

On the evolutionary origins of the egalitarian syndrome

Sergey Gavrilets¹

Department of Ecology and Evolutionary Biology, Department of Mathematics, and National Institute for Mathematical and Biological Synthesis, University of Tennessee, Knoxville, TN 37996

Edited by Peter J. Richerson, University of California, Davis, CA, and accepted by the Editorial Board July 2, 2012 (received for review January 30, 2012)

The evolutionary emergence of the egalitarian syndrome is one of the most intriguing unsolved puzzles related to the origins of modern humans. Standard explanations and models for cooperation and altruism—reciprocity, kin and group selection, and punishment—are not directly applicable to the emergence of egalitarian behavior in hierarchically organized groups that characterized the social life of our ancestors. Here I study an evolutionary model of group-living individuals competing for resources and reproductive success. In the model, the differences in fighting abilities lead to the emergence of hierarchies where stronger individuals take away resources from weaker individuals and, as a result, have higher reproductive success. First, I show that the logic of within-group competition implies under rather general conditions that each individual benefits if the transfer of the resource from a weaker group member to a stronger one is prevented. This effect is especially strong in small groups. Then I demonstrate that this effect can result in the evolution of a particular, genetically controlled psychology causing individuals to interfere in a bully–victim conflict on the side of the victim. A necessary condition is a high efficiency of coalitions in conflicts against the bullies. The egalitarian drive leads to a dramatic reduction in within-group inequality. Simultaneously it creates the conditions for the emergence of inequity aversion, empathy, compassion, and egalitarian moral values via the internalization of behavioral rules imposed by natural selection. It also promotes widespread cooperation via coalition formation.

despotic | hawk–dove | bystander

Humans exhibit a strong egalitarian syndrome, i.e., the complex of cognitive perspectives, ethical principles, social norms, and individual and collective attitudes promoting equality (1–9). The universality of egalitarianism in mobile hunter-gatherers suggests that it is an ancient, evolved human pattern (2, 5, 6). Political egalitarianism of contemporary foragers is accomplished by a variety of cultural practices (leveling mechanisms) aiming at controlling overassertive, dominant, or very successful individuals who might wish to monopolize resources (3–5, 10). Although conscious, intentional choice of leveling behaviors can explain how egalitarianism is sustained (3–6), the question of how those cognitive and motivational processes evolved in the first place remains open (11).

In group-living organisms with strong social hierarchies, high-rank individuals get more and better resources than low-rank individuals (12). Although under certain conditions both monkeys and apes demonstrate some respect for ownership (13, 14), the dominant individuals can take food and other resources from subordinates especially if the value of the contested item is high, it can be easily taken over, and the owner's rank is low (13, 15, 16). Similar behavior is observed in experiments with young children (17, 18). In hierarchical groups, resisting high-rank bullies alone is dangerous and unlikely to be successful. The only feasible way to successful resistance is via using help of other group members.

Helping behavior is a form of altruism that can evolve by kin or group selection or if there is reciprocity and/or punishment of noncooperators (19–21). However, the first three mechanisms ignore any social role asymmetry, so that help would be as likely

directed toward the bully as toward the victim. The fourth mechanism does explicitly account for the social role (e.g., cooperator or not). However, in primates, policing and punishment are typically administered by the most dominant individuals (22–24), that is, by the bullies themselves. Therefore, although it can account for the maintenance of egalitarian behavior, the role of punishment in its origin is less clear. Helping behavior and its feasibility, dynamics, and patterns are studied by the theory of coalition and alliance formation (25–28). However, existing models typically assume that behavioral rules are fixed rather than evolvable (but see refs. 29 and 30) and that the number of interacting individuals is small (typically three, but see ref. 27) and do not consider explicitly the social role (e.g., bully or victim). Here I expand the theory of coalitions and alliances to study the evolutionary origins of egalitarian tendencies in a despotic society. Rather than focusing on a small number of fixed strategies (e.g., cooperate, defect, punish, etc), I study the evolution of genetically controlled situation-dependent behavioral rules used by individuals for making different decisions. This approach both accounts for the differences in individual personalities (31) and simultaneously describes how individual psychology (32) evolves by selection.

Models and Results

Owner–Bully Interactions. I start by considering agonistic dyadic interactions within a group of n individuals ranked according to their strengths $s_1 > s_2 \dots > s_n$. I assume that during their lifetime individuals come into possession of certain resource items of value b . Whenever one individual (referred to as “owner”) finds an item, another one (referred to as “bully”) may try to take it from the owner. These interactions can be described by a hawk–dove-type game (33) with two strategies: “display” (i.e., do not fight over the resource) and “escalate”. If the bully does not escalate, the owner keeps the item. If the bully escalates, the owner may give up without fighting. If both the owner and the bully escalate, a fight occurs. Whoever wins the fight, gets the resource item. I allow for differences in fighting costs between the winner (c_w) and the loser (c_l). I assume that the probability of winning a fight is an increasing S-shaped function of the ratio of the strengths of the two opponents and use a parameter σ_v to scale the steepness of this function. During the lifetime, each individual finds himself on average K times in each role (e.g., owner or bully). The overall amount of the resource accumulated, R_i , defines the reproductive success w_i (e.g., the proportion of the group's offspring fathered by the individual) according to a generalization of the Tullock contest success function,

Author contributions: S.G. designed research, performed research, analyzed data, and wrote the paper.

The author declares no conflict of interest.

This article is a PNAS Direct Submission. P.J.R. is a guest editor invited by the Editorial Board.

¹E-mail: gavrilas@tiem.utk.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1201718109/-DCSupplemental.

$$w_i = \frac{f(R_i)}{\sum_j f(R_j)}, \quad [1]$$

where $f(R)$ is a monotonically increasing function and the sum is taken over all competitors. In the standard Tullock contest success function, which is extensively used in economics (34, 35) and evolutionary biology (36–40) and which I use in the simulations below, $f(R) = R^\beta$. Parameter β measures the contest intensity: With $\beta = 0$ everybody gets an equal share; $\beta = \infty$ describes the “winner takes all” case.

I am interested in the situations when fights are rather costly relative to the benefit contested. In this case, if individuals have complete information about the conflict, the classical theory of evolutionarily stable strategies in asymmetric conflicts (33, 41) predicts that the conflicts will usually be settled without the actual fight but according to the ownership and/or strength asymmetry. With large differences in strengths, the stronger individual will escalate while the weaker one will display. With small differences in strengths, both the payoff-relevant (i.e., strength) and payoff-irrelevant (i.e., ownership) asymmetries can be used to settle the conflict without an actual fight (33).

Assuming that individuals have complete information is obviously unrealistic. I assume that individual decisions to escalate are made using imperfect information and a simple heuristic rule: Escalate if the perceived ratio of the opponent’s and one’s own strengths is smaller than a certain threshold. The larger the threshold, the more aggressive is the individual. The strengths are estimated with an error the magnitude of which is scaled by parameter σ_e . The escalation thresholds are different for the two roles (e.g., owner and bully) and are controlled genetically. Specifically I assume two unlinked additive diploid loci (with a continuum of alleles), one of which controls the average aggressiveness x (i.e., the average of the escalation thresholds for the two roles) and the other the ownership effect y (i.e., the ratio between the escalation thresholds in the roles of the owner and the bully). Trait x corresponds to the confidence trait in ref. 42. I assume that the escalation threshold in the role of the owner is not smaller than that in the role of the bully (e.g., because of a preexisting loss-aversion syndrome) (43) so that trait $y \geq 1$.

What kind of “evolutionary psychology” will evolve in this system? I used numerical simulations describing a large number of groups each with n males and n females. Generations are discrete and nonoverlapping. Each female produces exactly one male and one female offspring; female offspring disperse randomly whereas male offspring stay in their natal groups. The results show that individuals evolve reduced aggressiveness so that the population average $x < 1$ and the average ownership effect is small (i.e., y is close to 1). This behavior is expected (33, 42). The escalation threshold equal to one means that the decision to escalate is based exclusively on the estimated probability of win. However, besides the probability of win, the corresponding costs and benefits are also important. With high benefits and low costs, increased aggressiveness (or overconfidence in terminology of ref. 42) will maximize the expected payoff. In contrast, with high costs as assumed here a reduced aggressiveness is expected so that each individual escalates only if he estimates that he is sufficiently stronger than the opponent. Increasing the conflict intensity (β) or the evaluation error (σ_e) makes individuals more aggressive whereas increasing group size (n) or costs (c_w and c_l) has opposite effects. The dominance structure emerging in such a population corresponds to a linear rank-resource relationship with weaker individuals usually giving up their resources without fighting. The inequality in resources results in strong inequality in reproductive success (Fig. S1). Describing the situations when only a few males father most of the group offspring (as happens in chimpanzees and other species living in hierarchically organized groups) (23, 44–47) requires one to assume that β is sufficiently larger than 1 (e.g., in

the range 2–4) so that $f(R)$ grows faster than linearly with R . For the rest of the paper, I make this assumption.

Egalitarian Drive. In my model, natural selection optimizes individual behavior in possible dyadic interactions. However, the logic of competitive coexistence in groups implies that the outcome of each particular dyadic interaction has fitness consequences for all other group members (*Methods*). Specifically if function $f(R)$ grows faster than linearly, then each cost-free (i.e., peaceful) transfer of resource from a “poorer” to a “wealthier” individual is detrimental for everybody else in the group. [If $f(R)$ grows slower than linearly with R , the effect is reversed and each other group member benefits when a wealthier individual robs a poorer one.] From one’s perspective, one wants to maximize the amount of resource owned (which will increase the numerator of Eq. 1) and at the same time one wants everybody else to have an equal amount of resources (which will decrease the numerator of Eq. 1). This unappreciated feature of competitive coexistence in groups means that each observer of a conflict has an incentive in helping the poorer side (who, as a rule, is also the weaker). Is the egalitarian drive powerful enough to have evolutionary consequences, when helping is costly? To start attacking this question, I next generalize the model.

Owner–Bully–Helper Interactions. I assume that each owner–bully interaction is observed by a third individual (“bystander”) who may decide to help the victim. If an owner–helper coalition forms, its strength is defined as (27) $S = n^\alpha \bar{s}$, where \bar{s} is the average strength of the coalition partners, $n = 2$ is the coalition size, and α is a positive parameter. ($\alpha = 1$ means the coalition strength is determined additively; $\alpha > 1$ means there is some synergy, so that the coalition is stronger than the sum of the strengths of its members.) This formulation follows the classical Lanchester laws used to describe military conflicts as well as conflicts in animals (48–50). The decision to help is based on the same heuristic rule (i.e., help if the perceived ratio of the opponent’s and the coalition’s strengths is sufficiently small), but the corresponding escalation threshold z is controlled by an independent third locus. In making his decision, the bystander assumes that the owner will escalate, which, however, is not guaranteed as the owner’s decision depends on his own escalation threshold and also is probabilistic. The conflict develops in the following order: First, the bully escalates against a single victim, then the bystander decides whether or not to help (by escalating against the bully), and then the bully and the owner simultaneously decide whether to back down or escalate, respectively, given the bystander’s decision. If the bystander does not escalate, the situation considered above ensues. The payoffs for the case when the helper does escalate are shown in Fig. 1. Note that in the case of a fight, the costs are assumed to be split equally between the coalition partners and that the helper’s payoff is never positive.

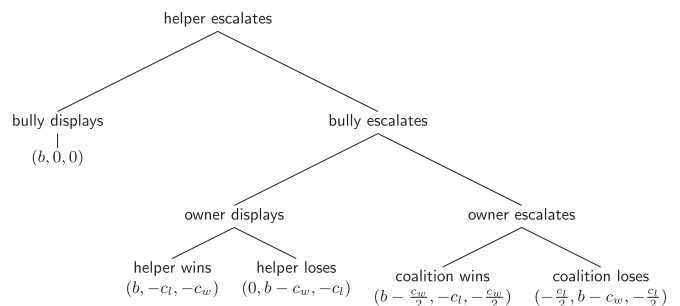


Fig. 1. The payoffs to the owner, the bully, and the helper in different situations if the helper escalates.

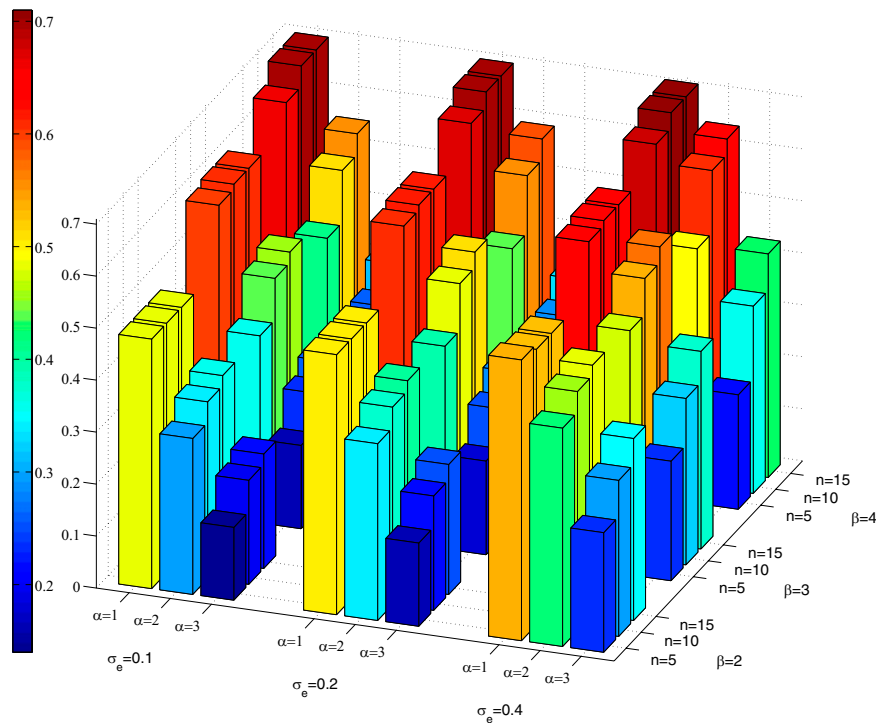


Fig. 3. Effects of α , β , n , and σ_e on the Gini inequality index for fertilities at generation 20,000. Other parameters: $c = 4$, $\gamma = 0.2$, $\sigma_v = 0.4$.

Was the transition to helping behavior and a more egalitarian social structure a relatively sudden or a protracted process? This question is difficult to answer. One can speculate, however, on the reasons for its onset that also offer some insights on why it did not happen in other animals. In terms of the model studied, these reasons could be an increased efficiency of coalitionary aggression (larger α) and decreased uncertainty in evaluating fighting skills of group members (smaller σ_e). Both could have followed the evolution of better cognitive abilities (53, 54) and the development of better coordination skills and weapons as a result of cooperative big-game hunting.

There are several additional factors not considered here that are expected to promote the effect described. In particular, allowing for multiple helpers would increase the effectiveness and decrease the cost of helping. The presence of winner-loser effects (i.e., a correlation between past and future performance in fights) (55) creates additional motivation for helping (because defeating the bully decreases the chance he will successfully attack the helper in the future). Differential group fertility (assuming that groups with fewer internal conflicts will produce more offspring) will accelerate spreading the genes for helping across the whole population.

A few additional comments are in order. In the mechanism advanced here, bullying behavior is the reason for the evolution of helping. However, bullying tendencies remain present; they are not expressed only because of counterdominant helping. It works because it explores the concurrence of the interests of many in the face of the exploitation by a few. The helping benefits are direct but delayed. In the end, it is pure selfish tendencies that could drive the emergence of helping behavior, empathy, and moral values. I was focusing on helping the “poor” against the aggression of the “wealthy”. However, the general mechanism should operate any time one individual can facilitate the transfer of resources from a dominant to a weaker individual. The mechanism studied here is very powerful in that it does not require relatedness, group selection, reciprocity, or reputation. It also promotes widespread cooperation via the formation of coalitions, which in humans occur at many different levels (ranging from within-family to be-

tween-nation states) and represent the most dominant factor in social interactions that has shaped human history (56–58).

The origins of moral values have intrigued scholars for millennia. Darwin saw human morality as derived from animal “social instincts” (59) that transform to a “moral sense or conscience as soon as . . . intellectual powers become . . . well developed” (ref. 59, p. 8). In a modern perspective, viewing human conscience as a mere by-product of intelligence is an oversimplification. Boehm (6) convincingly argues that additional processes and factors such as moralistic punishment, internalization of culturally enforced norms, symbolic language and gossiping, and social selection for altruism and self-restraint applied by groups to its members need to be considered. That notwithstanding, identifying evolutionary roots for and the dynamics of genetically controlled egalitarian social instincts is a necessary step in getting a better understanding of the origins of a uniquely human sense of right and wrong.

Methods

Egalitarian Drive. Consider the fitness of a focal individual i before (w_i) and after (w'_i) a transfer of $\delta > 0$ units of the resource from individual ω to individual κ . Using Eq. 1,

$$w_i = \frac{f(x_i)}{f(x_\omega) + f(x_\kappa) + \sum f(x)}$$

$$w'_i = \frac{f(x_i)}{f(x_\omega - \delta) + f(x_\kappa + \delta) + \sum f(x)}$$

where the sums are over all individuals except for κ and ω . From here, the transfer decreases the fitness of the focal individual (i.e., $w'_i < w_i$) if

$$f(x_\omega - \delta) + f(x_\kappa + \delta) > f(x_\omega) + f(x_\kappa),$$

which can be rearranged as

$$f(x_\kappa + \delta) - f(x_\kappa) > f(x_\omega) - f(x_\omega - \delta).$$

A simple graphical argument (Fig. 4) shows that this inequality is satisfied if f grows faster than linearly with x provided that $x_\kappa > x_\omega - \delta$. The

52. Vehrencamp SL (1983) A model for the evolution of despotic versus egalitarian societies. *Anim Behav* 31:667–682.
53. Byrne RW, Whiten A (1988) *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans* (Clarendon, Oxford, UK).
54. Gavrilets S, Vose A (2006) The dynamics of Machiavellian intelligence. *Proc Natl Acad Sci USA* 103:16823–16828.
55. Hsu Y, Earley RL, Wolf LL (2006) Modulation of aggressive behaviour by fighting experience: Mechanisms and contest outcomes. *Biol Rev Camb Philos Soc* 81:33–74.
56. Wright HT (1977) Recent research on the origin of the state. *Annu Rev Anthropol* 6: 379–397.
57. Johnson AW, Earle T (1987) *The Evolution of Human Societies. From Foraging Group to Agrarian State* (Stanford Univ Press, Stanford, CA).
58. Turchin P (2003) *Historical Dynamics: Why States Rise and Fall* (Princeton Univ Press, Princeton).
59. de Waal F (2009) *The Age of Empathy: Nature's Lessons for a Kinder Society* (Harmony Books, New York).