Advances in Complex Systems, Vol. 13, No. 5 (2010) 635–641

DOI: 10.1142/S0219525910002785

World Scientific
www.worldscientific.com

THE JERUSALEM GAME: CULTURAL EVOLUTION OF THE GOLDEN RULE

JON F. WILKINS* and STEFAN THURNER*,†,‡

*Santa Fe Institute, Santa Fe, NM 87501, USA

†Section for Science of Complex Systems,
Medical University of Vienna,
Spitalgasse 23, A-1090 Vienna, Austria

‡IIASA, Schlossplatz 1, A-2361 Laxenburg, Austria

Received 18 June 2010
Revised 21 July 2010

It has often been noted that most of the major world religions espouse a version of the "golden rule." In this paper we consider the cultural evolution of such a doctrine, where the responsibility to act altruistically towards others applies universally, not just to other members of the same society. Using a game-theoretical model, we find that societies over a critical size benefit from adopting a mode of universal altruism. These "golden-rule societies" must justify violence against outsiders by formulating exceptions to this universal rule. For smaller groups, it is more efficient to adopt a rule that simply requires cooperation within the group. Data from the ethnographic record supports a correlation between group size and societal norms of universal cooperation. Our results provide an explanation for the prevalence of the golden rule among contemporary cultures. We find that universal altruism arises due to cultural selection for greater ingroup bias, and is a natural byproduct of the emergence of large-scale societies.

Keywords: Social norms; game theory; interaction of ethnic groups; social tension.

1. Introduction

On November 25, 1095, Pope Urban II summoned the first crusade, and declared that there are exceptions to the commandment prohibiting Christians from killing [1, 2]. In this statement, the Pope defined special circumstances in which violence and war were not just acceptable, but actually favored by the church. The introduction of this exception did not, however, constitute a reversal of the church's position opposing violence. Dissemination of the doctrine of "universal" love continued alongside exhortations to participate in the crusade. Significant theological efforts were required to reconcile these two contradictory values, and yet many of the leaders of the first crusade expressed confusion about the morality of the enterprise [3].

The edict of Pope Urban II does not represent the first attempt to disseminate a justification of warfare and violence while simultaneously espousing a pacifist

morality. A similar situation had occurred in the Muslim world three centuries earlier, when theologians sought to justify the violent expansion of Islam in light of the Koranic condemnation of warfare. The result of these efforts is the theology of *jihad*, which formalized an exception to the morality of peace [4]. Other historical examples include the exemption from prosecution for the killing of gypsies in the Diets of Freiburg and Augsburg in 1498 and 1500 [5], and the papal justification for the extermination of Native Americans [6].

It is common for humans to treat members of their own group more altruistically than outsiders, a phenomenon referred to as "ingroup bias" [7, 8]. Here we are interested in the cultural evolution of group-level institutions that reinforce this bias, and how this reinforcement might lead to the widespread adoption of some version of the golden rule. We define a *golden-rule society* as one that advocates altruistic behavior towards everyone, regardless of their group membership. A norm of universal altruism may benefit the group by promoting robust cooperation among its members. However, this same moral norm can be a hindrance when the group attempts to pursue political, economic, or territorial goals through violence against outsiders. In each of our historical examples, the existence of a rule for universal cooperation meant that special justification was needed to permit attacks on other groups.

We contrast this *golden-rule structure* — universal cooperation with exceptions made to justify violence against outsiders — with the *social-identity structure*. A *social-identity society* is one that lacks any universal rule for cooperation or altruism. In these societies, behavioral norms require that one behave altruistically only towards members of the group. Each of these rule structures will have the effect of reinforcing ingroup bias. Our purpose here is to ask under what conditions one of these rule structures is favored over the other.

Altruistic behavior and ingroup bias have a basis in the behavior of individuals, which is the product of both genetic and cultural evolution. Individual-based explanations of altruism have included models of kin selection [9] and group selection [10], or have incorporated strategic behavior [11], as in models of direct [12–14] or indirect [15, 16] reciprocal altruism. We focus here on group-level behavioral norms that modify or reinforce individual behaviors. We assume that these norms are the products of cultural evolution, with selection occurring at the level of the group [17–20].

2. The Jerusalem Game

We imagine a population divided into some number of groups. Each individual belongs to one of the groups. Interactions occur between pairs of individuals. Individuals engage in two types of interaction: interactions with members of their own group (ingroup interactions), and interactions with members of other groups (outgroup interactions). Interactions between individuals take the form of a standard Prisoner's Dilemma, where each individual pursues one of two strategies, conventionally referred to as "cooperate" and "defect." We use a payoff matrix

where payoffs to individual players are given by

Payoff for player 1	Player 1 cooperates	Player 1 defects
Player 2 cooperates	r	$r+i$
Player 2 defects	0	i

The term r is the reward for the other player's cooperation, and i is the incentive to defect (with $r > i > 0$). The game is symmetric, so that the payoff for player 2 is found by reversing the rows and columns. Note that here we set the sucker's payoff to zero to avoid an extra parameter.

We now define the Jerusalem Game at the level of the group, where the payoff to a group is the average payoff to its members. We assume that cultural selection favors institutions that maximize this group mean payoff. The mechanism of cultural evolution might be conflict between groups [19], or strategic construction of social norms by leaders [17]. We emphasize that this is not an evolutionary game but a game played at the individuals' level and evaluated at the group level.

If $r > i$, it is immediately clear that the mean group payoff is maximized if members of the group always cooperate in ingroup interactions, but defect in outgroup interactions. In this setup, cultural evolution will favor structures that reinforce and amplify ingroup bias. In the following, we examine two rule structures which enforce ingroup cooperation in a different way. Both reinforce ingroup bias. The two schemes are the golden-rule and the social-identity scheme.

Under the golden-rule structure, a universal rule — i.e., a rule valid for all interactions between individuals, regardless of their group membership — requires cooperation. An exceptional rule permits defection in outgroup interactions. Under the social-identity structure, this is reversed: a universal rule permitting defection is supplemented by an exceptional rule requiring ingroup cooperation. If group members complied perfectly with both group-level rules, these two formulations would be completely equivalent. Individuals would always cooperate in ingroup interactions, and always defect in outgroup interactions.

In practice, we do not expect compliance with any group-level behavioral prescription to be perfect. The efficacy of a particular rule could be expressed as a rate of compliance, or alternatively by an error rate. For example, the success of the rule "*cooperate within the group*" could be characterized by the fraction of ingroup interactions in which members actually cooperate. We have proposed two categories of rules: universal rules, which describe an unconditional mode of behavior ("*cooperate with everybody*" or "*defect against everybody*"), and exceptional rules, which require a violation of the universal rule under specific circumstances ("*cooperate within the group*" or "*defect against outsiders*").

Individuals find themselves in two situations: (i) situations to which only the universal rule applies, and (ii) situations in which the universal rule is contradicted by the exceptional rule. Let the error rates in these two situations be e_u and e_x, respectively. That is, when only the universal rule applies, individuals comply with

probability $1 - e_u$. Where the exceptional rule contradicts the universal one, individuals comply with the exceptional rule with probability $1 - e_x$.

We expect that compliance will be greater for a universal rule than for an exceptional one ($e_x > e_u$) based on three considerations. First, a universal rule will have a simpler formulation than an exceptional rule, and is therefore likely to be more cognitively salient. Second, interpretation of an exceptional rule requires the reconciliation of two contradictory rules, and is therefore more likely to result in an error of implementation. Finally, universal rules lend themselves to formulation in terms of absolute imperatives, which will make them more difficult to overcome by the formulation of a contradictory exceptional rule.

Let f be the fraction of within-group interactions with respect to all interactions between individuals. Let c be the fraction of outgroup interactions in which the other player cooperates. In a golden-rule society, the universal rule for cooperation is followed with probability $1 - e_u$, and the exceptional rule for defection against outsiders is followed with probability $1 - e_x$. The mean group payoff is then

$$\begin{aligned}\bar{p}_{\text{golden rule}} &= f[r(1-e_u)^2 + (r+i)(1-e_u)e_u + ie_u^2] + (1-f)(cr + (1-e_x)i) \\ &= f(r - re_u + ie_u) + (1-f)(cr + (1-e_x)i). \end{aligned} \tag{1}$$

Likewise, the mean group payoff for a social-identity society is

$$\bar{p}_{\text{social identity}} = f(r - re_x + ie_x) + (1-f)(cr + (1-e_u)\, i). \tag{2}$$

We define Δp_{switch} as the average benefit to a group switching from social identity to the golden rule:

$$\begin{aligned}\Delta p_{\text{switch}} &= \bar{p}_{\text{golden rule}} - \bar{p}_{\text{social identity}} \\ &= fe_u(i-r) - (1-f)e_x i - fe_x(i-r) + (1-f)e_u i \\ &= (e_x - e_u)(fr - i). \end{aligned} \tag{3}$$

If $e_x > e_u$ as we have argued, the sign of Δp_{switch} depends only on the sign of $fr - i$. If $r > i$, for fixed values of r and i, there exists a critical value of $f = i/r$ that determines which of the two implementations of our behavioral rule is more beneficial to the group. If $f < i/r$, maximal group benefit is achieved through a rule stipulating cooperation specifically in intra-group interactions. If $f > i/r$, it is better to require universal cooperation and to rationalize defection against outsiders on an exceptional basis. Note that this result is independent of c, and so does not depend on the behavior of other groups. It is also independent of the exact values of e_x and e_u, requiring only that e_x be greater than e_u.

3. Effect of Group Size

The parameters i and r are not easily related to measurable quantities. The parameter f could be measured in principle, but this would be difficult in practice. However, group size is a quantity that is easily estimated, and that can be related to f.

Because the preferred rule structure (determined by the sign of Δp_{switch}) has a simple threshold behavior in f, we do not need to relate group size to f quantitatively. It is enough to recognize that, under certain assumptions, larger groups will be associated with larger values of f.

As an example, consider the effect of group size on the value of f. Let us focus on two groups, A and B. Let f_{AA} be the fraction of interactions by members of group A that occur with other group A members and f_{AB} be the fraction that occur with members of group B. Then $f_{\text{AO}} = 1 - f_{\text{AA}} - f_{\text{AB}}$ is the fraction that occur with all other groups. Likewise, let us denote interactions by members of group B which occur with group A members, other group B members, or members of other groups with $f_{\text{BA}}, f_{\text{BB}}$, and $f_{\text{BO}} = 1 - f_{\text{BA}} - f_{\text{BB}}$, respectively. If groups A and B merge to form a common single group, the fraction of intra-group interactions will necessarily increase for the former members of each group. The value of f for former group A members increases from f_{AA} to $f_{\text{AA}} + f_{\text{AB}}$, and for former group B members from f_{BB} to $f_{\text{BB}} + f_{\text{BA}}$.

4. Experimental Evidence

To ask whether there is a possibility of detecting a correlation between community size and cultural norms, we have analyzed data from the Standard Cross Cultural Sample (SCCS), in which 186 cultures around the world have been coded for approximately 2000 variables based on data collected by ethnographers [21]. The 186 cultures included in the SCCS have been selected with an effort to minimizing correlations due to recent shared common ancestry.

We used the "mean size of local community" variable (v235) [21] as a measure of scale, assuming that the fraction f of within-group interactions increases with this size. Due to small numbers of samples at the largest sizes, we collected values from the three largest size categories, all corresponding to a mean community size of 1000 or greater, into a single category.

We used a combination of three variables as an indication of the nature of a universal moral prescription. These variables code the acceptability of violence within communities (v781), between communities but within the same society (v782), and towards other societies (v783) [22]. We included only those cultures for which at least two of these three variables were assigned the same value. Our reasoning is that when the same standard of behavior is applied to at least two different scales, the standard reflects a universal value. If the degree of acceptability of violence differs at one scale, we take this to represent an exceptional rule, which applies only at that scale. When behavioral standards differ at all scales, the data provide no basis for determining which among the standards is universal.

By these criteria, 42 of the cultures in the survey had an identifiable universal rule regarding attitude towards violence, coded here as "disapproved," "tolerated," "approved," or "valued." These data are presented in Fig. 1. We calculated the gamma statistic [23], used for comparisons of two ordinal variables, and found a

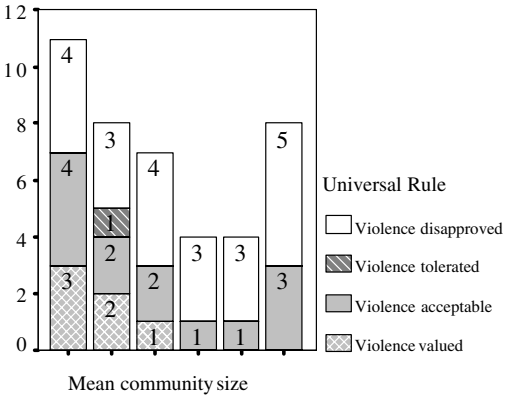

Fig. 1. Ethnographic data was used to infer the nature of universal behavioral norms in 42 societies. These norms, coded in terms of attitudes towards violence, correlate with mean community size ($g = 0.367, p = 0.03$). Local community size (variable 235 in the SCCS [21]) is indicated on the horizontal axis. Bars indicate the number of societies in each size category for which we were able to infer a universal behavioral norm. Small-scale societies were more often characterized by universal norms tolerant of violence, as in the social-identity structure. Large-scale societies more often have behavioral norms universally opposing violence, as in the golden-rule structure.

significant positive correlation between mean size of the local community and the extent to which the universal rule prohibits violence ($g = 0.367$, $p = 0.03$).

5. Conclusion

Despite the simplicity of the model developed here, it makes a clear and testable prediction. Specifically, increasing the fraction of group members' interactions that take place within the group favors group-level institutions that promote universal cooperation. This result is robust to the details of the formulation of the model, and permits insight into group-level behavioral prescriptions. Our model also suggests that the cultural evolution of the golden rule may actually arise from cultural selection for greater ingroup bias. Consolidation into larger organizational units makes a norm of universal altruism a more efficient mechanism of ingroup-bias reinforcement. This could potentially relate the establishment of the major world religions to the emergence of larger political units. It also suggests the possibility of conflict in hierarchically organized societal groups, where the optimal form of the behavioral prescriptions might differ between the higher and lower levels of organization.

Acknowledgments

We thank E. Foley, A. Gingrich, F. Marlowe, and K. Sigmund for the stimulating discussions and helpful comments on earlier versions of the manuscript. Author Stefan Thurner would like to thank the Santa Fe Institute, and in particular J. D. Farmer, for their great hospitality and support.

References

[1] Krey, A. C., *The First Crusade: The Accounts of Eyewitnesses and Participants* (Princeton University Press, Princeton, 1921).

[2] Fulcher of Chartres: *Gesta Francorum Jerusalem Expugnantium*, in Bongars, *Gesta Dei per Francos*, 1, pp. 382 f., translation in O. J. Thatcher and E. Holmes McNeal (eds.), *A Source Book for Medieval History* (New York, Scribners, 1905).

[3] *Gesta Tancredi* in J. Riley-Smith, *The First Crusade and the Idea of Crusading*, (London, 1986).

[4] Watt, W. M., Islam and the holy war, T. P. Murphy (ed.), *The Holy War* (Columbus, 1974).

[5] R. Vossen: Zigeuner: Roma, Sinti, Gitanos, Gypsies zwischen Verfolgung und Romantisierung (Frankfurt, Berlin and Vienna, 1983).

[6] Pope Alexander VI, *Bull Inter Caetera***,** Rome May 4, 1493**;** Bartolome de las Casas, *The Devastation of the Indies: A Brief Account* (translated by Herma Briffault), (Johns Hopkins University Press, Baltimore, 1992).

[7] Billig, M. and Tajfel, H., Social categorization and similarity in intergroup behaviour, *Eur. J. Soc. Psychol.* **3** (1973) 27–55.

[8] Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D. and Wetherell, M. S., *Rediscovering the Social Group: A Self-Categorization Theory* (Blackwell, Oxford, 1987).

[9] Hamilton, W. D., The genetical evolution of social behavior, *J. Theor. Biol.* **7** (1964) 1–16.

[10] Sober, E. and Wilson, D. S., *Unto Others, The Evolution and Psychology of Unselfish Behavior* (Harvard University Press, Cambridge, 1998).

[11] Maynard, S. J., *Evolution and the Theory of Games* (Cambridge University Press, Cambridge, UK, 1982).

[12] Trivers, R. L., The evolution of reciprocal altruism, *Q. Rev. Biol.* **46** (1971) 35–37.

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

[14] Hofbauer, J. and Sigmund, K., *Evolutionary Games and Population Dynamics* (Cambridge University Press, Cambridge, 1998).

[15] Alexander, R. D., *The Biology of Moral Systems* (Aldine de Gruyter, New York, 1987).

[16] Nowak, M. A. and Sigmund, K., Evolution of indirect reciprocity by image scoring, *Nature* **393** (1998) 793–799.

[17] von Hayek, F. A., *Law, Legislation and Liberty: Vol 1: Rules and Order* (University of Chicago Press, Chicago, 1973).

[18] Sartorius, C., The relevance of the group for the evolution of social norms and values, *Constit. Polit. Econ.* **13** (2002) 149–172.

[19] Boyd, R. and Richerson, P. J., *Culture and the Evolutionary Process* (University of Chicago Press, Chicago, 1985).

[20] Henrich, J., Cultural group selection, coevolutionary processes and large-scale cooperation, *J. Econ. Behav. Organ.* **53** (2003) 3–35.

[21] Murdock, G. P. and White, D. R., The standard cross-cultural sample, *Ethnology* **8** (1969) 329–369.

[22] Ross, M., Political decision making and conflict: additional cross-cultural codes and scales, *Ethnology* **22** (1983) 169–192.

[23] Sheskin, D. J., *The Handbook of Parametric and Nonparametric Statistical Procedures* (Chapman & Hall/CRC, 2007).