Evolution of In-Group Favoritism

Citation
Fu, Feng, Corina E. Tarnita, Nicholas A. Christakis, Long Wang, David G. Rand, and Martin A. Nowak. 2012. Evolution of in-group favoritism. Scientific Reports 2:460.

Published Version
doi:10.1038/srep00460

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:10520827

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story
The Harvard community has made this article openly available. Please share how this access benefits you. Submit a story.

Accessibility
In-group favoritism is a central aspect of human behavior. People often help members of their own group more than members of other groups. Here we propose a mathematical framework for the evolution of in-group favoritism from a continuum of strategies. Unlike previous models, we do not pre-suppose that players never cooperate with out-group members. Instead, we determine the conditions under which preferential in-group cooperation emerges, and also explore situations where preferential out-group helping could evolve. Our approach is not based on explicit intergroup conflict, but instead uses evolutionary set theory. People can move between sets. Successful sets attract members, and successful strategies gain imitators. Individuals can employ different strategies when interacting with in-group versus out-group members. Our framework also allows us to implement different games for these two types of interactions. We prove general results and derive specific conditions for the evolution of cooperation based on in-group favoritism.

In-group bias is a central aspect of human behavior. Across a variety of scenarios, people tend to be more helpful to members of their own group rather than to those of other groups. In-group favoritism has been shown to occur based on real-world salient groupings, such as ethnicity, religiosity and political affiliation, and has also been artificially manufactured in the laboratory using trivial groupings. Discrimination based on group identity is a powerful force for both good and ill in human societies. Therefore understanding the evolutionary dynamics of group identity and discrimination is of primary significance. Evolutionary game theory offers a powerful framework for exploring such issues.

In-group bias is common, yet the implementation of that bias is dynamic and flexible. Thus culture and cultural evolution must play an important role in the evolution of bias. The dynamic nature of bias results from complex social network interactions which play a central role in human societies, with genetic as well as social components affecting network formation. Such network dynamics can turn yesterday’s allies into today’s competitors, and drive former enemies together in the face of a common threat. This is true of many hunter-gather societies which are characterized by fission-fusion dynamics, suggesting the importance of flexible groupings for early human societies. Group identities are also flexible in today’s modern world: the dynamic remodeling of in-group bias was recently demonstrated in a behavioral study using economic games. During the 2008 US presidential election, supporters of Democratic primary candidates Barack Obama and Hillary Clinton formed two separate groups, and showed strong bias against each other. After the transition from the primary to the general election, the two groups fused and intra-Democrat bias disappeared.

Thus group membership is itself an evolving feature of humans. If membership in a particular group is associated with a particular strategy, it can be beneficial to act differently towards members of different groups. This was demonstrated in a recent behavioral experiment involving a series of coordination games. In each round, subjects chose both a cultural group (represented by one of two arbitrary markers) and an action in a coordination game. Subjects tended to change tag and action together, or to change neither. This created an association between group membership and strategy, and in-group favoritism emerged: subjects preferred to interact with members of their own group.

The multi-faceted, dynamic and emergent nature of group identity is central to bias. Its evolution, however, has not yet been explained, as previous models considered either competition between stable fixed groups, or a priori imposed in-group bias on agents. In this paper we introduce a novel model to address this issue (Fig. 1, see model...
Results

Although our framework allows us to study many different evolutionary games, in the main part of this paper we focus on cooperative dilemmas, where individuals choose between cooperation and defection. A strategy is given by two parameters, \((p, q)\), where \(p\) denotes the probability to cooperate with another individual in the same set and \(q\) denotes the probability to cooperate with an individual in another set. If \(p > q\), then cooperation is biased towards in-group members. If \(p < q\), then cooperation is biased towards out-group members. Our strategy space consists of a continuum of strategies represented by the unit square, \([0, 1]^2\). It includes ‘always defect’ \((0, 0)\) and ‘always cooperate’ \((1, 1)\), strategies which either do not recognize group membership or choose to ignore it. The strategy \((1, 0)\), conversely, shows maximum in-group favoritism: it always cooperates with in-group members but never with out-group members.

In Fig. 2 we show computer simulations of the stochastic evolutionary dynamics of our system. We use agent based simulations for intermediate intensities of selection. Each individual is characterized by a particular \((p, q)\) strategy. The figure shows the stationary distribution of strategies that is reached in this mutation-selection process. We observe that the distribution is either peaked around ‘always-defect’ \((0, 0)\), or maximum in-group favoritism \((1, 0)\). We are interested in finding the conditions for evolution of in-group favoritism (Fig. 3).

Using the analytical theory for weak selection we calculate the region of the strategy space that is favored by natural selection. In the online material we perform this calculation for all cooperative dilemmas, but here we present the result for the simplified Prisoner’s Dilemma, where cooperation implies a cost, \(c\), for the donor and a benefit, \(b\), for the recipient, where \(b > c > 0\). Defectors pay no cost and distribute no benefit. We find that the \((p, q)\)-strategy space is intersected by a straight line, given by the equation \(q = K(p - 1/2) + 1/2\), where \(K\) is the slope of the line. The explicit expression of \(K\) depends on the benefit-to-cost ratio, \(b/c\), the number of sets, \(M\), and the rescaled mutation rate, \(\mu\), and the rescaled set mutation rate, \(\tilde{\beta}\) (see the online material for derivations). Strategies below the line (lower \(q\) values) are favored by selection, and strategies above the line (higher \(q\) values) are opposed by selection (Fig. 2). The line always intersects the strategy \((1/2, 1/2)\) which is neutral for all parameter choices. The maximum slope of the line is \(+1\) and the minimum slope is \(-\infty\). If the slope is negative \((K < 0)\) then the most favored strategy is always-defect \((0, 0)\). If the slope is positive \((K > 0)\) then the most favored strategy is maximum in-group favoritism \((1, 0)\).

There is another interesting threshold. If \(K < -1\) then the average \(p\) value in the population is less than the average \(q\) value. In this (paradoxical) case, individuals are more likely to cooperate with out-group members rather than in-group members. The reason is the following: for these population structures cooperation is not selected and, moreover, most interactions occur within sets; it is therefore more costly to cooperate with individuals in the same set rather than with individuals from another set. Hence, for \(K < -1\) our system predicts some level of out-group favoritism. For \(K > -1\) the average \(p\) value in the population is greater than the average \(q\) value. Thus \(K > -1\) is the crucial condition for the evolution of at least some degree of in-group favoritism, while \(K > 0\) is the condition for selection to favor minimal in-group favoritism \((1, 0)\).

Furthermore, these two critical thresholds of \(K\) also determine the population average levels of in-group helping \(p\) and out-group helping \(q\). The average \(p\) is greater than \(1/2\) greater than the average \(q\) for \(K > 0\), the average \(q\) is smaller than the average \(p\) if \(-1 < K < 0\), and the average \(p\) is less than the average \(q\) less than \(1/2\) for \(K < -1\). Note that natural selection always disfavors out-group helping (the average \(q\) is less than \(1/2\) for all parameter choices), whereas in-group cooperation can be favored by selection (the average \(p\) is greater than \(1/2\)) if \(K > 0\).

For large population size and in the limit of low strategy mutation, \(\mu \to 0\), the condition \(K > -1\) leads to

\[
\frac{b}{c} > \frac{3(\tilde{\beta} + 2)^2 - (M/2)(9/2 + 3\tilde{\beta}^2 + 9/2)}{(M - 1)(9/2 + 2)}
\]

(1)

while \(K > 0\) leads to

\[
\frac{b}{c} > \frac{3(\tilde{\beta} + 2)^2 + M(2/3 + 3\tilde{\beta})}{(M - 1)(9/2 + 2)}
\]

(2)

We observe that increasing the number of sets, \(M\), reduces both critical benefit-to-cost ratios. Increasing the set mutation rate, \(\tilde{\beta}\), always reduces the critical benefit-to-cost ratio that is needed for the evolution of some in-group favoritism \((1, 0)\), but for evolution of maximum in-group favoritism \((2)\) there is an intermediate optimum set mutation rate (order of \(\tilde{\beta} = \sqrt{M}\)) which minimizes the critical benefit-to-cost ratio. We also note that condition \((2)\) is the same condition that is needed for evolution of cooperation in set struc-
tured populations\textsuperscript{23}. Noteworthy, the evolution of some in-group favoritism can be promoted for any benefit-to-cost ratios if the set mutation rate \(q\) is above a certain threshold depending on the number of sets \(M\) and the strategy mutation \(\mu\). In the online material we provide further discussion of these conditions and also study the effect of the strategy mutation rate, \(\mu\). Our analytical results are in excellent agreement with the agent based simulations.

**Figure 2 | Evolutionary panorama of in-group favoritism.** Shown are the stationary distributions of strategies over the unit square with respect to varying (a) cost of cooperation, \(c\), (b) number of groups, \(M\), (c) strategy mutation rate, \(u\), and (d) migration rate between groups, \(v\). The color bar indicates the equilibrium abundance of strategies: red means high while blue low. The straight lines, below which strategies are favored by natural selection, represent our analytical predictions for weak selection. A positive slope suggests that the lower-right corner (1, 0), maximum in-group favoritism, is most favored. We find excellent agreement between our analytical theory and computer simulations. Increasing \(c\) and \(u\) values each disfavor the emergence of in-group favoritism, while increasing \(M\) helps its establishment. An intermediate optimal migration rate \(v\) most promotes the evolution of in-group favoritism. Parameters: \(N = 100, \beta = 0.005, \beta = 1\), (a) \(M = 4, u = 0.04, v = 0.06\), (b) \(c = 0.1, u = 0.04, v = 0.06\), (c) \(c = 0.08, M = 4, v = 0.06\), (d) \(c = 0.08, M = 4, u = 0.06\). Results are averaged over \(T = 2 	imes 10^9\) time steps.

**Discussion**

We have demonstrated how individual-level selection can favor the emergence of in-group bias in a world of dynamic social ties and group identities. We consider the full continuum of strategies for both in-group and out-group interactions, not just those that show in-group favoritism. We also include strategies which do not discriminate between in-group and out-group members, as well as strategies which give preferential treatment to out-group members. We derive the conditions for natural selection to favor the recognition of group membership, and for the ensuing success of in-group favoritism (Fig. 2).

Previous theoretical studies have examined the thought experiment of the green-beard effect (or tag-based cooperation), finding that cooperation can evolve when similar others receive preferential treatment\textsuperscript{15,18,19,21,22}. Phenotypic tags in these models can be regarded as a sort of minimal grouping. In contrast to our present model, these previous models assume \textit{a priori} that individuals never cooperate with out-group members, and therefore group structure (as in an island model) supports the evolution of within-group helping\textsuperscript{48,49}. Yet empirical evidence indicates that although people exhibit more in-group cooperation than out-group cooperation, cooperation with out-group members is still far above zero\textsuperscript{8,44}. Thus unlike previous
models, we allow the possibility of both in-group and out-group cooperation, and determine the conditions under which preferential in-group cooperation is favored by selection. As seen experimentally, we find the emergence of in-group bias together with substantial levels of out-group cooperation.

A second set of previous models consider in-group helping and out-group harming in the context of explicit intergroup conflict with discrete (binary) strategies\(^{26,24}\). Two of these models use simulations to explore the evolution of such 'parochial altruism', and find that explicit inter-group conflict is needed for selection to favor bias\(^{26,24}\): groups with more in-group helpers and/or out-group harmers battle and replace other groups. A third model also considers intergroup conflict as the driving force behind bias, and derives analytical results. However, we note that their approach requires infinitely many groups of fixed size, applies only to additive games, and only gives the direction of selection (rather than any actual evolutionary dynamics) for mutant invasion attempts at a locus without polymorphism at the other locus. Unlike these models, we show that sensitivity to group identity need not be the result of such intergroup conflict, but instead can result from particular population structures. Furthermore, our approach yields general analytical results, characterizing the full evolutionary dynamics of our system, and can easily be applied to a range of "in-group/out-group" interactions (including, but not limited to, additive games).

Our mutation-selection analysis shows that increasing the strategy mutation rate \(\mu\) always makes it harder for cooperation to predominate (Fig. 3c), but that this is not the case for the set mutation rate \(\nu\): there is an optimal level of \(\nu\) for promoting in-group cooperation (Fig. 3d). In our model, prosperous groups tend to attract new individuals, which can be seen as "payoff-biased migration"\(^{46-47}\). However, this non-random migration cannot promote the success of in-group solidarity if the set mutation rate, or "confusion about the other's set", is too low. Without flexibility in group identity (\(\nu \rightarrow 0\)) all individuals eventually end up in the same group, and then the system is prone to exploitation of in-group members. This is why some out-group favoritism can evolve under this circumstance. On the other hand, when the set mutation rate is excessively high (\(\nu \rightarrow \infty\)) and thus group identities are overly fluid, the association between strategy and group membership breaks down. In our model, as in behavioral experiments\(^6\), this correlation between strategy and group membership is essential for the establishment and maintenance of in-group favoritism. Thus evolution cannot favor in-group cooperation if \(\nu\) is too large. An intermediate set mutation rate is optimal, maintaining a delicate balance between mobility and consistency.

Our work generates testable predictions about evolved human behavior. For example, we find that a larger number of possible groups leads to stronger selection for in-group bias (Fig. 3b). This has implications both for laboratory experiments and field studies. Most existing laboratory experiments with economic games use binary groupings\(^{48}\), but our theory predicts stronger in-group bias if a richer set of groupings was used. Psychological research is consistent with this prediction\(^6\). Our results also suggest that people who developed their intuitions in more diverse communities, where a wider range of possible social groups exist, will demonstrate stronger in-group bias. Evaluating these predictions using quantitative empirical analysis is a promising research direction.

The framework we introduce here is not limited to the study of cooperation. It is a general theory for studying situations where one game is played with in-group members while another game is played with out-group members. For example, our framework can explain the evolution of costly in-group helping and costly out-group harming\(^{26}\) (see the online material). Applying this model to a variety of other social behaviors such as coordination, trust and bargaining\(^{50}\) is an important topic for future study, as is exploring situations where individuals are members of multiple groups with competing allegiances. Extending this framework to coalition formation\(^{53,54}\), which cannot always be reduced to pairwise interactions, will be useful.

In-group bias is a central aspect of human behavior\(^1,2,3,5,44\). This sensitivity to group membership may seem to imply a key role for inter-group conflict in human evolution. Yet, while inter-group conflict can promote the evolution of in-group bias\(^{49}\), to do so it requires stable groups whose membership remains largely unchanged over time. Contrary to this vision of stable groups, experiments\(^5,25\) and anthropological evidence\(^{51}\) have demonstrated that group identity and membership is often flexible. Over short timescales, the relevant grouping can be remodeled, leading to dramatic changes in behavior towards the same set of others. Here we show how in-group bias can evolve and thrive when group compositions are constantly shifting. It can be advantageous to love your friends, even when today’s comrades might be tomorrow’s enemies.

**Methods**

**Model.** We study the evolutionary dynamics of a population of finite size, \(N\), which is distributed over \(M\) sets. Each individual belongs to one set. Without constituting more complicated group culture or identity, this minimal group setting in our model is similar with trivial groupings in previous empirical studies\(^3,5\). Individuals engage in evolutionary games and accumulate payoff from pairwise interactions. Note that the payoff of an individual depends both on its strategy and on the set it belongs to.
They reproduce proportional to payoff and inherit both strategy and set membership. Reproduction can be genetic or cultural. That said, successful strategies spawn imitators and successful sets attract more members. Therefore, the population structure is dynamic and driven by the social interactions. To study the limiting distribution of strategies under mutation-selection equilibrium, we include a strategy mutation rate, μ, and a set mutation rate, ν. The latter represents random migration between sets. It is useful to consider the rescaled mutation rates, μ = Nμ and ν = Nν, in the theoretical analysis. At times some sets can become empty to be re-populated later. It is also possible that occasionally the whole population clusters in one set, although increases in set mutation tend to make the population dwell in a mixture of different sets most of the time.

Each individual plays with all others: interactions occur both between individuals in the same set and between individuals in different sets. Here we consider equal interaction rates within and between sets, an assumption that can be easily modified by introducing additional model parameter (see the online material). What varies between in-group and out-group interactions is the strategy used by each individual, and potentially the nature of the game that is played. At any one time the interaction structure of the population is given by a two-colored graph (Fig. 1). The individuals represent the vertices of the graph. The colors signify the two types of interactions: within sets (in-group) and between sets (out-group). The graph changes as individuals move between sets. Individuals may use different strategies for in-group and out-group games. We can also consider situations where different games are being played within and between sets. For example, the game between sets could be an aggression game, but the game within the set could be a Prisoner’s Dilemma.

Analysis. In the online material we develop a general analytical theory for such “in-group/out-group” evolutionary processes. We consider two different games for the two types of interactions and each game can have any number of strategies. The behavioral rule of an individual is given by a pair of strategies determining the two types of interactions and each game can have any number of strategies. The population structure is described by a two-colored graph (Fig. 1). The individuals represent the vertices of the graph. The colors signify the two types of interactions: within sets (in-group) and between sets (out-group). The graph changes as individuals move between sets. Individuals may use different strategies for in-group and out-group games. We can also consider situations where different games are being played within and between sets. For example, the game between sets could be an aggression game, but the game within the set could be a Prisoner’s Dilemma.

Analysis. In the online material we develop a general analytical theory for such “in-group/out-group” evolutionary processes. We consider two different games for the two types of interactions and each game can have any number of strategies. The behavioral rule of an individual is given by a pair of strategies determining the two types of interactions and each game can have any number of strategies. The population structure is described by a two-colored graph (Fig. 1). The individuals represent the vertices of the graph. The colors signify the two types of interactions: within sets (in-group) and between sets (out-group). The graph changes as individuals move between sets. Individuals may use different strategies for in-group and out-group games. We can also consider situations where different games are being played within and between sets. For example, the game between sets could be an aggression game, but the game within the set could be a Prisoner’s Dilemma.
48. Uyenoyama, M. & Feldman, M. W. Theories of kin and group selection: A population genetics approach. *Theor. Pop. Biol.* 17, 380–414 (1980).
49. Rogers, A. R. Group selection by selective emigration: The effects of migration and kin structure. *Am. Nat.* 135, 398–413 (1990).
50. Lehmann, L. & Feldman, M. War and the evolution of belligerence and bravery. *Proc. R. Soc. Lond. B. Biol. Sci.* 275, 2877–2885 (2008).
51. Axtell, R. L., Epstein, J. M. & Young, H. P. in *Social Dynamics* (eds. Durlauf, S. & Young, H. P.) 191–211 (MIT Press, 2001).
52. Hewstone, M., Rubin, M. & Wills, H. Intergroup bias. *Annu. Rev. Psychol.* 53, 575–604 (2002).
53. Milchtaich, I. & Winter, E. Stability and segregation in group formation. *Games Econ. Behav.* 38, 318–346 (2002).
54. Goldstone, R. L., Roberts, M. E. & Gureckis, T. M. Emergent processes in group behavior. *Curr. Dir. Psychol. Sci.* 17, 10–15 (2008).

**Acknowledgements**

We are grateful for support from the Harvard Milton fund, the Harvard Society of Fellows, the John Templeton Foundation, the National Science Foundation/National Institute of Health joint program in mathematical biology (NIH grant no. R01GM078986), the Bill and Melinda Gates Foundation (Grand Challenges grant 37874), NSFC (grant nos. 10972082, 60736022 and 61020106005), the National Institute on Aging (P-01 AG031093 and P30 AG034420) and the National Science Foundation (SES-0719404) and J. Epstein.

**Author contributions**

F.F. and C.E.T. contributed to this work equally. F.F., C.E.T., N.A.C., L.W., D.G.R. and M.A.N. performed analyses, discussed the results, and contributed to the text of the manuscript.

**Additional information**

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

**Competing financial interests:** The authors declare that they have no competing financial interests.

**License:** This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivative Works 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/3.0/

**How to cite this article:** Fu, F. et al. Evolution of in-group favoritism. *Sci. Rep.* 2, 460; DOI:10.1038/srep00460 (2012).