RESEARCH ARTICLE

Socially Enforced Nepotism: How Norms and Reputation Can Amplify Kin Altruism

Doug Jones*

Department of Anthropology, University of Utah, Salt Lake City, Utah, United States of America

* douglas.jones@anthro.utah.edu

Abstract

Kin selection, which can lead organisms to behave altruistically to their genetic relatives, works differently when—as is often the case in human societies—altruism can be boosted by social pressure. Here I present a model of social norms enforced by indirect reciprocity. In the model there are many alternative stable allocations of rewards ("distributional norms"); a stable norm is stable in the sense that each player is best off following the norm if other players do the same. Stable norms vary widely in how equally they reward players with unequal abilities. In a population of mixed groups (some group members follow one norm, some follow another, and some compromise) with modest within-group coefficients of relatedness, selection within groups favors those who compromise, and selection between groups favors generous generalized reciprocity rather than balanced reciprocity. Thus evolved social norms can amplify kin altruism, giving rise to a uniquely human mode of kin-based sociality distinct from spontaneous altruism among close kin, or cooperation among non-kin.

Introduction: Beating Hamilton's Rule

This article presents a souped-up version of the theory of kin selection. In the standard version, an organism has a chance to provide benefit b to one of her kin at cost c to herself. Natural selection favors this altruistic act as long as Hamilton's Rule is satisfied, $r \cdot b > c$, where r is the coefficient of relatedness of recipient to donor [1]. But in the theory presented here, altruism toward kin is influenced by a wider social context, as members of a group play a game in which each player can reward others for behaving generously toward their mutual kin. I show that this can favor the evolution of high levels of altruism, in the form of unbalanced generalized reciprocity among kin, even when r is small.

The theory of socially enforced nepotism should be of interest to scholars in several disciplines [2, 3]. It bears on arguments among evolutionary theorists about how useful inclusive fitness theory is in accounting for social behavior, among social scientists about the relationship between rational choice and social norms, and among cultural anthropologists about how far their theories of kinship can be reconciled with evolutionary theories.

In the rest of this introduction, I summarize theoretical and empirical reasons for wanting to add a model of social norms and their enforcement to the theory of kin selection. On the

theoretical side: Kin selection theory is sometimes equated with the proposition that organisms are shaped by natural selection to maximize their inclusive fitness. But this result rests on the assumption that for each actor in a group, we can separate out an independent contribution that she makes to the well-being of her kin [4]. This amounts to treating kin selection as a collection of one-person games.

However kin selection gets more complicated when there is more in the way of strategic interaction among actors. Consider a simple model which combines game theory and population genetics, the Brothers Karamazov Game [5]. The game is played among three brothers, with a coefficient of relatedness $r = 1/2$ for each pair. One brother is a potential recipient of altruism, and the other two are potential donors, so we are looking at a two-person game between the two donors regarding how they treat the third. To establish a benchmark for comparison, we begin with each of the two potential helpers deciding independently of the other one whether to help their needy sibling. In this case, the evolution of altruism is governed by Hamilton's Rule. Each brother should help as long as his cost c and the needy brother's benefit b satisfy $1/2 \cdot b > c$. But now consider the case where the two potential helpers adopt a strategy of *conditional nepotism*: each will provide extra help as long as the other does the same. It is tempting to analyze this package deal using inclusive fitness theory, with costs and benefits to one's siblings counting for half as much as costs to oneself. By this accounting, with two brothers taking turns, each on his turn paying cost c and providing benefit b to the third, the rule for altruism is $2/3 \cdot b > c$. But a more careful analysis using gene and genotype frequencies gives a different answer: the condition for the evolution of altruism is $r_C \cdot b > c$, where $r_C = (1 + 6p)/(2 + 8p)$, with p being the frequency of the allele for conditional altruism. The conditional coefficient of relatedness r_C increases from 1/2 to 7/10 as p increases from 0 to 1. A conditional nepotism gene cannot invade a population when very rare, but can spread once it gains a foothold.

The Brothers Karamazov Game demonstrates several things. First, it shows the limitations of inclusive fitness accounting. The calculation based on the genealogical coefficient of relatedness of 1/2 between siblings doesn't work, because the probability that the two helpers are identical-by-descent for the conditional nepotism allele is related in a complicated way to the strategies each chooses. While inclusive fitness theory can offer qualitative insights, it takes an explicit population genetic calculation to get the correct answer. Second, the model provides a proof-of-concept that a socially enforced rule can amplify kin altruism. The conditional strategy can lead each donor brother to behave *as if* he were closer kin than a brother. One way to understand this is to note that altruism toward kin is a public good. When one brother provides help, he is also giving a free boost to the inclusive fitness of the other potential helper, and vice versa. As with other public goods, more is provided when players reduce free-riding by enforcing agreed-upon rules. In this case, if we gauge kin altruism by c/b, then conditional nepotism beats individual nepotism following Hamilton's Rule by up to 40 percent.

The Brothers Karamazov Game is perhaps the simplest possible model of socially enforced nepotism. The bulk of this article explores a more general model of an n-person game that differs from the Brothers Karamazov Game in several respects. First, the model here is not built around a simple two-fold division of players into those who help and those who are helped. Instead it includes players with a range of helping abilities. Second, players in the model do not rely on a simple conditional strategy to boost altruism to kin. Such a strategy ("I'll help if you do too") works for two helpers. But the n-person conditional strategy ("I'll help if everyone else does too") breaks down for n much greater than two, because it imposes an impossibly stringent unanimity requirement [6]. Instead, players rely on indirect reciprocity and honest advertising of ability to enforce norms of sharing. The model lets us explore the contrast, and the evolutionary contest, between norms of balanced reciprocity (where

player *j* helps player *i* as much as *i* helps *j*) and generalized reciprocity (where *j*, concerned with her reputation among other players, helps *i* even if *i* cannot reciprocate). Many anthropologists regard generalized reciprocity, or communal sharing [7], as a hallmark of relations among kin.

This brings us to the empirical motivation for the present work. Long before evolutionary biologists devised the theory of kin selection, cultural anthropologists developed their own theory of kinship—perhaps the discipline's single greatest contribution to the human sciences. The study of kinship is important because, as one scholar writes, "In many societies, both primitive and sophisticated, relationships to ancestors and kin have been the key relationships in the social structure; they have been the pivots on which most interaction, most claims and obligations, most loyalties and sentiments, turned" [8]. The study of kinship has resulted in a number of important findings; of special interest here is a generalization regarding the connection between kinship and altruism: "Kindred and kindness go together—two words whose common derivation expresses . . . one of the main principles of social life" [9]. In other words,

> the central value premise associated with [the notion of kinship] is uniform. Kinship predicates the axiom of amity, the prescriptive altruism exhibited in the ethic of generosity. . . . [T]he rule posits . . . that "kinsfolk" have irresistible claims on one another's support and consideration . . . simply by reason of the fact that they are kin. Kinsfolk must ideally share . . . and they must, ideally, do so without putting a price on what they give. [10]

The observation that people act altruistically toward their kin even when they can expect nothing in return sounds like what the theory of kin selection predicts. But there is a crucial complication: human kinship has a normative side, as implied by the words "prescriptive," "ethic," "rule," and "ideally" above. How one person treats others, including her kin, is partly guided by her personal feelings toward them but also partly dictated by social conventions. This means, according to many anthropologists, that human kin altruism comes in two varieties. (Much of the evidence regarding the two varieties of nepotism comes from a large body of qualitative ethnographic studies, but the distinction has also been upheld by recent quantitative research [11, 12].) One variety (call it *individual nepotism*) is motivated by benefactors' feelings of attachment to their beneficiaries. It is largely directed to close kin (consistent with inclusive fitness theory, since coefficients of relatedness fall off rapidly with genealogical distance). The second variety (call it *socially enforced nepotism*) is governed by conventions about how kin ought to be treated, and is motivated by social pressure and concern for one's standing in a moral community. It can extend even to distant kin. "[T]he most convincing evidence . . . of kinship amity [is found in] situations where kinship is so tenuous as to be only nominal, as when persons seek out remote clansfolk . . . and without further ado claim and receive hospitality and protection." [10] It is socially imposed nepotism that helps make kinship important not just for family life, but for social organization on a larger scale.

Individual nepotism looks a lot like kin altruism in other species [13]. Socially enforced nepotism is different. It depends on the moral regulation of behavior according to socially transmitted norms—something regarded by many social scientists, including some evolution-minded ones, as a distinguishing feature of our species [14–16]. Hence the model below, combining game theory and population genetics, in which the starting point is not an individual deciding on her own how much to help her kin, but a group of unequal individuals settling on and enforcing a norm determining who gets how much.

Results

Distributional norms, reputation, and incentives

Imagine social life as a drama [17]. Each individual plays the role(s) appropriate to her status, treating other players as directed by her role and theirs, with all players acting from a shared script. As part of the drama, real costs are paid and benefits received, so a player has an extrinsic motive to stick to the script: if she breaks character, other players react so as to leave her with lower net benefits. Here I translate this dramaturgical metaphor into a game-theoretical model in which players use indirect reciprocity and reputation to get one another to play their parts. In the model, players differ in their ability to help others (= status); they help and are helped based on their advertised abilities (= assumed roles); and each has an incentive to honestly advertise her ability by following a distributional norm (= shared script).

Consider a game among a group of n players. In each round of play, each player has the opportunity to help some randomly selected target players. Costs and benefits of helping are related by the production function

$$b_{ij} = \alpha_j^{1-\tau} c_{ji}^{\tau} \qquad (1)$$

where b_{ij} is the benefit that i receives from j, c_{ji} is the cost that j incurs in helping i (both averaged over a long span of play), α_j is a measure of j's productive ability, and τ generates diminishing marginal returns to helping, with b_{ij}, c_{ji}, and $\alpha_j \geq 0$, and $0 < \tau < 1$.

We are interested in indirect reciprocity, where player j produces benefits for player i based on the benefits i has produced for other players in earlier rounds of play [18–25]. Who produces how much for whom—the value of b_{ij}—depends on *distributional norms* and *reputations*. Let's assume initially that members of a group agree on a single distributional norm, call it A, defined by a function $b^A[x, y]$ over $x \geq 0$, $y \geq 0$. This function dictates how much benefit a player with ability x ought to get from a player with ability y. If i and j have abilities α_i and α_j, and j follows the norm, then $b_{ij} = b^A[\alpha_i, \alpha_j]$. Suppose however that j cannot assess i's ability directly, but must act according to an ascribed ability, or reputation, for i. To achieve this, the function b^A must made to operate in two directions, determining both the benefits j produces for other players *and* the reputation others assign her. First, if j knows her own ability, α_j, and ascribes ability s_i to player i, she can reward i according to $b_{ij} = b^A[s_i, \alpha_j]$. Second, if some player k knows what benefits j has produced for other players, and has ascribed abilities for them, she can work backward to assign a new reputation, s_j, to j, where s_j is implicitly given by b^A (as spelled out below). This means that reputation is defined recursively: the reputation of each player depends on the benefits she bestows on others and on their reputations. The result, at equilibrium, is that reputation equals ability ($s_j = \alpha_j$) and players receive benefits according to $b_{ij} = b^A[\alpha_i, \alpha_j]$ and pay costs according to $c_{ji} = \alpha_j^{-\frac{1-\tau}{\tau}} b^A[\alpha_i, \alpha_j]^{1/\tau}$, for all i and j.

In principle, any norm with non-negative $b^A[x, y]$ is allowed, including a norm of pure altruism, "from each according to his ability, to each according to his need" (Fig 1A). But here we are especially interested in the class of incentive-compatible stable norms, which motivate each self-interested player to honestly advertise her ability by following the norm, provided others do likewise. A stable norm and associated reputation function must satisfy several conditions.

(i) *Positive net benefits*. For each active player j, summed benefits minus summed costs averaged over many rounds of play must be greater than zero.

$$\sum_k b^A[\alpha_j, \alpha_k] - \alpha_j^{-\frac{1-\tau}{\tau}} \sum_i b^A[\alpha_i, \alpha_j]^{1/\tau} > 0 \qquad (2)$$
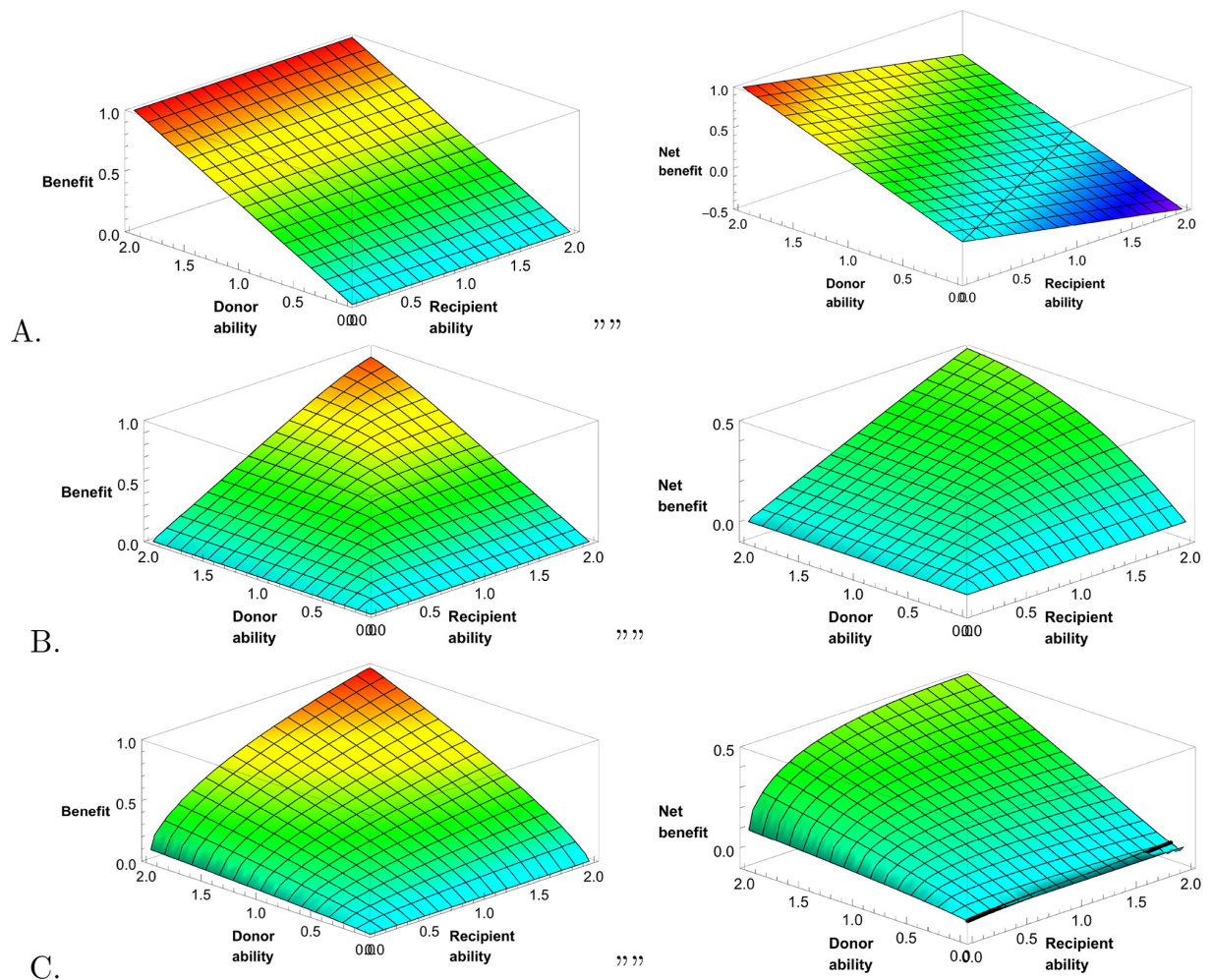
Fig 1. Three distributional norms. How benefits should be distributed as a function of recipient (x) and donor (y) ability, according to norms of (A) Pure Altruism, (B) Almost Balanced Reciprocity, and (C) Generalized Reciprocity. On left, benefits, $b[x, y]$; on right, recipient net benefits, $b[x, y] - c[x, y]$. Colors scale with benefits or net benefits, 1.0 is red, 0.0 is light blue. Black lines show where net benefits equal zero.

doi:10.1371/journal.pone.0155596.g001

(ii) *Penalty for impersonation*. For each active player j, her summed net benefits must be greater if she honestly advertises her ability, α_j, than if she "impersonates" a player of lesser or greater ability, y:

$$\sum_k b^A[\alpha_j, \alpha_k] - \alpha_j^{-\frac{1-\tau}{\tau}} \sum_i b^A[\alpha_i, \alpha_j]^{1/\tau}$$
$$> \sum_k b^A[y, \alpha_k] - \alpha_j^{-\frac{1-\tau}{\tau}} \sum_i b^A[\alpha_i, y]^{1/\tau} \qquad (3)$$

for $y \neq \alpha_j$. So summed net benefits must increase, neither too slowly nor too quickly, with donor ability.

(iii) *Penalty for inconsistency*. Player j might be less generous to some players and more generous to others than is consistent with honestly advertising her ability a_j or impersonating a player with ability y: there might not be any one y such that $b_{ij} = b^A[s_i, y]$ for all i. In this case, j is assigned an intermediate weighted average reputation, s_j, chosen so that $b_{ij} < b^A[s_i, s_j]$ for

some i and $b_{ij} > b^A[s_i, s_j]$ for other i. The aim is not to make s_j an accurate estimate of α_j, but to penalize j for behaving inconsistently. For example, let

$$0 = \sum_i \left(b_{ij}^{1/\tau} - b^A[s_i, s_j]^{1/\tau} \right) \omega^{(b_{ij}^{1/\tau} - b^A[s_i, s_j]^{1/\tau})} \tag{4}$$

This gives s_j as an implicit function of the benefits j produces, b_{ij}, and the reputations of her beneficiaries, s_i. Suppose $\omega = 1$. Then the ω term drops out, and j's actual cost, $\alpha_j^{-\frac{1-\tau}{\tau}} \sum_i b_{ij}^{1/\tau}$, is equal to the cost, $\alpha_j^{-\frac{1-\tau}{\tau}} \sum_i b^A[s_i, s_j]^{1/\tau}$, she would pay if she consistently impersonated a player with ability s_j. So inconsistent play gets her the same reputation, for the same cost, as consistent play. But suppose instead that $0 < \omega < 1$. Then j pays the same cost as before, but her reputation is now lower, because subpar b_{ij}'s are more heavily weighted, and reduce s_j more than excessive b_{ij}'s increase it. So we make $0 < \omega < 1$: either j is consistent, and gets the appropriate s_j for honest advertising or consistent impersonation, or j is inconsistent, and pays a reputational penalty.

With a distributional norm and a reputation function that satisfy these conditions, consistent honest advertising is the best strategy for each player, as long as others follow the norm. So the honest advertising constraints set limits on what distributional norms are stable among self-interested players—for example, the norm matching the payoffs under pure altruism is *not* stable. But within these limits, there is a wide range of stable norms, including the two below, which differ in how well they treat weak players, and are meant to capture the anthropologist's distinction between balanced and generalized reciprocity.

Almost Balanced Reciprocity

One family of distributional norms assigns equal payoffs in both directions, $b^A[x, y] = b^A[y, x]$ for all x and y. Players get only as much they give, and "there ain't no such thing as a free lunch." We can pick one member of this family by allowing the weaker player to set $b^A[x, y] = b^A[y, x]$ to her preferred value, $\tau^{-\frac{1-\tau}{\tau}} x$ for $x \leq y$. This norm is stable under direct reciprocity, and also under indirect reciprocity where players estimate reputations according to Condition 4.

There's a hitch, however. Exact balanced reciprocity, direct or indirect, is not *dynamically* stable. If players start somewhere other than at equilibrium, and estimate one another's abilities based on past play, then reputations and benefits produced don't converge on the equilibrium. Even small perturbations are destabilizing. The underlying problem is familiar from the Iterated Prisoner's Dilemma game: two tit-for-tat players don't get to the cooperative equilibrium if they aren't already there. In the prisoner's dilemma case, this can be fixed with a forgiving amendment to the tit-for-tat strategy [26]. In the present case, the analogous fix is to set $b^A[x, y]$ slightly higher than $b^A[y, x]$ for $x < y$, with the difference adjusted to satisfy Condition 3. We call the resulting stable distributional norm Almost Balanced Reciprocity, or B, with defining function b^B (Figs 1B and 2).

Generalized Reciprocity

With Almost Balanced Reciprocity, a strong player could provide a weak player with greater benefits at little added cost to herself, but doesn't do so because it would be too costly for the weak player to (almost) match the extra benefits. Discarding the principle that exchanges must balance allows for a more even distribution of benefits and greater total net benefits. Consider the norm that maximizes group net benefits, subject to the honest advertising constraints. With this norm, a strong player gives substantial help to weak players, even entirely helpless
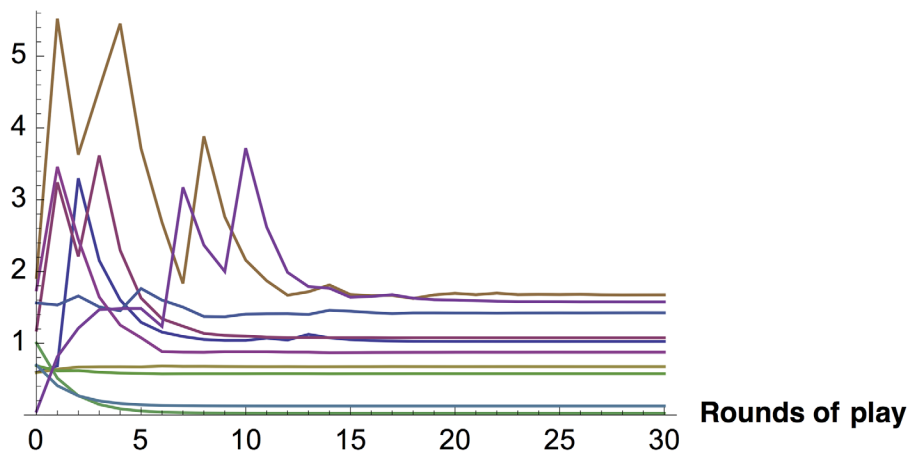
Estimated reputations

Fig 2. Reaching equilibrium with Almost Balanced Reciprocity. In a group of 40 players, each player j produces benefits for a random 5 players i in each round, following b^B and beginning in round 1 using random private estimates of s_i. For each j, every other player k calculates a revised s_j of her own, after observing a random 3 of j's 5 random acts of kindness, by plugging the 3 b_{ij}'s and her 3 unrevised s_i's into the implicit function given by Condition 4. For each j, player k then averages her revised and unrevised s_j's to generate new s_j's that will determine how much she helps a random 5 players in the next round. Chart shows a typical result for a random player's private estimates of s_j's for 10 other random players j: her s_j's converge on α_j's. (Generalized Reciprocity gives similar results.)

players with $\alpha = 0$, in order to uphold her reputation and be rewarded by other strong players. This is socially enforced altruism. It is less generous than pure altruism, because strong players need to do better than weak players to encourage honest advertising, but more generous than Almost Balanced Reciprocity.

An even more generous distribution is possible if we maximize total net benefits, but allow j some knowledge, not just of the benefits i produces, but of the costs i pays. In this case, a weak i helping others at some cost to herself can receive more credit than a stronger player providing the same benefits at less cost [27], making it possible to enforce more generosity toward weak players without motivating shirking among strong players. Here we use a version of a distributional norm with some knowledge of costs, calling it Generalized Reciprocity, or G, with defining function b^G (Fig 1C).

Mixed groups, compromise, and kin selection

Either norm, Almost Balanced Reciprocity (B) or Generalized Reciprocity (G), is stable if everyone agrees to it. But what happens if players don't agree? Evolutionary considerations can help answer this question.

Consider a mixed group, in which some players follow norm B and some follow G. In such a group, each player j can be assigned two reputations, s_j^B and s_j^G, depending on the benefits she produces and the reputations, s_i^B and s_i^G, of her beneficiaries. We can numerically calculate an equilibrium, at which type B players reward other players according to their s^B reputations and type G players reward according to s^G (Table 1). We can further calculate the mean net benefits accruing to each type. Simulations show that in groups with a given frequency of types B and G, the more common type usually gets greater mean net benefits; but in a collection of groups with different frequencies of B and G, both types usually get greater mean net benefits in those groups where B is less frequent and G more frequent (Fig 3, lower portion).

Table 1. Reputation in a mixed group. Equilibrium reputations, s^B and s^G, calculated for a mixed group of 40 players of varying ability (α). A random 20 players follow norm B and 20 follow G; results are given for three representative players of each type. Players accurately assess abilities of those of their own type, but under- or overestimate abilities of those of unlike type.)

Type	α	s^B	s^G
B	1.925	1.925	1.156
B	1.025	1.025	.934
B	.325	.325	.229
G	1.725	1.958	1.725
G	.925	.968	.925
G	.125	.178	.125

doi:10.1371/journal.pone.0155596.t001

Now consider a large population of groups, each of size n and average within-group relatedness r, where the overall frequency of type B in the population is p_B. On average, a B player finds herself in a group where type B numbers $1 + (n-1)(r + p_B - r \cdot p_B)$, and a G player finds herself in a group where type B numbers $(n-1)(p_B - r \cdot p_B)$ (lower arrows in [Fig 3](#)). If n is not very small, and r is not very large, either number is close to $n \cdot p_B$. If net benefits are proportional to increments in fitness, then the common type (away from the equal payoff crossover point) is expected to have higher fitness in most groups. Selection within groups, favoring the common type, is stronger than selection between groups, favoring G. Whichever distributional norm starts out at high frequency goes to fixation.

Yet the survival of the commonest is not the end of the evolutionary story. The equilibrium where some group members play B and others G is not a Nash equilibrium. Individuals of
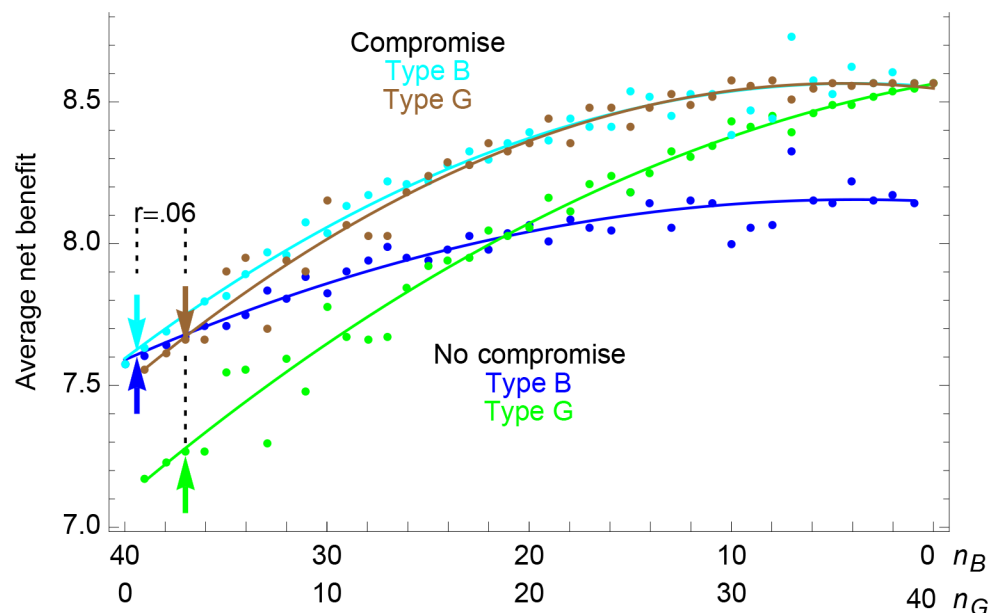

Fig 3. Mixed groups. Average net benefits in mixed groups of 40 players, with n_B playing Almost Balanced Reciprocity and n_G playing Generalized Reciprocity—or with the two types compromising on an intermediate norm—based on random sampling of players with differing abilities. Lines show quadratic best fits to points. Arrows show that with a modest coefficient of relatedness ($r = .06$, suggested by [[5](#), [28](#)]) and no compromise, a rare type G does worse than a common type B, but with compromise, a rare type G does better.

doi:10.1371/journal.pone.0155596.g003

either type can do better if they don't stick to their preferred strategy, but adapt to the presence of the other type. Such accommodation within a mixed group can take various forms.

First, all players can compromise on an intermediate strategy. A compromise distributional norm (call it W) can be defined by a parameter w_0, with $0 < w_0 < 1$, such that the summed net benefit received by a player with ability y is a weighted mean of the summed net benefits received by a y player under norms B and G, with weights w_0 and $1-w_0$. For a given w_0, a benefit function b^W can be defined such that the corresponding norm W is stable, giving each player an incentive to honestly advertise her abilities as long as others follow the norm. Here we allow each group to settle on the w_0 for which $b^{W1/\tau}$ gives a least-squares best fit to $b_{ij}^{1/\tau}$ for type B and type G players.

When players compromise like this, simulations give different results than when they don't (Fig 3). Both types do better when they settle on an intermediate norm W. Furthermore, among compromising groups with a given frequency of types B and G, B usually does slightly better than G. This is because strong players have more influence on the value of w_0 than weak players. In groups where more of the strong players are type B, the B's "vote" greater net benefits for the strongest players, themselves; where more of the strong players are type G, the G's "vote" greater net benefits for the average player. But G can still prevail: with positive within-group relatedness, type G players are more likely to find themselves in groups with extra type G players (upper arrows in Fig 3). Strong G players' support for higher mean net benefits in these groups can give G a greater average fitness.

In an alternative form of compromise, type B and G players persist in playing their respective strategies, while a third type of player, type W, adopts the compromise strategy corresponding to the group's least-squares w_0. Experimentation with a range of parameter values shows that in groups where B and G players refuse to compromise, type W—a moral chameleon who takes on the average color of her surroundings—generally does better (Fig 4). Once type W reaches high frequency, selection changes, and the rarest type is no longer at such a disadvantage. In a population with a very common type W, rare type G, less rare type B, and moderate within-group relatedness, a type G player may often find herself in a group containing many type W's, several type G kin, and fewer type B's. Such a group will lean toward enforcing a generous, G-weighted norm, to the advantage of type G.

So combining a small measure of principled rule-following with a large measure of self-interested compromise opens up new space for kin selection. When types B and G compromise on norm W, variance within groups decreases. When B and G don't compromise and the compromising type W becomes more frequent, variance between groups increases. Either way, compromise strengthens selection between groups relative to selection within them. With socially enforced altruism, a little relatedness can go a long way.

Discussion

A big motive for developing the theory of kin selection was the problem of *altruism*: Why does one organism sometimes help another when that help will not be reciprocated? The analogous motive for developing a theory of socially enforced nepotism is the problem of *distribution*: How are the gains from cooperation distributed? [29] Why does social exchange sometimes shade into uni-directional generosity—balanced reciprocity into generalized reciprocity?

Previous theory [19] and evidence [30, 31] show that indirect reciprocity can motivate cooperation in the form of contributions to public goods. The model developed here goes further. It is simultaneously about motivating cooperation *and* about distributing the gains from cooperation among players of unequal abilities. Kinship is part of the story because even modest coefficients of relatedness—too low to motivate much individual nepotism—can turn the problem of
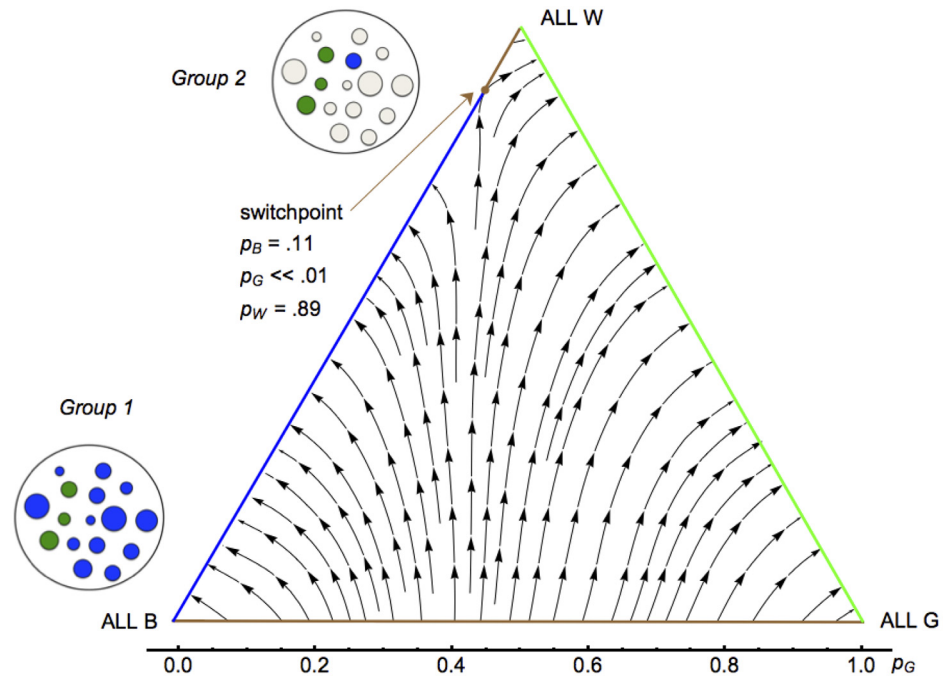
Fig 4. Three types. Stream plot of selection vectors for populations with fractions p_B, p_G, and p_W playing Almost Balanced Reciprocity, Generalized Reciprocity, and a Weighted Compromise. Fitnesses are proportional to net benefits, averaged over random samples of mixed groups of 40 players with coefficient of relatedness $r = .06$. Blue line shows a neutrally stable mix of B and W players, green line a neutrally stable mix of G and W players. At switchpoint, the mixture of B and W is unstable: population leaves B/W equilibrium for G/W, never to return. *Group 1*: Typical group composition experienced by a rare G player and her relatives when type W is absent. Type G (green circles; size stands for ability) is outnumbered by type B (blue circles). *Group 2*: Typical group composition experienced by a rare G player and her relatives when type W is common. G player and her G relatives outnumber type B; type W players (white circles) skew their play toward Generalized Reciprocity, amplifying small r.

doi:10.1371/journal.pone.0155596.g004

distribution from a constant sum game into a public goods game: all are better off evolutionarily if strong players give up potential net benefits to help weak players. (A rule of thumb: if $r \gg 1/n$, then a player can expect that other group members carry more copies of her genes than she does, and she should "vote" for the norm that is best for the average player, even if it's not best for her. [5])

The work here extends earlier findings that indirect reciprocity depends on reputation and reputation must be defined recursively, so that each player's reputation depends both on the benefits she provides and on the reputations of the beneficiaries. In the model here, reputation is an index of ability, and, like ability, can vary continuously. So the model faces the problem of assigning a numerical reputation to each player based on the reputations of other players. This is more complicated than simply choosing to cooperate or not based on whether the other player is in good or bad standing. It is close to the problem of designing a page-ranking algorithm that ranks a collection of web pages by assigning a score to each page based on the scores of the pages that link to it [32]. Some of the same issues arise in both cases: for example, whether a given distributional norm or page-ranking algorithm is dynamically stable and converges on a consistent result.

The model must satisfy a further set of conditions. We require that a distributional norm be enforceable (or *incentive compatible*): a player of whatever ability should do better sticking to the norm, rather than impersonating a stronger or weaker player or playing inconsistently, as

long as other players do the same. All these conditions rule out many distributional norms. For example, strict balanced reciprocity is dynamically unstable, and pure altruism is unenforceable. But the conditions also allow a wide range of distributional norms with very different distributions of benefits, including the two we concentrate on here, Almost Balanced Reciprocity and Generalized Reciprocity.

One way to think about how reputation works in enforcing one norm or another is to compare it to another sort of currency, money. Either reputation or money can be used for indirect reciprocity: one person helps another in order to gain reputational or monetary credit, not for its own sake, but to "buy" favors from third parties. Yet a moral economy based on reputation is different from a money economy. Money makes indirect reciprocity work by reducing it to direct reciprocity, to dyadic money-for-benefits trades. A money-based economy can scale up indefinitely, coordinating exchanges among millions of strangers. But it has a hard time harnessing these transactions to a common social purpose; money-based economies normally rely on auxiliary political institutions to produce desired public goods. Reputation, by contrast, doesn't scale up so easily, and works most readily in face-to-face groups, where reputations can be a matter of common knowledge—known to each, and known to be known to each [33]. Because the community decides, based on shared values, how much credit one player gains for helping another, reputation is more easily enlisted for social purposes.

How one norm or another is selected from among those allowed is a further question. A feature of many models of social enforcement through indirect reciprocity is that once a norm is in place it is locked in, and resistant to change. This makes it hard to explain how groups might shift from a merely adequate norm to a better one. In the present model, lock-in occurs in the simplest case. With just two norms competing in a population of groups, the winner is the norm that starts out at higher frequency, not necessarily the best. But the analysis here also shows how lock-in can be avoided: in mixed groups, those who compromise on an intermediate norm, rather than sticking to either pure norm, do better. Compromise, which is favored by within-group selection, also acts as a catalyst, lowering the barrier to a transition between norms and facilitating between-group selection in favor of rare pro-social norms, like Generalized Reciprocity.

Like most models of human sociality, the present model of socially enforced nepotism concentrates on just one part of the process of social evolution, and treats other parts more cursorily. In this spirit, we make the simplest possible assumption about how group membership translates into fitness: groups that produce more offspring make a proportionately greater contribution to the whole population in the next generation. This is multi-level selection, but not group selection in the strong sense—differential multiplication and extinction of groups—which would require additional assumptions about population processes. We also implicitly adopt a *weak selection* assumption: we ignore the effects of selection between the time a gene is present in an ancestor and the time it acts in her descendants, which can make genealogical relatedness a poor indicator of identity by descent for genes under strong selection. This assumption is probably roughly correct for groups on the scale of local groups, lineages, and clans, but not on the scale of tribes and ethnic groups.

These simple assumptions about population dynamics and population genetics might be revised in later work. But here the focus has been on an aspect of sociality that cultural anthropologists past and present have highlighted: norms and roles in relation to kin altruism. Some anthropologists argue that human kinship, insofar as it is socially enforced, is divorced from biology. The argument here, on the contrary, is that kinship is uniquely elaborate and important in our species because norms that push people to treat distant kin like close kin have been favored by natural selection.

Supporting Information

S1 File. Supporting information. Further information about mathematical results and numerical simulations, including references to S1 and S2 Figs.
(PDF)

S1 Fig. Not reaching equilibrium with exact balanced reciprocity. As in Fig 2, but with exact (not almost) balanced reciprocity. Estimated reputations do not converge on α_j. (Many estimated reputations are wildly exaggerated, but adding a ceiling on estimates does not result in convergence.)
(PDF)

S2 Fig. Weights for a compromise norm. For different values of w_0, this shows how a player with ability α_j should weight her costs between b^B and b^G to achieve a compromise distributional norm satisfying Condition 13 for stability (See S1 File).
(PDF)

Acknowledgments

Author Contributions

Analyzed the data: DJ. Contributed reagents/materials/analysis tools: DJ. Wrote the paper: DJ.

References

1. Hamilton WD. The evolution of altruistic behavior. American Naturalist. 1963; 97:354–356. doi: 10.1086/497114

2. Nowak MA. Five rules for the evolution of cooperation. Science. 2006; 314(5805):1560–1563. doi: 10.1126/science.1133755 PMID: 17158317

3. Taylor C, Nowak MA. Transforming the dilemma. Evolution. 2007; 61(10):2281–2292. doi: 10.1111/j.1558-5646.2007.00196.x PMID: 17711471

4. Nowak M, Wilson E, Tarnita C. The evolution of eusociality. Nature. 2010; p. 1057–1062. doi: 10.1038/nature09205 PMID: 20740005

5. Jones D. Group nepotism and human kinship. Current Anthropology. 2000; 41:779–809. doi: 10.1086/317406

6. Gintis H. Game Theory Evolving: A Problem Centered Introduction to Modeling Strategic Interaction. 2nd ed. Princeton University Press; 2009.

7. Fiske AP. Structures of Social Life: The Four Elementary Forms of Human Relations. Free Press; 1991.

8. Fox R. Kinship and Marriage: An Anthropological Perspective. Cambridge University Press; 1967.

9. Tylor EB. Anthropology: An Introduction to the Study of Man and Civilization. Routledge (MacMillan); 1994 (1881:).

10. Fortes M. Kinship and the Social Order: The Legacy of Lewis Henry Morgan. Aldine Transaction; 1969.

11. Alvard M. Genetic and cultural kinship among the Lamaleran whale hunters. Human Nature. 2011; 22:89–107. doi: 10.1007/s12110-011-9104-x PMID: 22388802

12. Nolin D. Kin preference and partner choice: patrilineal descent and biological kinship in Lamaleran cooperative relationships. Human Nature. 2011; 22(1):156–176. doi: 10.1007/s12110-011-9113-9 PMID: 22388806

13. Chapais B. Primeval Kinship: How Pair-Bonding Gave Birth to Human Society. Harvard University Press; 2008.

14. Haidt J. The Righteous Mind: Why Good People Are Divided by Politics and Religion. Random House; 2012.

15. Hill K. Animal "culture"? In: Laland KN, Galef BG, editors. The Question of Animal Culture. Harvard University Press; 2009.

16. Gintis H. The Bounds of Reason: Game Theory and the Unification of the Social Sciences. Princeton University Press; 2014.

17. Goffman E. The Presentation of Self in Everyday Life. Doubleday; 1959.

18. Alexander RD. The Biology of Moral Systems. Aldine Transaction; 1987.

19. Panchanathan K, Boyd R. Indirect reciprocity can stabilize cooperation without the second-order free rider problem. Nature. 2004; 432:499–502. doi: 10.1038/nature02978 PMID: 15565153

20. Nowak MA, Sigmund K. Evolution of indirect reciprocity by image scoring. Nature. 1998; 393:573–577. doi: 10.1038/31225 PMID: 9634232

21. Leimar O, Hammerstein P. Evolution of cooperation through indirect reciprocity. Proceedings of the Royal Society B. 2001; 268:745–753. doi: 10.1098/rspb.2000.1573 PMID: 11321064

22. Panchanathan K, Boyd R. A tale of two defectors: the importance of standing for evolution of indirect reciprocity. Journal of Theoretical Biology. 2003; 224:115–126. doi: 10.1016/S0022-5193(03)00154-1 PMID: 12900209

23. Nowak MA, Sigmund K. Evolution of indirect reciprocity. Nature. 2005; 437:1291–1298. doi: 10.1038/nature04131 PMID: 16251955

24. Ohtsuki H, Iwasa Y. The leading eight: social norms that can maintain cooperation by indirect reciprocity. Journal of Theoretical Biology. 2006; 239:435–444. doi: 10.1016/j.jtbi.2005.08.008 PMID: 16174521

25. Sigmund K. Moral assessment in indirect reciprocity. Journal of Theoretical Biology. 2012; 299:25–30. doi: 10.1016/j.jtbi.2011.03.024 PMID: 21473870

26. Nowak MA, Sigmund K. Tit for tat in heterogeneous populations. Nature. 1992;. doi: 10.1038/355250a0

27. Mark 12:41–44;.

28. Bowles S. Group competition, reproductive leveling, and the evolution of human altruism. Science. 2006; 304:1569–1572. doi: 10.1126/science.1134829

29. Du WB, Cao XB, Hu MB, Wang WX. Asymmetric cost in snow drift game on scale-free networks. EPL. 2009; 87(6004):1–6.

30. Lyle HF III, Smith EA. The reputational and social network benefits of prosociality in an Andean comuunity. Proceedings of the National Academy of Sciences. 2014; 111:4820–4825. doi: 10.1073/pnas.1318372111

31. Milinski M, Semman D, Krambeck HJ. Reputation helps solve the'tragedy of the commons'. Nature. 2002; 415:424–426. doi: 10.1038/415424a PMID: 11807552

32. Langville AN, Meyer CD. Google's PageRank and Beyond: The Science of Search Engine Rankings. Princeton University Press; 2012.

33. Chwe MSY. Rational Ritual: Culture, Coordination and Common Knowledge. Princeton University Press; 2001.