Evolving institutions for collective action by selective imitation and self-interested design

Sergey Gavrillets¹ and Mahendra Duwal Shrestha²

¹Department of Ecology and Evolutionary Biology, Department of Mathematics, National Institute for Mathematical and Biological Synthesis, Center for the Dynamics of Social Complexity, University of Tennessee, Knoxville, TN 37996 USA

²Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN 37996 USA

August 26, 2019

Acknowledgements. We thank L.Glowacki, L. Perry, P. J. Richerson, M. Singh and reviewers for comments and suggestions. Supported by the U. S. Army Research Office grants W911NF-14-1-0637 and W911NF-17-1-0150, the Office of Naval Research grant W911NF-18-1-0138, the National Institute for Mathematical and Biological Synthesis through NSF Award #EF-0830858, and by the University of Tennessee, Knoxville.

Evolving institutions for collective action by selective imitation and self-interested design

Abstract

Human behavior and collective actions are strongly affected by social institutions. A question of great theoretical and practical importance is how successful social institutions get established and spread across groups and societies. Here, using institutionalized punishment in small-scale societies as an example, we contrast two prominent mechanisms - selective imitation and self-interested design - with respect to their ability to converge to cooperative social institutions. While selective imitation has received a great deal of attention in studies of social and cultural evolution, the theoretical toolbox for studying self-interested design is limited. Recently Perry *et al.* (2018) expanded this toolbox by introducing a novel approach, which they called foresight, generalizing standard myopic best response for the case of individuals with a bounded ability to anticipate actions of their group-mates and care about future payoffs. Here we apply this approach to two general types of collective action – “us vs. nature” and “us vs. them” games. Our results show that foresight increases leaders’ willingness to punish free-riders. This, in turn, leads to increased production and the emergence of an effective institution for collective action. We also observed that largely similar outcomes can be achieved by selective imitation, as argued earlier. Foresight and selective imitation can interact synergistically leading to a faster convergence to an equilibrium. Our approach is applicable to many other types of social institutions and collective action.

1 Introduction

Cooperating human groups can acquire material benefits that would be completely out of reach (or too costly) for single individuals. For this to happen, however, group members have to be able to effectively coordinate their actions, resolve potential conflicts, and eliminate or minimize free-riding. Collective action problem (i.e., free-riding of group members) is generic for both human and non-human animal groups and can easily undermine within-group cooperation (Olson, 1965, Hardin, 1982, Sandler, 1992, Pecorino, 2015). Collective action problems can be (partially) resolved by several mechanisms including kin cooperating with each other (and gaining compensatory benefits through indirect fitness), direct and indirect reciprocity, punishment, group selection, selective incentives, within-group heterogeneity as well as social norms and social institutions regulating individual and group behavior (Olson, 1965, Nowak, 2006, McElreath and Boyd, 2007, Gavrillets, 2015b).

Collective action problems can also be solved via leadership (Hooper *et al.*, 2010a, Glowacki and von Rueden, 2015, Smith *et al.*, 2016a, Gächter and Renner, 2018, Dogan *et al.*, 2018, Garfield *et al.*, 2019, Wiessner, 2019). Leaders can coordinate the actions of group members making their efforts more efficient, monitor and punish free-riders, reward contributors, and foster pro-social norms and values. Leaders and followers emerge naturally as a result of heterogeneity in preferences, motivation, personality, physical characteristics, information available, and other features affecting individual performance in different activities (Smith *et al.*, 2016a, Gavrillets *et al.*, 2016, Perry *et al.*, 2018). In some small-scale societies, leaders get an equal share of the collective goods produced by the group while in others they get extra benefits (Glowacki and von Rueden, 2015, Smith *et al.*, 2016a, Garfield *et al.*, 2019). Leadership can be informal or institutionalized, e.g., when there are rules or shared expectations establishing who is a leader, what they do, and what their privileges are.

A question of crucial theoretical and practical importance is how social institutions for collective action become effective and stable. Institutions that regulate social life, including those that

reconcile disputes, manage the commons, and ensure norms compliance are ubiquitous and a key feature enabling the success of our species (North, 1990, Richerson and Boyd, 2005, Alesina and Giuliano, 2015, Powers *et al.*, 2016, Singh *et al.*, 2017). Yet, we know little about how these institutions develop. To what extent do they reflect the interests and intentionality of their members? Do the design features reflect just transmission processes such as selective imitation or are they better explained by individual wielding agency over their shape through expectations about the future and the behavior of others?

From the theoretical point of view, game theory methods can often identify certain, often multiple, states (Nash equilibria) at which nobody benefits from changing their current actions/strategies. However, how individuals and institutions find the appropriate strategies and how these strategies spread across groups and societies in situations without a full knowledge of the total structure of the corresponding game, or the ability and inclination to go through any complex reasoning process is much less clear (Sandholm, 2010).

One powerful method of optimizing individual strategies is random innovation coupled with selective imitation by payoff-biased social learning. Under this method, individuals observe and evaluate actions and payoffs of others and adapt strategies resulting in a higher payoff. Selective imitation is mathematically analogous to natural selection acting in biological systems but it can operate on much faster time-scales (Richerson *et al.*, 2016). Selective imitation can also drive cultural group selection, resulting in the spread of beneficial institutions across different groups (Richerson and Boyd, 2005, Richerson *et al.*, 2016, Turchin, 2016). Some researchers view cultural group selection as the most important (or even the only) mechanism that can account for institutionalized cooperation in human societies (Chudek *et al.*, 2013, Richerson *et al.*, 2016, Turchin, 2016). However, the power and usefulness of selective imitation within the context of collective action can be questioned. At the individual level, because free-riders often end up with a higher payoff than cooperators, their strategies are more likely to be copied which would undermine cooperation (Molleman *et al.*, 2014, van den Berg *et al.*, 2015, Burton-Chellew *et al.*, 2017). Moreover, because individuals differ in a variety of characteristics, a strategy that is advantageous for one will not necessarily be beneficial or even feasible for another. At the group level, selective imitation of institutions requires a flow of information between (potentially competing) groups and the intimate knowledge of relevant details. Even if these are readily available, institutions might not be transferable “off the shelf” because of differences among groups in their social or ecological environment (Aoki, 2001, Powers *et al.*, 2016, Singh *et al.*, 2017). Additionally, models of selective imitation typically fail to more broadly consider how within-group variation in interests and power constrains the form of emerging institutions (Singh *et al.*, 2017, Cofnas, 2018) and pay only cursory attention to how new rules emerge treating innovation as pretty much a random process.

An alternative view emphasizes the power of within-group design processes driven by the motivation of the whole group or some of its segments to increase their material payoffs or some more general utility. For example, Ostrom (1990) has identified a number of “design principles” for stable and successful management of common resources by local communities. Early eighteenth-century pirates designed democratic institutions (with constitutions, separation of power, and checks and balances) making pirate predatory groups very efficient (Defoe, 1724, Leeson, 2009). Similar examples are known among contemporary prison gangs (Skarbek, 2012). Singh *et al.* (2017) argue for the importance of self-interested design in the creation of institutions and put forward a “self-interested enforcement” hypothesis, which proposes that many observed group-level traits and institutions reflect the differences in relative enforcement capabilities of different group segments. We note that the idea of self-interested design also captures key aspects of human sociality – that we can in fact take guesses about the future and the future behavior of our peers.

One way to contrast selective imitation and self-interested design as the mechanisms of social

evolution is through mathematical modeling. Inspired by the pioneering work of Cavalli-Sforza and Feldman (1981) and Boyd and Richerson (1985), there is now a great diversity of mathematical models and theoretical approaches dealing with social learning and imitation. Some of these focus on the evolution of leadership in small-scale societies (Hooper *et al.*, 2010a, Powers and Lehmann, 2013, 2014) and institutionalized punishment in large-scale societies (Isakov and Rand, 2012, Roithmayr *et al.*, 2015) within the context of collective action. In contrast, there is no established mathematical framework for modeling the evolution of group-level traits and institutions by self-interested design. A powerful and well studied method of optimizing individual behavior is myopic best response (Sandholm, 2010), which is an example of the bounded rationality approach (Gigerenzer and Selten, 2001). Under this method, individuals attempt to optimize their behavior under the assumption that everybody else keeps their strategies. If each group member is using myopic optimization, the group can end up at a Nash equilibrium (Hofbauer and Sandholm, 2002, Xu, 2016). However, myopic best response can fail in social dilemmas or when there is a collective action problem, because self-interested individuals will be motivated to free-ride on the effort of others.

Recently Perry *et al.* (2018) introduced a novel strategy updating method which generalizes myopic best response for individuals with a bounded ability to i) anticipate future actions of their group-mates and ii) consider their effects on future payoffs. The method, which we called (one-step) foresight, attempts to capture some aspects of the “theory of mind”, i.e. the ability to reason about the knowledge and thought processes of others in the social context (Premack and Wodruff, 1979). The “theory of mind” is well-established in humans (Tomasello *et al.*, 2005) and also appears to exist in great apes (Krupenye *et al.*, 2016, de Waal, 2016). Foresight with respect to the effects of punishment is also well established in experimental studies of cooperation as a very powerful driver of individual behavior: the threat of punishment immediately makes subjects more cooperative (Fehr and Gächter, 2002, Spitzer *et al.*, 2007) and subjects expect that individuals who were punished earlier will be more cooperative in the future (Krasnow *et al.* 2012; see also Axelrod 1986’s discussion of “deterrence” as a mechanism for maintaining cooperative norms). Our earlier work focused on collective action in heterogeneous groups in the presence of peer punishment (Perry *et al.*, 2018). In particular, we showed that foresight can allow groups to overcome the first- and second-order free-riding problems leading to successful cooperation.

We also demonstrated the emergence of a division of labor in which some individuals (e.g., more powerful) specialized in punishment while others (e.g., less powerful) mostly contributed to the production of collective goods. Our first major goal here is to extend our approach to the case of collective actions under institutionalized punishment in small-scale societies. Specifically, we will assume that the division of labor between leaders who punish cheaters and the rest of the group who produce collective goods is already established and collectively endorsed (Garfield *et al.*, 2019, Wiessner, 2019) and will study its evolution. Our second goal is to compare selective imitation and self-interested design with respect to their ability to identify and converge to cooperative social institutions.

2 Models and Results

Consider a population comprised of a number of groups. Each group has n individuals which we will call “commoners” and an additional entity which we will call a “leader”. (The “leader” does not have to be a single individual but can be a group, such as elders in a small-scale society, or a formal institution.) Commoners have an opportunity to participate in collective actions producing shared benefits. Leaders monitor commoners’ efforts, punish free-riders, and collect tax.

We will consider separately and contrast two types of mathematical models aiming to describe

two general kinds of collective action that our ancestors were most definitely engaged in: “us vs. nature” games and “us vs. them” games (Gavrilets and Fortunato, 2014, Gavrilets, 2015a,b, Whitehouse *et al.*, 2017, Gavrilets and Richerson, 2017). The former describe collective actions such as defense from predators, cooperative hunting, cooperative breeding, habitat improvements, building dams or fences to drive animals, etc. The success of a particular group in solving these problems does not depend much on the actions of neighboring groups. In contrast, “us vs. them” games describe direct conflicts and/or other costly competition with other groups over territory, mating opportunities, access to trade routes, etc. The success of one group in an “us vs. them” game means failure or reduced success for other competing groups.

Basic model: Groups without leaders. To get a better intuition about our general model, we first consider acephalous groups (i.e. groups without leaders). We assume a commoner’s effort in a collective action (specified by a binary variable $x = 0$ or 1) is costly while any benefit produced and retained by a group of commoners is shared equally among all of them; this creates an incentive to free-ride (Olson, 1965). Without leaders, the payoff of a commoner making effort x in a collective action is

$$\pi_c(x) = bP - cx, \quad (1)$$

where b and c are the benefit and cost parameters. The function P gives the normalized value of the resource produced or secured by the group. Let $X = \sum x$ be the total group effort. In “us vs. nature” games, we define $P = X/(X + X_0)$, where X_0 is a half-success parameter (Gavrilets, 2015a,b). If $X = X_0$, the probability of group success P is equal to one half. The larger X_0 , the more group effort is required to secure the reward. [A linear public goods game is a special case of model (1) with $P = X/X_0$, where $X_0 = n$.] In “us vs. them” games, we define $P = X/\bar{X}$, where $\bar{X}$ is the average group effort over all G competing groups in the system. Note that “us vs. nature” games are a special case of the generalized Volunteer’s Dilemma (Diekmann, 1985, Archetti, 2009) while “us vs. them” games are common in the theory of between-group contests (Konrad, 2009, Rusch and Gavrilets, 2017).

Equilibria. For both types of games, Nash equilibria can be found in a straightforward way (see the Supplementary Information, SI); there can be multiple Nash equilibria. One can also find the corresponding Quantal Response Equilibria (QRE) by evaluating the corresponding integrals numerically (Goeree *et al.*, 2016) or by performing stochastic agent-based simulations. In both cases, we use the logit QRE approach in which the probability that an agent updating its strategy chooses strategy x is proportional to the exponential function of the expected payoff $\exp(\lambda\pi_c(x))$, where λ is a (nonnegative) precision parameter. [If $\lambda = 0$, the agent chooses $x = 0$ or $x = 1$ with equal probabilities. If $\lambda \rightarrow \infty$, the agent always chooses the best response and the dynamics converges to a Nash equilibrium (Hofbauer and Sigmund, 1998, Sandholm, 2010).] From previous work (Gavrilets and Fortunato, 2014, Gavrilets, 2015a), we also know the ESS-equilibria predicted when individual efforts x are continuous rather than binary. In “us vs. nature” games, the corresponding prediction for the group effort is

$$X^* = X_0(\sqrt{R} - 1), \quad (2a)$$

if $R \equiv b/(cX_0) > 1$, and is 0 otherwise. Note that R is the ratio of the individual benefit b to the group cost cX at half-success effort $X = X_0$. In “us vs. them” games,

$$X^* = \frac{G-1}{G} \frac{b}{c}, \quad (2b)$$

so that the group effort is always positive.

Figure S1 in the Supplementary Information (SI) shows the average group effort X in “us

vs. nature” games observed for different parameters in stochastic agent-based simulations (more details on simulations are given below). Decreasing precision parameter λ naturally increases the contributions while increasing the half-effort parameter X_0 decreases them. The group size n has no effect (data not shown). Also shown is the ESS equilibrium (2a) which approximates the Nash equilibrium (corresponding to the infinite precision case, $\lambda = \infty$) quite well. Convergence to an equilibrium is very fast as illustrated in Figure S2 in the SI.

Figure S3 in the SI shows the average group effort X in “us vs. them” games with two groups. Figure S4 in the SI illustrates the corresponding transient dynamics. [In contrast to “us vs. nature” games, the model now can exhibit nonequilibrium dynamics.]

Full model: Institutionalized punishment. Next we consider the full model with leaders added to groups. The collectively endorsed role of leaders is to identify and punish free-riders. That is, punishment in our model works via the institute of leadership (Hooper *et al.*, 2010a, Isakov and Rand, 2012, Roithmayr *et al.*, 2015, Glowacki and von Rueden, 2015) rather than been administered by peers (Boyd and Richerson, 1992, Hauert *et al.*, 2007).

We assume that a leader makes a costly monitoring effort y ($0 \leq y \leq 1$). As a result of this effort each free-rider in the group is identified with probability y . The leader then punishes each identified free-rider by reducing their payoff by κ at a cost δ to the leader. We define the leader’s cost of monitoring as $c_y n y$, i.e. it grows linearly with the group size; c_y is a cost of monitoring parameter. [We note that realistically in small-scale societies the costs of monitoring and collectively endorsed punishment can be low.] The leader’s benefit comes from a constant tax ρ which they collect from the group’s production. In some of our graphs below, the tax will be specified by parameter $\theta = \rho n / (1 - \rho)$ which is the leader-to-commoner share ratio (e.g., with $\theta = 1$, the leader’s and a commoner’s shares of the reward are the same; with $\theta = 2$, the leader gets twice as much as a commoner). Our model of institutionalized punishment can be viewed as a multi-player extension of the inspection game (Fudenberg and Tirole, 1992).

In this model, the expected payoff of a commoner is

$$\pi_c(x, y, X) = (1 - \rho)bP(X) - cx - \kappa y(1 - x), \quad (3a)$$

where the last term is the expected cost of being punished. The expected payoff of the leader is

$$\pi_l(y, X) = \rho n b P(X) - c_y n y - \delta(n - X)y, \quad (3b)$$

where the first term is the tax collected from n commoners and the last term is the expected cost of punishing $(n - X)y$ identified free-riders.

Given a fixed level of monitoring y , Nash equilibria for the commoners’ group effort X can be predicted from the results on acephalous groups as the presence of a leader merely decreases the benefit and cost terms from b and c in eq. (1) to $\tilde{b} = b(1 - \rho)$ and $\tilde{c} = c - \kappa y$ (compare eq. 1 and 3a). That is, the presence of leaders effectively decreases the benefit and cost terms for commoners. Therefore the latter can be motivated to produce given a sufficiently high level of monitoring y . In contrast, for leaders, π_l always decreases with y , so they will chose a zero effort. As a result, the only Nash equilibrium in this model is the state with no production and no monitoring.

To study the dynamics of the full model with changing efforts x and y we use stochastic agent-based simulations. We assume time to be discrete and focus on a sequence of collective actions occurring synchronously in all groups.

Strategy revision. After each collective action, each commoner and each leader is independently given an opportunity to revise their efforts x and y with probability q . We allow for random innovation (analogous to mutation in biology), selective imitation, and self-interested optimization

at rates E_1, E_2 , and E_3 per time step per individual, respectively ($E_1 + E_2 + E_3 = q$). To deal with errors in decision-making, we use a QRE-like approach (Goeree *et al.*, 2016) with logit errors and with nonnegative precision parameter λ (as outlined above). We assume that under selective imitation commoners can imitate their group-mates as well as commoners from other groups; leaders can imitate other leaders. We consider two self-interested optimization strategies. The first is the standard myopic best response (Hofbauer and Sandholm, 2002, Sandholm, 2010) with errors. While under myopic best response commoners can volunteer some effort (eq. 2 and Fig. S1 and S3), leaders will make an effort only by error because it would result in immediate costs but no immediate benefits. The latter effect is analogous to a well-known second-order free-rider problem in models of peer-punishment (Boyd and Richerson, 1992, Boyd *et al.*, 2003). How then can leaders' effort increase by self-interested design?

Foresight. Our solution is a novel self-interested optimization method of strategy revision which we call foresight (Perry *et al.*, 2018). Here for simplicity we will apply foresight only to leaders. Our justification is that leaders have more information and more power in using it than commoners. The foresight mechanism includes two components: i) consideration of future benefits and ii) the forecast of actions of others.

With respect to the former, the idea is that a major goal of punishment is often to modify the transgressor's future behavior (Ellsworth and Ross, 1983, Axelrod, 1986, Krasnow *et al.*, 2012, Cushman, 2015). This suggests that expected *future* payoffs are usually a part of the punisher's utility function. In our implementation of (one-step) foresight for leaders, the leaders attempt to maximize their utility function u_l , which we define as a sum of the expected payoff $\pi_l(y, X)$ after the current round of strategy updates and the forecasted payoffs after the next round of strategy updating $\pi_l(y', X')$:

$$u_l = \pi_l(y, X) + \pi_l(y', X'). \quad (4a)$$

The leader expects that their action y this round will affect the subordinates' effort X' in the next round. At the same time, their y has no effect on the benefit $\rho nbP(X)$ to be produced by the subordinates this round, or the cost of the inspection in the next round, $[c_y n + \delta(n - X')]y'$ (see equation 3b). Therefore the leader's utility function (4a) reduces to a sum of the costs of inspection and punishment this round and the benefit next round

$$u_l = -c_y n y - \delta(n - X)y + \rho nbP(X'). \quad (4b)$$

Forecasting commoners' effort. To evaluate utility function (4b), we need to specify the leaders' forecast for the group effort X and X' . Our assumption is that leaders know (from previous experience or previous leaders) how the group of commoners typically behave in response to a given level of monitoring y . To capture this assumption mathematically we have used two approaches. In the first approach, leaders predict X on the basis of the ESS equations (2) appropriately adjusted for the corresponding level of monitoring. [That is, to predict X and X' , we use equations (2) with b and c substituted for $(1 - \rho)b, c - \kappa y$ and $(1 - \rho)b, c - \kappa y'$, respectively.] In the second approach, instead of using equations (2), we pre-compute the average total group effort X as observed in numerical simulations for different values of parameters b, X_0, n, λ and $c = 1$ in the model of acephalous groups. We interpret these "empirical" functions as capturing the leader's knowledge of commoners' group behavior. We then adjust X values appropriately for the corresponding level of monitoring (as described above). The numerical results observed were similar for both approaches; the results shown below correspond to the second approach. Note that both the leaders' and commoners' attempts to optimize their actions are subject to stochasticity as described above.

Given a certain level of monitoring y , the commoners will attempt to optimize their behavior

which will lead to a certain group effort $X^*(y)$. (As stated above, there can be multiple equilibria for commoners' effort.) Then, a leader capable of predicting the commoners' behavior is expected to make a minimum effort still assuring that the current commoners' group effort X^* is stable. We can then expect multiple Nash equilibria differing in the amount of production, monitoring, and payoffs (see below).

Simulations. We considered different combinations of various strategy revision methods for commoners and leaders. Each agent updated its strategy randomly and independently with probability $q = 0.25$ per time step. This was comprised by innovation probability fixed at $E_1 = 0.01$ and various combinations of selective copying E_2 and self-interested optimization E_3 probabilities such that $E_2 + E_3 = 0.24$. Innovation for commoners meant flipping x (between 0 and 1). Innovation for leaders meant perturbing their y by a random number from a truncated normal distribution (so that y stays between 0 and 1) with standard deviation σ . Selective imitation was implemented by comparing the payoff of the focal agent with that of a randomly chosen “model” and choosing the corresponding strategy with a probability proportional to $\exp(\lambda\pi)$ (where λ is the precision parameter as in the QRE approach). Myopic optimization for commoners was based on evaluating the payoffs $\pi_c(0)$ and $\pi_c(1)$ and then choosing $x = 0$ and $x = 1$ with probabilities proportional to $\exp(\lambda\pi_c)$. To implement optimization for leaders, we first generated K “candidate strategies” using the same approach as for innovation. Then we evaluated the associated $K + 1$ expected payoffs π_l or utilities u_l (including those for the original strategy). The leader was then assigned one of these strategies with a probability proportional to $\exp(\lambda\pi_l)$ (or $\exp(\lambda u_l)$). [In the terminology of evolutionary game theory, this is a *direct* strategy revision protocol (Sandholm, 2010).] Simulations were run for 5,000 time steps. Convergence to a stochastic equilibrium was confirmed by visual inspection of trajectories. The equilibrium values were estimated as the averages over the last 50 samples; samples were made every 10 time steps; 20 independent runs for each parameter combination. Characteristic time-scale τ for convergence to an equilibrium was evaluated as the time for the average y to reach half of its equilibrium value for the first time.

To better see the effects of leadership and institutionalized punishment, we focused on relatively small values of the benefit parameter b under which acephalous groups would not make any effort in “us vs. nature” games and a relatively low effort in “us vs. them” games.

Results for the full model We observed that if both commoners and leaders used only innovation and selective imitation, cooperation and punishment were practically absent. As discussed in the Introduction, with selective imitation, commoners tend to imitate higher-fitness defectors which suppresses group production and removes the incentives for leaders to contribute. If leaders used innovation and myopic best response, not much punishment happened and the levels of cooperation were relatively low independently of how commoners updated their strategies. This happens because leaders reduce monitoring and punishment to avoid associated costs. High levels of monitoring and cooperation were observed when commoners used myopic best response and leaders employed either selective imitation or foresight. Next we focus on and contrast these two scenarios with respect to equilibrium levels of production by commoners x , monitoring (and punishment) by leaders y , their average payoffs π_c and π_l as well as the time τ to reach an equilibrium.

Figure 1 illustrates the effects of relative frequencies E_2 and E_3 of selective imitation and foresight in the leader's decision making for different levels of taxation (and, consequently, inequality). [We also allowed for random innovation at a constant small rate $E_1 = 0.01$.] In each graph, the leftmost set of bars (labeled $\theta = 0$) corresponds to the case of acephalous groups with no leaders. In this case, group efforts are practically absent in “us vs. nature” games and are relatively low in “us vs. them” games due to the collective action problem (Gavrilets and Fortunato, 2014, Gavrilets, 2015a,b). Adding leaders capable of foresight and allowing for institutionalized punishment leads to the establishment of a certain level of monitoring and punishment accompanied by a significant

increase in commoners' production in both games. In the graphs shown, relative frequencies of selective imitation E_2 and foresight E_3 have weak effects. This is a typical result (see the SI.) Note that an increase in cooperation is observed even when $\theta = 1$, so that the leader and commoners get an equal share of the reward. Increasing θ does not necessarily increase production but does affect the payoff in an obvious way (i.e. decreases it for commoners and increases it for leaders). The graphs in the lowest rows of Fig. 1 show that an equilibrium is reached the fastest if leaders use selected imitation and foresight at similar frequencies. This is a typical result (see the SI).

Figure 2 illustrates the effects of some parameters in “us vs. nature” games. (Similar results are observed for “us vs. them” games.) Increasing benefit b and punishment κ and decreasing group size n , the half-effort parameter X_0 , and precision parameter λ increase cooperation and punishment. These results are intuitive and similar to those in the models of peer punishment (Boyd and Richerson, 1992, Boyd *et al.*, 2003, Traulsen *et al.*, 2012). The payoffs of leaders increase with the group size n (because more commoners means larger overall tax). The same happens for commoners in “us vs. nature” games because with a fixed X_0 , a larger group size means a smaller individual effort will be sufficient to result in a specific group effort. In “us vs. them” games, the payoff of commoners only weakly depends on the group size. Increasing punishment κ increases payoffs in “us vs. nature” games but decreases it in “us vs. them” games where it causes “overproduction” (Figs.S6-S16; cf. Konrad (2009), Gavrillets (2015b)). Increasing taxation θ decreases commoners' production and payoff; decreasing precision λ causes higher efforts (by errors) which is expected.

Overall, across a range of parameters, selective imitation and foresight in leaders result in comparable levels of cooperation and punishment (see the SI). Under some relatively narrow sets of parameters, these two strategy revision methods lead to different levels of monitoring and cooperation. When this happens, in “us vs. nature” games, foresight leads to higher monitoring and cooperation than selective imitation if the cost of punishment κ is low and the group is egalitarian (i.e. $\theta = 1$). With larger κ and θ , selective imitation results in higher monitoring and cooperation. In “us vs. them” games, if there are differences between selective imitation and foresight, they are manifested in higher monitoring and cooperation under selective imitation. It thus appears that when leaders are more powerful, selective imitation is a better approach to finding appropriate strategies.

Intuitively, our results can be understood in the following way. Consider first the case of selective imitation and innovation with no foresight (i.e., $E_3 = 0$). Earlier Isakov and Rand (2012), Roithmayr *et al.* (2015) studied a similar models of institutionalized punishment. They showed that selective imitation can lead to the evolution of punishment if leaders update their strategy at a much slower rate than subordinates. A low update rate prevents leaders from abandoning a costly punishment strategy before subordinates have learned to contribute to avoid punishment. In contrast, in our model of selective imitation, subordinates do not have to learn from others via incremental improvements to adapt but rather they use the best response to the current strategy of the leader. But the overall effect is similar - monitoring and punishment evolve by selective imitation in leaders. Let $X_{BR} = X_{BR}(y)$ be the best response value of the total commoners' effort to a given y . Then selective imitation in leaders will optimize the leader's payoff given by equation (3b) with X substituted for X_{BR} :

$$\pi_l(y, X_{BR}) = \rho n b P(X_{BR}) - c_y n y - \delta(n - X_{BR})y, \quad (5)$$

Next consider the case of foresight with no selective imitation (i.e., $E_2 = 0$). For a leader using strategy y who is able to predict the subordinates' total production $X_{FR}(y)$, the leaders utility

function u_l is given by equation (4b) with X and X' substitute for $X_{FR}(y)$:

$$u_l = -c_y n y - \delta(n - X_{FR})y + \rho n b P(X_{FR}). \quad (6)$$

Both $\pi_l(y, X_{BR})$ and u_l as functions of y are represented by a difference between a benefit term which asymptotically approaches a fixed limit ($\rho n b$ and $\omega \rho n b$, respectively) and a cost terms which increases linearly with y . Both functions can have multiple local maxima. In the case of selective imitation, a local maximum can be found by trial-and-error and then it can spread across system by imitation. In the case of foresight, a maximum can be discovered by leaders via a process of mental scenario building by considering several candidate strategies and comparing their expected utilities. If the leader's prediction X_{FR} is the same as the commoners' best response X_{BR} , the functions in the right-hand side of equations (5) and (6) are identical. This explain why our numerical results do not show much difference between selective imitation and foresight in leaders.

Usually the speed of convergence to a stochastic equilibrium is the fastest when both mechanisms are in operation at comparable frequencies. One consistent difference is that under selective imitation the actual spread of innovations across the whole population in a particular run can happen more rapidly but there is more variation in the onset of the transition to higher monitoring across different runs (Fig. 3).

Figure 4 illustrates the possibility of multiple equilibria for “us vs. nature” games. In the simulations shown, the initial values of y were chosen in a uniform way across the range of values between 0 and 1 and the effects of stochasticity were reduced by making precision parameter λ infinite.

3 Discussion

Here, we modeled social dynamics in groups composed by a number of commoners investing in the production of collective goods and a leader whose collectively endorsed role was to identify and punish free-riders. We postulated that the institution of punishment by leaders is already established and studied how it could become efficient. Earlier work has shown that leaders emerge naturally under many circumstances including leaders who specialize in punishment (Smith *et al.*, 2016b, Perry *et al.*, 2018, Garfield *et al.*, 2019). Our main goals were to, first, investigate the effects of a novel strategy updating method - foresight - on the evolution of cooperation by institutionalized punishment, and, second, compare foresight with selective imitation with respect to their ability to identify and converge to cooperative social institutions. Foresight generalizes standard myopic best response for the case of individuals caring about their future payoffs and capable of anticipating to a certain extent future actions of their group-mates (Perry *et al.*, 2018). Relative to the myopic best response, foresight is more realistic and, simultaneously, not too taxing on cognitive abilities of individuals and required information. In our model, leaders are only able to probabilistically forecast commoners' response to punishment, but cannot immediately identify long-term optimum strategies. As a result, convergence to such strategies happens only asymptotically.

We have shown that foresight makes monitoring and punishment an utility-increasing option (cf. Perry *et al.* 2018). This, in turn, leads to increased production, cooperation, and the emergence of an effective institution for collective action by self-interested design (Singh *et al.*, 2017). Richerson *et al.* (2016) questioned the existence of “the alternatives to [cultural group selection that] can easily account for the institutionalized cooperation that characterizes human societies” (p.16). Our results here offer one such alternative. Many of our conclusions rely on agent-based simulations but we were able to obtain some analytical results on Nash, QRE, and ESS equilibria.

Our results thus provide theoretical support to empirical research in small-scale societies showing

that leadership in the form of institutionalized punishment can solve the collective action problem (Glowacki and von Rueden, 2015, Garfield *et al.*, 2019). We also note that recent experiments strongly points at institutionalized punishment as a more efficient and preferred form of punishment than peer punishment and pool punishment (Fehr and Williams, 2017). In small-scale societies leadership may or may not be associated with receiving a higher proportion of the group-produced collective goods (Glowacki and von Rueden, 2015, Smith *et al.*, 2016a, Garfield *et al.*, 2019). In our models, the relevant parameter was θ - the leader’s share of the reward relative to that of a commoner. The only noticeable effect of θ was that increasing it increased the payoffs of leaders and decreased it for commoners. In contrast, there was not much effect on cooperation, and the groups in which leaders were getting the same share of the reward as commoners were quite successful in collective actions. We have not considered other potentially beneficial effects of leadership such as coordination, norm promotion, or engineering specific religious doctrines incentivizing collective actions – leaving this for future work.

We studied the *evolution* of the institute of leadership rather than its *emergence*. In our paper, the evolving part of the institution was the level of monitoring and punishment levels as in Isakov and Rand (2012), Roithmayr *et al.* (2015). In Hooper *et al.* (2010b), Powers and Lehmann (2014) the evolving trait was the tax imposed by the leaders while in Powers and Lehmann (2013) it was the proportion of public goods invested into the group’s growth rate. Hooper *et al.* (2010b), Powers and Lehmann (2013, 2014) assumed that players inherited their strategies directly from parents (subject to rare random mutation). In Isakov and Rand (2012), Roithmayr *et al.* (2015) players used payoff-biased imitation. In contrast, we have considered and compared a number of different strategy update methods.

Strategy revision method foresight is an example of self-interested behavior. We also used our model to contrast two major mechanisms that have been proposed to explain the evolution of institutions for collective action: self-interested design and selective imitation. In modeling selective imitation, we used standard approaches developed in the emerging theory of cultural evolution (Richerson and Boyd, 2005). The social institution we focused on was institutionalized punishment. Our goal was to put the ongoing debate on the relevance and generality of cultural group selection and self-interested design in the evolution of social institutions (Richerson *et al.*, 2016, Singh *et al.*, 2017) on more solid, quantitative grounds.

Our main and unexpected conclusion is that both mechanisms generally lead to very similar outcomes with respect to the levels of cooperation, punishment, and payoffs. The intuition behind is relatively simple, at least in hind-side. Assuming best response in commoners, both strategies update protocols in leaders optimize mathematically similar payoff and utility functions. As a result, “optimum” solutions are similar. (These optimum solutions are different from the Nash equilibria in the corresponding one-shot games as both selective imitation and foresight change the structure of the corresponding game.) We did observe that under some relatively narrow ranges of parameters there were differences between the mechanisms. Specifically, in “us vs. nature” games, when leaders are not too powerful, foresight can lead to higher monitoring and cooperation whereas with more powerful leaders, selective imitation can outperform foresight. In “us vs. them” games, selective imitation can lead to higher monitoring and cooperation. These differences however are not translated into differences in payoffs. We remind that our conclusions are based on the assumption that commoners use myopic best response. As discussed above, if commoners rely exclusively on selective imitation, not much cooperation happens (unless they update their strategies much more often than leaders (Isakov and Rand, 2012, Roithmayr *et al.*, 2015)).

Although both selective imitation and foresight can result in similar outcomes, their prerequisites differ. Selective imitation is a cognitively simple optimization method based on learning from others with whom the focal agent (i.e. an individual or a group) shares important characteristics

(so that the strategy used by the “model” remains feasible and successful for the “mimic”). The agent using selective imitation aims to be as successful as its model. Foresight and, more generally, self-interested design also use social information and learning about behavior of others. However they are not restricted to interactions with similar agents, and agents using them can become more successful than their social partners. Cognitive skills needed for foresight, as modeled here, are not too demanding. Predicting others’ behavior requires some “theory of mind” which can be formed on the basis of previous observations or just by asking a question “what would I do if I were in their place”. With respect to group traits (such as social institutions), foresight could “work” within a single group, that is, without (cultural) group selection, which would not be possible in the case of selective imitation. Perry and Gavrillets (2019) take a closer look at the relative efficiency of foresight and selective imitation in some simple models allowing for analytical, rather than just numerical, investigation.

Our results also show that foresight and selective imitation can interact synergistically leading to faster convergence to an equilibrium if leaders use both of them. What seems to happen is that self-interested design will lead to a faster establishment of a social innovation in a single group while selective imitation will speed up its spread across other groups.

In our simulations, we assumed the same rates of strategy revision for both mechanisms. However imitation of institutions from other groups is likely to be a rarer event than attempts to improve poorly functioning institutions by local “means”. This implies that the relative rate of social evolution by cultural group selection will likely be slower than that by self-interested design. If however selective imitation is unconstrained, the timing of adoption of a new effective institution by different groups will be more similar than that under self-interested design because it will spread in an infection-like fashion. (See Fig. 3 showing that transitions from low to high values of inspection frequency y in individual runs are much more rapid under selective imitation.)

Our comparison of “us vs. them” and “us vs. nature” games parallel earlier conclusions: the former games are more conducive for the evolution of cooperation than the latter but they can also easily lead to over-production of individual efforts and wasted payoffs. The effects of parameters on the dynamics of punishment and cooperation are also pretty much in line with intuition and earlier results (Gavrillets and Fortunato, 2014, Gavrillets, 2015a,b, Gavrillets and Richerson, 2017).

An interesting (but not surprising) feature of our models is the existence of multiple equilibria which appear under both selective imitation and self-interested design. Multiple equilibria imply significant influence of initial conditions and stochastic exogenous or endogenous (e.g., due to errors in payoff or utility evaluation) events on both transient and long-term dynamics of social institutions.

Our strategy updating method foresight is related to the level- k and cognitive hierarchy modeling (Stahl and Wilson, 1995, Nagel, 1995) as well as to models of (limited) forward-looking players (Jehiel, 1995, 2001). The former assume that players follow a particular system of reasoning which forms a hierarchy. In a simple case, level-0 individuals change their behavior completely randomly. Level-1 individuals assume that all others are of level-0 type and optimize their behavior accordingly. Level-2 individuals assume that all others are of level-1, etc. Such models typically focus on the coexistence of different types of individuals, often within the context of dyadic interactions. Foresight is different from it in two aspects. First, in our model, level-0 individuals do not change their strategies between the rounds and level-1 types use standard myopic optimization. [We note that randomly uniform choice of strategies is not an appropriate level-0 model in studies of cooperation as it would predict 50% cooperation rate.] Individuals employing foresight then corresponds to level-2 reasoning types as they assume that others use myopic best response. Second and most important, we explicitly include future benefits (which are completely ignored in level- k models) in the utility function of level-2 individuals.

In classical game theory, players have complete information and are fully rational. This implies that they should be able to predict what exactly will happen (say, plays and payoffs) in any future moment. Relaxing this assumption, Jehiel (1995, 2001) assumed that each player in a dyadic game can predict what exactly will happen over a limited number of steps in the future. A player’s prediction of what is to come beyond his horizon of foresight is given by an exogenous noise. In contrast, in our approach players use the theory of mind in an attempt to predict what will happen rather than know this for sure. We also focus on group level behavior and collective actions rather than on dyadic games.

There is a number of important directions for extending our work such as explicitly considering the dynamics of population densities (as in Powers and Lehmann (2013, 2014)), allowing for the simultaneous presence of competition of acephalous and hierarchical groups (as in Hooper *et al.* (2010a), Powers and Lehmann (2013, 2014)), and allowing for changeable rather than fixed taxes as well as for a market for leaders (as in Hooper *et al.* (2010a)). Foresight can be applied to any other game with repeated interactions (dyadic or group-level). In our model, the current and future payoffs weighted equally in the leaders’ utility function. Perry *et al.* (2018) studied the effects of varying these weights in a model of peer punishment. While we have only considered selective imitation by payoff-biased social learning, cooperation can also be maintained by conformity- and prestige-biased social learning (Henrich and Boyd, 1998, Henrich and Gil-White, 2001). These mechanisms could act synergistically with those studied here.

Human capacity for cultural learning and selective imitation has no doubt greatly contributed both to our uniqueness as a species (Boyd *et al.*, 2011, Henrich, 2016) and to cooperative social institutions we have built (Richerson *et al.*, 2016). However our abilities to innovate and to design and enforce certain rules and social institutions benefiting our societies or some of their segments have also left a significant footprint in our history and will certainly continue to be important in the future (Singh *et al.*, 2017).

References

- Alesina, A. and Giuliano, P. (2015). Culture and institutions. *Journal of Economic Literature*, **53**, 898–944.
- Aoki, K. (2001). Theoretical and empirical aspects of gene-culture coevolution. *Theoretical Population Biology*, **59**, 253–261.
- Archetti, M. (2009). Cooperation as a volunteer’s dilemma and the strategy of conflict in public goods games. *Journal of Evolutionary Biology*, **22**, 2192–2200.
- Axelrod, R. (1986). An evolutionary approach to norms. *American Political Science Review*, **80**(4), 1095–1111.
- Boyd, R. and Richerson, P. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology*, **13**, 171–195.
- Boyd, R. and Richerson, P. J. (1985). *Culture and the evolutionary process*. University of Chicago Press, Chicago.
- Boyd, R., Gintis, H., Bowles, S., and Richerson, P. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences USA*, **100**, 3531–3535.

Boyd, R., Richerson, P. J., and Henrich, J. (2011). Rapid cultural adaptation can facilitate the evolution of large-scale cooperation. *Behavioral Ecology and Sociobiology*, **65**(3, SI), 431–444.

Burton-Chellaw, M. N., El Mouden, C., and West, S. A. (2017). Social learning and the demise of costly cooperation in humans. *Proceedings of the Royal Society London B*, **284**(1853).

Cavalli-Sforza, L. L. and Feldman, M. W. (1981). *Cultural Transmission and Evolution: A Quantitative Approach*. Princeton University Press, Princeton, NJ.

Chudek, M., Zhao, W., and Henrich, J. (2013). Culture-gene coevolution, large-scale cooperation and the shaping of human social psychology. In R. Joyce, K. Sterelny, and B. Calcott, editors, *Signaling, Commitment, and Emotion*, pages 425–458. MIT Press, Cambridge.

Cofnas, N. (2018). Power in cultural evolution and the spread of prosocial norm. *The Quarterly Review of Biology*, **93**, 297–318.

Cushman, F. (2015). Punishment in humans: From intuitions to institutions. *Philosophy Compass*, **10**(2), 117–133.

de Waal, F. (2016). *Are we smart enough to know how smart animals are?* W W Norton and Company, New York.

Defoe, D. (1724). *A general history of the pyrates*. Dover, Mineola.

Diekmann, A. (1985). Volunteer’s dilemma. *Journal of Conflict Resolution*, **29**, 605–610.

Dogan, G., Glowacki, L., and Rusch, H. (2018). Spoils division rules shape aggression between natural groups. *Nature Human Behavior*, **2**, 322–326.

Ellsworth, P. C. and Ross, L. (1983). Public opinion and capital punishment: A close examination of the views of abolitionists and retentionists. *Crime & Delinquency*, **29**(1), 116–169.

Fehr, E. and Gächter, S. (2002). Altruistic punishment in human. *Nature*, **415**, 137–140.

Fehr, E. and Williams, T. (2017). Creating an efficient culture of cooperation. *Department of Economics, University of Zurich*, page Working Paper No. 267.

Fudenberg, D. and Tirole, J. (1992). *Game Theory*. The MIT Press, Cambridge, MS.

Gächter, S. and Renner, E. (2018). Leaders as role models and “belief managers” in social dilemmas. *Journal of Economic Behavior and Organization*, **154**, 321–334.

Garfield, Z. H., Hubbard, R. L., and Hagen, E. H. (2019). Evolutionary models of leadership: Tests and synthesis. *Human Nature*, **30**, 23–58.

Gavrilets, S. (2015a). Collective action and the collaborative brain. *Royal Society Interface*, **12**, article 20141067.

Gavrilets, S. (2015b). Collective action problem in heterogeneous groups. *Philosophical Transactions of the Royal Society London B*, **370**, 20150016.

Gavrilets, S. and Fortunato, L. (2014). A solution to the collective action problem in between-group conflict with within-group inequality. *Nature Communications*, **5**, article 3526 (doi:10.1038/ncomms4526).

- 588 Gavrillets, S. and Richerson, P. J. (2017). Collective action and the evolution of social norm
589 internalization. *Proceedings of the National Academy of Sciences of the United States of America*,
590 **114**, 6068–6073.
- 591 Gavrillets, S., Auerbach, J., and van Vugt, M. (2016). Convergence to consensus in heterogeneous
592 groups and the emergence of informal leadership. *Scientific Reports*, **6**, article number: 29704.
- 593 Gigerenzer, G. and Selten, R. (2001). *Bounded Rationality : The Adaptive Toolbox*. MIT Press,
594 Cambridge, Mass.
- 595 Glowacki, L. and von Rueden, C. (2015). Leadership solves collective action problems in small-scale
596 societies. *Philosophical Transactions of the Royal Society London B*, **370**, 20150010.
- 597 Goeree, J., Holt, C., and Palfrey, T. (2016). *Quantal Response Equilibrium: A Stochastic Theory*
598 *of Games*. Princeton University Press, Princeton, NJ.
- 599 Hardin, R. (1982). *Collective action*. John Hopkins University Press, Baltimore.
- 600 Hauert, C., Traulsen, A., Brandt, H., Nowak, M. A., and Sigmund, K. (2007). Via freedom to
601 coercion: The emergence of costly punishment. *Science*, **316**(5833), 1905–1907.
- 602 Henrich, J. (2016). *The secrets of our success*. Princeton University Press, Princeton, NJ.
- 603 Henrich, J. and Boyd, R. (1998). The evolution of conformist transmission and the emergence of
604 between-group differences. *Evolution and Human Behavior*, **19**, 215–241.
- 605 Henrich, J. and Gil-White, F. J. (2001). The evolution of prestige: Freely conferred deference as a
606 mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior*,
607 **22**, 165–196.
- 608 Hofbauer, J. and Sandholm, W. (2002). On the global convergence of stochastic fictitious play.
609 *Econometrica*, **70**(6), 2265–2294.
- 610 Hofbauer, J. and Sigmund, K. (1998). *Evolutionary games and population dynamics*. Cambridge
611 University Press, Cambridge.
- 612 Hooper, P. L., Kaplan, H. S., and Boone, J. L. (2010a). A theory of leadership in human cooperative
613 groups. *Journal of Theoretical Biology*, **265**(4), 633–646.
- 614 Hooper, P. L., Kaplan, H. S., and Boone, J. L. (2010b). A theory of leadership in human cooperative
615 groups. *Journal of Theoretical Biology*, **265**, 633–646.
- 616 Isakov, A. and Rand, D. (2012). The evolution of coercive institutional punishment. *Dynamic*
617 *Games and Applications*, **2**, 97–109.
- 618 Jehiel, P. (1995). Limited horizon forecast in repeated alternate games. *Journal of Economic*
619 *Theory*, **67**(2), 497–519.
- 620 Jehiel, P. (2001). Limited foresight may force cooperation. *Review of Economic Studies*, **68**(2),
621 369–391.
- 622 Konrad, K. (2009). *Strategy and dynamics in contests*. Oxford University Press, Oxford.
- 623 Krasnow, M. M., Cosmides, L., Pedersen, E. J., and Tooby, J. (2012). What are punishment and
624 reputation for? *PLOS ONE*, **7**(9).

- Krupenye, C., Kano, F., Hirata, S., Call, J., and Tomasello, M. (2016). Great apes anticipate that other individuals will act according to false beliefs. *Science*, **354**, 110–114.
- Leeson, P. (2009). *The invisible hook: the hidden economics of pirates*. Princeton University Press, Princeton.
- McElreath, R. and Boyd, R. (2007). *Mathematical models of social evolution. A guide for the perplexed*. Chicago University Press, Chicago.
- Molleman, L., van den Berg, P., and Weissing, F. J. (2014). Consistent individual differences in human social learning strategies. *Nature Communications*, **5**.
- Nagel, R. (1995). Unraveling in guessing games: an experimental study. *The American Economic Review*, **85**(5), 1313–1326.
- North, D. C. (1990). *Institution, institutional change and economic performance*. Cambridge University Press, NY.
- Nowak, M. (2006). *Evolutionary dynamics*. Harvard University Press, Harvard.
- Olson, M. (1965). *Logic of collective action: Public goods and the theory of groups*. Harvard University Press, Cambridge, MA.
- Ostrom, E. (1990). *Governing the commons. The evolution of institutions for collective action*. Cambridge University Press, Cambridge.
- Pecorino, P. (2015). Olson’s Logic of Collective Action at fifty. *Public Choice*, **162**, 243–262.
- Perry, L. and Gavrillets, S. (2019). Foresight in a game of leadership.
- Perry, L., Shrestha, M. D., Vose, M. D., and Gavrillets, S. (2018). Collective action problem in heterogeneous groups with punishment and foresight. *Journal of Statistical Physics*, pages <https://link.springer.com/article/10.1007/s10955-018-2012-2>.
- Powers, S. and Lehmann, L. (2013). The co-evolution of social institutions, demography, and large-scale human cooperation. *Ecology letters*, **16**, 1356–1364.
- Powers, S. and Lehmann, L. (2014). An evolutionary model explaining the neolithic transition from egalitarianism to leadership and despotism. *Philosophical Transactions of the Royal Society London B*, **281**, 20141349.
- Powers, S. T., van Schaik, C. P., and Lehmann, L. (2016). How institutions shaped the last major evolutionary transition to large-scale human societies. *Philosophical Transactions of the Royal Society London B*, **371**, 20150098.
- Premack, D. and Wodruft, G. (1979). Does the chimpanzee have a theory of mind. *Behavioral and Brain Sciences*, **1**, 515–526.
- Richerson, P., Baldini, R., Bell, A. V., Demps, K., Frost, K., Hillis, V., Mathew, S., Newton, E. K., Naar, N., Newson, L., Ross, C., Smaldino, P. E., Waring, T. M., and Zefferman, M. (2016). Cultural group selection plays an essential role in explaining human cooperation: A sketch of the evidence. *Behavioral and Brain Sciences*, **39**, article number UNSP e30.

- Richerson, P. J. and Boyd, R. (2005). *Not by genes alone. How culture transformed human evolution*. University of Chicago Press, Chicago.
- Roithmayr, D., Isakov, A., and Rand, D. (2015). Should law keep pace with society? Relative update rates determine the co-evolution of institutional punishment and citizen contributions to public goods. *Games*, **6**(2), 124–149.
- Rusch, H. and Gavrillets, S. (2017). The logic of animal intergroup conflict: A review. *Journal of Economic Behavior & Organization*.
- Sandholm, W. H. (2010). *Population Games and Evolutionary Dynamics*. MIT Press, Cambridge, Massachusetts.
- Sandler, T. (1992). *Collective Action: Theory and Applications*. University of Michigan Press, Ann Arbor.
- Singh, M., Wrangham, R., and Glowacki, L. (2017). Self-interest and the design of rules. *Human Nature*, **28**, 45–480.
- Skarbek, D. (2012). Prison gangs, norms, and organizations. *Journal of Economic Behavior & Organization*, **82**, 702–716.
- Smith, J. E., Gavrillets, S., Mulder, M. B., Hooper, P. L., ElMouden, C., Nettle, D., Hauert, C., Hill, K., Perry, S., Pusey, A. E., van Vugt, M., and Smith, E. A. (2016a). Leadership in mammalian societies: Emergence, distribution, power, and payoff. *Trends in Ecology and Evolution*, **31**, 54–66.
- Smith, J. E., Gavrillets, S., Mulder, M. B., Hooper, P. L., El Mouden, C., Nettle, D., Hauert, C., Hill, K., Perry, S., Pusey, A. E., *et al.* (2016b). Leadership in Mammalian Societies: Emergence, Distribution, Power, and Payoff. *Trends in Ecology & Evolution*, **31**(1), 54–66.
- Spitzer, M., Fischbacher, U., Herrnberger, B., Grön, G., and Fehr, E. (2007). The neural signature of social norm compliance. *Neuron*, **56**, 185–196.
- Stahl, D. O. and Wilson, P. W. (1995). On players’ models of other players: theory and experimental evidence. *Games and Economic Behavior*, **10**(1), 218–254.
- Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, **28**(05), 675–691.
- Traulsen, A., Röhl, T., and Milinski, M. (2012). An economic experiment reveals that humans prefer pool punishment to maintain the commons. *Proceedings of the Royal Society London B*, **279**, 3716–3721.
- Turchin, P. (2016). *Ultrasociety: How 10,000 Years of War Made Humans the Greatest Cooperators on Earth*. Beresta Books.
- van den Berg, P., Molleman, L., and Weissing, F. J. (2015). Focus on the success of others leads to selfish behavior. *Proceedings of the National Academy of Sciences USA*, **112**(9), 2912–2917.
- Whitehouse, H., Jong, J., Buhrmester, M. D., Ángel Gómez, Bastian, B., Kavanagh, C. M., Newson, M., Matthews, M., Lanman, J. A., McKay, R., and Gavrillets, S. (2017). The evolution of extreme cooperation via shared dysphoric experiences. *Scientific Reports*, **7**, article 44292.

- 699 Wiessner, P. (2019). Collective action for war and for peace. a case study among the enga of papua
700 new guinea. *Current Anthropology*, **60**.
- 701 Xu, Z. (2016). Convergence of best-response dynamics in extensive-form games. *Journal of Eco-*
702 *nomics Theory*, **162**, 21–54.

Figure 1: Average values of efforts x, y , payoffs π_c, π_l and the time to equilibrium τ for 5 different combinations of frequencies of selective imitation E_2 and foresight E_3 in leaders. The frequencies are $E_2 : E_3 = 0.23 : 0.01$ (leaders mostly use selective imitation, left-most bar in each set of five bars), $0.16 : 0.08$, $0.12 : 0.12$, $0.08 : 0.16$ and $0.01 : 0.23$ (leaders mostly use foresight, right-most bar in each set of five bars). The frequency of random mutation $E_1 = 0.01$. (a) “Us vs. nature” games for $K = 1, \lambda = \infty, \kappa = 0.5, n = 24, b = 17, X_0 = 24$. (b) “Us vs. them” games for $K = 1, \lambda = \infty, \kappa = 0.5, n = 16, b = 4$. Other parameters are at default values (see the SI). Commoners always use innovation at rate $E_1 = 0.01$ and myopic best response at rate $E_3 = 0.24$. Initial values of x and y are assigned randomly and independently; for commoners, $x = 0$ or 1 with equal probabilities, for leaders, y is chosen from a uniform distribution on $[0, 0.05]$.

Figure 2: Effects of parameters benefit b , group size n , half-effort X_0 , tax θ , punishment strength k , precision λ , and the number of candidate strategies K on the average efforts of commoners x and leaders y and their payoffs π_c and π_l in “us vs. nature” games. Parameters are changed one at a time relative to a “default” set with $b = 16, n = 24, X_0 = 24, \theta = 2, \kappa = 0.5, \lambda = \infty$ and $K = 2$. Standard deviation in strategy innovation $\sigma = 0.25$. Frequencies of updating events: $E_1 = 0.01, E_2 = E_3 = 0.12$ in leaders and $E_1 = 0.01, E_2 = 0, E_3 = 0.24$ in commoners.

Figure 3: The dynamics of average values of x, y, π_c and π_l in 20 independent runs. (a) “Us vs. nature” for $K = 1, \lambda = \infty, \kappa = 0.75, b = 7, n = 24, X_0 = 16, \theta = 2$ and mostly selective imitation in leaders ($E_2 = 0.23, E_3 = 0.01$). (b) Same as (a) but with mostly foresight in leaders ($E_2 = 0.01, E_3 = 0.23$), (c) “Us vs. them” for $K = 1, \lambda = 32, \kappa = 0.5, b = 2, n = 16, \theta = 4$ and mostly selective imitation in leaders ($E_2 = 0.23, E_3 = 0.01$). (d) Same as (c) but with mostly foresight in leaders ($E_2 = 0.01, E_3 = 0.23$). Other parameters: $E_1 = 0.01, \sigma = 0.2$. Initial values of x (0 or 1) are drawn randomly and independently with equal probabilities. Initial values of y of all leaders are drawn randomly and independently from a uniform distribution on $[0.00, 0.05]$ in 20 runs. Red lines show the averages over 20 runs.

Figure 4: An example of multiple equilibria in “us vs. nature” games. Shown are the dynamics of average values of x, y, π_c and π_l in 20 independent runs. Parameters: $K = 1, \lambda = \infty, \kappa = 0.75, b = 7, n = 24, X_0 = 16, \theta = 2$ and mostly selective imitation in leaders ($E_2 = 0.23, E_3 = 0.01$). Initial values of x (0 or 1) are drawn randomly and independently with equal probabilities. In each run, initial values of y of leaders are drawn randomly and independently from a uniform distribution within a narrow band; the bands for different runs did not overlap and covered the whole range of possible y values. Red lines show the averages over 20 runs. In the case shown, there are four different locally stable equilibria.

Supplementary Information for
“Evolving institutions for collective action by selective imitation
and self-interested design”

immediate

August 26, 2019

Contents

1	Model’s variables, functions and parameters	1
2	Nash, best response, ESS, and QRE equilibria in “us vs. nature” games without leaders	1
3	Nash equilibria in “us vs. nature” games with leaders	5
4	Equilibria in “us vs. them” games without leaders	6
5	Additional agent-based simulations	10

1 Model’s variables, functions and parameters

Table S1 summarizes the model’s variables, functions and parameters.

2 Nash, best response, ESS, and QRE equilibria in “us vs. nature” games without leaders

We consider a model in which the payoff to a commoner making an effort x ($= 0$ or 1), who belongs to a group making the total effort X , is

$$\pi = b \frac{X}{X + X_0} - cx, \quad (\text{S1})$$

where X_0 is the half-effort parameter.

Nash equilibria. Consider a state where the total group effort $X = 0$. The payoff to each individual is $\pi_0 = 0$. If a single individual switches to $x = 1$, his payoff will be $\pi_{0 \rightarrow 1} = b \frac{1}{1 + X_0} - c$. Therefore the state $X = 0$ is a strict Nash equilibrium if $\pi_0 > \pi_{0 \rightarrow 1}$ or

$$b/c < 1 + X_0. \quad (\text{S2})$$

Table S1: Model variables, functions and parameters.

	Symbols	Their meaning
Variables	x	commoner's production effort ($x = 0$ or 1)
	y	leader's monitoring effort ($0 \leq y \leq 1$)
Functions	X	total effort of the group of commoners, $X = \sum x$
	$P(X)$	normalized value of the resource produced or secured by the group: $P = X/(X + X_0)$ in "us vs. nature" games; $P = X/\bar{X}$ in "us vs. them" games
	π_c	commoner's expected payoff, $\pi_c = (1 - \rho)bP(X) - cx - \kappa y(1 - x)$
	π_l, π'_l, π''_l	leader's expected payoffs, $\pi_l = \rho nbP(X) - c_y ny - \delta(n - X)y$, $\pi'_l = -c_y ny' - \delta(n - X')y'$, $\pi''_l = \rho nbP(X'')$
	u_l	leader's utility with foresight, $u_l = (1 - \omega)\pi'_l + \omega\pi''_l$
Parameters	n	number of commoners per group
	b, c	benefit and cost parameters for commoners
	X_0	half-effort parameter in "us vs. nature" games
	ρ, θ	tax rate and the leader-to-commoner share ratio; $\theta = \rho n/(1 - \rho)$
	c_y	leader's cost of monitoring parameter
	κ	commoner's cost of being punished
	δ	leader's cost of punishing a commoner
	ω	weight of future payoffs in leaders' utility function
	λ	precision parameter in the QRE approach
	K, σ	number of candidate strategies and standard deviation in innovation

23 Consider a state where the total group effort $X = n$. The payoff to each individual is $\pi_1 =$
24 $b \frac{n}{n+X_0} - c$. If a single individual switches to $x = 0$, his payoff will be $\pi_{1 \rightarrow 0} = b \frac{n-1}{n-1+X_0}$. Therefore
25 the state $X = n$ is a strict Nash equilibrium if $\pi_1 > \pi_{1 \rightarrow 0}$ which simplifies to

$$2n - 1 + X_0 + \frac{n(n-1)}{X_0} < b/c. \quad (\text{S3})$$

26 Consider a state where the total group effort $0 < X < n$. The payoff to an individual contribut-
27 ing 0 is $\pi_0 = b \frac{X}{X+X_0}$. If this individual switches to $x = 1$, his payoff will be $\pi_{0 \rightarrow 1} = b \frac{1+X}{1+X+X_0} - c$.
28 The individual will not be interested in switching if $\pi_0 > \pi_{0 \rightarrow 1}$. The payoff to an individual cur-
29 rently contributing 1 is $\pi_1 = b \frac{X}{X+X_0} - c$. If this individual switches to $x = 0$, his payoff will be
30 $\pi_{1 \rightarrow 0} = b \frac{X-1}{X-1+X_0}$. The individual will not be interested in switching if $\pi_1 > \pi_{1 \rightarrow 0}$. Solving the two
31 inequalities above, a state with total group effort X is a strict Nash equilibrium if

$$\frac{(X + X_0)(X + X_0 - 1)}{X_0} < b/c < \frac{(X + X_0)(X + X_0 + 1)}{X_0} \quad (\text{S4a})$$

32 An alternative way to express this result is to say that if conditions (S2) and (S3) are not satisfied,
33 then there exists a unique strict Nash equilibrium at which the group effort X is an integer within

an unit-length interval $(I_c - 1/2, I_c + 1/2)$ centered on the value

$$I_c = \sqrt{\frac{1}{4} + rX_0} - X_0 \approx X_0 \left(\sqrt{\frac{r}{X_0}} - 1 \right), \quad (\text{S4b})$$

where the benefit-to cost ratio $r = b/c$ and the approximation assumes that $r \gg 1/(4X_0)$. [This follows from the fact that, using variable $u = X + X_0$, conditions (S4a) can be rewritten as $u(u - 1) < rX_0 < u(u + 1)$, or equivalently, $\sqrt{\frac{1}{4} + rX_0} - \frac{1}{2} < u < \sqrt{\frac{1}{4} + rX_0} + \frac{1}{2}$.] Note that the approximate expression above is exactly the same as the ESS solution 2 in the main text.

Best response dynamics in stochastic agent-based simulations. In numerical implementation, we assumed that each agent updates its strategy independently with probability q . Each updating agent evaluates the expected payoffs π_0 and π_1 if choosing $x = 1$ and $x = 0$ under the assumption that all group-mates keep their strategies. Then the agent chooses to cooperate ($x = 1$) rather than defect ($x = 0$) with probability

$$p = \frac{1}{1 + \exp[\lambda(\pi_0 - \pi_1)]}, \quad (\text{S5})$$

where λ is a non-negative precision parameter. This formulation follows the QRE approach with logit errors (Goeree *et al.*, 2016). If $\lambda \rightarrow \infty$, the agent always chooses the best response, if $\lambda = 0$, the agent choses $x = 0$ or $x = 1$ with equal probabilities. If $\lambda \rightarrow \infty$, the dynamics converges to a Nash equilibrium.

Figure S1: Average effort X of an acephalous group in stochastic simulations in “us vs. nature” games for different values of parameters X_0, b and λ . The dashed lines show the ESS prediction (2a). Other parameters: $n = 24, c = 1$, probability of strategy updating $q = 0.25$. The simulations were run for 2,000 times steps; 20 runs for each parameter combinations; the averages were evaluated over the last 1,000 time steps.

Figure S1 shows the equilibrium values while Figure S2 illustrates the transient dynamics.

Mixed Nash equilibria. Assume that each commoner in a group contributes independently an effort 1 to a collective action with probability $0 \leq p \leq 1$. Following Archetti (2009), the probability

Figure S2: Dynamics of group efforts X under stochastic best response in “us vs. nature” games for different X_0 and b values. $G = 32, n = 24$. Single run for each parameter combination. Each graph shows G different lines representing efforts of G groups. Probability of updating is 0.25; $\lambda = \infty$.

that there are j contributors among $n - 1$ group mates of the focal individual is

$$f_j = \binom{n-1}{j} p^j (1-p)^{n-1-j}, \quad (\text{S6})$$

where $\binom{n-1}{j}$ is the corresponding binomial coefficient. Then the expected payoff to a focal individual if he cooperates is

$$\pi_1 = \sum_{j=0}^{n-1} f_j \left(b \frac{j+1}{j+1+X_0} - c \right). \quad (\text{S7})$$

If the focal individual defects, his expected payoff is

$$\pi_0 = \sum_{j=0}^{n-1} f_j b \frac{j}{j+X_0}. \quad (\text{S8})$$

The mixed Nash equilibrium for p can be found from equality $\pi_1 = \pi_0$. Using a symbolic

manipulation program and the Pfaff transformation, the latter can be simplified to

$$\frac{F(2, 1 - n; X_0 + 2; p)}{1 + X_0} = \frac{c}{b}, \quad (\text{S9})$$

where $F(\dots)$ is the hyper-geometric function. Equation (S9) can be solved for p numerically (or graphically). One can show graphically that a positive solution exists only if $b/c > 1 + X_0$, i.e. if the benefit to cost ratio is sufficiently large.

QRE equilibria. Within the realm of the QRE approach with logit errors (Goeree *et al.*, 2016), an individual chooses to cooperate ($x = 1$) rather than defect ($x = 0$) with probability p given by equation (S5) above. The QRE solution for p satisfies the equality

$$\frac{\ln(p(1 - p))}{\lambda} = \pi_0 - \pi_1 \quad (\text{S10})$$

(Goeree *et al.*, 2016). Note that as $\lambda \rightarrow \infty$, the QRE solution converges to the mixed Nash equilibrium considered above. The QRE values can be found by numerically solving the above equation.

Potential games and stochastic equilibria. A game is an exact potential game if there is a real-valued function $\psi(z)$ defined on the space of strategies $z = (z_1, \dots, z_n)$ such that whenever player i unilaterally changes its strategies from z_i to z'_i , the corresponding change in his payoffs $\pi_i(z) - \pi_i(z')$ is equal exactly to the change in potential function, $\psi(z) - \psi(z')$ (Monderer and Shapley, 1996). The players of a potential game act as if they are jointly attempting to maximize the potential function. In any finite potential game, best response dynamics always converge to a Nash equilibrium (Hofbauer and Sandholm, 2002, Xu, 2016).

Consider the general public goods game with payoffs in the form $\pi_i(z) = G(z) - c_i(z_i)$. [Note that our “us vs. nature” game is a special case of this game.] Then the potential function

$$\psi(z) = G(z) - \sum c_i(z_i) \quad (\text{S11})$$

(Myatt and Wallace, 2009).

With multinomial-logit quantal response updating, the ergodic probability that the system is found in a state z is

$$\Pr(z) \big|_{t \rightarrow \infty} = \frac{\exp(\lambda \psi(z))}{\sum_{z'} \exp(\lambda \psi(z'))} \quad (\text{S12})$$

(Myatt and Wallace, 2009). The above equation allows one to find the equilibrium stochastic distribution of different strategies in “us vs. nature” games numerically.

3 Nash equilibria in “us vs. nature” games with leaders

Assume first that the monitoring effort of a leader is fixed at y . Comparing the expected payoffs of commoners given by eq. (1) and (3a) of the main text, we see that a constant level of monitoring (and punishment) effectively means that the commoners are engaged in a collective action with the benefit and cost parameters adjusted to $b(1 - \rho)$ and $c - \kappa y$, respectively, so that the relevant benefit-to-cost ratio is

$$r = b(1 - \rho)/(c - \kappa y).$$

As inequalities (S4a) show, each value of the group effort X is stable for a range of values of the benefit-to-cost ratio r . Specifically, as one increases r , the value of r at which the state with

88 $X = i - 1$ becomes unstable and the state with $X = i$ becomes a Nash equilibrium is

$$r_i = \frac{(X_0 + i)(X_0 + i - 1)}{X_0}.$$

89 Because the leaders' effort is costly, the (Nash) equilibrium value of y will be a minimum value still
90 compatible with stability of state $X = i$. This minimum value is

$$y_i = \frac{1}{\kappa} \left(c - \frac{b(1 - \rho)}{r_i} \right).$$

91 This strict Nash equilibrium state with $X = i, y = y_i$ is meaningful if $0 \leq y_i \leq 1$. The leader's
92 payoff at such a state is

$$\pi_{l,i} = bn\rho \frac{i}{X_0 + i} - c_y n y_i - \delta(n - i)y_i.$$

93 For a given set of parameters, there can be multiple Nash equilibria (i, y_i) with different expected
94 payoffs to the leader $\pi_{l,i}$ and different domains of attraction. The corresponding stochastic dynamics
95 are expected to wander among these equilibria perhaps settling predominantly on one of them. We
96 have explored these dynamics numerically in the main text.

97 4 Equilibria in “us vs. them” games without leaders

98 Consider first dyadic between-group conflicts. With two groups making efforts X and Y , respec-
99 tively, in a conflict over a resource of value $2b$, the expected payoff to a member of the first group
100 can be written as

$$\pi = \begin{cases} 2b \frac{X}{X+Y} - cx, & \text{if } X + Y > 0, \\ b, & \text{if } X + Y = 0. \end{cases}$$

101 To identify Nash equilibria in this model using the results from the previous section (specifically,
102 inequalities S4a) we need to consider all combinations of X and Y where each variable takes values
103 from $0, \dots, n$.

104 The state $X = Y = 0$ is a Nash equilibrium if $b < c$ (because in this case, b is larger than
105 $2b - c$).

106 From inequalities (S4a) with $X_0 = X$ and b substituted for $2b$, a symmetric state with $Y = X$
107 is a Nash equilibrium if

$$X - 1/2 \leq b/(2c) \leq X + 1/2. \quad (\text{S13})$$

108 That is, $X = Y = 1$ is stable for $1 < b/c < 3$, $X = Y = 2$ is stable for $3 < b/c < 5$, and so on.
109 Alternative, we can say that at the symmetric Nash equilibrium X is an integer closest to $b/(2c)$:

$$\frac{b}{2c} - \frac{1}{2} < X < \frac{b}{2c} + \frac{1}{2}.$$

110 One can also show that for $b/c = 2k + 1$ where k is an integer, there are also additional Nash
111 equilibria in the form $X = k, Y = k + 1$ and $X = k + 1, Y = k$ for $k = 0, 1, 2, \dots$. One can show
112 there are no other asymmetric Nash equilibria (i.e., with $|Y - X| > 1$).

113 In the case of G competing groups, let us define

$$\pi = \begin{cases} Gb \frac{X}{X + \sum Y} - cx, & \text{if } X + \sum Y > 0, \\ b, & \text{if } X + \sum Y = 0, \end{cases}$$

114 where $\sum Y$ is the sum of efforts of other $G - 1$ groups.
 115 State $X = 0$ is a Nash equilibrium if $Gb - c < b$, i.e. if

$$b/c < \frac{1}{G-1}.$$

116 From eq. (S4a) and b substituted for Gb , the symmetric state $X > 0$ with $X_0 = (G-1)X$ is a Nash
 117 equilibrium if

$$X + \frac{X-1}{G-1} < b/c < X + \frac{X+1}{G-1}.$$

118 Alternatively, we can write the above inequalities as

$$\left(1 - \frac{1}{G}\right) \frac{b}{c} - \frac{1}{G} < X < \left(1 - \frac{1}{G}\right) \frac{b}{c} + \frac{1}{G}.$$

119 As G increases, X becomes close to b/c . That is, with large G such an equilibrium exists only if
 120 b/c is an integer. With large G , the ranges of stability of these symmetric equilibria become very
 121 narrow.

122 There are also many asymmetric equilibria. For example, consider a focal group with $X = 0$
 123 in system where all other groups are making nonzero efforts. Nobody in the focal group will be
 124 willing to make an effort unless $b > c(\sum X + 1)$, where $\sum X$ is the sum of efforts over all groups.

125 Numerical results suggest there is a very large number of asymmetric equilibria with the average
 126 group effort $\bar{X}$ close to b/c . As the number of groups in the system grows, the system exhibits non-
 127 equilibrium dynamics (at least at the time scale of our simulations). See Figure S3 and Figure S4
 128 here.

Figure S3: Average effort X of an acephalous group in stochastic simulations in “us versus them” games for different values of parameters n, b and λ . The dashed lines show the ESS prediction (2b). Other parameters: $c = 1$, probability of strategy updating $q = 0.25$, number of groups $G = 2$. Note that X levels off at the maximum possible size $X = n$.

129 **Mixed Nash equilibria.** Assume that each individual in a group contributes independently

Figure S4: Dynamics of group efforts X under stochastic best response in “us versus them” games for different G and b values. $n = 24$. Single run for each parameter combination. Each graph shows G different lines representing efforts of G groups. Probability of updating is 0.25; $\lambda = \infty$.

an effort 1 to a collective action with probability $0 \leq p \leq 1$. The (mixed) Nash equilibrium value of p satisfies the equation for expectations

$$E(\pi_0) = E(\pi_1), \quad (\text{S14a})$$

where $E(\pi_x)$ is the expected payoff of an individual making effort x . The expected payoff of a defector can be written as

$$E(\pi_0) = Gb E\left(\frac{i}{i+j}\right), \quad (\text{S14b})$$

where $i \sim \text{Bin}(n-1, p)$ and $j \sim \text{Bin}(n-1 + (G-1)n, p)$ are independent random variables drawn from the corresponding binomial distributions. The expected payoff of a cooperator can be written as

$$E(\pi_1) = Gb E\left(\frac{1+i}{1+i+j}\right) - c. \quad (\text{S14c})$$

While the exact calculation of the expectations of the two ratios above does not seem to be possible, we can approximate these expectations using a formula based on the second order Taylor expansion

139 for the expectation of a ratio of two random variable (here x and y):

$$E(x/y) \approx E(x)/E(y) - \text{cov}(x, y)/E(y)^2 + \text{var}(y)E(x)/E(y)^3, \quad (\text{S15})$$

where $\text{cov}(x, y)$ is the covariance of x and y , and $\text{var}(y)$ is the variance of y (Stuart and Ord, 2010). This approach is justified if the variance $\text{var}(y)$ and covariance $\text{cov}(x, y)$ are both much smaller than $E(y)^2$. In our case, this assumption is satisfied if the group size n is not too small. The relevant expectations are:

$$\begin{aligned} E(i) &= (n-1)p, \\ E(i+j) &= [n-1 + (G-1)n]p, \\ \text{var}(i+j) &= [n-1 + (G-1)n]p(1-p), \\ \text{cov}(i, i+j) &= (n-1)p(1-p). \end{aligned}$$

140 Making appropriate substitutions, one ends up with a cubic equation for p which can be solved
141 numerically. Figure S5 shows the corresponding solutions for the total group effort $X^* = np^*$ as a
142 function of the benefit b . These solutions are well approximated by the ESS values refESSb given
143 in the main text.

Figure S5: Mixed Nash equilibrium values in “us vs. them” games with $G = 2$ and 100 for $n = 24, c = 1$. Equilibrium values are found numerically by solving equations (S14) with help of approximation (S15).

144 **Effects of ω on commoners’ payoffs.** The leaders’ effort y always increases with ω . To
145 understand the effect of ω on commoners in “us vs. nature” games, assume that commoners use
146 best response and that the commoners effort is described by the ESS solution (2a) of the main text

$$X^* = X_0(\sqrt{R} - 1), \quad (\text{S16})$$

147 if $R > 1$ and $X^* = 0$ otherwise. With punishment by leaders

$$R \equiv \frac{(1-\rho)b}{(c-\kappa y)X_0}.$$

148 Using equation (3a) of the main text, the total payoff of n commoners is

$$\pi_c(y, X) = (1-\rho)nbP(X) - cX - \kappa y(n-X), \quad (\text{S17})$$

Assume that $R > 1$. Taking the derivative, we find that

$$\frac{d\pi_c(y, X)}{dy} = \left(\sqrt{R} \frac{(n+1)X_0}{2} - n - X_0 \right) \kappa$$

The expression in the right-hand side of the last equation is always positive if $X_0 > 2n/(n-1)$. Therefore if commoners best response is positive, increasing y typically further increases their effort X and their payoffs.

If $R < 1$ and the best response for commoners $X^* = 0$, then increasing y always decreases their payoffs.

5 Additional agent-based simulations

.

The graphs S4-S15 below aim to show the transition from the state with no punishment and production (with smaller values of b) to punishment and production (with larger values of b). The values of b for which such a transition happens depend on the punishment parameter κ , with stronger punishment allowing for the transition at smaller benefits b . To make the patterns clearer, we only show the case of perfect precision $\lambda = \infty$. With smaller λ , there are naturally more noise in the data.

References

- Archetti, M. (2009). Cooperation as a volunteer’s dilemma and the strategy of conflict in public goods games. *Journal of Evolutionary Biology*, **22**, 2192–2200.
- Goeree, J., Holt, C., and Palfrey, T. (2016). *Quantal Response Equilibrium: A Stochastic Theory of Games*. Princeton University Press, Princeton, NJ.
- Hofbauer, J. and Sandholm, W. (2002). On the global convergence of stochastic fictitious play. *Econometrica*, **70**(6), 2265–2294.
- Monderer, D. and Shapley, L. S. (1996). Potential games. *Games and Economic Behavior*, **14**, 124–143.
- Myatt, D. P. and Wallace, C. (2009). Evolution, teamwork, and collective action: production targets in the private provisioning of public goods. *The Economic Journal*, **119**, 61–90.
- Stuart, A. and Ord, K. (2010). *Kendall’s Advanced Theory of Statistics. Volume 1*. Wiley, London.
- Xu, Z. (2016). Convergence of best-response dynamics in extensive-form games. *Journal of Economic Theory*, **162**, 21–54.

Figure S6: Equilibrium values of x, y, π_c, π_l, τ in “us vs. nature” games for different tax θ , benefit b , and cost of punishment κ parameters. Within each set, different bars correspond to different combinations of the frequencies of selective imitation E_2 and foresight E_3 in leaders. Specifically, from the left-most bar to the right-most bar the ratio $E_2 : E_3$ is equal to 0.23 : 0.01 (i.e, predominantly, selective imitation), 0.16 : 0.08, 0.12 : 0.12, 0.08 : 0.16, 0.01 : 0.23 (predominantly foresight). The frequency of random mutation $E_1 = 0.01$. Other parameters: $\lambda = \infty, n = 16, X_0 = 16, \sigma = 0.25, K = 1$ and initial values $y = 0$. With $\theta = 0$, the leader’s effort is set to 0, so the groups are acephalous. τ is measured in units of 10 time steps.

Figure S7: Equilibrium values of x, y, π_c, π_l, τ in “us vs. nature” games for different tax θ , benefit b , and cost of punishment κ parameters. Within each set, different bars correspond to different combinations of the frequencies of selective imitation E_2 and foresight E_3 in leaders. Specifically, from the left-most bar to the right-most bar the ratio $E_2 : E_3$ is equal to 0.23 : 0.01 (i.e, predominantly, selective imitation), 0.16 : 0.08, 0.12 : 0.12, 0.08 : 0.16, 0.01 : 0.23 (predominantly foresight). The frequency of random mutation $E_1 = 0.01$. Other parameters: $\lambda = \infty, n = 24, X_0 = 16, \sigma = 0.25, K = 1$ and initial values $y = 0$. With $\theta = 0$, the leader’s effort is set to 0, so the groups are acephalous. τ is measured in units of 10 time steps.

Figure S8: Equilibrium values of x, y, π_c, π_l, τ in “us vs. nature” games for different tax θ , benefit b , and cost of punishment κ parameters. Within each set, different bars correspond to different combinations of the frequencies of selective imitation E_2 and foresight E_3 in leaders. Specifically, from the left-most bar to the right-most bar the ratio $E_2 : E_3$ is equal to 0.23 : 0.01 (i.e, predominantly, selective imitation), 0.16 : 0.08, 0.12 : 0.12, 0.08 : 0.16, 0.01 : 0.23 (predominantly foresight). The frequency of random mutation $E_1 = 0.01$. Other parameters: $\lambda = \infty, n = 32, X_0 = 16, \sigma = 0.25, K = 1$ and initial values $y = 0$. With $\theta = 0$, the leader’s effort is set to 0, so the groups are acephalous. τ is measured in units of 10 time steps.

Figure S9: Equilibrium values of x, y, π_c, π_l, τ in “us vs. nature” games for different tax θ , benefit b , and cost of punishment κ parameters. Within each set, different bars correspond to different combinations of the frequencies of selective imitation E_2 and foresight E_3 in leaders. Specifically, from the left-most bar to the right-most bar the ratio $E_2 : E_3$ is equal to 0.23 : 0.01 (i.e, predominantly, selective imitation), 0.16 : 0.08, 0.12 : 0.12, 0.08 : 0.16, 0.01 : 0.23 (predominantly foresight). The frequency of random mutation $E_1 = 0.01$. Other parameters: $\lambda = \infty, n = 16, X_0 = 24, \sigma = 0.25, K = 1$ and initial values $y = 0$. With $\theta = 0$, the leader's effort is set to 0, so the groups are acephalous. τ is measured in units of 10 time steps.

Figure S10: Equilibrium values of x, y, π_c, π_l, τ in “us vs. nature” games for different tax θ , benefit b , and cost of punishment κ parameters. Within each set, different bars correspond to different combinations of the frequencies of selective imitation E_2 and foresight E_3 in leaders. Specifically, from the left-most bar to the right-most bar the ratio $E_2 : E_3$ is equal to 0.23 : 0.01 (i.e, predominantly, selective imitation), 0.16 : 0.08, 0.12 : 0.12, 0.08 : 0.16, 0.01 : 0.23 (predominantly foresight). The frequency of random mutation $E_1 = 0.01$. Other parameters: $\lambda = \infty, n = 24, X_0 = 24, \sigma = 0.25, K = 1$ and initial values $y = 0$. With $\theta = 0$, the leader’s effort is set to 0, so the groups are acephalous. τ is measured in units of 10 time steps.

Figure S11: Equilibrium values of x, y, π_c, π_l, τ in “us vs. nature” games for different tax θ , benefit b , and cost of punishment κ parameters. Within each set, different bars correspond to different combinations of the frequencies of selective imitation E_2 and foresight E_3 in leaders. Specifically, from the left-most bar to the right-most bar the ratio $E_2 : E_3$ is equal to 0.23 : 0.01 (i.e, predominantly, selective imitation), 0.16 : 0.08, 0.12 : 0.12, 0.08 : 0.16, 0.01 : 0.23 (predominantly foresight). The frequency of random mutation $E_1 = 0.01$. Other parameters: $\lambda = \infty, n = 32, X_0 = 24, \sigma = 0.25, K = 1$ and initial values $y = 0$. With $\theta = 0$, the leader’s effort is set to 0, so the groups are acephalous. τ is measured in units of 10 time steps.

Figure S12: Equilibrium values of x, y, π_c, π_l, τ in “us vs. nature” games for different tax θ , benefit b , and cost of punishment κ parameters. Within each set, different bars correspond to different combinations of the frequencies of selective imitation E_2 and foresight E_3 in leaders. Specifically, from the left-most bar to the right-most bar the ratio $E_2 : E_3$ is equal to 0.23 : 0.01 (i.e, predominantly, selective imitation), 0.16 : 0.08, 0.12 : 0.12, 0.08 : 0.16, 0.01 : 0.23 (predominantly foresight). The frequency of random mutation $E_1 = 0.01$. Other parameters: $\lambda = \infty, n = 16, X_0 = 32, \sigma = 0.25, K = 1$ and initial values $y = 0$. With $\theta = 0$, the leader's effort is set to 0, so the groups are acephalous. τ is measured in units of 10 time steps.

Figure S13: Equilibrium values of x, y, π_c, π_l, τ in “us vs. nature” games for different tax θ , benefit b , and cost of punishment κ parameters. Within each set, different bars correspond to different combinations of the frequencies of selective imitation E_2 and foresight E_3 in leaders. Specifically, from the left-most bar to the right-most bar the ratio $E_2 : E_3$ is equal to 0.23 : 0.01 (i.e, predominantly, selective imitation), 0.16 : 0.08, 0.12 : 0.12, 0.08 : 0.16, 0.01 : 0.23 (predominantly foresight). The frequency of random mutation $E_1 = 0.01$. Other parameters: $\lambda = \infty, n = 24, X_0 = 32, \sigma = 0.25, K = 1$ and initial values $y = 0$. With $\theta = 0$, the leader’s effort is set to 0, so the groups are acephalous. τ is measured in units of 10 time steps.

Figure S14: Equilibrium values of x, y, π_c, π_l, τ in “us vs. nature” games for different tax θ , benefit b , and cost of punishment κ parameters. Within each set, different bars correspond to different combinations of the frequencies of selective imitation E_2 and foresight E_3 in leaders. Specifically, from the left-most bar to the right-most bar the ratio $E_2 : E_3$ is equal to 0.23 : 0.01 (i.e, predominantly, selective imitation), 0.16 : 0.08, 0.12 : 0.12, 0.08 : 0.16, 0.01 : 0.23 (predominantly foresight). The frequency of random mutation $E_1 = 0.01$. Other parameters: $\lambda = \infty, n = 32, X_0 = 32, \sigma = 0.25, K = 1$ and initial values $y = 0$. With $\theta = 0$, the leader’s effort is set to 0, so the groups are acephalous. τ is measured in units of 10 time steps.

Figure S15: Equilibrium values of x, y, π_c, π_l, τ in “us vs. them” games for different tax θ , benefit b , and cost of punishment κ parameters. Within each set, different bars correspond to different combinations of the frequencies of selective imitation E_2 and foresight E_3 in leaders. Specifically, from the left-most bar to the right-most bar the ratio $E_2 : E_3$ is equal to 0.23 : 0.01 (i.e, predominantly, selective imitation), 0.16 : 0.08, 0.12 : 0.12, 0.08 : 0.16, 0.01 : 0.23 (predominantly foresight). The frequency of random mutation $E_1 = 0.01$. Other parameters: $\lambda = \infty, n = 16, \sigma = 0.25, K = 1$ and initial values $y = 0$. With $\theta = 0$, the leader’s effort is set to 0, so the groups are acephalous. τ is measured in units of 10 time steps.

Figure S16: Equilibrium values of x, y, π_c, π_l, τ in “us vs. them” games for different tax θ , benefit b , and cost of punishment κ parameters. Within each set, different bars correspond to different combinations of the frequencies of selective imitation E_2 and foresight E_3 in leaders. Specifically, from the left-most bar to the right-most bar the ratio $E_2 : E_3$ is equal to 0.23 : 0.01 (i.e, predominantly, selective imitation), 0.16 : 0.08, 0.12 : 0.12, 0.08 : 0.16, 0.01 : 0.23 (predominantly foresight). The frequency of random mutation $E_1 = 0.01$. Other parameters: $\lambda = \infty, n = 24, \sigma = 0.25, K = 1$ and initial values $y = 0$. With $\theta = 0$, the leader’s effort is set to 0, so the groups are acephalous. τ is measured in units of 10 time steps.

Figure S17: Equilibrium values of x, y, π_c, π_l, τ in “us vs. them” games for different tax θ , benefit b , and cost of punishment κ parameters. Within each set, different bars correspond to different combinations of the frequencies of selective imitation E_2 and foresight E_3 in leaders. Specifically, from the left-most bar to the right-most bar the ratio $E_2 : E_3$ is equal to 0.23 : 0.01 (i.e, predominantly, selective imitation), 0.16 : 0.08, 0.12 : 0.12, 0.08 : 0.16, 0.01 : 0.23 (predominantly foresight). The frequency of random mutation $E_1 = 0.01$. Other parameters: $\lambda = \infty, n = 32, \sigma = 0.25, K = 1$ and initial values $y = 0$. With $\theta = 0$, the leader’s effort is set to 0, so the groups are acephalous. τ is measured in units of 10 time steps.

References

- Archetti, M. (2009). Cooperation as a volunteer’s dilemma and the strategy of conflict in public goods games. *Journal of Evolutionary Biology*, **22**, 2192–2200.
- Goeree, J., Holt, C., and Palfrey, T. (2016). *Quantal Response Equilibrium: A Stochastic Theory of Games*. Princeton University Press, Princeton, NJ.
- Hofbauer, J. and Sandholm, W. (2002). On the global convergence of stochastic fictitious play. *Econometrica*, **70**(6), 2265–2294.

- 184 Monderer, D. and Shapley, L. S. (1996). Potential games. *Games and Economic Behavior*, **14**,
185 124–143.
- 186 Myatt, D. P. and Wallace, C. (2009). Evolution, teamwork, and collective action: production
187 targets in the private provisioning of public goods. *The Economic Journal*, **119**, 61–90.
- 188 Stuart, A. and Ord, K. (2010). *Kendall’s Advanced Theory of Statistics. Volume 1*. Wiley, London.
- 189 Xu, Z. (2016). Convergence of best-response dynamics in extensive-form games. *Journal of Eco-*
190 *nomics Theory*, **162**, 21–54.