Collective action and the evolution of social norm internalization

Sergey Gavrilets[a,b,c,1] and Peter J. Richerson[d]

[a]Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996; [b]Department of Mathematics, University of Tennessee, Knoxville, TN 37996; [c]National Institute for Mathematical and Biological Synthesis, University of Tennessee, Knoxville, TN 37996; and [d]Department of Environmental Science and Policy, University of California, Davis, CA 95616

Edited by Simon A. Levin, Princeton University, Princeton, NJ, and approved May 4, 2017 (received for review March 7, 2017)

Human behavior is strongly affected by culturally transmitted norms and values. Certain norms are internalized (i.e., acting according to a norm becomes an end in itself rather than merely a tool in achieving certain goals or avoiding social sanctions). Humans' capacity to internalize norms likely evolved in our ancestors to simplify solving certain challenges—including social ones. Here we study theoretically the evolutionary origins of the capacity to internalize norms. In our models, individuals can choose to participate in collective actions as well as punish free riders. In making their decisions, individuals attempt to maximize a utility function in which normative values are initially irrelevant but play an increasingly important role if the ability to internalize norms emerges. Using agent-based simulations, we show that norm internalization evolves under a wide range of conditions so that cooperation becomes "instinctive." Norm internalization evolves much more easily and has much larger effects on behavior if groups promote peer punishment of free riders. Promoting only participation in collective actions is not effective. Typically, intermediate levels of norm internalization are most frequent but there are also cases with relatively small frequencies of "oversocialized" individuals willing to make extreme sacrifices for their groups no matter material costs, as well as "undersocialized" individuals completely immune to social norms. Evolving the ability to internalize norms was likely a crucial step on the path to large-scale human cooperation.

cooperation | conflict | modeling | evolution | values

Human social behavior is controlled by many interacting factors including material cost–benefit considerations, genetically informed social instincts, personality, and culturally transmitted norms, values, and institutions (1–5). A social norm is a behavior that one is expected to follow and expects others to follow in a given social situation (6, 7). Humans learn norms from parents, through educational and religious practices, and from friends and acquaintances, books, and media. The adherence to norms is socially reinforced by the approval of, and rewards to, individuals who follow them and punishment of norm violators. Certain norms are internalized, that is, acting according to a norm becomes an end in itself rather than merely a tool in achieving certain goals or avoiding social sanctions (1, 2, 8–11). For individuals who have strongly internalized a norm, violating it is psychologically painful even if the direct material benefits for the violation are positive. Many individuals and groups are willing to pay extremely high costs to enact, defend, or promulgate norms that they consider important (12). At the same time, virtually all norms can be violated by individuals under some conditions (e.g., if the costs of compliance are too high). Norms thus can be viewed as one of the arguments in the utility function that each individual maximizes (9).

Internalizing a norm has two significant effects upon human behavior: People who have internalized a norm follow it even when doing so is personally costly, and they will tend to criticize or punish norm violators (13). Norm internalization allows individuals to reduce the costs associated with information gathering, processing, and decision making (11) and the costs of monitoring, punishments, or conditional rewards that would otherwise be necessary to ensure cooperation (9, 14). Internalization of norms allows individuals and groups to adjust their utility functions in situations with a rapidly changing environment when genetic mechanisms would be too slow to react (9). A society's values are transmitted through the internalization of norms (15), with some societies being more successful than others due to their norms and institutions (16). The presence of both costs and benefits of norm internalization suggests that the human ability to internalize norms has been subject to natural, sexual, and social selection for as long as human culture has been in existence.

Norm internalization is an elaboration of imitation and imprinting found in various species of birds and mammals (17). Plausibly, then, a propensity to follow norms is at least partly an innate feature of our social psychology, whereas the substantive content of the norms of a given society are largely cultural (18). Our models below are designed to explore these features of norms. There is a rapidly growing body of theoretical work on gene–culture coevolution and its effects on human behavior (3, 19). Some approaches take the capacity for internalized norms as a given and study the cultural evolution of specific norms that are internalized (e.g., ref. 20). However, except for an attempt in ref. 9 that equated internalization of norms with blind copying of a behavior from others (21, 22) while largely ignoring their associated material payoff or normative value to individuals, the question of how humans evolved to internalize norms has apparently not received much attention from theoreticians. We aim to fill this important gap in our knowledge. The key question we ask

Significance

People often ignore material costs they incur when following existing social norms. Some individuals and groups are often willing to pay extremely high costs to enact, defend, or promulgate specific values and norms that they consider important. Such behaviors, often decreasing biological fitness, represent an evolutionary puzzle. We study theoretically the evolutionary origins of human capacity to internalize and follow social norms. We focus on two general types of collective actions our ancestors were regularly involved in: cooperation to overcome nature's challenges and conflicts with neighboring groups. We show that norm internalization evolves under a wide range of conditions, making cooperation "instinctive." We make testable predictions about individual and group behavior.

Author contributions: S.G. and P.J.R. designed research; S.G. performed research; S.G. contributed new reagents/analytic tools; S.G. and P.J.R. analyzed data; and S.G. and P.J.R. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. Email: gavrila@tiem.utk.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1703857114/-/DCSupplemental.

is, "How could ... norm-using types of players have emerged and survived in a world of rational egoists?" (ref. 23, p.143).

Models and Results

We consider two general kinds of collective action problems our ancestors might have evolved to solve. The first includes group activities such as defense from predators, cooperative hunting, cooperative breeding, and so on. The success of a particular group in solving these problems does not depend very much on the actions of neighboring groups. We refer to such collective actions as "us-vs.-nature" games. The second kind of collective actions, which we refer to as "us-vs.-them" games, include direct conflicts and/or other costly competition with other groups over territory, mating opportunities, access to trade routes, and so on. The success of one group in an us-vs.-them game means failure or reduced success for other groups albeit at a cost to the winner as well. In both of these types of models, group success is an important component of fitness. Much previous modeling of group competition has not modeled this distinction but has leaned on us-vs.-them games when interpreting empirical examples just because of the costly self-sacrifice often displayed in violent conflict (3, 19). Both us-vs.-nature and us-vs.-them models will generate cooperation in the right circumstances (19, 24, 25), but modeling them in a comparative framework is instructive because the fitness payoffs to solving these two kinds of cooperative dilemmas are very different. In particular, escalation of efforts due to an intergroup arms race is common in the latter but absent in the former (*SI Appendix*). We will consider separately and contrast these two games.

We consider a population of individuals living in a large number of groups of constant size n. Generations are discrete and nonoverlapping. During their lifetime, group members have an opportunity to participate in a number of collective actions. Individual participation in collective actions is costly although any benefit is shared equally among all group members; this creates an incentive to free ride (26). An effective mechanism to reduce free riding is punishment (27–29). Therefore, we assume that individuals can punish their free-riding groupmates at a cost to themselves. Identifying free riders requires the individual to pay additional costs of monitoring the group. The costs of monitoring and punishing others, and being punished by them, increase linearly with group size n (which will vary between different simulations). Individual efforts in a collective action and in punishing free riders will be described by variables x and y, respectively, each equal to 0 or 1. As a result of participating in collective actions and punishment, individuals accumulate material payoff π. At the end of each generation, groups survive and duplicate with probabilities dependent on their success in collective actions; in surviving groups, individuals reproduce with probabilities proportional to their accumulated material payoffs. Some offspring disperse randomly to different groups.

We extend the standard approach outlined above for the case of norm internalization. We assume that the society has a prosocial (injunctive) norm in the sense that individuals learn (e.g., from parents, elders, or peers) that they are expected to contribute to collective actions and punish free riders (i.e., choose $x = y = 1$). However, individuals' decisions are controlled by both the ability to internalize the norm η and material payoff considerations. We treat η as a continuous trait controlled genetically ($0 \le \eta \le 1$). We postulate that any individual updating its behavior attempts to maximize the utility function

$$u_\eta(x, y) = (1 - \eta) \times \pi(x, y) + \eta \times (v_1 x + v_2 y). \qquad [1]$$

The two terms in Eq. **1** capture the effects of nature and culture, respectively. Individuals with $\eta = 0$ are "undersocialized" (i.e., they do not care about the norm and only want to maximize their material payoff π) (1, 2). If $\eta > 0$, following the norm is a part of the individual's preference. Individuals with $\eta = 1$ are "oversocialized" [i.e., they do not care about the material payoff and always follow the norm (1, 2) by choosing $x = y = 1$]. Nonnegative parameters v_1 and v_2 measure the maximum value of following the norms of contributing (i.e., choosing $x = 1$) and punishing free riders (i.e., choosing $y = 1$) for oversocialized individuals. These parameters increase with the strength of "social pressure" to follow the norm; we assume that they are exogenously specified. We wish to understand the evolution of η, starting with very low values, and its effects on individual and group behavior.

To study our models we used agent-based simulations. Figs. 1 and 2 illustrate observed evolutionary dynamics. In both cases shown, norm internalization trait η evolves after some time; its evolution results in increasing within-group cooperation and punishment. In the us-vs.-nature game (Fig. 1), there is an increase in material payoffs and fitness. In contrast, in the us-vs.-them game (Fig. 2), material payoffs substantially decline as group members put increasingly more effort in between-group competition. In the example of the us-vs.-nature game shown, the population becomes dimorphic in η with approximately two-thirds of the population having large values of η and the rest with very small values of η. These dynamics are analogous to those

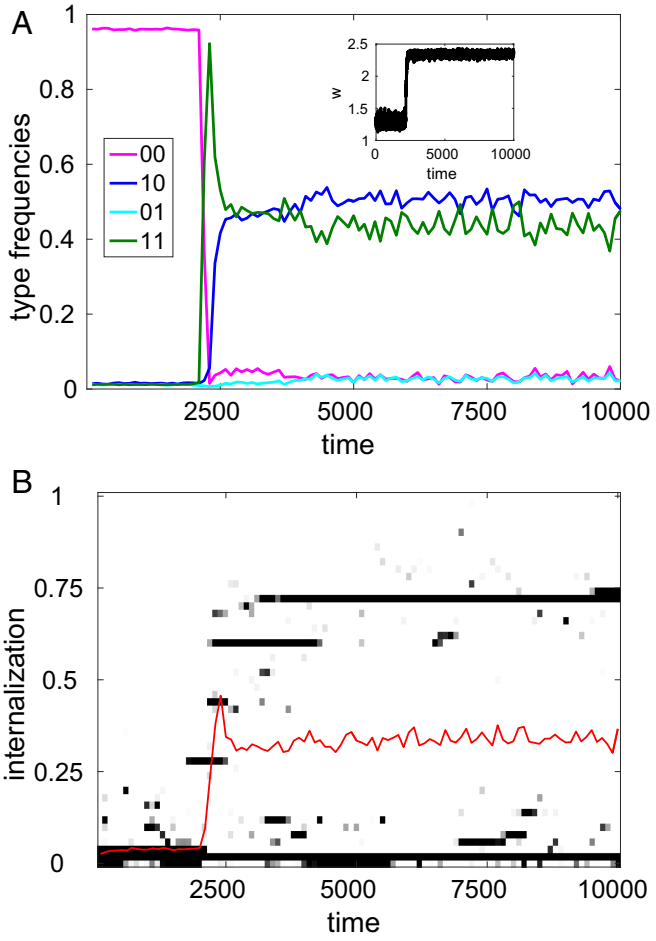

Fig. 1. Examples of evolutionary dynamics. Us-vs.-nature game with $n = 16$, $b = 4$, $v_x = 0$, $v_y = 0.5$, $X_0 = 8$, $\delta = 0.50$, $K = 4$. (*A*) Frequencies of individuals using different combinations of strategies (x, y). (*Inset*) The average fitness. (*B*) The dynamics of the distribution of the internalization trait η. The intensity of the black color is proportional to the number of individuals with the corresponding trait values present at a given time. The red line shows the mean value of η. See *Methods* and *SI Appendix* for exact definitions of parameters.

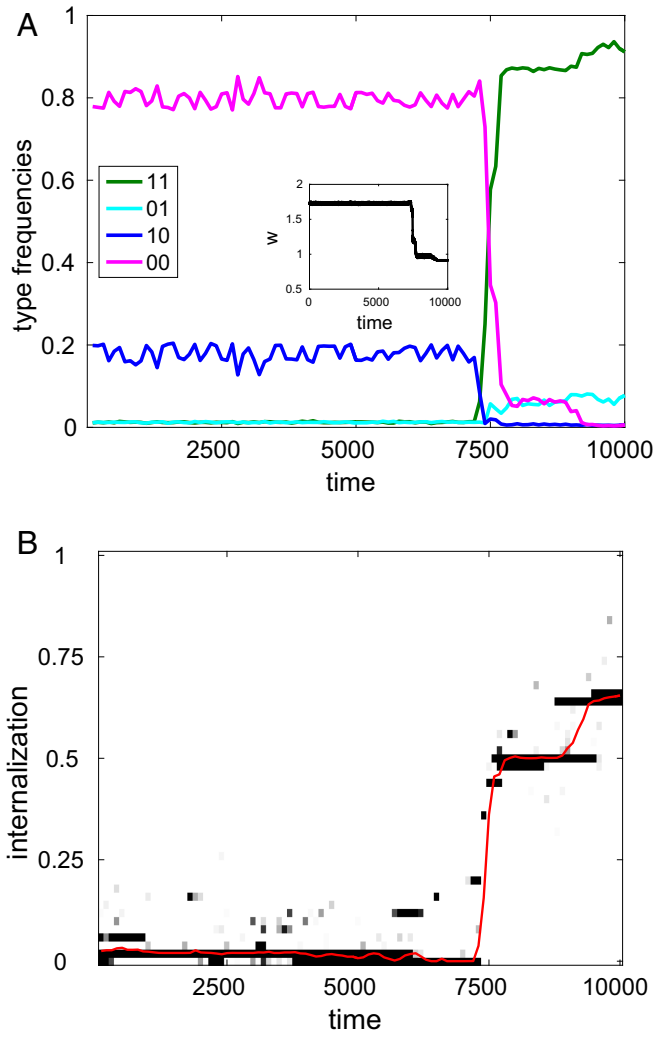

Fig. 2. Examples of evolutionary dynamics: us-vs.-them game with $n = 8$, $b = 1$, $v_x = 0$, $v_y = 0.5$, $\delta = 0.50$, $K = 2$. (*A*) Frequencies of individuals using different combinations of strategies (x, y). (*Inset*) The average fitness. (*B*) The dynamics of the distribution of the internalization trait η. The intensity of the black color is proportional to the number of individuals with the corresponding trait values present at a given time. The red line shows the mean value of η. See *Methods* and the *SI Appendix* for exact definitions of parameters.

coming under the rubric of "evolutionary branching" identified by adaptive dynamics methods (30, 31).

Overall, simulations show that if the norm internalization trait η remains small, individuals make no effort ($x = y = 0$). If norm internalization trait η instead evolves to a large value, individuals both contribute to collective action and monitor and punish occasional free riders ($x = y = 1$). With intermediate values of η, the population represents a mix of free riders ($x = y = 0$) and contributors who either punish ($x = y = 1$) or do not ($x = 1, y = 0$). The frequency of "selfish punishers" ($x = 0, y = 1$) typically remains very low. If norm internalization evolves, it often emerges quite rapidly after some waiting time, suggesting a transition between alternative quasi-stable states. Occasionally, rapid reverse transitions to low-internalization states are also observed.

The effects of parameters on evolutionary dynamics are illustrated in Figs. 3 and 4 (see also *SI Appendix*). Consider first what happens if no norm internalization is allowed to evolve [or production and peer punishment are not promoted (i.e., $v_x = v_y = 0$; left part of the first column in each set of graphs). For parameters used here, in us-vs.-nature games individuals make no effort (Fig. 3). In us-vs.-them games (Fig. 4), some individuals volunteer for production; the average group effort is small and decreases with group size. In either type of games there is no punishment.

If groups encourage their members to contribute to production but peer punishment is not promoted (i.e., $v_x > 0, v_y = 0$; left columns of graphs in Figs. 3 and 4), there is no punishment but some norm internalization evolves. In us-vs.-nature games, its effects on production and material payoffs are insignificant. In us-vs.-them games, there is a small increase in production accompanied by decreasing payoffs.

If group members are encouraged to punish free riders but production itself is not promoted (i.e., $v_x = 0, v_y > 0$; left part of the second and third columns of graphs in Figs. 3 and 4), then strong norm internalization can evolve simultaneously with a dramatic increase in production and punishment. In us-vs.-them games, increasing group size n has a strong negative effect on internalization, production, and punishment. The effects of costs of punishing and of being punished on η are relatively small (*SI Appendix*). There is a decline in fitness in us-vs.-them games but an increase in us-vs.-nature games.

If both production and punishment of free riders are encouraged (i.e., $v_x, v_y > 0$), all efforts generally increase in both games. There are nonadditive interactions of parameters but the general patterns are hard to see. Increasing the costs of being punished increases production efforts. In us-vs.-them games, smaller groups typically have higher η and production and punishment efforts than larger groups; larger groups can have higher fitness if they evolve no internalization and make fewer or no efforts in production and punishment. In us-vs.-nature games, the situation is more complex: If the cost of being punished is small, larger groups can evolve higher η (because the total effect of punishment is higher). If this cost is moderate or large, smaller groups evolve higher η. Larger groups can also have relatively high production effort. Interestingly, increasing the difficulty of us-vs.-nature games results in more norm internalization (*SI Appendix*, Table S1 and Figs. S14–S16). Increasing normative values v_x and v_y speeds up the evolution of norm internalization. Increasing v_y typically increases η, x and y, but increasing v_x can decrease η, x and y, especially in large groups. Overall, in

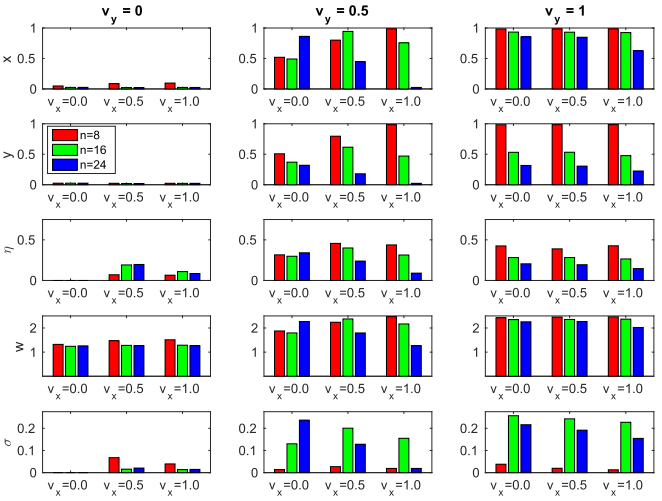

Fig. 3. Summary graphs for us-vs.-nature games: efforts x, punishment y, internalization η, fitness w, and SD σ in internalization trait η for different normative values of production v_x and punishment v_y, and group size n. Other parameters: $X_0 = n/2$, $\delta = 0.5$, $b = 4$, $K = 3$. Shown are averages based on 10 runs for each parameter combination.

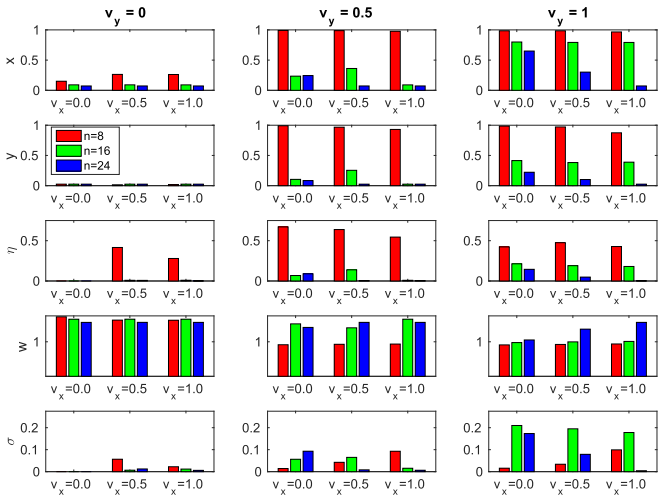

Fig. 4. Summary graphs for us-vs.-them games: efforts x, punishment y, internalization η, fitness w, and SD σ in internalization trait η for different normative values of production v_x and punishment v_y, and group size n. Other parameters: $\delta = 0.5$, $b = 1$, $K = 3$. Shown are averages based on 10 runs for each parameter combination.

us-vs.-nature games, the observed patterns in x and y resemble each other, so that cooperation and punishment come hand in hand. In contrast, in us-vs.-them games, cooperation requires less punishment. Unexpectedly, simulations often show high genetic variation in η and the emergence of different clusters of individuals with high and low η values similar to those in Fig. 1 (last row of graphs in Figs. 3 and 4).

Discussion

The collective action models considered above belong to a general class of the volunteer dilemmas (32), where individuals prefer to free ride on the effort of their groupmates but if nobody else is willing to contribute it may become advantageous to volunteer despite the costs involved. Standard theory predicts that under moderate benefit-to-cost ratios individual contributions will be completely absent in us-vs.-nature games and relatively low in us-vs.-them games (24, 33, 34).

In our extension of the standard theory, we allow for the possibility of norm internalization that dramatically changes these predictions. Norm internalization does evolve under a range of conditions greatly favoring cooperation in both types of collective action. The driving force of this evolution is both individual and group selection working to increase benefits (via increased acquisition of collective goods and group survival) and decrease associated costs of optimization, monitoring, and punishment.

From both our everyday experience and experimental work we know that people often behave in prosocial ways, despite the associated individual costs. This can happen for a number of reasons including selfless concern for the well-being of others, a desire to improve one's reputation or other strategic reason, expressing biological or social instincts evolved under ancestral conditions, or because they make errors in evaluating expected payoffs. Here we studied an additional, well-established mechanism: People behave prosocially because they duly follow or are strongly affected by an internalized social norm. Our theoretical results show that internalized norms can under some conditions trump material payoff considerations in human decision making.

Our results show that promoting costly punishment of free riders is more efficient in causing norm internalization than promoting production. Our prediction is thus that groups and societies promoting disapproval/punishment of free riders will have stronger norms and be more successful in collective actions than those promoting approval/reward of participation in production of collective goods. This can be tested using laboratory studies or ethnographic data.

Experimental public goods games with punishment show that in many populations a significant minority of people act as altruistic punishers whereas a majority of people will cooperate if there are punishers who lead the way. There is also great variation in how people react to the presence of punishers; a small minority of people are selfish maximizers, who take advantage of any cooperation unless the penalty is assured and severe (35). Recent work suggests that such between-individual differences are domain-general and temporally stable (36). Our theoretical prediction of significant genetic variation in the ability to internalize norms is compatible with these empirical results. In particular, under some conditions our models predict a relatively small frequency of oversocialized individuals—"true believers" or "heroes"—willing to make sacrifices whereas the masses express only a limited norm internalization. This modeling prediction has a simple explanation: Mixed equilibria arise when groups benefit if some but not all individuals deeply internalize very costly norms. Suicide bombers, and other displays of extreme self-sacrificial behavior, may be another example of oversocialization. One could also view oversocialized individuals as leaders who organize cooperation, along the lines of the "big men" of some small-scale societies (37). Our models also predict the maintenance of some proportion of individuals who are completely unable to internalize prosocial norms [as is observed in some psychopaths (38, 39)].

Of course, the extent of norm internalization is also affected by cultural and social factors and may change during the individual's life span (6, 7, 20). Although here we have neglected these effects for simplicity, our modeling framework is flexible enough and can be extended for the case when η depends both on genetic factors and, say, on the frequency of particular behaviors in the population.

The group sizes used here ($n = 8, 16, 24$) are within the range of those for both chimpanzees and extant hunter-gatherer bands (40, 41). Norm internalization and cooperation readily evolve if the effects of punishment are strong enough; for larger groups conditions are strict. Group sizes compatible with cooperation were much larger (up to $n = 64$) in ref. 27. However, they assumed that the costs of punishment did not depend on the group size and used very small migration rates (only up to a few percent of each group). In contrast, in our model the costs of punishment and monitoring are proportional to the group size and migration rates are realistically high [as expected with males' philopatry and random female dispersal (42)]. Evolving norm-based cooperation in large units such as tribes requires additional mechanisms, e.g., cultural group selection (25).

The extent of norm internalization depends on various parameters including the benefits, costs, group size, and intensity of between-group competition. Our models thus predict considerable variation in the strength of cultural norms across cultures that differ in their ecological and social/cultural environments (43, 44). Cultural variation in social norm internalization translates in variation in the decision-making processes. For example, the use of rule-, role-, or case-based decision making may be more common in collectivist than in individualist cultures, as evidenced by different frequencies of their use found in Chinese vs. American novels (4).

Previous direct comparisons of us-vs.-nature and us-vs.-them games show that the latter are more conducive for the evolution of cooperation (24). This conclusion was in line with recent theoretical work arguing that within-group cooperation is much easier to evolve if groups are involved in intergroup conflict and warfare (19, 24, 45). However, a case can be made that the human

population was so low up until about 50,000 years ago that intergroup conflict was rare. Perhaps only in the Holocene (i.e., over the last 12,000 years) did violent intergroup conflict become an important selective force. For example, depictions of group fighting are essentially absent in Gravettian cave art, as are depictions of defensive weapons such as shields (46). Both are common in Holocene art. Thus, human cooperation may have mainly evolved under a regime of us-vs.-nature games rather than us-vs.-them games. Our results show that cooperation via norm internalization readily evolves in us-vs.-nature games. Once the ability to internalize norms is in place, it will simplify various other types of social interactions and cooperation.

Although evolving norm internalization increases individual efforts in both types of games, its effects on material payoffs and fitness varies. The effect is predicted to be positive in us-vs.-nature games (because norm internalization implies less free riding and reduced costs of optimization). The effect is predicted to be negative in us-vs.-them games because norm internalization furthers "overproduction" and "rent dissipation" [which are well-known characteristics of contests in economics literature (47, 48)]. This negative effect will be especially large in the cases where intergroup competition takes the form of feud and warfare. Violent intergroup competition often leads to high death rates and much destruction of property. The rule of law evolved to limit the costs of such conflicts (49, 50). Thus, there is a paradox here. In us-vs.-them games, norm-driven cooperation tends to evolve more easily than in us-vs.-nature games even though the evolution of cooperation in the former drives down mean fitness and in the latter increases it. As is very often the case when strategic interactions are important, naive adaptationism would lead to the wrong conclusion here.

Institutions are the human-devised constraints and rules that structure political, economic, and social interactions (51). Examples include family, economy, education, religion, property rights, and democratic institutions. A growing number of studies show that social norms can define the success of establishment and the direction of evolution of various social institutions (52, 53). As a result, the same institutions may function differently in different cultures. Simultaneously, social institutions affect the development or nature of norms and other cultural traits (53). Building a predictive theory of social institutions is hardly possible without explicitly considering an evolved human ability to internalize social norms.

Methods

Each generation consists of Q rounds with three stages: a collective action stage, a punishment stage, and a strategy revision stage.

Collective Actions. The variable x specifies the effort of a focal individual from a particular group in a specific collective action: $x = 1$ (cooperation) and $x = 0$ (defection). The material payoff of the individual from a collective action is

$$\pi_{CA} = bP - cx, \qquad [2]$$

where b and c are constant benefit and cost parameters. Function P gives the normalized value of the resource produced or secured by the group. In us-vs.-nature games, we define $P = X/(X+X_0)$. Here $X = \sum x$ is the total group effort and X_0 is a half-success parameter (24, 34). (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.) In us-vs.-them games, we define $P = X/\overline{X}$, where $\overline{X}$ is the average group effort over all groups in the system; this is the Tullock contest success function (24, 33, 34, 47).

Punishment. Let variable y be equal to 1 if the focal individual punishes all defectors in the group. Otherwise, $y = 0$. Let δ be the cost of punishing $n-1$ free riders, κ the cost of being punished by $n-1$ groupmates, and c_{mon} the cost of monitoring $n-1$ groupmates. Then, assuming additive accumulation of benefits and costs, the material payoff of an individual using a pair of strategies (x, y) can be written as

$$\pi(x, y) = \pi_{CA} - y\,[(1-\tilde{p})\delta + c_{mon}] - (1-x)\kappa\tilde{q}. \qquad [3]$$

Here $\tilde{p}$ and $\tilde{q}$ are the frequencies of cooperators and punishers among $n-1$ other group members. The terms in the right-hand side of Eq. **3** are the net payoff of the collective action (given by Eq. **2**), the cost of punishing and monitoring others, and the cost of being punished. Note that defectors in production can still punish other defectors. Such individuals using strategy $(x = 0, y = 1)$, or (0, 1) for short, are "selfish punishers" in the terminology of ref. 28.

Norm Internalization Trait and Utility Function. We assume that the norm internalization trait η is controlled genetically by a single locus with a continuum of alleles. η remains constant during the individual's life; it is changed by random mutation during reproduction. The individual utility function $u_\eta(x, y)$ is given by Eq. **1**. Our approach is related to the one used for modeling the evolution of preferences in economics and biological literature (54–57).

Strategy Revision. After each collective action and punishment stages, with probability ν each individual updates his strategy using myopic optimization (58). Specifically, first, the individual computes four values $u_\eta(0, 0)$, $u_\eta(0, 1)$, $u_\eta(1, 0)$, $u_\eta(1, 1)$ under the assumption that all other individuals keep their strategies. Then, the individual chooses a combination of (x, y) values giving the largest utility value $u_\eta(x, y)$ with probability $1-e$, or, with probability e, chooses a random combination of (x, y). Parameter e is the error rate of optimization. This approach implies each individual knows/estimates the total contribution of his peers and the total number of punishers among them. Nonupdating individuals keep their strategies for the next round.

Biological Fitness, Group Survival, and Individual Reproduction. We define biological fitness as

$$w = 1 + \overline{\pi} - c_{opt}(1-\eta) - c_{int}\eta, \qquad [4]$$

where $\overline{\pi}$ is the average material payoff of the individual across all Q rounds. In the right-hand side of Eq. **4**, the last two terms describe the cost of finding a strategy optimizing material payoff and the genetic/physiological cost of the ability to internalize norms; c_{opt} and c_{int} are the corresponding parameters. Note that we assume that the cost of optimization decreases as the strength of norm internalization grows.

To implement selection, we use a two-level Fisher–Wright framework. Group selection is captured by making each group in the new generation independently descend from a group in the previous generation with probability proportional to their average success in collective actions $\overline{P}$ across Q rounds. Individual selection within each group is implemented by first independently choosing n parents from the group members with probabilities proportional to biological fitness w and then producing n offspring subject to random mutation. Offspring production is followed by random dispersal of half of the offspring (interpreted as females, ref. 42).

ACKNOWLEDGMENTS. We thank K. Rooker, J. van Cleve, and reviewers. This work was supported by the National Institute for Mathematical and Biological Synthesis through NSF Award EF-0830858 (to S.G.), The University of Tennessee, Knoxville, and by US Army Research Office Grant W911NF-14-1-0637 (to S.G.).

1. Wrong D (1961) The oversocialized concept of man in modern sociology. *Am Socio Rev* 26:183–193.
2. Granovetter M (1985) Economic action and social structure: The problem of embeddedness. *Am J Sociol* 91:481–510.
3. Richerson PJ, Boyd R (2005) *Not by Genes Alone: How Culture Transformed Human Evolution* (Univ of Chicago Press, Chicago).
4. Weber EU, Ames D, Blais AR (2005) How do I choose thee? Let me count the ways: A textual analysis of similarities and differences in modes of decision making in China and the United States. *Manag Organ Rev* 1:87–118.
5. Simpson B, Willer R (2015) Beyond altruism: Sociological foundations of cooperation and prosocial behavior. *Annu Rev Sociol* 41:43–63.
6. Lapinski MK, Rimal RN (2005) An explication of social norms. *Comm Theor* 15:127–147.
7. Bicchieri C (2006) *The Grammar of Society: The Nature and Dynamics of Social Norms* (Cambridge Univ Press, Cambridge, UK).
8. Axelrod R (1986) An evolutionary approach to norms. *Am Polit Sci Rev* 80:1095–1011.
9. Gintis H (2003) The hitchhiker's guide to altruism: Gene-culture coevolution, and the internalization of norms. *J Theor Biol* 220:407–418.

10. Gintis H (2003) Solving the puzzle of prosociality. *Ration Soc* 15:155–187.
11. Henrich J, Ensminger J (2014) Theoretical foundations: The coevolution of social norms, intrinsic motivation, markets, and the institutions of complex societies. *Experimenting with Social Norms: Fairness and Punishment in Cross-Cultural Perspective*, eds Ensminger J, Henrich J (Russel Sage Foundation, New York), pp 19–44.
12. Atran S, Ginges J (2013) Religious and sacred imperatives in human conflict. *Science* 336:855–857.
13. Cooter R (2000) Do good laws make good citizens? An economic analysis of internalized norms. *Va Law Rev* 86:1577–1601.
14. Buckholtz JW, Marois R (2012) The roots of modern justice: Cognitive and neural foundations of social norms and their enforcement. *Nat Neurosci* 15: 655–661.
15. Grusec JE, Kuczynski L (1997) *Parenting and Children's Internalization of Values: A Handbook of Contemporary Theory* (Wiley, New York).
16. Morris I (2015) *Foragers, Farmers, and Fossil Fuels: How Human Values Evolve* (Princeton Univ Press, Princeton).
17. Whiten A (1992) On the nature and evolution of imitation in the animal kingdom: Reappraisal of a century of research. *Adv Stud Behav* 21:239–283.
18. Chudek M, Henrich J (2011) Culture-gene coevolution, norm-psychology and the emergence of human prosociality. *Trends Cognit Sci* 15:218–226.
19. Bowles S, Gintis H (2011) *A Cooperative Species: Human Reciprocity and Its Evolution* (Princeton Univ Press, Princeton).
20. Tabellini G (2008) The scope of cooperation: Values and incentives. *Q J Econ* 123: 905–950.
21. Cavalli-Sforza LL, Feldman MW (1981) *Cultural Transmission and Evolution: A Quantitative Approach* (Princeton Univ Press, Princeton).
22. Boyd R, Richerson PJ (1985) *Culture and the Evolutionary Process* (Univ of Chicago Press, Chicago).
23. Ostrom E (2000) Collective action and the evolution of social norms. *J Econ Perspect* 14:137–158.
24. Gavrilets S (2015) Collective action problem in heterogeneous groups. *Phil Trans Roy Soc Lond B* 370:20150016.
25. Richerson P, et al. (2016) Cultural group selection plays an essential role in explaining human cooperation: A sketch of the evidence. *Behav Brain Sci* 39:e30.
26. Olson M (1965) *Logic of Collective Action: Public Goods and the Theory of Groups* (Harvard Univ Press, Cambridge, MA).
27. Boyd R, Gintis H, Bowles S, Richerson P (2003) The evolution of altruistic punishment. *Proc Natl Acad Sci USA* 100:3531–3535.
28. Eldakar OT, Wilson DS (2008) Selfishness as a second order altruism. *Proc Natl Acad Sci USA* 105:6982–6986.
29. Hauert C, Traulsen A, Brandt H, Nowak MA, Sigmund K (2007) Via freedom to coercion: The emergence of costly punishment. *Science* 316:1905–1907.
30. Geritz SAH, Kisdi E, Meszéna G, Metz JAJ (1998) Evolutionary singular strategies and the adaptive growth and branching of the evolutionary tree. *Evol Ecol* 12: 35–57.
31. Waxman D, Gavrilets S (2005) 20 questions on adaptive dynamics. *J Evol Biol* 18: 1139–1154.
32. Archetti M (2009) Cooperation as a volunteer's dilemma and the strategy of conflict in public goods games. *J Evol Biol* 22:2192–2200.
33. Gavrilets S, Fortunato L (2014) A solution to the collective action problem in between-group conflict with within-group inequality. *Nat Commun* 5:3526.
34. Gavrilets S (2015) Collective action and the collaborative brain. *J R Soc Interface* 12:20141067.
35. Kurzban R, Houser D (2005) Experiments investigating cooperative types in humans: A complement to evolutionary theory and simulations. *Proc Natl Acad Sci USA* 102:1803–1807.
36. Peysakhovich A, Nowak MA, Rand DG (2014) Humans display a "cooperative phenotype" that is domain general and temporally stable. *Nat Commun* 5:4939.
37. Smith JE, et al. (2016) Leadership in mammalian societies: Emergence, distribution, power, and payoff. *Trends Ecol Evol* 31:54–66.
38. Mealey L (1995) The sociobiology of sociopathy: An integrated evolutionary model. *Behav Brain Sci* 18(3):523–541.
39. Boddy CR (2014) Corporate psychopaths, conflict, employee affective well-being and counterproductive work behaviour. *J Bus Ethics* 121:107–121.
40. Mitani JC (2006) Reciprocal exchange in chimpanzees and other primates. *Cooperation in Primates and Humans*, eds Kappeler PM, van Schaik CP (Springer, Berlin), pp 107–119.
41. Hill KR, et al. (2011) Co-residence patterns in hunter-gatherer societies show unique human social structure. *Science* 331:1286–1289.
42. Chapais B (2008) *Primeval Kinship: How Pair-Bonding Gave Birth to Human Society* (Harvard Univ Press, Cambridge, MA).
43. Gelfand MJ, et al. (2011) Differences between tight and loose cultures: A 33-nation study. *Science* 332:1100–1104.
44. Ensminger J, Henrich J (2014) *Theoretical Foundations: The Coevolution of Social Norms, Intrinsic Motivation, Markets, and the Institutions of Complex Societies* (Russel Sage Foundation, New York).
45. Turchin P (2016) *Ultrasociety: How 10,000 Years of War Made Humans the Greatest Cooperators on Earth* (Beresta Books, Chaplin, CT).
46. Guthrie RD (2005) *The Nature of Paleolithic Art* (Univ of Chicago Press, Chicago).
47. Konrad K (2009) *Strategy and Dynamics in Contests* (Oxford Univ Press, Oxford).
48. Abbink K, Brandts J, Herrmann B, Orzen H (2010) Intergroup conflict and intra-group punishment in an experimental contest game. *Am Econ Rev* 100:420–447.
49. North DC, Wallis JJ, Weingast BR (2009) *Violence and Social Orders: A Conceptual Framework for Interpreting Recorded Human History* (Cambridge Univ Press, Cambridge, UK).
50. Fukuyama F (2009) *The Origins of Political Order: From Prehuman Times to the French Revolution* (Farrar, Straus and Giroux, New York).
51. North DC (1990) *Institution, Institutional Change and Economic Performance* (Cambridge Univ Press, New York).
52. Nunn N (2012) Culture and the historical process. *Econ Theor Dev* 27:S108–S126.
53. Alesina A, Giuliano P (2015) Culture and institutions. *J Econ Lit* 53:898–944.
54. Bester H, Guth W (1998) Is altruism evolutionarily stable? *J Econ Behav Organ* 34: 211–221.
55. Heifetz A, Shannon C, Spiegel Y (2007) The dynamic evolution of preferences. *Econ Theor* 32:251–286.
56. Akcay E, Van Cleve J, Feldman MW, Roughgarden J (2009) A theory for the evolution of other-regard integrating proximate and ultimate perspectives. *Proc Natl Acad Sci USA* 106:19061–19066.
57. Norman TWL (2012) Equilibrium selection and the dynamic evolution of preferences. *Game Econ Behav* 74:311–320.
58. Sandholm WH (2010) *Population Games and Evolutionary Dynamics* (MIT Press, Cambridge, MA).