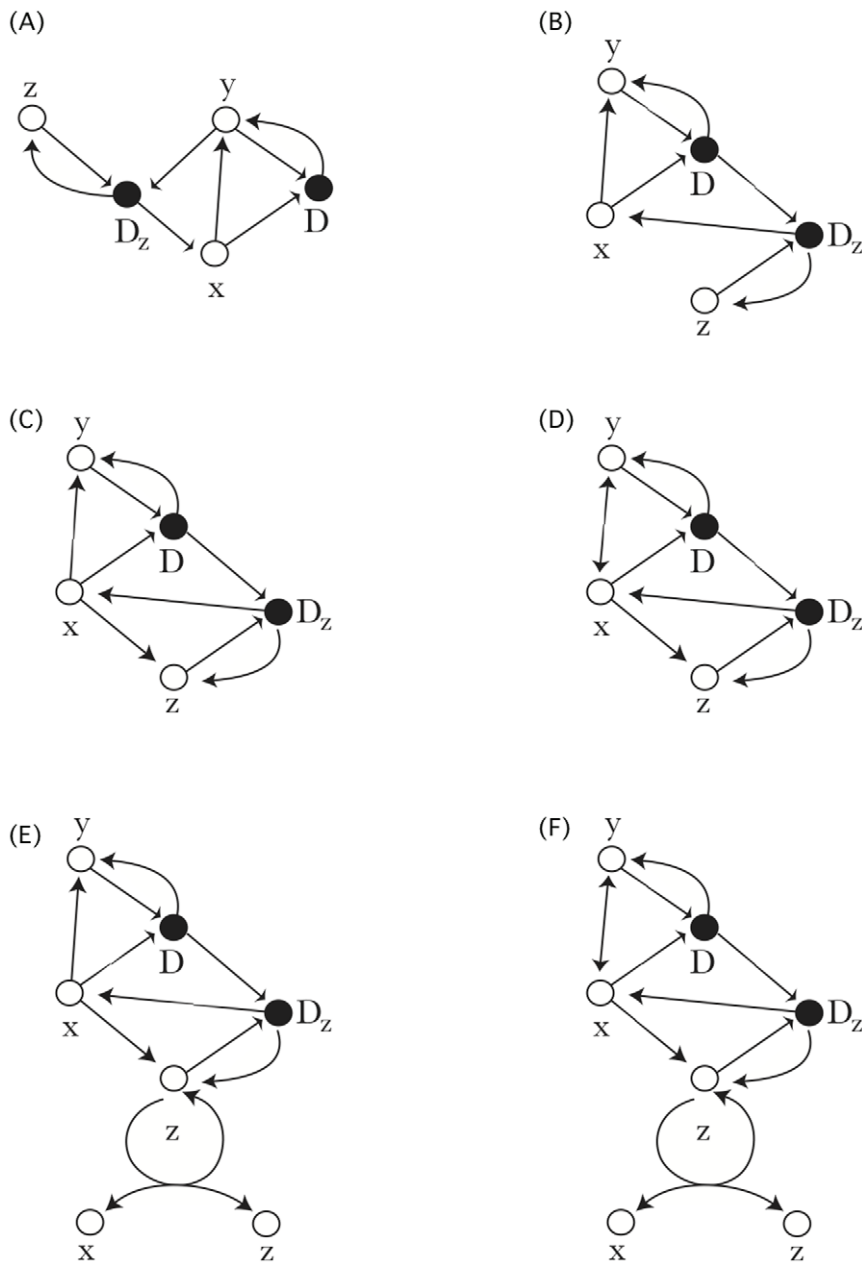
$$\dot{x} = -cx - sxy + 2f(z_0 - z) \quad (10)$$

$$\dot{y} = cx - sy(x + 2y) + f(K - 3z_0 - x - y + 2z) \quad (11)$$

$$\dot{z} = -sz \frac{1}{2}(K - 3z_0 - x - y + 2z) + f(z_0 - z) \quad (12)$$

This dynamic has a steady state with  $y \approx 0$ . In this case, however, for critical values of the parameter set, this state is stable. In particular the state has  $z \approx z_0$ ,  $x \approx K - z_0$  and  $y \approx 0 \approx D \approx D_z$  and is stable iff  $z_0$  is greater than approximately  $f/s$ . Thus we have a state in which there are almost no aggressors or fights being stable provided there are initially more than a threshold number of pacifiers. This threshold declines as the duration of fights increases and as the interaction rate (which is both the rate at which fights are initiated and the rate at which police intervene in fights) increases. When these values are high there are few free aggressors. Interestingly the policing threshold appears to be independent of the population size  $K$ , provided that this size is not too large ( $K \ll f^2/(sc)$ ). Thus unlike in the B cell model, the T cell model is capable of leading to stable societies in which there is no unrest with only small rates of conflict management. This modeling finding is consistent with the data from our study group in which 17% of the  $\approx 1100$  conflicts observed received effective policing interventions. This level of policing has been shown empirically in a behavioral knockout experiment to be sufficient to reduce general levels of aggression [23].

We consider a larger family of models expanding on our basic policing and pacifying structures to cover and analyze a richer space of strategic permutations. These are illustrated using conflict reaction graphs in Figure 4 (full mathematical description in Methods). These include cases in which passives can transform into police (spontaneous policing); aggressives can transform into the passive state spontaneously (temporary aggression), and where the pacifiers switch to non-policing when policing is common (conditional policing – negative frequency dependence). The results of all of these models are summarized in Table 1. Each model possesses multiple stationary states (a maximum of three) and we indicate those that are stable. Of greatest social interest are those strategies where multiple equilibria exist. These are models that allow transitions between strategy classes, such as switching from a pacifier to a pacifier. In the case where pacifiers become police spontaneously at a low rate, and where policing interventions can fail inducing the police to become aggressive, this results



**Figure 4. Conflict networks illustrated as reaction graphs.** Open nodes are single individuals. Closed nodes are groups of 2 or more individuals. Directed edges represent transformations but not stoichiometry. Merging input edges into a single closed node corresponds to the formation of a complex (dyad or triad). Cycles represent transformations that yield one starting state and one alternative. The six conflict networks are: (A) B-cell inspired pacification. (B) T-cell inspired policing. (C) Spontaneous policing. (D) Temporary aggression. (E) Conditional Policing. (F) Conditional, Temporary Policing. Hence in (A)  $x$  becomes aggressive and transforms to  $y$ . The aggressive individual  $y$  interacts with  $x$  to form a fight  $D$  which then resolves into aggressive individuals  $y$ . The strategy  $z$  interacts with  $y$  to form the pacifying dyad  $D_z$  which then resolves into  $z$  and  $x$ . In the most complicated example (F),  $x$  becomes aggressive and transforms to  $y$ , and  $y$  can revert to  $x$ . The aggressive individual  $y$  interacts with  $x$  to form a fight  $D$  which then resolves into aggressive individuals  $y$ . The strategy  $x$  can spontaneously transform into  $z$ . The strategy  $z$  can interact with  $D$  to form the triadic complex  $D_z$ , which resolves into  $x$  and  $z$ . The strategy  $z$  can spontaneously transform into  $x$ .  
doi:10.1371/journal.pone.0022709.g004

in a solution with multiple equilibria. These include a population that engages in chronic violence (only aggressive individuals and aggressive dyads exist), a ‘civil-solution’ in which policing effectively controls aggression and passive individuals dominate, and a ‘police-state’ in which the entire population is driven to become police. The most harmonious case arises when we assume that aggressive individuals can transform spontaneously to passive individuals. In this case, the combination of policing and a

tendency towards peace, generates a population with minimal aggression, dominated by a few police and many passives.

We also consider models that allow for a form of proto-clonal expansion through negative frequency dependence. In the class of strategies that we refer to as *conditional*, police transform into a passive state when the frequency of police in the population is high. Hence when two police meet, one will transform into a passive. This generates two stable equilibria. One, the ‘Utopian



**Table 1.** The composition of multiple, stable equilibria under six models of third party intervention.

Model	Equilibrium 1	Equilibrium 2	Equilibrium 3
1. Pacifying	$x,y,z,D,D_z$	–	–
2. Policing	$x,y,z,D,D_z$	$z,x$	–
3. Spontaneous policing	$y,D$	$x,z$	$z$
4. Temporary aggression	$x$	$Z$	–
5. Conditional police	$x$	$y,D$	–
6. Conditional, temporary	$x,y$	–	–

Models represent variation in pacifying and policing strategies. Models are described in the text and full mathematical details are provided in Methods 2. Those configurations for which there exist stable solutions are marked by the variables present in their locally stable equilibria. Empty columns (marked by a dash) are not stable or do not exist. Hence in the pacifying model (model 1) all strategies are stable and present at equilibrium. In the model of spontaneous policing (model 3) there are three stable equilibria, a violent equilibrium in which there are only fights and aggressive individuals present, a civil equilibrium in where there are passive individuals and police, and a police-state equilibrium in which all individuals become police.

doi:10.1371/journal.pone.0022709.t001

equilibrium<sup>2</sup>, in which the entire population becomes passive. And another “Dystopian equilibrium” in which the population descends into violence.

These more complex models of conflict management illustrate the double-edged sword of policing. If conditions are favorable, then policing can be very effective at reducing or eliminating conflict, with a small population of police. But if policing can fail or policing ceases in response to the presence of others (a form of free-loading), policing can lead to deleterious outcomes in which populations become police states -everyone polices - or violence becomes chronic and ubiquitous.

## Discussion

We have investigated mechanisms that minimize the contagion of deleterious conflict in a social system. We described the mechanics of two fundamental classes of conflict management mechanism empirically, using data collected from a primate society model system, and mathematically, using dynamical models inspired by the structure of immune systems. We observe a dyadic (two-way) class in which individuals preempt aggression thereby preventing the contagion of aggression (pacifying), and a triadic class (three way) in which individuals directly manage ongoing conflicts (policing), minimizing the redirection and propagation of aggression.

We observe empirically that there is no individual specialization for pacification behavior. Pacification is performed by many individuals. In our model simulations and analysis aggression declines monotonically with an increasing frequency of pacification. We observe empirically that there is however specialization by a few individuals for policing interventions. In our simulation and analytical results policing but not pacifying is capable of almost completely eliminating conflict, and does so above a critical-threshold proportion of police. These results are consistent with a previous experimental study in which it was found that pacifying interventions alone were not sufficient to maintain low levels of aggression [23].

When policing interactions can fail – aggravating aggression and inducing non-policers to switch to policing in compensation – multiple stable equilibria are observed (Model 3 Table 1).

The population can either occupy a highly aggressive state, the population of policers can grow to take over the population in order to control aggression, or policing and passive individuals can coexist at comparable numbers. The solutions are determined by the balance between spontaneous policing and the corruption of police following failed interventions. Our empirical data suggest that policing is a *relative* state dependent strategy, in so far as it appears to require a high variance in the distribution of power, or some other analog measure of resource disparity, to arise. Hence the second and third of the equilibrium states of this model are unlikely to be realized as these states assume that many individuals assume a policing role - albeit somewhat ineffectively. The first equilibrium state –a transition to high aggression when policing fails– is consistent with experimental findings showing that when policing is disabled, aggression increases, and hence that effective policing is critical to preventing social destabilization in systems in which policing is the primary conflict management mechanism [7,23].

Finally, we have shown analytically that policing (T cell strategy) is the more *efficient* strategy, as conflict can be eliminated with only a small proportion of policers. We find no such threshold for pacifying interventions, nor is there any empirical support suggesting such a threshold. The advantage of pacifying interventions is that they are not dependent on power but can be used by anyone, and thus might provide a first line of defense against contagion of aggression in the absence of complex social structure. More generally, our results here and in previous work [6] suggest that social structure is a constraint on conflict management, with different distributions of power, for example, favoring different conflict management mechanisms.

## Comparison to Cellular Immunity

The immuno-dynamics approach and its results reveal critical similarities between social and cellular immunity. Both can be thought of as evolved mechanisms for minimizing the costs of contagion. Similarities include: (1) Dedicated agents adapted to preventing propagation of deleterious or dangerous states. (2) An ability to recognize, engage and clear deleterious factors. (3) A pairwise mechanism and a triadic mechanism. (4) Role differentiation for some aspects of conflict control. (5) Thresholds in the response associated with clearance of danger.

Critical differences include: (1) Non-destructive interactions in the social mechanisms and destructive mechanisms in the cellular mechanisms. (2) Direct activation in the social case and fourth-party mediated activation in the cellular case. (3) Selective proliferation of distinct cell-types in the cellular case, and a steady state response or generic proliferation in the social case. (4) An ability to assume both pacifying and policing roles in the social case, but with only a subpopulation effective in the policing role. (5) Differentiation into management roles in primate social systems acquired through a comparatively fast learning mechanism rather than genetic mechanisms such as somatic hypermutation.

In much the same way that immuno-dynamics grew out of the application of epidemic models to the cells of a single individual, and prion dynamics grew from immuno-dynamics in the absence of pathogens assuming only protein mis-folding, here we have considered the immuno-dynamics of a state of behavior without an extrinsic pathogen-like entity. As with prions in which misfolded proteins induce further misfolding, aggressive individuals induce further aggression. The key analogy across all of these systems is our ability to describe them using dynamical systems that take account of contagion and evolved mechanisms of mitigation.

## Evolutionary-ecological Issues

In emphasizing mechanisms for the containment of aggression within a single generation, the approach we have adopted is more akin to ecological models of microbial infection and clearance than to evolutionary models in social evolution of punishment (e.g. [15–17]) and policing (e.g. [3,5,15]). The goals of models of punishment and policing are to identify optimal or stable parameter values that through fixed payoffs facilitate cooperative evolution. Punishment is typically defined functionally, with the term punishment applied when an individual, at a cost to itself, inflicts a cost on another individual for failing to cooperate. Policing is typically defined as the repression of competition, and sometimes mechanistically as impartial or indiscriminate conflict intervention that can lead to the repression of competition, as in this paper. Much work on punishment assumes policing has the same basic payoff structure to punishment and hence is a subclass of punishment.

One advantage of studying ecological effects is that doing so can reveal a complex strategy space lurking behind functional assumptions. It is well known in the empirical community that there are multiple behavioral strategies for managing conflict [2,6,18,21]. Neither impartial intervention (policing) nor pacifying – two of these strategies – falls easily under the supposed catch-all cost-benefit definition of punishment. Policing and pacifying vary in cost to performer and in cost to target. In both cases, the direct cost (e.g. aggression received during the intervention) to target and performer can be close to zero, and the average cost of policing paid by powerful individuals is nearly zero [6], which violates common assumptions in punishment models. Policing and pacifying also vary on three rarely considered factors relating to benefits. They vary in their effectiveness at controlling the proliferation of aggression and reducing the frequency with which conflicts are expressed as fights. They also vary in terms of the demands made on social structure, with policing requiring rather special resource distributions or power structures to be accessible, regardless of whether individuals acquire them through learning or genetic inheritance. They are also likely to vary in indirect effects that operate over longer timescales, but little is known about this.

The theory literature on punishment makes a distinction between *peer punishment* and *pool punishment* [29–31]. In peer punishment individuals impose a fine on defectors at a cost to themselves. In pool punishment individuals contribute, prior to the joint effort, resources to a pool that funds defector control mechanisms like punishment. A “police force” paid for by taxes is an example. Hence pool punishment would appear to be distinguished from peer punishment by two factors: the assignment of conflict management roles to specific individuals or subgroups and a tax on the population to support these roles.

Our data, however, suggest that a tax levied specifically to support the role division is not necessary. To understand why this is the case, first consider how the policing role is assigned in our study system. The high variance power structure that supports policing emerges from status signaling network in which individuals give subordination signals to others they perceive to be more capable of using force [26]. The decision to signal is the outcome of an agonistic interaction history between the signal receiver and sender in which the sender has learned it is likely loose with that particular receiver. Conceding to the subordinate role by signaling is costly, but it is *less costly* than not signaling at all when an asymmetry in fighting ability is apparent [32]. It also has benefits as pairs with subordination contracts show increased socio-positive interactions over those without subordination contracts [32,33]. Hence subordination signal exchange is in the interest of sender as well as receiver. As such the signaling

dynamics are *cost-free* [34] as long as the subordination contract can be reversed if the underlying asymmetry in fighting ability is reversed, or terminated if the underlying asymmetry shrinks [27,32].

Receivers by tracking the total number of signals they receive as well as how much agreement there is the number of signals sent by each sender can estimate how much power group members perceive them to have [6], which in turn tells them about the cost they will pay for intervening and engaging in social interactions more generally [26]. In many respects the signaling dynamics underlying the emergence of policing are like voting dynamics [32], as group members are, by virtue of how they distribute their signals, effectively determining whether there will be a “police force” and who should be on it. The policing role is in essence assigned to individuals through this voting scheme. This means that a police force can arise naturally if there already are pre-existing underlying heterogeneities in state (in our case, fighting ability or resource holding potential) that support signaling patterns leading to a heavy-tailed distribution of power [6]. No additional taxation is required beyond that which weaker individuals pay to maintain subordination contracts.

Looking forward, evolutionary models focusing on the function of policing will ideally derive optimum parameter values for intervention and switching based on empirically observed strategies. In this way frequency-dependent decisions involving policing will build upon a demonstrated density dependent dynamics of contagion.

## Future Work

The creation of an immuno-dynamic theory of conflict raises many issues for social systems. In much the way that there are optimal schedules for delivering drugs [35] that minimize opportunities for the evolution of resistance, are there schedules of intervention behavior that reduce co-evolutionary-escalation? Analyses quantifying strategic periodicities in conflict dynamics show that in our study group policing occurs on the hour timescale [24]. This suggests that the concept of intervention schedules in social systems is not farfetched. Often disease is not caused by infection but by an over-reactive immune system - immuno-pathology [36,37]. This is reminiscent of the response of states to terrorism in which the principal damage is achieved by a response incommensurate with the magnitude of the attack. When policing has a high failure rate, this is what our models predict – anarchy or a police state. And there is in the intriguing phenomenon of immune memory, whereby, chronic low level infection might be required to ensure long term resistance to infection [38]. Does society require an analogous chronic conflict of low magnitude to maintain effective responses to rare, high-magnitude assaults? Whereas the analogy to the cellular immune system is only approximate, broad patterns of behavior associated with mechanisms for containing infection are expected to be rather general. By incorporating observations relating to individual variation, and individually-targeted responses, we foresee further parallels with the theory of clonal selection and expansion.

## Methods

### Ethics Statement

All data were collected in compliance with the ethical standards set by the Emory University animal care and welfare committee and IACUC approval (proposal 216-97). was obtained to conduct the study. As this was an observational study, the only change to the daily routine of the animals that was required to collect the



data was that the animals had to be confined to their outdoor housing during each observation period. Water, monkey chow (remaining from morning feeding), enrichment (e.g. toys, climbing structures, etc.) and substantial space were available continuously throughout all observation periods. On very hot or rainy days, observations were terminated and the monkeys were given access to their indoor housing. As part of standard Yerkes management protocol, the animals were routinely subject to medical examination and care.

## Model System

Macaque societies are characterized by social learning at the individual level, social structures that arise from nonlinear processes and feedback to influence individual behavior, frequent non-kin interactions and multiplayer conflicts, the cost and benefits of which can be quantified at the individual and social network levels [7,21–23,39]. These properties coupled to highly resolved data make this system an excellent one for drawing inferences about critical processes in social evolution as well as for developing new modeling approaches that are intended to apply more broadly.

In this study we focus on one species in the genus, the pigtailed macaque (*Macaca nemestrina*). The data set, collected by J.C. Flack, is from a large, captive, breeding group of pigtailed macaques that was housed at the Yerkes National Primate Research Center in Lawrenceville, Georgia. Pigtailed macaques have frequent conflict and employ targeted intervention and repair strategies for managing conflict [23]. The study group had a demographic structure approximating wild populations. Subadult males were regularly removed to mimic emigration occurring in wild populations. The group contained 84 individuals, including 4 adult males, 25 adult females, and 19 subadults (totaling 48 socially-mature individuals used in the analyses). All individuals, except 8 (4 males, 4 females), were either natal to the group or had been in the group since formation. The group was housed in an indoor-outdoor facility, the outdoor compound of which was 125×65 ft.

Pigtailed macaques are indigenous to south East Asia and live in multimale, multifemale societies characterized by female matrilineal and male group transfer upon onset of puberty [40]. Pigtailed macaques breed all year. Females develop swellings when in  $\text{Estrus}$ .

## Data Collection Protocol

During observations all individuals were confined to the outdoor portion of the compound and were visible to the observer. The 156 hours of observations occurred for up to eight hours daily between 1,100 and 2,000 hours over a twenty-week period from June until October 1998 and were evenly distributed over the day. Provisioning occurred before observations, and once during observations. The data were collected over a four-month period during which the group was stable (defined as no reversals in status signaling interactions resulting in a change to an individual's power score [26]).

Conflict and power (subordination signal) data were collected using an all-occurrence sampling procedure [41] in which the compound was repeatedly scanned from left to right for onset of conflict or the occurrence of silent-bared teeth displays. The entire conflict event was then followed and data collected included start time, end time, the identity of individuals involved as aggressors, recipients, or interveners and their behavior.

## Operational Definitions

**Conflict:** any interaction in which one individual threatens or aggresses a second individual. A conflict was considered

terminated if no aggression or withdrawal responses (fleeing, crouching, screaming, running away, submission signals) occurred for two minutes from the last such event. A conflict can involve multiple pairs if pair-wise conflicts result in aggressive interventions by third parties or redirections by at least one conflict participant.

**Intervention:** Third-party to conflict approaches with 3 m and directs aggressive behavior, affiliative behavior, submissive behavior, interposes itself between/equidistant to conflict participants, or approaches in a directed manner, looking at the conflict, but showing no other behavior.

**Pacifying Intervention:** Third-party to conflict approaches within 3 m and directs non-aggressive behavior at one conflict participant within 5 s of conflict. Nonaggressive behavior can include grooming, lip-smacking, puckering, presenting (directing hindquarters at another individual), or emitting a silent bared-teeth display.

**Policing Intervention:** Third-party to conflict approaches within 3 m and impartially threatens all conflict participants or interposes itself approximately equidistant to conflict participants within 5 s of conflict.

**Cost:** highest level of aggression received by an individual from any conflict participants in response to its intervention regardless of whether the conflict participant was the target of the intervention. Aggression varies from facial threat to severe bites (threat = 1 point, lunge or brief chase = 2 points, long chase or slap = 3 points, grapple or wrestle = 4 points, bite less than 5 seconds = 5 points, and bite greater than 5 s = 6 points).

**Effectiveness:** Interventions were considered effective if within 5 seconds of the intervention the entire fight was terminated, meaning that all conflict participants dispersed, or the intensity of aggression used by any of the participants was reduced (and remained reduced for the duration of the fight) without a concomitant increase in aggression or agitation (screaming) by any other participants.

**Power:** Degree of consensus among group members than an individual can use force successfully. Power scores for each individual were calculated using a procedure described in [26]. In brief, the total frequency of peacefully-emitted subordination signals (which reflect perception by the sender that the receiver can successfully use force [32]) received by an individual over a given duration (in this case, the study duration, which was approximately four months) is corrected for the uniformity (measured using Shannon entropy) of its distribution of signals received from its population of potential senders (all socially-mature individuals).

## Calculation of Corrected Frequencies

Raw data for all dependent variables (policing and pacifying frequency, cost, and effectiveness) were processed into deviation from expected frequencies (called observed minus expected scores previously, for full exposition see: [6]. This approach controls for underlying variation in the tendency to intervene.

## Model Details

In the following 6 models we introduce the stoichiometry and dynamical systems describing conflict management introduced in the paper. For the basic antibody model, and T cell model, we only present the stoichiometry since the analysis appears in the paper. For the remaining models, we provide details of stoichiometry, dynamics and dynamical stability. Throughout, parameters are as described in the paper.

### Antibody Model



The initial conditions corresponding to the start of observations of behavior are,  $x(0)=x_0$ ,  $y(0)=y_0$ ,  $z(0)=z_0$  and  $D(0)=D_z=0$ .

### T cell Model

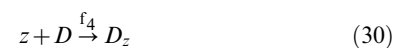
In the T cell model the conflict managers, called policers, do not target aggressives  $y$  before they engage in conflict as in the antibody model, but resolve disputes directly, by intervening and eliminating fights between pairs forming in this case a triadic variable  $D_z$ . The kinetics of T cell conflict management are described by the following scheme:



### Spontaneous Policing

Now instead of assuming a constant population of policers, we assume that passives can become police at some small rate. In

addition, there is a small probability that policing fails to resolve fights in a satisfactory manner. In that case, instead of two passives and a police emerging from the interaction, all participants become aggressive. Thus the number of police is determined by the balance between the spontaneous new policing activity and the corruption of police in failed interventions. The system can be represented by the following “chemical” scheme and differential equations:



Assume,  $f_1 = c$ ,  $f_6 = c'$ ,  $f_2 = s$ ,  $f_4 = s\tilde{O}$ ,  $f_5 = f(1 - q)$ ,  $f_7 = fq$  and  $p + q = 1$ . Here  $p$  is the probability that a policing is successful.

$$\dot{x} = -(c + c')x - sxy + 2fpD_z \tag{33}$$

$$\dot{y} = cx - sy(x + 2y) + 2fD + 3fqD_z \tag{34}$$

$$\dot{z} = c'x - s'zD + fpD_z \tag{35}$$

$$\dot{D} = sxy + sy^2 - s'zD - fD \tag{36}$$

$$\dot{D}_z = s'zD - fD_z. \tag{37}$$

At steady state, we have  $D_z = c'x/(fq)$ ,  $zD = c'x/(s'q)$ ,  $fD = sxy + sy^2 - c'x/q$ . Substituting into the equation for  $x$ , we get  $y = 1/s(2pc'/q - c - c')$ . This is true if  $p > q(c + c')/2c'$ , otherwise  $x \rightarrow 0$  and hence  $D_z \rightarrow 0$  and  $z \rightarrow 0$  and the population consists entirely of  $ys$  and  $Ds$ .

We then obtain  $z=f/[s'(1-3q-cq/c')]$ , for which we require  $1 > 3q+qc/c'$ . If this is the case then, since  $c,c' < < s, y < < 1$  and  $D, D_z < < x$  and hence  $x \approx K-z$ . Otherwise, we get  $z \approx K$ .

Thus there are three regimes: 1) If  $q > 2c'/(3c'+c)$ , then the steady state population consists almost entirely of  $ys$  and  $Ds$ . 2) If  $q < c'/(3c'+c)$ , then the steady state population consists of  $xs$  and  $zs$  with  $z=f/[s'(1-3q-cq/c')]$ . 3) If  $c'/(3c'+c) < q < 2c'/(3c'+c)$  then the steady state populations consists almost entirely of  $zs$ .

Since we assume that  $c' < < c$ , the value of  $q$  must be small in order to avoid conflict.

**Temporary Aggression**

We now study a variant of the previous model in which aggressives become passive at some slow rate. The model can be represented:

$$x \xrightarrow{f_6} z \tag{38}$$

$$x \xrightarrow{f_1} y \tag{39}$$

$$y \xrightarrow{f_8} x \tag{40}$$

$$x+y \xrightarrow{f_2} D \tag{41}$$

$$y+y \xrightarrow{f_2} D \tag{42}$$

$$D \xrightarrow{f_3} 2y \tag{43}$$

$$z+D \xrightarrow{f_4} D_z \tag{44}$$

$$D_z \xrightarrow{f_5} z+2x \tag{45}$$

$$D_z \xrightarrow{f_7} 3y \tag{46}$$

Assume,  $f_1=c, f_6=c', f_2=s, f_4=s\tilde{O}, f_5=f(1-q), f_7=fq, f_8=c''$  and  $p+q=1$ .

$$\dot{x} = -(c+c')x - sxy + 2fpD_z + c''y \tag{47}$$

$$\dot{y} = cx - sy(x+2y) + 2fD + 3fqD_z - c''y \tag{48}$$

$$\dot{z} = c'x - s'zD + fpD_z \tag{49}$$

$$\dot{D} = sxy + sy^2 - s'zD - fD \tag{50}$$

$$\dot{D}_z = s'zD - fD_z. \tag{51}$$

This system has two steady states, namely when the whole population consists of passives and when it consists of police. The all  $x$  state is unstable and the all  $z$  state is stable. Thus in this version of the model for which aggressives spontaneously become passive at some slow rate, and passives spontaneously become police at some slow rate, the population evolves to a state in which almost the whole population consists of police.

**Conditional Policing**

We now remove the assumption that the aggressive population spontaneously become passive and add the assumption that if police encounter other police, one of them stops policing and becomes passive. The rationale is that police only want to incur the cost of policing if they don't believe that another individual will police. This is a form of negative frequency-dependence.

The system can be represented:

$$x \xrightarrow{f_6} z \tag{52}$$

$$x \xrightarrow{f_1} y \tag{53}$$

$$x+y \xrightarrow{f_2} D \tag{54}$$

$$y+y \xrightarrow{f_2} D \tag{55}$$

$$D \xrightarrow{f_3} 2y \tag{56}$$

$$z+D \xrightarrow{f_4} D_z \tag{57}$$

$$D_z \xrightarrow{f_5} z+2x \tag{58}$$

$$D_z \xrightarrow{f_7} 3y \tag{59}$$

$$z+z \xrightarrow{f_9} z+x \tag{60}$$

Assume,  $f_1=c, f_6=c', f_2=s, f_4=s\tilde{O}, f_5=f(1-q), f_7=fq, f_9=f'$  and  $p+q=1$ .

$$\dot{x} = -(c+c')x - sxy + 2fpD_z + f'z^2 \tag{61}$$

$$\dot{y} = cx - sy(x + 2y) + 2fD + 3fqD_z \quad (62)$$

$$\dot{z} = c'x - s'zD + fpD_z - f'z^2 \quad (63)$$

$$\dot{D} = sxy + sy^2 - s'zD - fD \quad (64)$$

$$\dot{D}_z = s'zD - fD_z. \quad (65)$$

This has two steady states, namely all  $x$  and all  $y$  and  $D$ . This is true unless  $q < c'$ , in which case the all  $x$  state is replaced by a state in which  $x, y$  and  $D$  are all present. The all  $x$  state is unstable while the all  $y$  and  $D$  state is stable.

### Conditional, Temporary Policing

In this case aggressive individuals can spontaneously become passive and two police interact to give one police and one passive. The system is represented:

$$x \xrightarrow{f_6} z \quad (66)$$

$$x \xrightarrow{f_1} y \quad (67)$$

$$y \xrightarrow{f_8} x \quad (68)$$

$$x + y \xrightarrow{f_2} D \quad (69)$$

$$y + y \xrightarrow{f_2} D \quad (70)$$

$$D \xrightarrow{f_3} 2y \quad (71)$$

$$z + D \xrightarrow{f_4} D_z \quad (72)$$

$$D_z \xrightarrow{f_5} z + 2x \quad (73)$$

$$D_z \xrightarrow{f_7} 3y \quad (74)$$

$$z + z \xrightarrow{f_9} z + x \quad (75)$$

Assume,  $f_1 = c, f_6 = c', f_2 = s, f_4 = s\tilde{O}, f_5 = f(1 - q), f_7 = fq, f_8 = c'', f_9 = f'$  and  $p + q = 1$ .

$$\dot{x} = -(c + c')x - sxy + 2fpD_z + f'z^2 + c''y \quad (76)$$

$$\dot{y} = cx - sy(x + 2y) + 2fD + 3fqD_z - c''y \quad (77)$$

$$\dot{z} = c'x - s'zD + fpD_z - f'z^2 \quad (78)$$

$$\dot{D} = sxy + sy^2 - s'zD - fD \quad (79)$$

$$\dot{D}_z = s'zD - fD_z. \quad (80)$$

This has a single steady state with

$$x \approx \frac{qfK}{qf + 2pc'} - \frac{2pc'f(q + 2pc')}{qs(fq + 2pc')} \quad (81)$$

$$y \approx \frac{2pc'}{qs} \quad (82)$$

and the rest adopt strategy  $D$ .

This assumes that  $c, c', c'' < f, f'$  but that  $c'/q$  may be reasonably large.

Thus in a society in which the police remain vigilant in the face of policing, the aggressive state is of finite duration, and the steady state consists of a mixture of passive, aggressive, and fighting populations with very few police.

### Acknowledgments

We thank Rustom Antia and Carl Bergstrom for helpful discussion and one anonymous reviewer for his comments.

### Author Contributions

Conceived and designed the experiments: JF. Performed the experiments: JF. Analyzed the data: JF. Wrote the paper: DK, JF, KP. Conceived models: DK, KP, JF. Analyzed models: DK, KP. Performed simulations: DK.

### References

- Buss LW (1987) The evolution of individuality. Princeton University Press., Princeton, NJ.
- de Waal FBM (2000) Primates - a natural heritage of conflict resolution. Science 289: 586–590.
- Frank S (2003) Repression of competition and the evolution of cooperation. Evolution 57: 693–705.
- Michod RE (2003) The group covariance effect and fitness trade-offs during evolutionary transitions in individuality. PNAS 103: 9113–9117.
- Frank S (2009) Evolutionary foundations of cooperation and group cohesion. In Games, Groups and the Global Good. Springer Verlag, In press.
- Flack JC, de Waal FBM, Krakauer DC (2005) Social structure, robustness, and policing cost in a cognitively sophisticated species. American Naturalist 165: E126–E139.
- Flack JC, Girvan M, de Waal FBM, Krakauer DC (2006) Policing stabilizes construction of social niches in primates. Nature 439: 426–429.
- Wahl LW (2002) Evolving the division of labor: generalists, specialists, and task allocation. Journal of Theoretical Biology 250: 135–145.
- Sallusto F, Reiner SL, Lanzavecchia A (2007) Division of labor with a workforce of one: Challenges in specifying effector and memory t cell fate. Science 217: 622–625.

10. Levin S, ed. (2009) *Games, Groups and the Global Good* Springer.
11. Pancer Z, Cooper M (2006) The evolution of adaptive immunity. *Annual Reviews* 24: 497–518.
12. Flajnik M, Du Pasquier L (2004) Evolution of innate and adaptive immunity: can we draw a line? *Trends in immunology* 25(12): 640–644.
13. Medzhitov R, Janeway C (1997) Innate immunity: Minireview the virtues of a nonclonal system of recognition. *Cell* 91: 295–298.
14. Ratnieks FLW, Foster KR, Wenseleers T (2006) Conflict resolution in insect societies. *Annual Review of Entomology* 51: 581–608.
15. El Mouden C, West SA, Gardner A (2010) The enforcement of cooperation by policing. *Evolution* 64: 2139–2152 2010.
16. Boyd R, Gintis H, Bowles S (2010) Coordinated punishment of defectors sustains cooperation and can proliferate when rare. *Science* 328: 617–620.
17. Ohtsuki H, Iwasa Y, Nowak MA (2009) Indirect reciprocity provides only a narrow margin of efficiency for costly punishment. *Nature* 457: 79–82 2009.
18. Aureli F, de Waal FBM, eds (2000) *Natural Conflict Resolution*. University of California Press, Berkeley.
19. Nowak MA, May RA (2000) *Virus dynamics: mathematical principles of immunology and virology*. Oxford University Press. Oxford. 237 p.
20. Perelson A (2002) Modeling viral and immune system dynamics. *Nat Rev Immunol* 2: 28–36.
21. Thierry B, Singh M, Kaumanns W, eds (2004) *Macaque Societies: A Model for the Study of Social Organization*. Cambridge University Press, Cambridge.
22. Flack JC, de Waal FBM (2004) Dominance style, social power, and conflict management: A conceptual framework. In B. Thierry, M. Singh, W. Kaumanns, eds. *Macaque Societies: A Model for the Study of Social Organization*. pp 157–181. Cambridge University Press, Cambridge.
23. Flack JC, Krakauer DC, de Waal FBM (2005) Robustness mechanisms in primate societies: A perturbation study. *Proceedings of the Royal Society of London, Series B* 272: 1091–1099.
24. DeDeo S, Krakauer DC, Flack JC (2011) Evidence of strategic periodicities in collective conflict dynamics. *Proceedings of the Royal Society, Interface*, doi: 10.1098/rsif.2010.0687.
25. Durkheim E (1893) *The Division of Labor in Society* Free Press.
26. Flack JC, Krakauer DC (2006) Encoding power in communication networks. *American Naturalist* 168: 97–102.
27. Boehm C, Flack JC (2010) The emergence of simple and complex power structures through social niche construction. In A. Guinote, ed. *The Social Psychology of Power*.
28. Karzem AJN, Aureli F (2005) Redirection of aggression: multiparty signaling within a network. In P. McGregor, ed. *Animal Communication Networks*, Cambridge University Press, Cambridge.
29. Snolnoki A, Szabo G, Perc M (2011) Phase diagrams for the spatial public goods game with pool punishment. *Phys Rev E* 83: 036101.
30. Sigmund K, De Silva H, Traulsen A, Hauert C (2010) Social learning promotes institutions for governing the commons. *Nature* 466: 861–863.
31. Yamagishi T (1986) The provision of a sanctioning system as a public good. *J Pers Soc Psychol* 51: 110–116.
32. Flack JC, de Waal FBM (2007) Context modulates signal meaning in primate communication. *PNAS* 104: 1581–1586.
33. de Waal FBM (1986) The integration of dominance and social bonding in primates. *Quarterly Review of Biology* 61: 459–479.
34. Lachmann M, Szamado S, Bergstrom C (2001) Cost and conflict in animal signals and human language. *PNAS* 98: 13189–13194.
35. Kirschner D, Lenhart S, Serbin S (1997) Optimal control of the chemotherapy of hiv. *Journal of Mathematical Biology* 35(7): 1432–1416.
36. Wodarz D, Krakauer DC (2000) Defining cdt-induced pathology: implications for hiv. *Virology* 274(1): 94–104.
37. Bergstrom C, Antia R (2005) How do adaptive immune systems control pathogens while avoiding autoimmunity? *Trends in Ecology and Evolution* 21: 22–28.
38. Ahmed R, Gray D (1996) Immunological memory and protective immunity: understanding their relation. *Science* 272(5258): 54–60 1996.
39. Thierry B (2000) Conflict management patterns across macaque species. In F. Aureli, FBM, de Waal, eds. *Natural Conflict Resolution* 106–128. University of California Press, Berkeley.
40. Caldecott JO (1986) *An Ecological and Behavioral Study of the Pigtailed Macaque*, volume 21 of *Contributions to Primatology*. S. Karger, Basel.
41. Altmann S (1974) Observational study of behavior: sampling methods. *Behaviour* 49: 227–267.