

6-18-2018

Hypocrisy and Corruption: How Disparities in Power Shape the Evolution of Social Control

Omar T. Eldakar

Nova Southeastern University, oeldakar@nova.edu

J. Oliver Kammeyer

Emerson College

Nikhil Nagabandi

Nova Southeastern University, nn318@nova.edu

Andrew C. Gallup

SUNY Polytechnic Institute - Utica

Follow this and additional works at: https://nsuworks.nova.edu/cnso_bio_facarticles

 Part of the [Biology Commons](#), and the [Psychology Commons](#)

NSUWorks Citation

Eldakar, Omar T.; J. Oliver Kammeyer; Nikhil Nagabandi; and Andrew C. Gallup. 2018. "Hypocrisy and Corruption: How Disparities in Power Shape the Evolution of Social Control." *Evolutionary Psychology* 16, (2): 1-12. doi:10.1177/1474704918756993.

This Article is brought to you for free and open access by the Department of Biological Sciences at NSUWorks. It has been accepted for inclusion in Biology Faculty Articles by an authorized administrator of NSUWorks. For more information, please contact nsuworks@nova.edu.

Hypocrisy and Corruption: How Disparities in Power Shape the Evolution of Social Control

Omar Tonsi Eldakar¹, J. Oliver Kammeyer², Nikhil Nagabandi¹, and Andrew C. Gallup³

Abstract

Altruism presents an evolutionary paradox, as altruistic individuals are good for the group yet vulnerable to exploitation by selfish individuals. One mechanism that can effectively curtail selfishness within groups is punishment. Here, we show in an evolutionary game-theoretical model that punishment can effectively evolve and maintain high levels of altruism in the population, yet not all punishment strategies were equally virtuous. Unlike typical models of social evolution, we explicitly altered the extent to which individuals vary in their power over others, such that powerful individuals can more readily punish and escape the punishment of others. Two primary findings emerged. Under large power asymmetries, a powerful selfish minority maintained altruism of the masses. In contrast, increased symmetry of power among individuals produced a more egalitarian society held together by altruism and punishment carried out by the collective. These extremes are consistent with the coercive nature of the powerful elites in social insects and egalitarian mechanisms of punishment in humans such as coalitional enforcement and gossip. Our overall findings provide insights into the importance of oversight, the consequences to changes in the power structure of social systems, and the roots of hypocrisy and corruption in human and nonhuman animal societies.

Keywords

altruism, punishment, corruption, multilevel selection, group selection, public good, hypocrisy

Date received: June 1, 2017; Accepted: January 7, 2018

Power tends to corrupt, and absolute power corrupts absolutely.
Great men are almost always bad men.

—John Emerich Edward Dalberg-Acton in letter to Bishop Mandell Creighton (1887, p. 504; Acton, 1907)

The evolution of altruism has remained a popular topic explored by biologists, psychologists, and philosophers alike. If selection favors individuals with traits that confer an advantage over their rivals, then helping others seems paradoxical. This notion, however, is an oversimplified and antiquated view of how natural selection acts on populations (Eldakar & Wilson, 2011). When comparing fitness differences within groups, selfishness (cheating) always beats altruism (helping). However, as selfishness increases in frequency within a group, the average fitness of all group members (including the selfish individuals) in the population decreases compared to individuals within more altruistic groups (e.g., the tragedy of the commons, Hardin, 1968; Rankin, Bargum, & Kokko, 2007). Therefore, if groups vary in the

proportion of altruists, then the differential contribution of groups to the total gene pool can favor altruism despite their selective disadvantage within groups (Eldakar & Wilson, 2011; Wilson, 1975; Wilson & Wilson, 2007).

One way to reduce or eliminate the advantage (or frequency) of selfishness within groups is through punishment. The

¹ Department of Biological Sciences, Nova Southeastern University, Fort Lauderdale, FL, USA

² Department of Writing, Literature, and Publishing, Emerson College, Boston, MA, USA

³ Department of Social and Behavioral Sciences, State University of New York Polytechnic Institute, Utica, NY, USA

Corresponding Author:

Omar Tonsi Eldakar, Department of Biological Sciences, Nova Southeastern University, 3301 College Ave., Fort Lauderdale, FL 33314, USA.
Email: oeldakar@gmail.com



concept of punishment is quite simple; individuals suffer fitness costs for behaving selfishly, thereby eliminating the local selective advantage of selfishness. The evolution of punishment, however, is not as simple. Individuals incur personal costs to inflict punishment on others; therefore, if members of the group benefit from punishment but do not share in the costs, punishment is subject to the same local disadvantage as altruism (Bowles & Gintis, 2002, 2004; Boyd, Gintis, Bowles, & Richerson, 2003; Eldakar, Gallup, & Driscoll, 2013; Eldakar & Wilson, 2008; Fehr, 2004; Panchanathan & Boyd, 2004; Yamagishi, 1986). There are two ways by which punishers can directly overcome this disadvantage within groups.

The first is for punishers to counterbalance punishment costs by themselves behaving selfishly. These *selfish punishers* incur the costs of punishment, yet recoup these losses by hypocritically behaving selfishly in other social interactions, particularly the very same behavior they themselves prohibit (Eldakar, Farrell, & Wilson, 2007; Eldakar et al., 2013; Eldakar & Wilson, 2008; Eldakar, Wilson, & O’Gorman, 2006; Nakamaru & Iwasa, 2006). For example, in the tree wasp (*Dolichovespula sylvestris*), individuals that police against worker-laid eggs are also those most likely to cheat and lay eggs themselves, appropriately deemed “corrupt policing” (Wenseleers, Tofilski, & Ratnieks, 2005). Similar hypocritical interactions occur even within multicellular organisms as cancerous tumors have been shown to inhibit the growth of other tumors (Camphausen et al., 2001; Guba et al., 2001). With regard to human social behavior, examples range from criminal organizations such as the Mafia reducing local crime rates (Fiorentini & Peltzman, 1995; Lupo, 2013) to socially dominant preschoolers compelling their peers to follow the rules while they themselves do not (Hawley & Geldhof, 2012).

The second mechanism in which punishers can overcome their local disadvantage is through power asymmetries. Natural selection is based on the premise that there is variation among individuals. While it is commonly assumed that individuals vary in their propensity for altruism and punishment, this is not widely incorporated into research studies (de Weerd & Verbrugge, 2011; Przepiorka & Diekmann, 2013; Úbeda & Duéñez-Guzmán, 2011). In other words, the costs of performing punishment are unlikely to be the same for all. Punishment is relatively more cost-effective for individuals that hold an asymmetric power advantage over their peers. Specifically, it has been shown that college students coming from families with greater financial and social resources are more likely to punish transgressors in experimental economic games (O’Brien, Gallup, & Eldakar, 2013). Furthermore, individuals in positions of power and high resource availability may also face relatively lower costs of committing transgressions. Thus, when coupled with the notion of selfish punishment, it is possible for punishment to actually be locally advantageous for some. For example, nonhuman primate dominance hierarchies are maintained through power asymmetries, and it is often the case that those with power both perform punishment and act selfishly (Boehm, 1999; de Waal, 1990; Flack, de Waal, & Krakauer, 2005; Flack, Girvan, de Waal, & Krakauer, 2006).

Specifically, in pigtailed macaques (*Macaca nemestrina*), powerful individuals are the most likely to perform types of punishment that both reduce within-group conflict (Flack et al., 2005) and stabilize social niches (Flack et al., 2006). Furthermore, in rhesus macaques (*Macaca mulatta*), high-ranking males punish individuals for failing to announce their discovery of food, yet avoid punishment for their own deception (Hauser & Marler, 1993). These implications extend to modern human society whereby hypocrisy in the form of corruption (abuse of authority for selfish gains) springs up among those in powerful positions of enforcement and/or governance (Rose-Ackerman, 1999a, 1999b, 2008).

Our previous studies have demonstrated that selfish punishment is more likely to evolve in populations than altruistic punishment (defined here as punishers that do not act selfishly) because individuals possessing the former strategy are able to recoup punishment costs (Eldakar et al., 2007, 2013; Eldakar & Wilson, 2008). However, we have also recently shown that selfish punishment transitions to altruistic punishment once selfishness is driven to exceedingly low levels and the cost of punishment is effectively reduced, allowing the population to attain greater levels of altruism (for an extended discussion, see Eldakar et al., 2013). Although human societies may at times resemble this overall transition (e.g., autocratic to democratic republics, Linz, 1990; from Mafia to government, Fiorentini & Peltzman, 1995; Lupo, 2013; from bands of thugs to honorable knights, Bisson, 1994), the same transition cannot be said for the individuals. Within populations, existing genetic variation in the proclivity to punish (McDermott, Tingley, Cowden, Frazzetto, & Johnson, 2009), as well as evidence of diverse punishment strategies in empirical studies (Eldakar et al., 2006; O’Brien et al., 2013), suggests an existing blend of punishment strategies. The maintenance of these heterogeneous punishment tactics is likely attributed to fluctuations in population density, cost/benefit structure, group size, group longevity, and power asymmetries.

Here, we investigate the role of power asymmetries in the evolution of altruism and punishment through the use of an agent-based simulation model. Our model highlights three important factors in the evolution of selfish punishment (i.e., hypocrisy and corruption): (1) the disparity in power among individuals, (2) the proportion of individuals in the population that possess power, and (3) the severity of consequences for committing social transgressions. Further understanding the role of power asymmetries in the evolution of enforcement strategies is crucial for untangling the evolutionary paradox of punishment across the animal kingdom as well as providing insights into the roots of hypocrisy and corruption in human society.

The Model

The agent-based simulation was constructed from the equations presented here in the analytical model. Although insightful, the equations illustrate what occurs from the perspective of an individual within a given group and does not account for the

average fitness of each strategy across groups and thus are not intended to represent population-level outcomes. The simulation occurs in a large interacting multigroup population. Therefore, the average fitness of each strategy is the result of the average experience of all individual agents of that strategy in the overall population. The simulation model was implemented using the agent-based modeling program Netlogo 6.0.2, which is freely available for download (for Netlogo program, see <http://ccl.northwestern.edu/netlogo/>; contact corresponding author for annotated code).

The model is an evolutionary game-theoretical model based on the standard public goods game often employed in the experimental economics literature (e.g., Fehr & Gächter, 2000). Specifically, this model is closely based on our previous model with the addition of power asymmetries and conditional strategies (Eldakar et al., 2013). The model begins with a population (size = N) with individuals that are either altruistic or selfish and either punish or do not punish, creating four possible strategies: altruistic nonpunishers (a_n), selfish nonpunishers (s_n), altruistic punishers (a_p), and selfish punishers (s_p). Individuals associate in a large number of groups (size = n) and interact for a number of iterations (I) in a two-phase public goods game. For the initial phase (cooperation), all individuals are endowed with a resource (value = R) with altruists contributing their entire share to the group while selfish individuals withhold the resource. Contributions to the group (by altruists within the group, $n_a R$) are multiplied by a value (m) and then distributed equally among all group members. This is biologically sound as the combined effort of the group yields a greater return than the sum of individual efforts, for instance, with group hunting, or the modification of the local environment by microbes (Driscoll, Espinosa, Eldakar, & Hackett, 2013).

After the cooperation phase, the fitness of altruists (w_a) and selfish individuals (w_s) is represented by Equations 1 and 2, respectively. Individuals maximize their fitness within groups by withholding the resource, while overall group fitness is maximized by altruism, producing the classic struggle between altruism and selfishness.

$$w_a = \frac{n_a R m}{n}. \quad (1)$$

$$w_s = R + \frac{n_a R m}{n}. \quad (2)$$

For each iteration, the cooperation phase is followed by the punishment phase in which punishers pay a cost (C) in an attempt to expel each selfish member of the group. Therefore, the total cost of punishment for individuals is the product of the cost of punishment (C) and the number of selfish individuals (n_s) in the group ($n_s - 1$; selfish punishers do not punish themselves).

Power asymmetry and conditional strategies. Before social interactions take place each generation, a proportion of the population ($P = .05$ to $.50$) is randomly attributed a power advantage over their peers. This advantage reduces the cost of punishment, as well as the chance of being punished by a factor (x)

that ranges between 0 and 1. Individuals endowed with this advantage experience (x), whereas the remaining $(1 - P)\%$ of the population experience $x = 1$. In essence (x) represents a symmetry of power and $(1 - x)$ represents the disparity in power in the population.

All individuals also possess a conditional strategy that is triggered in the powerful of the population. This conditional strategy is merely a genetic tendency to switch from their default strategy (one of the four strategies: a_n , a_p , s_n , and s_p) to an alternative strategy (individuals that possess the same default and conditional strategy experience no change). In other words, an individual with a_n as their default strategy and s_p as their conditional strategy will be born and behave as an a_n but switch to s_p if among the $P\%$ the population attributed power. While both default and conditional strategies are inherited from parent to offspring, power is entirely random, thus selection can favor one set of strategies among the powerful and a separate set within the rest of the population. Therefore, it can be observed that certain strategies are only adaptive to express when powerful or not powerful. At the conclusion of each round, overall fitness of altruistic and selfish nonpunishers is represented by Equations 1 and 2, respectively, and altruistic and selfish punishers represented by Equations 3 and 4, respectively.

$$w_{ap} = \frac{n_a R m}{n} - x C n_s. \quad (3)$$

$$w_{sp} = R + \frac{n_a R m}{n} - x C (n_s - 1). \quad (4)$$

The consequence of punishment is expulsion from the remaining social interactions of the group. The probability of punishment by a single punisher is represented by the term (D) which can vary from 0 to 1. Therefore, the chance of escaping a given punisher is $(1 - D)$, while the chance of escaping all punishers in the group (n_p) is $(1 - D)^{n_p}$, leaving the overall chance of punishment represented by Equation 5. Note the chance of punishment is reduced by (x) for powerful individuals, whereas $x = 1$ for nonpowerful individuals. In the mathematical sense, powerful individuals are assumed less likely to be punished. This evasion can be achieved biologically through a greater capability of evading detection, fending off punishers, or merely deterring potential punishers through fear of retribution.

$$(1 - (1 - D)^{n_p}) x. \quad (5)$$

Punished individuals are excluded from all subsequent iterations for that generation and replaced by individuals drawn randomly from a pool based on the initial frequencies at the start of the current generation. Although replacements play fewer rounds than the original members, they still contribute to fitness differentials of the overall strategies in the total population. In addition, individuals die if possessing a fitness of ≤ 0 at any point, in which case they are replaced in the same manner as those expelled.

Expelled individuals continue to receive the resource endowment for the remaining iterations (e), but this is reduced

by the cost of solitary behavior (S ; also considered the consequence of punishment) that ranges from 0.05 to 1, such that an S of 0.05 is extremely harsh while when $S = 1$ produces no real consequences from punishment. This reduced payoff is assumed as social animals forced into a solitary existence likely suffer significant fitness consequences. Therefore, the total fitness after all iterations (I) for selfish nonpunishers and selfish punishers is represented by Equations 6 and 7, respectively.

$$w_s = \left(R + \frac{n_a R m}{n} \right) (I - e) + R S e. \quad (6)$$

$$w_{sp} = \left(R + \frac{n_a R m}{n} - x C (n_s - 1) \right) (I - e) + R S e. \quad (7)$$

After completion of the iterations, all individuals (including those expelled) reproduce asexually in proportion to their relative fitness in the overall population and then die. Offspring assumes the strategy (default and conditional) of the parent with a chance of mutating to any strategies and then disperse to a random group to begin the next cycle of interactions. Figure 1 provides an illustrative summary of the model, and Table 1 shows the baseline conditions. Solitary (S) and symmetry of power (x) were varied from 0.05 to 1.0 in increments of 0.05. The distribution of power as measured by the proportion of powerful to not powerful individuals (P) was varied from 0.05 to 0.5. The proportion of powerful individuals was not extended beyond 50% of the population as then powerful becomes just the norm. Each variable set was replicated 10 times and compared for consistency. Output variables for statistical analysis were the averages across the replicate runs on the final frequencies of the default strategies, conditional strategies, and phenotypically expressed strategies. Furthermore, total altruism was considered as the phenotypic expression of altruism in the cooperation phase, which was the sum of both altruistic punisher and nonpunisher phenotypes.

Results

Enforcement and Altruism

Overall, simulated populations were attracted to one of the three equilibriums based on the combination of variables related to power. One of these equilibriums was pure selfishness (selfish nonpunishers) in both phenotypic expression and a conditional strategy. The other two equilibrium points involved a dimorphic population of altruistic nonpunishers along with either selfish punishers or altruistic punishers. Typical runs quickly settle on one of the three equilibriums within a few hundred generations and remain stable over time (Figure 2).

Total altruism in the population (as measured by the frequency of individuals expressing the altruism phenotype in the cooperation phase) evolves under a wide range of conditions so long as there are meaningful consequences for punishment ($S > 0.5$), such that cheating is not performed with impunity. However, the same cannot be said for the punishment strategies that maintain altruism. Overall, a minority of powerful selfish punishers both maintains altruism when there is low symmetry of power and when power is concentrated among a few

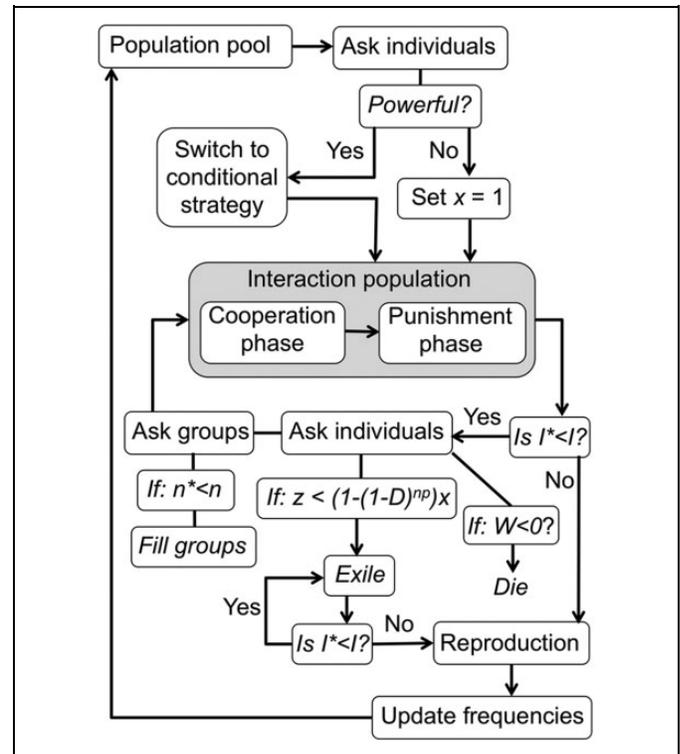


Figure 1. The population begins at set initial frequencies of strategies and individuals disperse randomly to groups (size n). Power is assigned randomly to $P\%$ of the population at the value (x). These individuals also employ their conditional strategy while the remainder of the population is set $x = 1$ and carry out their default strategy. Individuals interact within groups involving a cooperation phase and punishment for a set number of iterations (total number of iterations = I). Following each iteration less than (I), individuals may remain in groups, die due to a fitness (W) of ≤ 0 , or be expelled if the chance of being punished $(1 - (1 - D)^{np})x$ is greater than a random value between 0 and 1 (z). Once banished, the punished individuals acquire a fitness of R^*S (resources, reduced by solitary life) for the remaining iterations (e). Groups losing interacting members to either death and/or punishment are randomly filled to the set group size by individuals drawn from the population stock of the same frequencies from the beginning of the generation. At the conclusion of the iterations, individuals from both the interacting population and the punishment patch reproduce asexually (with a chance of mutation M) in proportion to their relative fitness in the overall population and then die, thus updating the population frequencies and beginning the next generation.

individuals (Figure 3). These equilibriums involving selfish punishers resulted in the selfish punishment conditional strategy, reaching fixation ($>99\%$ of the population), albeit its frequency of phenotypic expression was limited by the abundance of powerful individuals (P). In contrast, as power is essentially decentralized by increasing symmetry of power and spreading power among a greater portion of the population, altruistic punishers maintain altruism (Figure 3).

Despite both punishment strategies—whether altruistic or selfish—maintaining altruism in the model, the extent to which it is maintained is considerable. Populations under the governance of altruistic punishment flourish with the phenotypic expression of altruism attaining fixation ($>99\%$ of the

Table 1. Definition of Variables With Baseline Values.^a

Variable	Baseline	Definition
N	50,000	Total population size
n	5	Group size
l	10	Iterations played within each group per generation
R	10	Resources allocated each iteration
C	5	Cost of punishing each selfish individual
D	0.7	The probability of a single punisher detecting and punishing a given cheater
S	0.4	Consequence of punishment. Cost of reduced fitness due to solitary life each round
M	$l \times 10^{-3}$	Mutation rate
x	0.5	Proportion of punishment cost paid by powerful individuals and value of reduced probability of being punished
P	0.30	Distribution of power as measured by the frequency of powerful individuals in the population
d_{an}	0.45	Initial frequency of altruistic nonpunisher default trait
d_{ap}	0.05	Initial frequency of altruistic punisher default trait
d_{sn}	0.45	Initial frequency of selfish nonpunisher default trait
d_{sp}	0.05	Initial frequency of selfish punisher default trait
c_{an}	0.25	Initial frequency of altruistic nonpunisher conditional strategy
c_{ap}	0.25	Initial frequency of altruistic punisher conditional strategy
c_{sn}	0.25	Initial frequency of selfish nonpunisher conditional strategy
c_{sp}	0.25	Initial frequency of selfish punisher conditional strategy

^aSimulations Were Terminated After 3,000 Generations (30,000 Iterations) With Typical Simulations Stabilizing Well Within 300 Generations (see Figure 2).

population; Figures 3–5). Populations, and more specifically, groups, under the enforcement of selfish punishers are effective but not efficient at promoting high levels of altruism. Despite the benefit of prohibiting other selfish individuals, the presence of a selfish punisher guarantees the existence of at least one selfish group member (his/herself), a downside not experienced with altruistic punishment. Therefore, selfish punishment can maintain altruism but at the same time hold back overall group- and population-level productivity.

Power Corrupts

Factors that contribute to the emergence of one enforcement strategy over the alternative are those that influence power in some way. Both symmetry and the distribution of power as well as the consequences to punishment largely shape the mechanisms of enforcement in our model. Specifically, when holding all other factors constant, merely increasing the disparity in power (reducing symmetry) pushes the population from altruistic to selfish punishment (Figures 3 and 4). Furthermore, essentially centralizing power also favors selfish punishment (Figures 3 and 5). Lastly, weak consequences of punishment

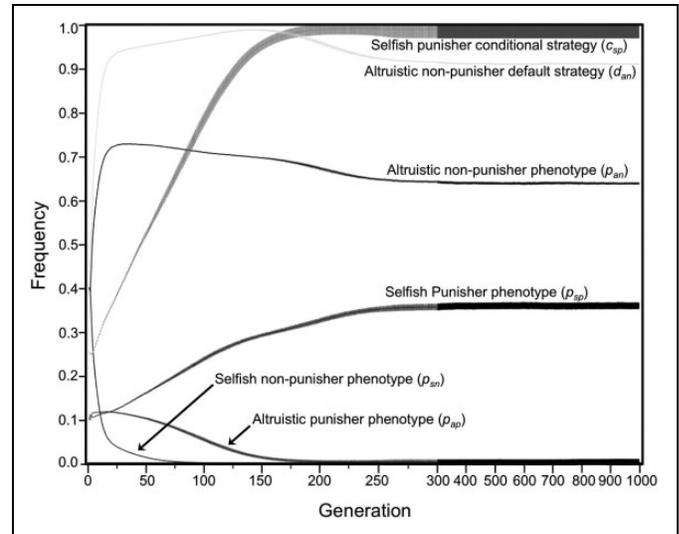


Figure 2. Time-series plot of populations in baseline conditions showing the frequency of certain traits over the course of 1,000 generations, with populations settling in under 300 generations and remaining stable thereafter (note that time is condensed on the x-axis after the 300th generation). Line widths represent the 95% CI intervals from 100 replications. Under baseline conditions, the selfish nonpunisher strategy readily becomes extinct and is soon joined by the altruistic punisher strategy. The remaining population consists of a mix of roughly 64% altruistic nonpunisher and 36% selfish punisher strategies. This almost exactly represents the mixture of nonpowerful (70%) and powerful individuals (30%) in the population. Furthermore, the distribution of cooperation and punishment strategies greatly coincides with power dynamics. The conditional selfish punisher trait, or in other words, *if you are powerful then switch your strategy to selfish punisher* evolves to fixation (99%), thus accounting for the overwhelming majority of selfish punishers in the population. Furthermore, the altruistic nonpunisher default trait, or otherwise the *if weak then play altruistic non-punisher*, evolves to ~91% of the population. Thus, while being phenotypically mixed, this dimorphic population is a result of the overwhelming majority of the population expressing a single strategy, that is, switch in the selfish punisher strategy when powerful but behave as an altruistic nonpunisher when not.

(e.g., lack of oversight) also favor selfish punishment as an enforcement strategy (Figures 4 and 5).

To further illustrate the role of power in favoring selfish punishment, multiple regression analyses were performed across the mean outcomes for all variable conditions (solitary S , power symmetry x , and the distribution of power P) with the average ending frequency of the selfish punisher conditional strategy as the dependent variable. Overall, decreasing symmetry of power (x ; $\beta = -0.727$, $df = 3,996$, $p < .001$) and reducing the distribution of power (P ; $\beta = -0.383$, $df = 3,996$, $p < .001$) both significantly favor the selfish punisher strategy. In other words, increasing power, whether it is through physical or other resource asymmetries, and concentrating and centralizing power lead to increased hypocrisy, corruption, and coercion. However, increasing solitary value (S)—or reducing the severity of punishment—is negatively associated with selfish punishment ($\beta = -0.517$, $df = 3,996$, $p < .001$). This relationship is misleading as once solitary (S) exceeds a value of 0.55, pure

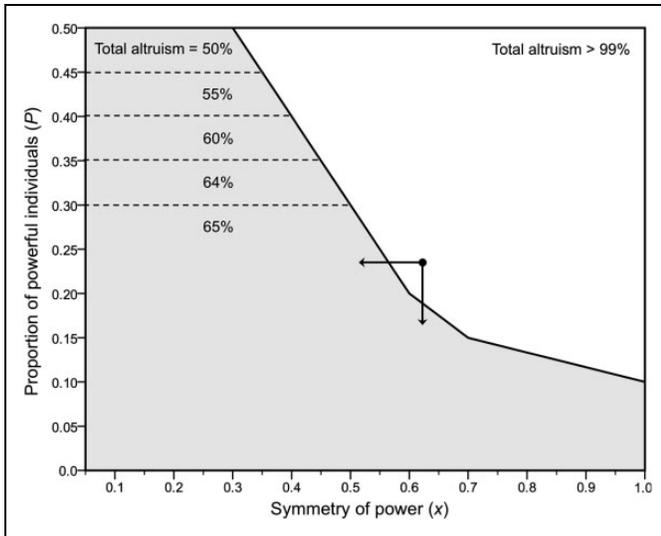


Figure 3. When holding all other variables at baseline conditions (most notably, $S = 0.4$), the variation of symmetry of power (x) and proportion of powerful individuals (P) produces one of the two mixed equilibria. The solid line indicates the boundary between conditions producing either equilibrium. Below the line (light gray background), in conditions consisting of low proportions of power individuals and low symmetry of power between individuals, selfish punishment became fixed (>99% of the population) as the preferred conditional strategy, leading the population to converge on dimorphic equilibrium comprised of a minority of powerful selfish punishers and a majority of altruistic nonpunishers. Above the line (white background), conditions of high proportions of powerful individuals and high symmetries of power among individuals favor altruistic punishment over selfish punishment, producing a dimorphic population of altruistic nonpunishers and altruistic punishers. Overall, total altruism (frequency of individuals expressing altruism in the cooperation phase) is fixed (>99%) under the governance of altruistic punishment but reduced in populations controlled by selfish punishers. Dashed lines within the light gray region delineate boundaries based on total altruism. While selfish punishment and altruistic nonpunishment form an equilibrium, the exact mixture varies slightly based on x and P . Arrows indicate the change in the outcome of a population when either increasing the disparity of power (moving left) or concentrating power in a few individuals (moving down).

selfishness is favored, leading to the overall extinction of all other strategies including selfish punishers (see Figures 4 and 5). An additional multiple regression analysis only including conditions up to $S = 0.55$ revealed a relationship with S that conformed to what is depicted in Figures 4 and 5 and expected overall. Reducing the severity of punishment up to a point favors selfish punishment ($\beta = 0.217$, $df = 2,196$, $p < .001$) but then leads to the collapse of altruism and enforcement thereafter. Furthermore, the relationships with symmetry of power (x) and distribution of power (P) become even more robust (x : $\beta = -1.001$, $df = 2,196$, $p < .001$; P : $\beta = -0.931$, $df = 2,196$, $p < .001$).

Power of the Masses

In contrast, under conditions of little to no difference between individuals in power, the population is comprised of a mixed

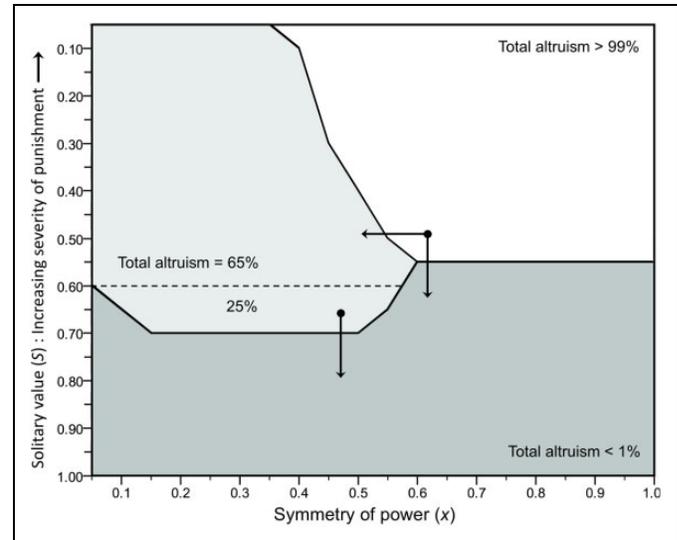


Figure 4. When holding all other variables at baseline conditions (most notably, $P = .3$), the variation of symmetry of power (x) and solitary value (S), or better understood as the consequences of punishment ($1 - S$), produces one of the three equilibria. The solid line indicates the boundary between conditions producing these equilibria. The white background delineates conditions of high symmetries of power among individuals and great consequences to punishment that favor altruistic punishment over selfish punishment, producing dimorphic populations of altruistic nonpunishers and altruistic punishers. The light gray background marks conditions of low symmetry of power and limited consequences of punishment which favor selfish punishment as it readily became fixed (>99% of the population) as the preferred conditional strategy. This leads the population to converge on a dimorphic equilibrium comprised of a minority of powerful selfish punishers and a majority of altruistic nonpunishers. The medium gray background reveals conditions whereby weak consequences of punishment, and/or reduced symmetry of power renders either form of punishment ineffective and the population falls to pure selfishness, attaining fixation in both phenotypic expression and conditional strategy. Dashed lines within the light gray region delineate boundaries based on total altruism. Arrows indicate the change in the outcome of a population when either increasing the disparity of power (moving left) or weakening the consequences for social transgression (moving down).

equilibrium of altruistic punishers and nonpunishers (see Figures 3 and 4). These equilibria experience oscillations of both strategies at $\sim 50\%$ in phenotypic expression and conditional traits. This was attributed to both drift and slight fitness differences between altruistic punishment and altruistic nonpunishment strategies resulting from the consistent reintroduction of selfish individuals through mutation (also, see Eldakar et al., 2007). Overall, increasing the symmetry and/or distribution of power favors altruistic punishment over selfish punishment in populations. Furthermore, increasing the severity of punishment also favors altruistic punishment over selfish punishment (Figures 4 and 5).

The conditions favoring altruistic punishment were further supported by a similar multiple regression analysis with the conditional strategy of altruistic punishment as the dependent variable. All things considered, altruistic punishment is favored

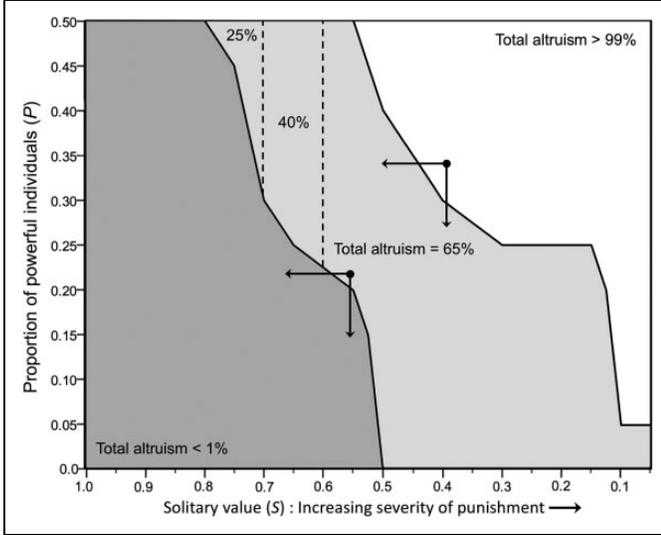


Figure 5. When holding all other variables at baseline conditions (most notably, $x = .5$), the proportion of powerful individuals (P) and solitary value (S), or otherwise understood as the consequences of punishment ($1 - S$), produces one of the three equilibriums. The solid line indicates the boundary between conditions producing these equilibriums. The white background indicates conditions of high distributions of power among individuals and severe consequences of punishment that favor altruistic punishment over selfish punishment and produce dimorphic populations of altruistic nonpunishers and altruistic punishers. The light gray background marks conditions of the concentration of power among relatively few individuals and limited consequences of punishment which favor selfish punishment as it readily became fixed (>99% of the population) as the preferred conditional strategy, leading the population to converge on a dimorphic equilibrium comprised of a minority of powerful selfish punishers and a majority of altruistic nonpunishers. The medium gray background reveals conditions again whereby weak consequences of punishment, and/or the concentration of power in a few individuals, render either form of punishment ineffective and the population falls to pure selfishness, attaining fixation in both phenotypic expression and conditional strategy. Dashed lines within the light gray region delineate boundaries based on total altruism. Arrows indicate the change in the outcome of a population when either weakening the consequences for social transgression (moving left) or reducing the distribution of power (moving down).

as a conditional strategy when symmetry of power (x) is increased ($\beta = 0.252$, $df = 3,996$, $p < .001$), power is more distributed (P ; $\beta = 0.372$, $df = 3,996$, $p < .001$), and solitary value (S) is reduced (severity of punishment, $1 - S$; $\beta = -0.403$, $df = 3,996$, $p < .001$). A separate multiple regression analysis on the subset of conditions only of up to $S = 0.55$ revealed that these relationships only strengthened (x : $\beta = 0.459$, $df = 2,196$, $p < .001$; P : $\beta = 0.673$, $df = 2,196$, $p < .001$; S : $\beta = -0.441$, $df = 2,196$, $p < .001$). In other words, leveling the playing field leads to less cheating, more cooperative forms of punishment when cheating does occur, and a generally more egalitarian society.

Analytically, the condition for which selfish punishment outcompetes altruistic punishment *within* groups (Equation 8)

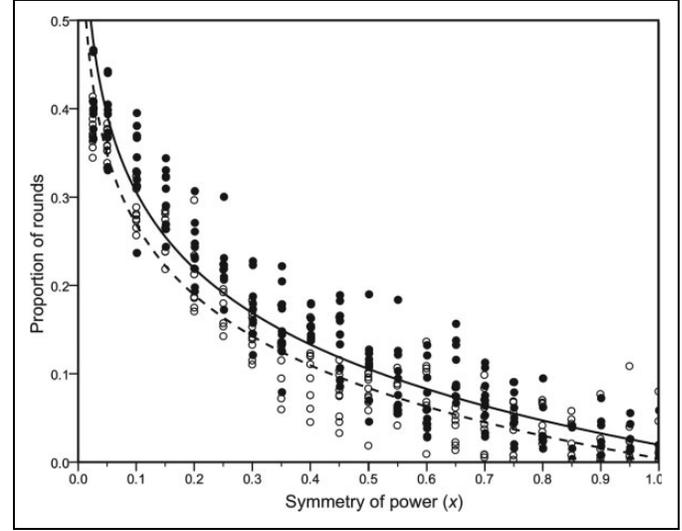


Figure 6. Overall, as symmetry of power decreases (power disparity increases), selfish individuals with greater power are less likely to be punished compared to their counterparts lacking a power advantage. As symmetry of power (S) decreases (or increasing power asymmetry), those with power are more likely to avoid punishment and remain part of group interactions compared to those without power. Open circles with the dashed quadratic fit line mark the difference in the proportion of rounds participating in group interactions between selfish nonpunishers with power and those without power in relation to symmetry of power ($F = 842.89$, $df = 209$, $p < .001$, $R^2 = .80$). Filled circles with the solid quadratic fit line mark the difference in the proportion of rounds participating in group interactions between selfish punishers with power and those without power in relation to symmetry of power ($F = 1,183.35$, $df = 209$, $p < .001$, $R^2 = .85$).

is dependent on the symmetry of power (x ; Equation 8 configured in terms of x to Equation 9).

$$w_{sp} > w_{ap}. \quad (8)$$

$$\left(R + \frac{n_a R m}{n} - x C (n_s - 1) \right) (I - e) + R S e > \left(\frac{n_a R m}{n} - x C n_s \right) I, \quad (9)$$

$$\frac{\frac{n_a m}{n} e - R(I - e + S e)}{C(I - e + n_s e)} < x.$$

When x is low (high disparity in power), selfish punishment is advantageous, and when x is high (individuals are more symmetric in power), altruistic punishment is advantageous. What is not shown analytically is the effect x has on the probability of being expelled and receiving the solitary payoff. A separate set of simulations varying x from 0.05 to 1 (at 0.05 increments and replicated 10 times) illustrated that as the disparity of power is increased (decreasing x), the probability of powerful selfish individuals avoiding punishment and remaining in the group increases compared to their less powerful counterparts (see Figure 6). In other words, as relative power over others is increased, the benefits of hypocrisy and corruption are also enhanced.

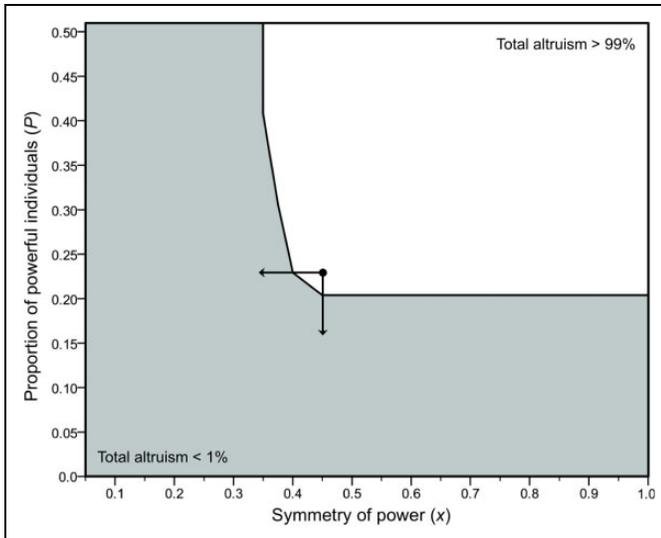


Figure 7. Outcome of simulations that prohibited the selfish punisher strategy. Similar to Figure 2, when holding all other variables at baseline conditions (most notably, $S = 0.4$), the variation of symmetry of power (x) and proportion of powerful individuals (P) produces one of the two equilibria. However, unlike with populations including selfish punishers (Figure 2), only one outcome includes altruism. Below the line (gray background), in conditions consisting of low proportions of power individuals and/or low symmetry of power between individuals, selfish nonpunishers quickly evolve to fixation in the population. Above the line (white background), conditions of high proportions of powerful individuals and high symmetries of power among individuals allow for altruistic punishment to effectively curtail selfishness, producing a dimorphic population of altruistic nonpunishers and altruistic punishers. Overall, in this region maintained by altruistic punishers, total altruism (frequency of individuals expressing altruism in the cooperation phase) is fixed ($>99\%$). Arrows indicate the change in the outcome of a population when either increasing the disparity of power (moving left) or concentrating power in a few individuals (moving down).

A World Without Corruption

To confirm that power disparity disproportionately favors selfish punishment (corruption and hypocrisy) more so than altruistic punishment, we ran the model across the previous parameter space but prohibiting the existence of the selfish punishment strategy. The initial population comprised of the same 50–50% mix of altruism and selfishness, while the entire initial frequency of punishers (10%) was of the altruistic type. In comparison, the previous conditions involved a population of 10% punishers, with 5% altruistic punishers and the other 5% as selfish punishers. Overall, results suggest that a world without corruption is actually more vulnerable to selfishness than one with it.

When precluding enforcement by power selfish punishers, populations readily transition directly from pure altruism to pure selfishness as any form of power disparity grows (see Figures 7–9). In comparison, when allowing for the possibility of corruption, altruism can be maintained by a minority of powerful selfish punishers in populations that

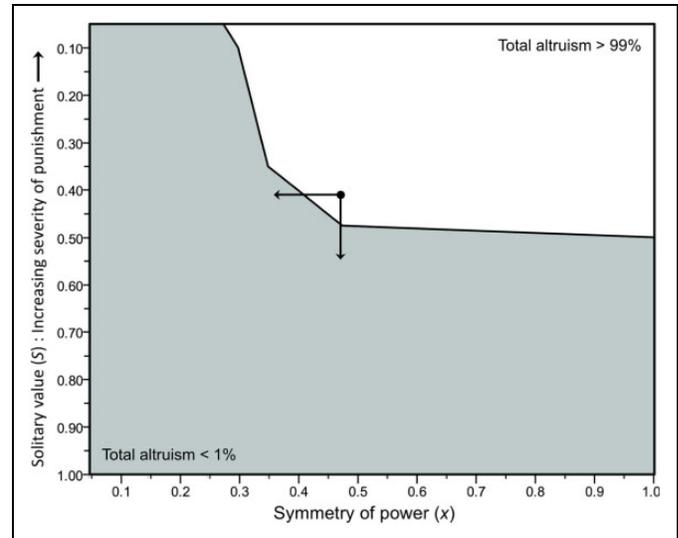


Figure 8. Outcome of simulations that prohibited the selfish punisher strategy. When holding all other variables at baseline conditions (most notably, $P = .3$), the variation in symmetry of power (x) and solitary value (S), or also considered as the consequences of punishment ($1 - S$), produces one of the two equilibria. The white background indicates conditions that favor the evolution and fixation of altruism in the population (altruism in the cooperation phase at $>99\%$). The gray background reveals conditions whereby weak consequences of punishment and/or reduced symmetry of power undermine punishment and the population falls to pure selfishness ($>99\%$ frequency for both phenotypic expression and conditional selfish nonpunisher strategies). Arrows indicate the change in the outcome of a population when either increasing the disparity of power (moving left) or weakening the consequences for social transgression (moving down).

would otherwise descend to pure selfishness (see Figures 3–5). This ability of selfish punishers to *save* a population that would otherwise collapse increases the parameter space in which altruism is viable. When directly comparing between the outcomes of simulations including (Figures 3–5) and excluding selfish punishment (Figures 7–9), altruism is maintained in a wider range of parameter space with their inclusion. However, a closer look revealed that despite the overall increase in the range in which altruism is viable, it also corresponds with a reduction in the parameter space whereby altruism evolves to fixation ($>99\%$). This is expected as the inherent selfish nature of selfish punishment ultimately limits the level of altruism that can be maintained in the population.

In summary, the power has quite different effects on the alternative punishment strategies. Increasing power (x), reducing the distribution of power (P), and reducing the consequences of punishment (S) all favor selfish punishment (corruption) over altruistic punishment. Even with selfish punishment excluded, increasing the disparity of power through any of the three variables (x , P , and S) serves to the detriment of altruism and altruistic punishment. Although this outcome might appear surprising at first, it has been demonstrated previously that selfish punishers, through directly offsetting of the costs of punishment, can better invade and

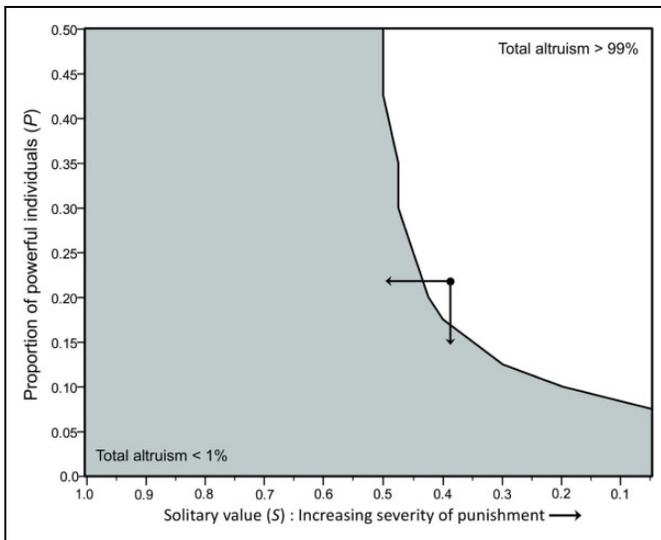


Figure 9. Outcome of simulations that prohibited the selfish punisher strategy. When holding all other variables at baseline conditions (most notably, $x = .5$), the proportion of powerful individuals (P) and solitary value (S), or otherwise understood as the consequences of punishment ($1 - S$), produces one of the two equilibriums. The white background indicates conditions of high distributions of power among individuals and severe consequences to punishment that favor the evolution of altruism maintained by altruistic punishment. The gray background reveals conditions whereby weak consequences of punishment and concentration of power in a few individuals favor pure selfishness, attaining fixation in both phenotypic expression and conditional strategy. Arrows indicate the change in the outcome of a population when either weakening the consequences for social transgression (moving left) or reducing the distribution of power (moving down).

maintain altruism in more selfish populations than their altruistic punisher counterparts (Eldakar et al., 2007). Power only exacerbates these effects. Although morally perplexing, corruption in the form of selfish punishment can both save and limit altruism in a society.

Discussion

Punishment, whether selfish or altruistic, has been shown to be effective in curtailing selfishness in populations (Fehr & Gächter, 2002). By directly reducing the fitness of defectors relative to altruists within groups, punishers diminish the within-group evolutionary advantage of selfishness. The punishment itself, however, is disadvantageous within groups due to the cost of punishment falling solely on the punishers, qualifying punishment as a form of altruism (Bowles & Gintis, 2002, 2004; Boyd et al., 2003; Eldakar & Wilson, 2008; Fehr, 2004; Panchanathan & Boyd, 2004; Yamagishi, 1986). Selfish punishment eliminates this disadvantage by hypocritically coupling selfish behavior (within-group advantage) with punishment, increasing its ability to restore altruism in predominantly selfish populations more so than altruistic punishment. Just as with pure selfishness, however, selfish punishment is disadvantageous at

the group level compared to altruistic punishment. Groups maintained by altruistic punishment are able to achieve greater degrees of altruism on average than groups governed by selfish punishers. This sentiment is illustrated in the field of economics, whereby it has been shown that hypocrisy in the form of corruption reduces economic growth (Mauro, 1995) and undermines criminal justice, where a “clean police force is a crucial barometer of a healthy society” (Punch, 2000, p. 301). This group-level advantage becomes paramount once selfishness is effectively reduced in the population, prompting a transition from selfish punishment to altruistic punishment in the overall population (Eldakar et al., 2013).

While examples of this transition are not uncommon (see Eldakar et al., 2013, for an extended discussion), this process may be interrupted by various factors including asymmetries in power among individuals. Here, we have shown that while an asymmetric power advantage allows some individuals to more effectively punish transgression, it also carries with it a greater temptation to defect. Although previous models have explored the influence of power asymmetries on the evolution of altruistic punishment (de Weerd & Verbrugge, 2011; Przepiorka & Diekmann, 2013), these do not include the possibility of selfish punishment and thus preclude the emergence of hypocrisy (e.g., corruption). As symmetry of power decreases, those with power are increasingly able to defect with impunity, favoring selfish punishment over altruistic punishment. With decreases in the symmetry of power, this dynamic leads to a slippery slope where those that provide protection from exploitation may slide down the path to corruption.

There is no shortage of examples where power asymmetries lead those with power to both exploit and limit the exploitation of others. In modern human societies, violent youth gangs have been shown to form in regions where governmental enforcement is absent, effectively providing a mechanism to reduce further violence (Sobel & Osoba, 2009). This is similar to members of the mafia, who partake in both committing and reducing overall crime in their local region and are believed to have emerged due to a lack of government protection (Fiorentini & Peltzman, 1995; Lupo, 2013). This concept is also common in the expression of dominance hierarchies that are ubiquitous in nature (Huntingford & Turner, 1987), ranging from humans (Boehm, 1999) to amoebas (Fortunato, Queller, & Strassmann, 2003), and are based on the premise of power asymmetries (Wilson, 1975). In nonhuman primates, leading members of dominance hierarchies maintain social order yet receive disproportionately greater benefits from the group (Boehm, 1999; de Waal, 1990; Flack et al., 2005, 2006) and may even directly perform the behaviors they are most vigilant against (Hauser & Marler, 1993).

While serving as inspiring examples of prosociality (Wilson, 1975), the social insects also provide a dark reflection of our own social institutions. Driven by power asymmetry and filled with tales of corruption (Wenseleers et al., 2005) and coercion (Ratnieks & Wenseleers, 2008), scientific observations of eusocial insect societies are more reminiscent of dramatic political novels than of utopian societies (Whitfield,

2002). In the queenless ants of the subfamily Ponerinae, reproductive dominance is maintained through the suppression of others and, in some cases, involves physical mutilation (*Diacamma australe*; Peeters & Higashi, 1989). While relatedness has been attributed as both a driver (Foster, Wenseleers, & Ratnieks, 2006) and byproduct (Wilson & Hölldobler, 2005) of eusociality in insects, it has also been shown to influence which individuals perform the policing of worker-laid eggs (Foster & Ratnieks, 2000). Based on genetic incentive (relatedness to eggs), queens of the wasp *Dolichovespula saxonica* that are singly mated perform policing in their respective colonies, whereas workers of multiply-mated queens assume the role of enforcement. Nevertheless, in social insects, the term “corrupt policing” has been applied to enforcers of tree wasp *Dolichovespula sylvestris* because those that limit cheating (i.e., laying eggs) were also the individuals most likely to cheat (Wenseleers et al., 2005). This may also be descriptive of general scenarios of policing by queens in which cooperation is induced by coercion (Ratnieks & Wenseleers, 2008). Not surprising, when the monarch falls, there is a power vacuum causing a loss of colony productivity (Strassmann et al., 2004). Furthermore, in some cases, power is transferred via asexual reproduction to keep ruling authority among those with highest genetic similarity despite the costs of reduced genetic diversity suffered by the colony (Matsuura et al., 2009). This is not unlike what occurs in human societies with succession through kin.

Interestingly, in our model, power asymmetries do not necessarily lead to the complete erosion of altruism, and in some conditions, the presence of corruption can prevent the failure of a society. Upon further consideration, coercion and corruption in the form of hypocrisy are quite stable in maintaining cooperation in social insects (Whitfield, 2002) and represent the foundation of social dominance relationships (Clutton-Brock & Parker, 1995), which exist due to their stability and reduction of conflict. However, as demonstrated in our model, overexploitation does run the risk of undermining group success to the point of failure, an observation recently echoed by the 43rd president of the United States, George W. Bush, stating “Power can be very addictive and it can be corrosive” (interview on NBC’s Today, on February 27, 2017).

While asymmetries in power favored selfish punishment, the opposite was true for altruistic punishment. This is not surprising when considering how one of the pitfalls of altruistic punishment is the ability to recover the costs of punishment. While disparity in power aids in the ability of selfish punishers to recoup these costs, altruistic punishment is effective when the costs of punishment are shared among many equally capable punishers. This distribution of costs is best achieved by mechanisms such as conformance bias transmission (Henrich & Boyd, 2001), coalitional enforcement (Bingham, 1999), coordination (Boyd, Gintis, & Bowles, 2010; Perc, 2012), or even by using cost-effective strategies such as gossip (Kniffin & Wilson, 2005; Wilson, Wilczynski, Wells, & Weiser, 2000).

It is straightforward to see instances whereby power corrupts, but the corruption of democratically elected officials

seems contradictory to general findings as these officials can be “punished” at the polls during reelection (Rose-Ackerman, 1999b). Thus, the relationship between corruption and democracy is not so straightforward. For example, narrow-interest groups can wield great political power in the presence of weak political parties, thus focusing power to a few. Corruption can infest systems between closely matched political parties, also concentrating power among the few (Rose-Ackerman, 1999b).

Our results also closely resemble hypothesized patterns of enforcement during hominid evolution. It has been proposed that linear dominance hierarchies were maintained through disparities in power between individuals but that the evolved ability to throw rocks and fashion weapons provided a leveling mechanism which led to the emergence of egalitarianism (Boehm, 1999). Coincidentally, it is not surprising that human societies are abound with examples of abuse of power as the result of failed leveling mechanisms. While enforcement of altruism exists across a variety of scenarios and social systems, disparities in power influence the nature of this enforcement. It can transform punishment from a mechanism that promotes altruism to one that exploits and potentially destroys it.

In this report, we set out to understand the role of power asymmetries in the evolution of altruism and punishment. Punishment provides a robust mechanism to promote altruism and keep selfishness at bay, but mechanisms that increase disparity in power bring along the seeds of corruption. Here, we provided a simplistic model to address a fundamental observation that power fuels self-interest. Our results, along with various observations from the animal kingdom and human societies, suggest that power is a principle factor in the structure of social systems and social control.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- Acton, D. J. L. (1907). Letter to Bishop Mandell Creighton, April 5, 1887. In John Neville Figgis and Reginald Vere Laurence (Eds.), *Historical Essays and Studies* (p. 504). London: Macmillan.
- Bingham, P. M. (1999). Human uniqueness: A general theory. *The Quarterly Review of Biology*, 74, 133–169.
- Bisson, T. N. (1994). The feudal revolution. *Past Present*, 142, 6–42.
- Boehm, C. (1999). *Hierarchy in the forest: The evolution of egalitarian behavior*. Cambridge, MA: Harvard University Press.
- Bowles, S., & Gintis, H. (2002). Behavioural science: Homo reciprocans. *Nature*, 415, 125–128.
- Bowles, S., & Gintis, H. (2004). The evolution of strong reciprocity: cooperation in heterogeneous populations. *Theoretical population biology*, 65, 17–28.

- Boyd, R., Gintis, H., & Bowles, S. (2010). Coordinated punishment of defectors sustains cooperation and can proliferate when rare. *Science*, *328*, 617–620.
- Boyd, R., Gintis, H., Bowles, S., & Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences USA*, *100*, 3531–3535.
- Camphausen, K., Moses, M. A., Beecken, W. D., Khan, M. K., Folkman, J., & O'Reilly, M. S. (2001). Radiation therapy to a primary tumor accelerates metastatic growth in mice. *Cancer Research*, *61*, 2207–2211.
- Clutton-Brock, T. H., & Parker, G. A. (1995). Punishment in animal societies. *Nature*, *373*, 209.
- de Waal, F. (1990). *Peacemaking among primates*. Cambridge, MA: Harvard University Press.
- de Weerd, H., & Verbrugge, R. (2011). Evolution of altruistic punishment in heterogeneous populations. *Journal of Theoretical Biology*, *290*, 88–103.
- Driscoll, W. W., Espinosa, N. J., Eldakar, O. T., & Hackett, J. D. (2013). Allelopathy as an emergent, exploitable public good in the bloom-forming microalga *Prymnesium parvum*. *Evolution*, *67*, 1582–1590.
- Eldakar, O. T., Farrell, D. L., & Wilson, D. S. (2007). Selfish punishment: Altruism can be maintained by competition among cheaters. *Journal of Theoretical Biology*, *249*, 198–205.
- Eldakar, O. T., Gallup, A. C., & Driscoll, W. W. (2013). When hawks give rise to doves: The evolution and transition of enforcement strategies. *Evolution*, *67*, 1549–1560.
- Eldakar, O. T., & Wilson, D. S. (2008). Selfishness as second-order altruism. *Proceedings of the National Academy of Sciences USA*, *105*, 6982–6986.
- Eldakar, O. T., & Wilson, D. S. (2011). Eight criticisms not to make about group selection. *Evolution*, *65*, 1523–1526.
- Eldakar, O. T., Wilson, D. S., & O'Gorman, R. (2006). Emotions and actions associated with altruistic helping and punishment. *Evolutionary Psychology*, *4*, 274–286.
- Fehr, E. (2004). Human behaviour: Don't lose your reputation. *Nature*, *432*, 449–450.
- Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, *90*, 980–994.
- Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*, 137–140.
- Fiorentini, G., & Peltzman, S. (1995). Introduction. In G. Fiorentini & S. Peltzman (Eds.), *The economics of organised crime* (pp. 1–32). Cambridge, MA: Cambridge University Press.
- Flack, J. C., de Waal, F. B., & Krakauer, D. C. (2005). Social structure, robustness, and policing cost in a cognitively sophisticated species. *The American Naturalist*, *165*, E126–E139.
- Flack, J. C., Girvan, M., de Waal, F. B., & Krakauer, D. C. (2006). Policing stabilizes construction of social niches in primates. *Nature*, *439*, 426–429.
- Fortunato, A., Queller, D. C., & Strassmann, J. E. (2003). A linear dominance hierarchy among clones in chimeras of the social amoeba *Dictyostelium discoideum*. *Journal of Evolutionary Biology*, *16*, 438–445.
- Foster, K. R., & Ratnieks, F. L. (2000). Social insects: Facultative worker policing in a wasp. *Nature*, *407*, 692–693.
- Foster, K. R., Wenseleers, T., & Ratnieks, F. L. (2006). Kin selection is the key to altruism. *Trends in Ecology & Evolution*, *21*, 57–60.
- Guba, M., Cernaianu, G., Koehl, G., Geissler, E. K., Jauch, K. W., Anthuber, M., . . . Steinbauer, M. (2001). A primary tumor promotes dormancy of solitary tumor cells before inhibiting angiogenesis. *Cancer Research*, *61*, 5575–5579.
- Hardin, G. (1968). The tragedy of the commons. *Science*, *162*, 1243–1248.
- Hauser, M. D., & Marler, P. (1993). Food-associated calls in rhesus macaques (*Macaca mulatta*): II. Costs and benefits of call production and suppression. *Behavioral Ecology*, *4*, 206–212.
- Hawley, P. H., & Geldhof, G. J. (2012). Preschoolers' social dominance, moral cognition, and moral behavior: An evolutionary perspective. *Journal of Experimental Child Psychology*, *112*, 18–35.
- Henrich, J., & Boyd, R. (2001). Why people punish defectors: Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *Journal of Theoretical Biology*, *208*, 79–89.
- Huntingford, F. A., & Turner, A. K. (1987). *Animal conflict*. London, England: Chapman.
- Kniffin, K. M., & Wilson, D. S. (2005). Utilities of gossip across organizational levels. *Human Nature*, *16*, 278–292.
- Linz, J. J. (1990). Transitions to democracy. *Washington Quarterly*, *13*, 143–164.
- Lupo, S. (2013). *History of the Mafia*. New York, NY: Columbia University Press.
- Matsuura, K., Vargo, E. L., Kawatsu, K., Labadie, P. E., Nakano, H., Yashiro, T., & Tsuji, K. (2009). Queen succession through asexual reproduction in termites. *Science*, *323*, 1687–1687.
- Mauro, P. (1995). Corruption and growth. *The Quarterly Journal of Economics*, *10*, 681–712.
- McDermott, R., Tingley, D., Cowden, J., Frazzetto, G., & Johnson, D. D. (2009). Monoamine oxidase A gene (MAOA) predicts behavioral aggression following provocation. *Proceedings of the National Academy of Sciences USA*, *106*, 2118–2123.
- Nakamaru, M., & Iwasa, Y. (2006). The coevolution of altruism and punishment: Role of the selfish punisher. *Journal of Theoretical Biology*, *240*, 475–488.
- O'Brien, D. T., Gallup, A. C., & Eldakar, O. T. (2013). The influence of real-world resource asymmetries on punishment in economic games. *Journal of Social, Evolutionary, and Cultural Psychology*, *7*, 3.
- Panchanathan, K., & Boyd, R. (2004). Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature*, *432*, 499–502.
- Peeters, C., & Higashi, S. (1989). Reproductive dominance controlled by mutilation in the queenless ant *Diacamma australe*. *Naturwissenschaften*, *76*, 177–180.
- Perc, M. (2012). Sustainable institutionalized punishment requires elimination of second-order free-riders. *Scientific Reports*, *2*, 1–6.
- Przepiorka, W., & Diekmann, A. (2013). Individual heterogeneity and costly punishment: A volunteer's dilemma. *Proceedings of the Royal Society of London B: Biological Sciences*, *280*, 20130247.

- Punch, M. (2000). Police corruption and its prevention. *European Journal on Criminal Policy and Research*, 8, 301–324.
- Rankin, D. J., Bargum, K., & Kokko, H. (2007). The tragedy of the commons in evolutionary biology. *Trends in Ecology & Evolution*, 22, 643–651.
- Ratnieks, F. L., & Wenseleers, T. (2008). Altruism in insect societies and beyond: Voluntary or enforced? *Trends in Ecology & Evolution*, 23, 45–52.
- Rose-Ackerman, S. (1999a). *Corruption and government: Causes, consequences, and reform*. Cambridge, MA: Cambridge university press.
- Rose-Ackerman, S. (1999b). Political corruption and democracy. *Connecticut Journal of International Law*, 14, 363.
- Rose-Ackerman, S. (2008). Corruption and government. *International Peacekeeping*, 15, 328–343.
- Sobel, R. S., & Osoba, B. J. (2009). Youth gangs as pseudo-governments: Implications for violent crime. *Southern Economic Journal*, 75, 996–1018.
- Strassmann, J. E., Fortunato, A., Cervo, R., Turillazzi, S., Damon, J. M., & Queller, D. C. (2004). The cost of queen loss in the social wasp *Polistes dominulus* (Hymenoptera: Vespidae). *Journal of the Kansas Entomological Society*, 77, 343–355.
- Úbeda, F., & Duéñez-Guzmán, E. A. (2011). Power and corruption. *Evolution*, 65, 1127–1139.
- Wenseleers, T., Tofilski, A., & Ratnieks, F. L. (2005). Queen and worker policing in the tree wasp *Dolichovespula sylvestris*. *Behavioral Ecology and Sociobiology*, 58, 80–86.
- Whitfield, J. (2002). Social insects: The police state. *Nature*, 416, 782–784.
- Wilson, D. S. (1975). A theory of group selection. *Proceedings of the National Academy of Sciences USA*, 72, 143–146.
- Wilson, E. O., & Hölldobler, B. (2005). Eusociality: Origin and consequences. *Proceedings of the National Academy of Sciences USA*, 102, 13367–13371.
- Wilson, D. S., Wilczynski, C., Wells, A., & Weiser, L. (2000). Gossip and other aspects of language as group-level adaptations. In C. Heyes & L. Huber (Eds.), *The evolution of cognition* (pp. 347–365). Cambridge, MA: MIT Press.
- Wilson, D. S., & Wilson, E. O. (2007). Rethinking the theoretical foundation of sociobiology. *The Quarterly Review of Biology*, 82, 327–348.
- Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology*, 51, 110.