A simple rule for the evolution of contingent cooperation in large groups

Robert Boyd *[1], Roberto H. Schonmann †[2], and Renato Vicente ‡[3]

[1]School of Human Evolution and Social Change, Arizona State University, Tempe, AZ 85287 USA
[2]Dept. of Mathematics,University of California at Los Angeles, CA 90095, USA
[3]Dept. of Applied Mathematics, Instituto de Matemática e Estatística,Universidade de São Paulo, 05508-090, São Paulo-SP, Brazil

March 27, 2014

Abstract

Humans cooperate in large groups of unrelated individuals, and many authors have argued that such cooperation is sustained by contingent reward and punishment. However, such sanctioning systems can can also stabilize a wide range of behaviors, including mutually deleterious behaviors. Moreover, it is very likely that large scale cooperation is derived in the human lineage. Thus, understanding the evolution of mutually beneficial cooperative behavior requires knowledge of when strategies that support such behavior can increase when rare. Here we derive a simple formula that gives the relatedness necessary for contingent cooperation in n-person iterated games to increase when rare. This rule applies to a wide range of payoff functions and assumes that the strategies supporting cooperation are based on the presence of a threshold fraction of cooperators. This rule suggests that modest levels of relatedness are sufficient for invasion by strategies that make cooperation contingent on previous cooperation by a small fraction of group members. In contrast, only high levels of relatedness allow the invasion by strategies that require near universal cooperation.

Keywords: cooperation, relatedness, n-person reciprocity, contingent punishment

---

*robert.t.boyd@gmail.com
†rhs@math.ucla.edu
‡rvicente@ime.usp.br

1 Introduction

Unlike other mammals, humans cooperate in large groups of unrelated individuals. Examples include warfare, the construction of roads, canals and other capital facilities, and risk buffering behaviors such as food sharing and mutual aid. It seems likely that our ability to cooperate played a crucial role in the rapid growth and spread of human populations over the last 50,000 years [1, 2]. Beginning with Trivers' seminal paper [3], many authors have argued that human cooperation is explained by reciprocity and other forms of contingent behavior. Because people can recognize a sizable number of individuals and remember their previous behavior, selection leads to a psychology in which the behavior of actors is contingent on the previous behavior of others. Individuals help only those who have helped them in the past, or punish those who don't cooperate in mutually beneficial activities. If, in the long run, benefits of sustained cooperation exceed the short term benefits of defection, then contingent strategies supporting cooperation can be evolutionarily stable. Such equilibria can explain the persistence of cooperation among unrelated individuals.

However, showing that cooperation can persist is not enough. Under plausible conditions, contingent strategies can stabilize virtually any behavior including nonadaptive and maladaptive behaviors [4]. A complete explanation must explain why contingent cooperation is a likely evolutionary outcome. Moreover, contingent cooperation, especially in sizable groups, appears to be very rare among primates [5], and thus it is very likely that the ancestral condition in the human lineage is noncooperative. This means it is not enough to explain the stability of contingent cooperation [6, 7, 8]; we must also explain how contingent strategies supporting cooperation can increase when rare. This is problematic because such strategies are altruistic when rare. Because other group members are unconditional defectors, rare contingent cooperators pay the cost of cooperation and benefit others, but do not gain any long run benefit. In a similar way, strategies that punish contingent on others punishing, must punish or make a costly signal of intent to punish in order to determine how many punishers there are in the group.

For reciprocity among pairs, kinship provides an easy solution to this problem. If interactions are repeated many times, the benefits to reciprocity can be very large. This means that rare reciprocators can increase even if they have only a small chance of interacting with another reciprocator, and thus even low levels of relatedness can allow reciprocating strategies to increase [9]. Since population structure often leads to low but positive background levels of relatedness, there is a plausible explanation for the evolution of pairwise reciprocity.

It is not clear whether relatedness can play a similar role in the evolution of contingent cooperation in larger groups. Boyd and co-authors [7, 10] have presented models which suggest that the effect of relatedness diminishes rapidly with group size. However, these models assumed that groups are formed by sampling individuals with a constant relatedness to each other. Basic models of population structure are not consistent with this assumption because the biological processes that generate relatedness lead to interdependencies so that knowing that two individuals share a gene by common descent increases the probability that other members of the group also share that gene by common descent. For a given relatedness, this increases the likelihood that groups will contain enough cooperators to sustain cooperation. As a result, existing work underestimates the possibility that contingent cooperation can increase when rare as a result of assortment due to population structure [11, 12].

Here we derive a rule that gives the relatedness necessary for contingent cooperation in n-person iterated games to increase when rare. This rule applies to a wide range of payoff functions, but requires that the strategies supporting cooperation are based on a threshold. Such strategies are common in the literature. For example, in the iterated public goods game, a plausible strategy is to cooperate during the first period, and then cooperate if a fraction θ of the $n-1$ other individuals in the group cooperated on the previous interaction, otherwise defect (e.g. [6, 7]). Plausible punishing strategies also incorporate thresholds [10, 4]. The derivation of this rule also assumes that interactions go on for a very long time, that groups are very large, and that relatedness is generated by island model population structure. We will present numerical results which suggest that this rule provides also useful estimates for the threshold relatedness when some of these assumptions are violated.

2 The Model

Individuals are drawn from a large population and interact T times in groups of size that may fluctuate, but is usually close to a common value n. During each interaction, they can express either an altruistic behavior A or a non-altruistic behavior N. Let δv_x be the incremental effect of an interaction on the fitness of an individual expressing A given that a fraction x of the individuals in the group express A. (By fitness we mean the expected number of adult offspring of an individual. The altruistic behavior affects also the fitness of individuals that do not express the altruistic behavior, but as we will see, for the purpose of deciding when the altruistic behavior can invade, the amount by which it happens is not relevant.) Here $\delta \geq 0$ is a constant that gives the strength of selection, and that we will always suppose to be small (weak selection). We make two assumptions about the payoff function v_x. First, $v_x < 0$ when $x = \frac{1}{n}$. This means that social interaction reduces the fitness of an individual behaving altruistically in a group in which no one else behaves altruistically. Second, v_x is non-decreasing in x which means that increasing the number of altruists in a group either leads to increased fitness of altruists or has no effect. Social interaction has no effect on the fitness of a non-altruist in a group in which all other individuals are also non-altruists. (For technical reasons we also assume that v_x is right-continuous and at most discontinuous on a finite number of points in $(0, 1)$.)

There are two heritable strategies. Cooperators express behavior A during the first interaction and continue to express A during future interactions if the fraction of individuals in the group expressing A during the previous interaction is greater or equal to θ. This means if the fraction of cooperators in the group is greater than θ, cooperators behave altruistically during all T interactions. We assume that $v_\theta > 0$ so that such sustained cooperation is mutually beneficial. Defectors never express the altruistic behavior. When cooperators are rare and groups are formed at random, virtually all cooperators are in groups without any other cooperators. Thus, cooperation cannot increase because cooperators experience a reduction in fitness in the first round compared to defectors, and thereafter the two types behave identically and receive no payoff.

Cooperators can increase when rare only if groups are formed assortatively so that there is some chance that they benefit from long term cooperation. However, knowing the coefficient of relatedness within groups (R) alone is not, in principle, enough to determine whether cooperation can increase because fitness is a nonlinear function of the frequency of cooperators in a group; to calculate the expected fitness of rare cooperators the entire probability distribution of frequencies is required [11, 12, 13]. Here we assume that groups are linked by migration in an island model so that each generation a fraction m of each group is replace by immigrants drawn at random from the population as a whole. We suppose that either the competition among individuals is mostly global, or if it has an important local component, then group size is sufficiently elastic to allow for average group fitness to increase with the number of cooperators in the group, and the excess production of groups with many cooperators to be released in the population at large, through migration. This is the regime that in [12] we called Hamilton regime, and which allows for the proliferation of costly cooperation, even when competition is purely local. As explained in [12], the needed variability in group size is of order δ, so that when selection is weak groups still have typical size close to n, and the computation of the distribution of alleles in groups can be performed assuming a fixed group size n, as we do next. (This differs from the inelastic island model [14] which assumes completely fixed group sizes and thus cannot accomodate average group fitness different from 1 and group-altruistic behavior.) If selection is weak, groups are large, migration rates low and cooperators are rare, we argued in [12] that the distribution of the frequency of cooperators in the group of a focal cooperator is approximately given by a beta probability density with parameters 1 and $l = 2n_{\mathrm{e}}m$, where the effective population size is given by $n_{\mathrm{e}} = n/\sigma^2$, and σ^2 is the variance of the number of adult offspring that each individual produces. (For the computation of the effective population size, see [15], pp.105,6 and note that he uses a group size $N = 2n$.) In the benchmark case of Fisher-Wright reproduction in the groups, the offspring distribution is Poisson with mean 1, and $n_{\mathrm{e}} = n$. The claim above, on the distribution of alleles in the group, follows from the fact that for the island model, the probability density for the fraction x of cooperators in a randomly chosen group is $f(x) = \mathrm{beta}(x|lp, l(1-p)) = Cx^{lp-1}(1-x)^{l(1-p)-1}$ where p is the frequency of cooperators in the population and C is a normalization constant [16, 17, 15, 18]. The joint density that the focal is a cooperator and the fraction of cooperators is x is $xf(x)$. Therefore, the conditional probability density that the focal is in a group with a fraction x of cooperators, given that the focal is a cooperator is proportional to $xf(x)$. When cooperators are rare $p \to 0$, and the probability density for the fraction of cooperators in the group of a focal conditioned to be a cooperator is therefore, with the proper normalization, $l(1-x)^{l-1} = \mathrm{beta}(x|1, l)$. This result holds for more general population structures [11, 12], with the island model as an important example, and thus the formulas derived here hold more widely.

The direction of selection is obtained by comparing the average fitness of a focal cooperator with that of a focal non-cooperator. When cooperators are rare, noncooperators are typically in groups with no cooperators, and hence, in the average, do not benefit from cooperation. Thus, the cooperative strategy can increase when rare, whenever the expected incremental fitness of cooperators due to social interaction is positive. This means that the minimum value of R for cooperation to spread when rare is implicitly defined by

$$\int_0^1 \mathrm{beta}(x|1, l)v_x dx + (T-1)\int_\theta^1 \mathrm{beta}(x|1, l)v_x dx = 0. \tag{1}$$

The first term gives the expected payoff to cooperators during the first interaction, and the second term the additional expected payoff to cooperators in groups in which the frequency of cooperators is greater than or equal to θ and therefore cooperation is sustained. When the left hand side of (1) is greater than zero, cooperators will increase in frequency, and thus the value of R that causes (1) to be satisfied is the minimum necessary for cooperation to increase.

Equation (1) can be used to compute the migration rate (m) needed for rare cooperators to proliferate, assuming a fixed incremental payoff per interaction v_x, expected number of interactions T and effective group size n_{e}. When these parameters yield a small value of R, then $l \approx \frac{1}{R} - 1$ is large and (1) simplifies considerably. To see why, recall that the beta distribution has the density $\text{beta}(x|1, l) = l(1-x)^{l-1}$. This means that the probability that the frequency of cooperators in a group is greater than θ is $(1-\theta)^l$, and thus we can rewrite (1) as

$$\begin{aligned}(T-1)(1-\theta)^l \int_\theta^1 \text{beta}(x|1, l, x \geq \theta) v_x dx \\ = - \int_0^1 \text{beta}(x|1, l) v_x dx, \qquad (2)\end{aligned}$$

where $\text{beta}(x|1, l, x \geq \theta)$ is the density conditioned on $x \geq \theta$. Observe that $[(1-\theta')/(1-\theta)]^l \to 0$ as $l \to \infty$ if $\theta' > \theta$. This implies that as l becomes large, the distribution of the unconditioned beta converges to that of a distribution concentrated on 0, and the distribution of the beta conditioned on being in $[\theta, 1]$ converges to a distribution concentrated on θ. Thus the integrals in (2) converge, respectively, to v_θ and v_0 as l becomes large. This implies then

$$(T-1)(1-\theta)^l v_\theta \approx -v_0. \qquad (3)$$

Solving (3) for l provides the following approximations for $m = \frac{l}{2n_\text{e}}$

$$m \approx \frac{\ln\left(\frac{(T-1)v_\theta}{-v_0}\right)}{-2n_\text{e} \ln(1-\theta)}. \qquad (4)$$

The critical value of R can then be computed using the expression $R = (1-m)^2/(n_\text{e} - (n_\text{e}-1)(1-m)^2)$, valid for arbitrary n_e and m, and that is approximately $1/(1+2n_\text{e}m)$ when m is small. When this approximation holds, we obtain from (4).

$$R \approx \frac{-\ln(1-\theta)}{\ln\left(\frac{(T-1)v_\theta}{-v_0}\right) - \ln(1-\theta)}. \qquad (5)$$

Error estimates (derived in the SI under mild additional conditions) indicate that the error is of order R^2. In the SI, we also derive results for more general situations, including the special case in which $v_\theta = 0$. Because v_x is non-decreasing in x, v_θ and v_0 are lower bounds on the integrals in (2), and therefore (5) gives an upper bound on the critical value of R.

From the arguments above, we expect (4) and (5) to be good approximations when n is large and R is small which means that $\frac{Tv_\theta}{-v_0}$ must be large. Numerical calculations using the methods outlined in [11] show that, somewhat surprisingly, these formulas (especially the former) also give reasonably accurate approximations for the critical values of m and R when n and T take on modest values. Assuming Fisher-Wright reproduction in the groups, and weak selection, the probability distribution π_k of the number k of cooperators in the group of a focal cooperator can be computed numerically for any values of n and m ([19] ((A17) in the Appendix), [11]), and then this distribution can be used to calculate the expected payoff $\sum_{k=1,\ldots,n} \pi_k v_{k/n} + (T-1) \sum_{k \geq \theta} \pi_k v_{k/n}$ to the focal cooperator, and the critical value of m that equates this expected payoff to 0. This is the analogue of (1) when n is finite. Figures 1 and 2 compare the approximation for the threshold value of R obtained from (4) to the exact values for two payoff functions: the linear public goods game in which $v_x = bx - c$ and a threshold function in which $v_x = -c$ for $x < \theta$ and $v_x = b - c$ for $x \geq \theta$. These two functions encompass a range of plausible payoff functions (e.g., [8, 10]). The approximation is very good for large n and T and provides a useful upper bound for smaller values. It is always better for the threshold payoff function than for the linear payoff function because the threshold function is the lowest possible payoff function compatible with a given value of v_0 and v_θ and a non-decreasing v_x, and therefore for a given θ, it maximizes the threshold R among these payoff functions. Figure 1 has $\theta = 0.4$, meaning that cooperation only continues after the first iteration if at least 40% of group members cooperate. Still the levels of relatedness required for cooperation to proliferate when rare are modest, even for large groups. In contrast, for large values of θ the threshold values of R are much larger. This is illustrated in Figure 2, where $\theta = 0.8$.

3 Discussion

Equation (5) gives a simple approximation for the level of relatedness necessary for cooperative strategies to increase when rare. It does not depend on the form of the underlying fitness function, but does depend on two key assumptions: (1) Contingent strategies that support cooperation lead individuals to cooperate if the number of cooperators exceeds a threshold, and (2) the assortment necessary for cooperative strategies to increase when rare results from island model population structure (or other population structures, such as the two level Fisher-Wright population structure studied in [11], for which (1) holds, [12]). The analytical derivation of (5) also depends on the assumption that groups are very large and interactions persist for a long time. However, numerical calculations suggest that (5) also gives a useful approximation for the necessary amount of relatedness when group size is modest and interactions only occur relatively infrequently. Moreover, the approximate values of relatedness obtained from (5) are always upper bounds on the level of relatedness needed for cooperation to spread when rare. This means that they represent worse case scenarios for the spread of cooperation.

The approximate expression (5) provides three insights about the evolution of contingent cooperation in sizable groups. First, notice that the first term in the denominator is the logarithm of the ratio of long run benefit to cooperators in groups with the threshold number of cooperators, $(T-1)v_\theta$, to the short term cost of cooperating in groups with fewer cooperators, $-v_0$. Social interactions which generate a substantial positive benefit at the threshold compared to the cost of unrewarded cooperation, or that are repeated frequently, can increase at lower levels of relatedness. Second, $\ln((T-1)v_\theta/(-v_0)) = \ln(T-1) + \ln(v_\theta/(-v_0))$. For interactions that are repeated many times it seems plausible that $\ln(T-1) \gg \ln(v_\theta/(-v_0))$, and when this is the case the threshold relatedness will be almost independent of the per interaction payoffs as long as $v_\theta > 0$. If also θ is small, then $R \approx \theta/\ln(T)$. Third, and most important, when $\theta \to 1$, $-\ln(1-\theta)$ grows rapidly with θ and diverges. This means that, all other things being equal, $R \to 1$ in this limit, and therefore strategies that require high levels of cooperation to persist will increase when rare only when relatedness is high. In contrast strategies that cooperate even when only a minority of other group members cooperate can increase at much lower levels of relatedness.

This last result suggest that high levels of cooperation are more likely to be supported by contingent punishment than by contingent cooperation. Strategies that continue cooperating even when only a small fraction of others cooperate typically reach a stable polymorphic equilibrium in which the population displays a mix of cooperative and noncooperative strategies [6, 7]. Strategies that tolerate more defectors achieve lower frequencies of cooperators at equilibrium. Thus reciprocating strategies that behave altruistically when a small fraction of the group also behave altruistically can support on going cooperation, but will produce equilibria in which most individuals in the group do not contribute. Cooperation of this kind is observed. For example, in the United States public radio is supported by voluntary contributions by a small fraction of listeners — most free ride. However such strategies cannot support the widespread cooperation observed in many contexts. For example, virtually all Turkana men participate in warfare, even though the Turkana lack formal coercive institutions [20]. Our result is consistent with the idea that wide spread cooperation is supported by punishment of noncooperators. In the model presented in [10], individuals punish non-cooperators if enough other individuals in the group are also willing to punish non-cooperators. Because even a modest fraction of punishers can motivate others to cooperate, such contingBoydGintisBowles10ent punishment strategies can increase when rare at relatively low levels of relatedness and still stabilize cooperation at a high level.

4 Acknowledgments

We are grateful to Clark Barrett, Marek Biskup, Sam Bowles, Nestor Caticha, Maciek Chudek, Daniel Fessler, Kevin Foster, Willem Frankenhuis, Herb Gintis, Bailey House, Anne Kandler, Laurent Lehmann, Glauco Machado, Sarah Mathew, Diogo Meyer,Cristina Moya, Peter Nonacs, Karthik Panchanathan,Susan Perry,Joan Silk, Jennifer Smith, Jeremy van Cleve, and Ming Xue for nice conversations and feedback on various aspects of this project and related subjects. This work was partially supported by CNAIPS-USP.

References

[1] Hill, K., Barton, M. and Hurtado, A. M. 2009. The emergence of human uniqueness: Characters underlying behavioral modernity, *Evolutionary Anthropology*, 18, 187–200.

[2] Wilson, E. O.2012. *The Social Conquest of Earth*, W.W. Norton, NY.

[3] Trivers, R. 1971. The evoluton of reciprocal altruism, *Quarterly Review of Biology*, 46, 35-57

[4] Boyd,R. and Richerson, P. J. 1992. Punishment allows the evolution of cooperation (or anything else) in sizable groups, *Ethology and Sociobiology*, 13, 171–195.

[5] Clutton-Brock, T. 2009. Cooperation among non-kin in animal societies, *Nature*, 462, 51–57.

[6] Joshi, N. V. 1987. Evolution of cooperation by reciprocation within structured demes. *Journal of Genetics* 66, 69-84.

[7] Boyd, R. and Richerson,P. J. 1988. The evolution of reciprocity in sizable groups, *Journal of Theoretical Biology*, 132, 337–356.

[8] Archetti, M. and Scheuring, I. 2010. The Coexistence of cooperation and defection in public goods games, *Evolution*, 65, 1140-1148.

[9] Axelrod, R. and Hamilton, W. D. 1981. The evolution of cooperation, *Science*, 211, 1390–1396.

[10] Boyd, R., Gintis,H. and Bowles, S. 2010. Coordinated punishment of defectors sustains cooperation and can proliferate when rare, *Science*, 328, 617–620.

[11] Schonmann, R.H., Vicente, R. and Caticha, N. 2013. Altruism can proliferate through population viscosity despite high random gene flow. *PLoS One* 8: e72043.

[12] Schonmann, R.H., Boyd, R. and Vicente, R. 2014. The evolution of cooperation under local regulation and non-additive gene action: an extension of Hamilton's fundamental ideas. Preprint.

[13] van Veelen, M. 2009. Group selection, kin selection, altruism and cooperation: when inclusive fitness is right and when it can be wrong. *Journal of Theoretical Biology*, 259, 589-600.

[14] Taylor, P.D. 1992. Altruism in viscous population – an inclusive fitness approach. *Evolutionary Ecology* 6, 352-356.

[15] Ewens, W.J. 1979. *Mathematical Population Genetics*, Springer Verlag (Berlin, Heidelber, New York).

[16] Wright, S. 1931. Evolution in Mendelian populations. *Genetics*, 16, 97-159.

[17] Crow, J.F. and Kimura, M. 1970. *A*n Introduction to Population Genetics Theory. Harper & Row (New York).

[18] Wakeley, J. 2002. Polymorphism and divergence for island-model species. *Genetics*, 163, 411-420.

[19] Gardner, A. and West, S.A. 2009. Greenbeards. *Evolution* 64, 25-38.

[20] Mathew, S. and Boyd, R. 2011. Punishment sustains large-scale cooperation in prestate warfare, *Proc. Nat. Acad. Sci. USA*, 108, 11375-11380.

Figure 1: The exact threshold relatedness necessary for cooperation to increase for a threshold payoff function and a linear payoff function (IPG) compared to the approximate value obtained from (4) and $R = (1-m)^2/(n-(n-1)(1-m)^2)$ as a function of group size for $\theta = 0.4$. On the left $T = 10$, in the middle $T = 100$, and on the right $T = 1000$. In all figures $v_\theta = 1$ and $v_0 = -1$. The values of R given by (5) are the limits of the approximating curves as n becomes large (right side of the graphs), and are, respectively, 0.19, 0.10 and 0.069. When (1) only a minority of cooperators is necessary to sustain cooperation, (2) groups are of modest size, and (3) interactions are repeated many times, only low levels of relatedness are necessary for coperative strategies to increase.

Figure 2: The exact threshold relatedness necessary for cooperation to increase for a threshold payoff function and a linear payoff function (IPG) compared to the approximate value obtained from (4) and $R = (1-m)^2/(n-(n-1)(1-m)^2)$ as a function of group size for $\theta = 0.8$. On the left $T = 10$, in the middle $T = 100$, and on the right $T = 1000$. In all figures $v_\theta = 1$ and $v_0 = -1$. The values of R given by (5) are the limits of the approximating curves as n becomes large (right side of the graphs), and are, respectively, 0.42, 0.26 and 0.19. When most individuals have to cooperate to sustain cooperation, substantially higher levels of relatedness are necessary for cooperative strategies to increase when rare.