

Jesus, Hillel and the Man of the Street. Moral and Social Norms in Heterogeneous Populations

Matteo Richiardi*
Florentin Paladi⁺

Version: 21/03/2005

Abstract

This paper studies how different behavioural norms affect the individual and social welfare in a population with heterogeneous preferences. We assume preferences are private information, and that interactions between individuals do not involve communication, nor bargaining. We first compare the so-called golden rules of Jesus (“do to your neighbours what you would like them do to you”) and Hillel (“don’t do to your neighbours what you would not like them do to you”). We consider them as an idealization of an imperative and a more liberal approach to social norms. We find that the aggregate welfare depends on the distribution of preferences in the society. In general, more polarized preferences lead to the supremacy of the Jesus rule, while the more liberal Hillel rule performs better when preferences are more dispersed in the society.

We then introduce a third, more realistic behavioural rule, a “tit-for-tat” strategy that prescribes “do to your neighbours what they have done to you” (keeping the biblical parallel, a reminder of the “eye for eye, tooth for tooth” rule) . We show that when this strategy is followed by everybody it leads to the selection of a single behaviour, which becomes established as a social norm. This behaviour leads in general to more inequality, with respect to the Jesus or Hillel rules. However, it is sufficient that a small group (about 1%) of the population keeps on playing one of the two moral norms to recover the same social welfare that we obtained when everybody played that moral norm.

Keywords : Liberalism, Tit-for-tat, non-market interaction, golden rule

JEL Classification : D63, D64, P50

Acknowledgements: A Lagrange fellowship by ISI Foundation is gratefully acknowledged. FP is grateful to the LABORatorio R. Revelli for the kind ospitality and financial support. MR wishes to thank Michele Sonnessa, author of JAS - the agent-based simulation platform used in the paper - for his software and programming assistance.

* LABORatorio R. Revelli Center for Employment Studies, Moncalieri (Torino), Italy

⁺ State University of Moldova and Free International University of Moldova

The Talmud tells that a gentile came to Hillel saying that he would convert to Judaism if Hillel could teach him the whole Torah in the time that he could stand on one foot. Hillel converted the gentile by telling him, “That which is hateful to you, do not do to your neighbor. That is the whole Torah; the rest is commentary. Go and study it.”

“And seeing the multitudes, Christ went up into a mountain. And when he was set, his disciples came unto him. And he opened his mouth, and taught them, saying – Do unto others as you would have them do unto you.” (Matthew 7:12)

“And if any mischief follow, then thou shalt give life for life, eye for eye, tooth for tooth, hand for hand, foot for foot, burning for burning, wound for wound, stripe for stripe.” (Exodus 21: 23-25)

1. Introduction

The emergence of pro-social behaviour in human societies has been the matter of thorough investigations. Two kinds of explanations have been advanced. One builds upon the hypothesis of rational behaviour of self-interested individuals, and stresses the importance of reciprocal altruism (Triver, 1971; Axelrod and Hamilton, 1981): individuals cooperate in exchange of other people’s cooperation. The other stresses the importance of cultural (Cavalli-Sforza et al., 1981; Boyd and Richerson, 1985) and genetic (Lumsden and Wilson, 1981; Simon, 1983; Wilson and Dugatkin, 1997; Sober&Wilson, 1998) evolution.

In particular, Fehr and Fischbacher (2004) review evidence that human behaviour is often based on *conditional* cooperation, i.e. cooperate if other group members cooperate, and defect if other group members defect. They stress the importance of mechanisms such as expectations, reputation and punishment in order to explain the emergence of reciprocal altruism. However, as Gintis (2000) argues, precisely when a group is threatened and is thus most in need of pro-social behaviour the probability of future interactions goes down, together with the incentives for reciprocal altruism.

It is no surprise then that many studies have shown¹ that people are not only motivated by economic self-interest but also by norms of fairness and reciprocity, that in turn could be explained in terms of evolutionary selection, as sketched above. Religion is one of the mechanisms for strengthening these social norms.

However, although in many cases it is straightforward to identify what is a pro-social behaviour, in general individual preferences are private information. Thus, if player A (the *active* player) wants to act in an altruistic way towards player B (the *passive* player), player A has to guess which action will please the most player B. This point has largely been neglected by the scientific literature, which assumes that the pro-social behaviour is always clearly identified. However, it is present in the religious literature, which generally makes the assumption that, not knowing what your neighbour likes, you should act as if your neighbour were not too different from yourself. This gave rise to a number of “golden rules”, of which two prototypes are the Christian and the Jewish golden rule quoted above. The rule stated by Jesus in his Mountain speech (hereafter, J-

¹ see the references in the review paper by Fehr and Fischbacher cited above.

rule) prescribes to do what you think is good; the rule stated by Hillel (hereafter, H-rule) prescribes not to do what you think is bad. In the history of philosophy there are many antecedents to both rules. On Jesus side we have the Greek philosophers Sextus, Aristotle, Aristippus and Isocrates, while on Hillel side we have Pittacus and Thales. The Chinese philosopher Confucius, at risk of not noting that the J-rule clearly implies the H-rule, looks more Solomonic and prescribes both².

It is easy to find a flavour of socialism in the J-rule, while the H-rule looks definitely more liberal. The purpose of this paper is to investigate their implications for aggregate welfare in the simplest possible model. The model is described in section 2, while the results are derived in section 3. Then, it is then interesting to see what happens if some part of the population departs from the moral norm and plays a tit-for-tat strategy (“what has been done unto you, do it to others”), which as we have seen is after all a very common behaviour.³ This extension is dealt with in section 4 and 5. Section 6 summarizes and concludes.

2. The model

There are N individuals, who can be in 3 different states (call them *Left*, *Center* and *Right*), and can play 3 actions (again *Left*, *Center* and *Right*). Interaction involves always one *active* and one *passive* player⁴. Individuals have preferences over their states: they love one state, they are neutral with respect to another state and they hate the remaining state. When two persons meet, the active player sets the passive player’s state according to his action, which in turn is determined by his moral norm.

This identifies only 6 possible combinations. Denote with $p_1 \dots p_6$ the shares of the population characterized by each combination of preferences, as in table 1. That is, drawing randomly one individual, she will be of type i with probability p_i .

Type	Loved state	Hated state	Share
1	Left	Center	p_1
2	Left	Right	p_2
3	Center	Left	p_3
4	Center	Right	p_4
5	Right	Left	p_5
6	Right	Center	p_6

Table 1: Distribution of preferences in the population

² “Do unto another what you would have him do unto you, and do not to another what you would not have him do unto you.”

³ This rule has also noble origins, reminding the “eye for eye, tooth for tooth” prescription of the Bible. However, as it will be clear later, the “tit-for-tat” rule used in the paper doesn’t allow to address the reaction only to the offender.

⁴ Agents can play both roles interchangeably.

After each interaction, the passive player gets a payoff of +1 if she is in her loved state, a payoff of 0 if she is in her neutral state, and a payoff of -1 if she is in her hated state. The active player does not get any feedback⁵.

If the active player follows the J-rule, he always plays the action corresponding to his loved state. If he follows the H-rule he randomise between the actions corresponding to his loved and neutral state. An example will clarify.

Suppose two individuals, A and B, meet. Player A is the active one. He hates *Left* and loves *Right* (he is thus neutral with respect to *Center*). Player B is the passive one. She loves *Left* and hates *Right* (she is neutral with respect to *Center*, like player A). Suppose A follows the J-rule. He will play *Right*, setting B's state to *Right*. B will then have a payoff of -1. Suppose on the other hand that A follows the H-rule. He will randomise between *Center* and *Right*. The payoff for B could then be either 0 or -1.

Note that there is no *strategic* interaction in the model: the passive player's payoff depends on the active player's choice, but the active player's choice does not depend on the passive player in any way. This implies that game-theoretic solution concepts like Nash equilibrium become useless.

Aggregate welfare is defined both in terms of the mean π and the variance σ^2 of the payoffs (which denote respectively how rich and how unequal the society is). However, in order to avoid arbitrary choices we do not specify a particular functional form, and report separately the results for the mean and the variance.

3. Results

It is straightforward to see that when all individuals share the same preferences (*polarization*) the J-rule is better. In the other extreme case, when preferences are equally distributed in the population (*dispersion*) and $p_1 = p_2 = \dots = p_6 = 1/6$, it is again straightforward to see that the two rules are equivalent, and lead to an average payoff $\pi = 0$. Should we infer that the J-rule always dominates the H-rule?

3.1 Average payoffs

Consider an active player of type 1 (he loves *Left* and hates *Center*), who meets in turn all other (passive) individuals, including himself. If he follows the J-rule, he will play *Left*, causing a payoff of +1 in $(p_1 + p_2)N$ individuals, and a payoff of -1 in $(p_3 + p_5)N$ individuals. Note that there are $(p_1 + p_2)N$ individuals like him in the population.

Suppose now that everybody meets everybody else both as active and as passive player⁶. The average payoff when everybody plays according to the J-rule is then

$$\pi_j = (p_1 + p_2)(p_1 + p_2 - p_3 - p_5) + (p_3 + p_4)(-p_1 + p_3 + p_4 - p_6) + (p_5 + p_6)(-p_2 - p_4 + p_5 + p_6) \quad (1)$$

⁵ We can suppose that he receives a positive payoff deriving from acting accordingly to his moral norm.

⁶ coupling individuals randomly and randomly choosing who is the active and who is the passive player only adds some noise to the results

Similarly, the average payoff with the H-rule is

$$\begin{aligned} \pi_H = & \frac{1}{2}(p_1(p_1 - p_3 - p_5 + p_6) + p_2(p_2 - p_3 + p_4 - p_5) + p_3(-p_1 + p_3 + p_5 - p_6) \\ & + p_4(-p_1 + p_2 + p_4 - p_6) + p_5(-p_2 + p_3 - p_4 + p_5) + p_6(p_1 - p_2 - p_4 + p_6)) \end{aligned} \quad (2)$$

To study the behaviour of $\pi_J - \pi_H$ it is convenient to set some of the probabilities to zero. It is straightforward to see that when there is just one probability different from 0 (and thus equal to 1) we have $\pi_J - \pi_H = \frac{1}{2} > 0$. This proves that if there is extreme polarization of the preferences the J-rule is always better than the H-rule. More in general, the number of different possible combinations of non-zero probabilities is given by the binomial coefficient ${}_n C_k \equiv \binom{n}{k} \equiv \frac{n!}{(n-k)!k!}$. Thus for two non-zero probabilities we obtain a set of ${}_6 C_2 = 15$ equations, which can be grouped in just 3 different functional forms⁷. Figure 1 graphs these 3 curves, for all values of p_i and $p_j = 1 - p_i$.

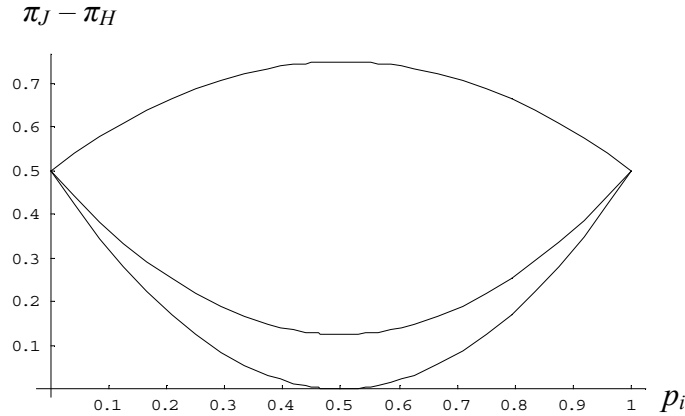


Figure 1: $\pi_J - \pi_H$ in the case of two non-zero probabilities, p_i and $p_j = 1 - p_i$

The J-rule still performs better in all cases but one, when the two rules are equivalent. However, from three non-zero probabilities onward things start to look differently. For three non-zero probabilities we have ${}_6 C_3 = 20$ equations, while for four non-zero probabilities we have ${}_6 C_4 = 15$ equations and for five non-zero probabilities ${}_6 C_5 = 6$ equations. These equations reduce to just three different functional forms in case of three and four non-zero probabilities, and to just one expression in case of five non-zero

⁷ see Appendix 1

probabilities⁸. Figure 2 shows the shape of $\pi_J - \pi_H$ in the case of three non-zero probabilities. A region where the H-rule is better has now appeared (see the last two graphs).

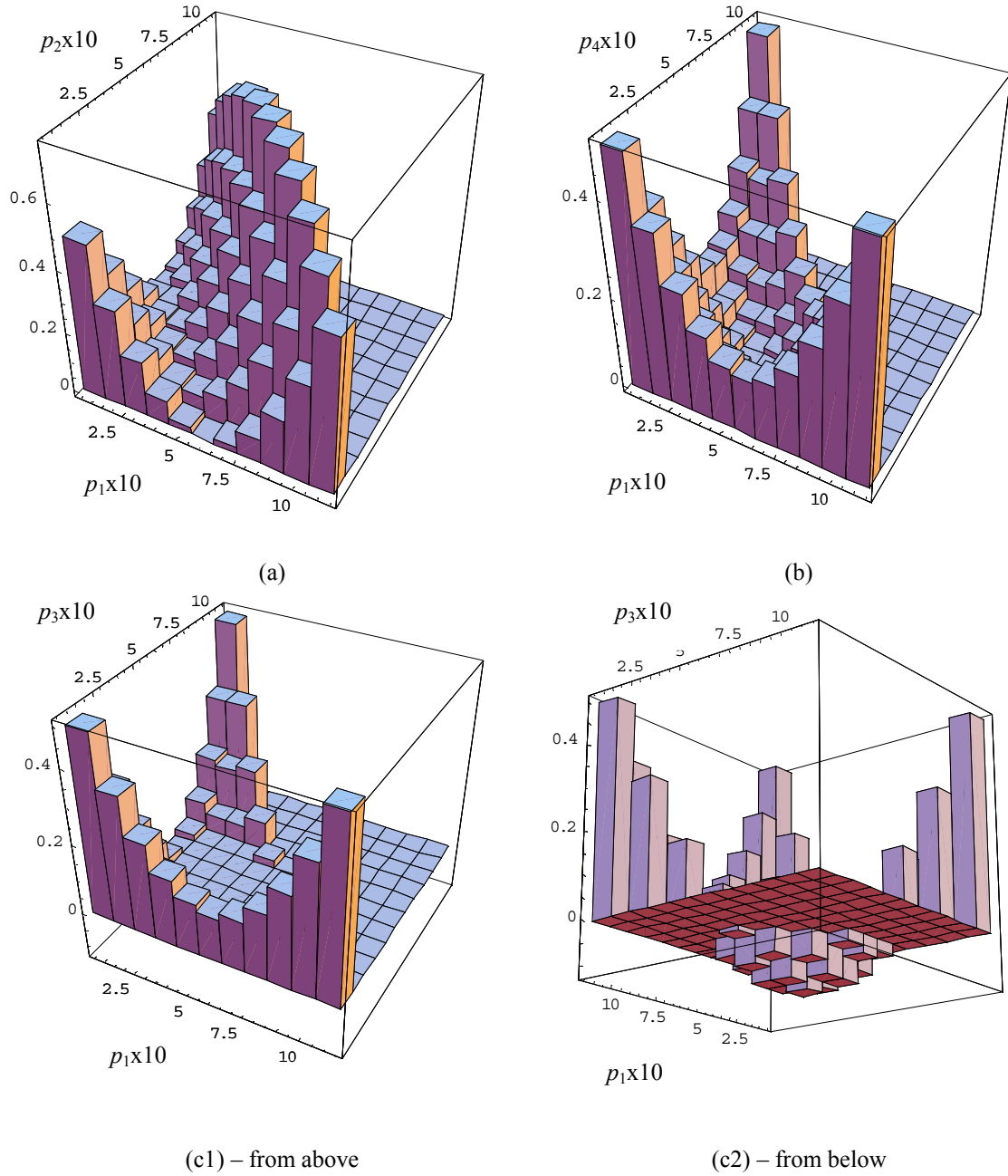


Figure 2: $\pi_J - \pi_H$ in the case of three non-zero probabilities. (a) corresponds to $p_1, p_2, p_3 \neq 0$, (b) to $p_1, p_4, p_5 \neq 0$, (c1) and (c2) to $p_1, p_3, p_5 \neq 0$. The H-rule outperforms the J-rule only for some regions in case (c).

⁸ see Appendix 1

A trend towards a better performance of the H-rule as the distribution of preferences in the society becomes less polarized is evident. However, in order to better investigate it a definition of how much preferences are polarized is needed. We represent the distribution of preferences in the society as a single point in a three dimensional space, where the axes are labeled l , c and r . The l coordinate is found by counting all individuals who love *Left*, and subtracting all individuals who hate *Left*. The result is then normalized to the size of the population. Similarly for the other two coordinates.

Hence,

$$\begin{aligned} l &= p_1 + p_2 - p_3 - p_5 \\ c &= p_3 + p_4 - p_1 - p_6 \\ r &= p_5 + p_6 - p_2 - p_4 \end{aligned} \quad (3)$$

and $l + c + r = 0$.

Note that different distributions of preferences can lead to the same point in the sphere. For instance, the point in the origin is given not only by $p_1 = p_2 = \dots = p_6 = \frac{1}{6}$, but by any combination of preferences such as $p_1 = p_3, p_2 = p_5, p_4 = p_6$.

Note also that this mechanism is very close to defining a Borda count social welfare function.

We can now define the polarization of the preferences in the society as the distance from the center of the sphere:

$$d(l, r, c) \equiv d(p_1, p_2, \dots, p_6) = \sqrt{l^2 + r^2 + c^2} \quad (4)$$

Note that $d \in [0, \sqrt{2}]$: all points thus lie inside a sphere around the origin.

Figure 3 explores how the outcome varies as a function of the distance d . The whole range $[0, 1]$ is sampled, for all probabilities $p_1 \dots p_6$ ⁹. When $\pi_J - \pi_H > 0$ a win is assigned to the J-rule; when $\pi_J - \pi_H < 0$ a win is assigned to the H-rule. For each bin¹⁰, the frequency of wins with each rule is computed (Figure 3a). The raw difference $\pi_J - \pi_H$ is shown in Figure 3b.

Exactly in the center of the sphere the two rules lead to the same payoff, independently of the underlying distribution of preferences. Close to the center, each rule wins in about 50% of the cases. Then, as we move away from the center the J-rule improves its performance, and is always better when the preferences are totally polarized.

⁹ The step considered for creating all combinations of probabilities is 0.025. See Appendix 2 for a general expression for the number of all combinations of probabilities in an N-sized population.

¹⁰ The bin width used in the figure is 0.025

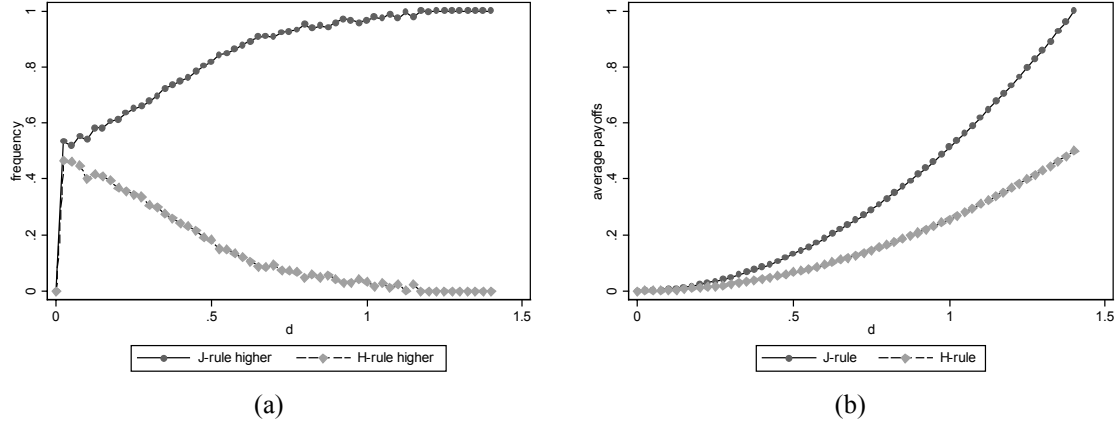


Figure 3: Frequency of negative and positive values of $\pi_J - \pi_H$ (a) and average values for π_J and π_H (b).

In conclusion, we have shown that, depending on the underlying distribution of preferences, both rules can be optimal. However, as the preferences become more polarized, the J-rule clearly takes the lead.

3.2 Variance

The variances σ_J^2 and σ_H^2 are defined for each discrete distribution $D \in J, H$ with the expectation (mean) value π_D as follows:

$$\sigma_D^2 = \sum_{i=1}^6 p_i (\pi_{i,D} - \pi_D)^2 \quad (5)$$

where

$$\begin{aligned} \pi_{1,J} &= p_1 + p_2 - p_3 - p_4 & \pi_{4,J} &= p_3 + p_4 - p_5 - p_6 \\ \pi_{2,J} &= p_1 + p_2 - p_5 - p_6 & \pi_{5,J} &= -p_1 - p_2 + p_5 + p_6 \\ \pi_{3,J} &= -p_1 - p_2 + p_3 + p_4 & \pi_{6,J} &= -p_3 - p_4 + p_5 + p_6 \end{aligned} \quad (5.1)$$

and

$$\begin{aligned} \pi_{1,H} &= (p_1 - p_3 - p_5 + p_6)/2 & \pi_{4,H} &= (-p_1 + p_2 + p_4 - p_6)/2 \\ \pi_{2,H} &= (p_2 - p_3 + p_4 - p_5)/2 & \pi_{5,H} &= (-p_2 + p_3 - p_4 + p_5)/2 \\ \pi_{3,H} &= (-p_1 + p_3 + p_5 - p_6)/2 & \pi_{6,H} &= (p_1 - p_2 - p_4 + p_6)/2 \end{aligned} \quad (5.2)$$

Figure 4a shows the frequency of time σ_J^2 is greater than σ_H^2 , as a function of the distance d^{11} . When everybody shares the same preferences, both variances go to zero¹².

¹¹ The step for increasing/decreasing the probabilities, and the bin width for computing frequencies are the same as above.

However, it can be seen that the variance with the J-rule is generally higher than the variance with the H-rule¹³, especially when the preferences are dispersed in the population, and when they are quite polarized (but not too much). If we assume that the variance in the payoffs enters a social welfare function in a negative way (as it is generally the case) the H-rule could easily become stochastically dominant over the H-rule for low values of the distance d , i.e. in heterogeneous (multiethnic) populations.

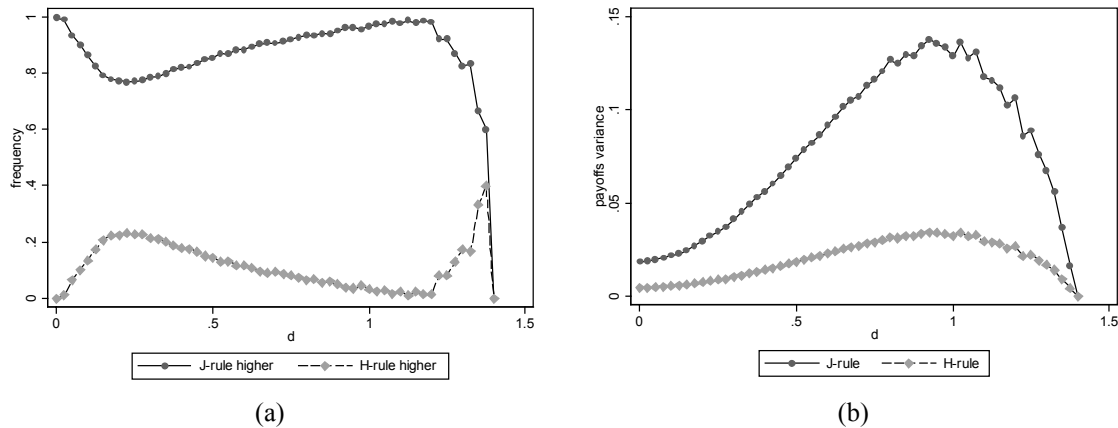


Figure 4: Frequency of negative and positive values of $\sigma_J^2 - \sigma_H^2$ (a) and average values of σ_J^2 and σ_H^2 (b).

4. From moral to social norms

Suppose now that all individuals act according to the following “tit-for-tat” strategy: “if nobody acted to you, play according to your moral norm; otherwise do what your last opponent did to you”. It is easy to see that this strategy always leads to the selection of a single action. Which action will actually be selected depends on the distribution of preferences in the population and on the (random) order of interactions. To investigate the selection process, we draw randomly 4 out of 6 probabilities, say $p_a \dots p_d$. We then set $p_e = 0$ and the remaining probability $p_f = 1 - (p_a + p_b + p_c + p_d + p_e)$. We consider, for computational reasons, a slightly modified version of the model, where each person interacts as active player with only one passive player, randomly chosen. We consider 600 individuals and simulate¹⁴ all interactions for 1,000 periods – an amount of time generally sufficient – given the population size – for the selection process to take place. We perform 50 runs with the same parameters, and then consider the average of the frequencies of each action being played at $t = 1,000$. We then update

¹² This is trivial in case of the J-rule, since everybody plays the same action and gets the same payoff ($+1 \cdot N$, where N is the population size). In case of the H-rule, two actions can be played, causing either a payoff of $+1$ or a payoff of 0 to the passive player, at each interaction. However, since everybody is playing against everybody else, as the population size gets larger everybody gets at the end of the day a payoff of $0.5N$.

¹³ In contrast with the call for equal societies by Jesus himself!

¹⁴ We develop an agent-based simulation using the open-source JAS platform (<http://jaslibrary.sourceforge.net>), see Sonnessa (2004).

the parameters by increasing p_e of a 0.01 step, and decreasing p_f accordingly. We repeat the process until $p_f = 0$. Figure 5 shows the results for $p_2 = 0.22\bar{3}$; $p_3 = p_6 = 0$; $p_4 = 0.1\bar{3}$; $p_1 \in [0, 0.76]$ and $p_5 = 1 - \sum_{i \neq 5} p_i$. As p_5 – which corresponds to people loving *Right* and hating *Left* – increases, the probability that *Left* is selected decreases and the probability that *Right* is selected increases. No threshold effects are present. Such smooth transitions are observed also for other combinations of the parameters. Note that a 50% probability that one action is selected does not mean that half of the population plays that action, while the other half plays something else. It means that in 50% of the runs, without changing the parameters, that action is played by *all* individuals, while in the other 50% of the runs some other action is selected as the only action being played.

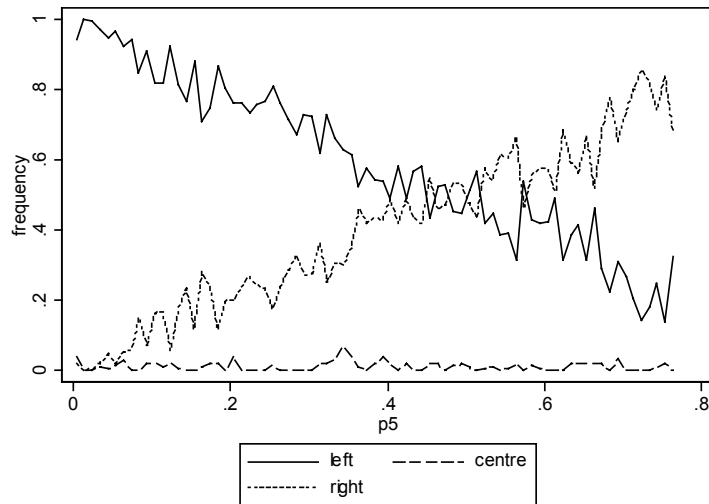


Figure 5: Frequency of each action being played after 1,000 periods, “tit-for-tat” strategy, four probabilities fixed, average of 50 runs for each set of the parameters.

Note that this process of *path dependency* closely resembles the creation of a social norm, which prescribes to play one single action, irrespective of individual preferences. Should we have two distinct populations with the same distribution of preferences, it is very likely that we could observe the selection of a different action within each population, as the social norm of that community. In fact, it is well known that the existence of social norms creates conformity within groups and heterogeneity across groups (Gintis, 2003).

An interesting question is whether the selection of a single action leads to higher or lower average payoff, and to higher or lower variance. Figure 6 plots the evolution of the average payoff, from a situation where everyone plays according to the H-rule (up to $t = 100$), to a situation where everyone plays according to the tit-for-tat strategy of the Man of the Street (from $t = 100$ onward). After a period of oscillations the system eventually settles down and a single action is played by all individuals in the population. In the particular case depicted in Figure 6, the average payoff actually increases, in the

stationary state (although during some parts of the transition process it is actually lower). The upper and lower bounds in the figure are computed as

$$\begin{aligned} \text{average payoff lower bound} &= m_{\pi} - 1.96s_{\pi} \\ \text{average payoff upper bound} &= m_{\pi} + 1.96s_{\pi} \end{aligned} \tag{6}$$

where m_{π} and s_{π} are respectively the mean and the standard deviation of the average payoff under the moral norm regime, *i.e.* in the first 100 periods.¹⁵

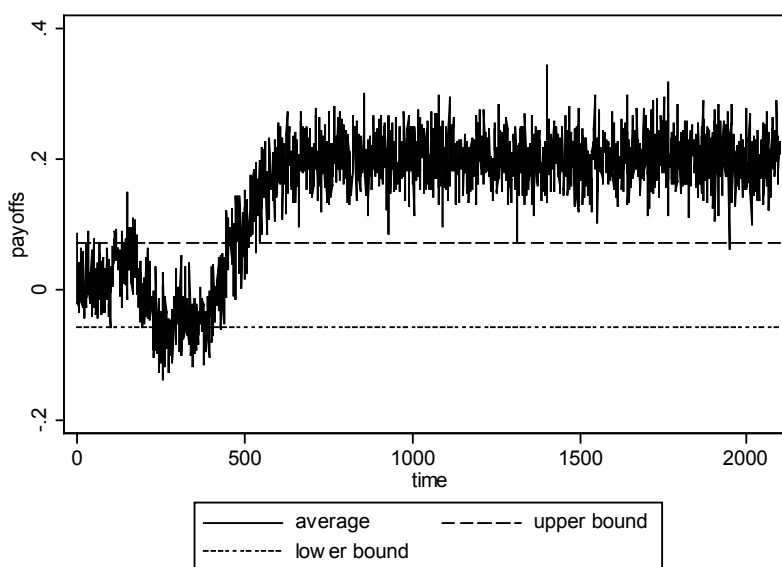


Figure 6: From Hillel to the Man of the Street. Up to $t = 100$ everybody plays according to the H-rule. From $t = 100$ onward everybody plays according to the tit-for-tat strategy. After a period of oscillations, one action is selected. Upper and lower bounds are computed using the average of the mean and the standard deviation of the payoffs during the H-rule regime.

Table 2 reports the frequency when the mean and variance under the tit-for-tat strategy are higher (lower) than under the J-rule and H-rule, respectively. 100 runs are performed starting with the J-rule, and 100 runs are performed starting with the H-rule. Preferences are distributed randomly. During the first 100 periods of each run, every individual in the population follows the moral norm (either the J-rule or the H-rule). From $t = 100$ onward, everybody plays according to the tit-for-tat strategy. The first 1,000 periods under the tit-for-tat regime are discarded¹⁶. Then, the average of the mean and variance of the payoffs in the last 1,000 periods are compared with the

¹⁵ By the law of large numbers any statistics on the population, for a given distribution of preferences, is a Gaussian random variable. Thus, approximately 95% of the observations should lie in the interval between the upper and the lower bound. If we observe a realization outside the interval after having changed the rules of behaviour, we can then conclude that we are sampling from a different distribution: the statistics (here, the average payoff or the variance) has significantly changed.

¹⁶ This is generally sufficient for the selection of one single action, which is played by everyone in the population.

corresponding significance interval computed when everybody played according to the moral norm. This interval is computed according to equation (6) for the mean, and to equation (7) for the variance:

$$\begin{aligned} \text{variance lower bound} &= m_{\sigma^2} - 1.96s_{\sigma^2} \\ \text{variance upper bound} &= m_{\sigma^2} + 1.96s_{\sigma^2} \end{aligned} \tag{7}$$

where m_{σ^2} and s_{σ^2} are respectively the mean and the standard deviation of the payoff variance under the moral norm regime, *i.e.* in the first 100 periods.

	Average payoff under Tit-for-tat (social norm established*)			Payoff variance under Tit-for-tat (social norm established*)	
	% higher than moral norm ⁽⁺⁾	% lower than moral norm ⁽⁻⁾		% higher than moral norm ⁽⁺⁾	% lower than moral norm ⁽⁻⁾
π_J	.58	.42	σ_J^2	.98	.02
π_H	.34	.39	σ_H^2	.98	.00

* after convergence to a single action

(+) mean above the confidence interval for the moral norm

(-) mean below the confidence interval for the moral norm

Table 2: Moral and social norms compared

Table 2 shows that almost can happen with respect to the average payoff. The selection of a single strategy under the tit-for-tat regime leads to a significant increase in the average payoff in 58% of the runs, and to a significant decrease in 42% of the runs, when compared with the J-rule. When compared with the H-rule, it leads to a significant increase in the average payoff in 34% of the cases, to a significant decrease in 39% of the cases, and to results that are roughly similar in 27% of the cases. However, playing tit-for-tat leads almost always to an increase in the variance of the payoffs, hence to an increase in the degree of inequality in the population. This is rather intuitive: when only a single action (the social norm) is played in a population with heterogeneous preferences, someone will be very happy, while someone else very unhappy.

5. From social norms to moral outcomes

So far, we have compared situations where everybody played according to the same strategy, *i.e.* either following a moral norm (the J-rule or the H-rule) or following the tit-for-tat rule. Now, it is interesting to see what happens when the “tit-for-tat” guys are mixed together with the “moral” individuals. Are we going to observe a proportionally “mixed” outcome? And if not, are a few Men of the Street enough to disrupt the moral order, or, conversely, a few fellows of Jesus and Hillel are sufficient to “redeem” the entire population? The observation that reciprocity is indeed one of the pillars of human societies suggests that the most relevant case is when a (possibly small) bunch of “moral” individuals are introduced in a tit-for-tat population.

We look at the fraction of the entire population that has to play according to the moral norm in order to have an outcome (in terms of average payoff and variance) not significantly different to the one obtained when everybody plays according to the moral norm. Table 3 reports, for different values of this fraction, the frequency when the outcome for the statistics considered is within the significance interval, as defined above (mean \pm 1.96 std. dev., computed when everybody plays the moral norm). About 2,000 simulation runs are performed. Preferences are randomly distributed, but are held constant while varying the fraction of the population that plays according to the moral norm.

Fraction of the population playing the moral norm	Runs	J-rule		H-rule	
		$H_0: \pi = \pi_J$	$H_0: \sigma^2 = \sigma_J^2$	$H_0: \pi = \pi_H$	$H_0: \sigma^2 = \sigma_H^2$
100.0%	560	97.7%	100.0%	100.0%	100.0%
5.0%	560	96.8%	99.8%	98.9%	99.6%
1.0%	560	78.3%	83.2%	77.0%	84.1%
0.5%	314	56.3%	65.9%	49.4%	63.1%

Table 3: Non-refusal of the hypothesis that the mean and variance in the payoffs are equal to the case when everybody plays according to the moral norm (H_0), different fractions of the population departing from the moral norm considered.

When only 5% of the population plays according to the moral norm, the outcome is not significantly different from that occurred when everybody shared the moral norm in 96.8% of the cases for the J-rule, and in 98.9% of the cases for the H-rule. A small fraction of 1% of “moral” guys is sufficient to guarantee the same result as in the “moral” society three quarters of the times!

6. Conclusions

Preferences lie at the foundations of economics. The literature on reciprocity and the emergence of social norms generally makes the assumptions that preferences, or at least some proxy, are observable. Individuals can thus decide whether to be keen toward their neighbours or not. Conversely, the case when preferences are not observable has received little or no attention at all in the scientific literature. This is surprising, especially because the theme is at the hearth of the western religious literature. This paper provides a very simple model of individual interaction, in order to test the implications in terms of aggregate welfare of two well-known moral norms: the so-called golden rules of Jesus (“do to your neighbours what you would like them do to you”) and the prescription by Hillel (“don’t do to your neighbours what you would not like them do to you”). We consider them as an idealization of an imperative and a more liberal approach to social norms. We find that the aggregate welfare depends on the distribution of preferences in the society. In general, more polarized preferences lead to the supremacy of the Jesus rule, while the more liberal Hillel rule performs better when preferences are more dispersed in the society.

We then introduce a third, more realistic behavioural rule, a “tit-for-tat” strategy that prescribes “do to your neighbours what they have done to you”. We show that when this strategy is followed by everybody it leads to the selection of a single behaviour, which becomes established as a social norm. This behaviour leads in general to more inequality, with respect to the Jesus or Hillel rules. However, it is sufficient that a small group (about 1%) of the population keeps on playing one of the two moral norms to recover the same social welfare that we obtained when everybody played that moral norm.

Appendix 1. Expressions of $\pi_J - \pi_H$ for different number of non-zero probabilities $p_i, i=1 \dots 6$

No. of non-zero p_i	Non-zero probabilities	Expressions for $\delta = \pi_J - \pi_H$
2	$(p_1, p_2) (p_3, p_4) (p_5, p_6)$	$\delta_1 = \frac{1}{2} + p_i - p_i^2$
	$(p_1, p_3) (p_1, p_6) (p_2, p_4)$ $(p_2, p_5) (p_3, p_5) (p_4, p_6)$	$\delta_2 = \frac{1}{2}(1 - 2p_i)^2$
	$(p_1, p_4) (p_1, p_5) (p_2, p_3)$ $(p_2, p_6) (p_3, p_6) (p_4, p_5)$	$\delta_3 = \frac{1}{2}(1 + 3p_i(p_i - 1))$
3	$(p_1, p_2, p_3) (p_1, p_2, p_4) (p_1, p_2, p_5)$ $(p_1, p_2, p_6) (p_1, p_3, p_4) (p_2, p_3, p_4)$ $(p_3, p_4, p_5) (p_3, p_4, p_6) (p_1, p_5, p_6)$ $(p_2, p_5, p_6) (p_3, p_5, p_6) (p_4, p_5, p_6)$ $(p_1, p_3, p_5) (p_1, p_3, p_6) (p_2, p_3, p_5)$ $(p_2, p_4, p_5) (p_1, p_4, p_6) (p_2, p_4, p_6)$	$\delta_1 = \frac{1}{2}(p_h^2 + p_i^2 + p_j^2 + 4p_h p_i - 2p_h p_j - p_i p_j)$
		$\delta_2 = \frac{1}{2}(p_h^2 + p_i^2 + p_j^2 - 2(p_h p_i + p_i p_j) - p_h p_j)$
	$(p_2, p_3, p_6) (p_1, p_4, p_5)$	$\delta_3 = \frac{1}{2}(p_h^2 + p_i^2 + p_j^2 - p_h p_i - p_h p_j - p_i p_j)$
4	$(p_1, p_2, p_3, p_4) (p_1, p_2, p_5, p_6)$ (p_3, p_4, p_5, p_6)	$\delta_1 = \frac{1}{2}(p_h^2 + p_i^2 + p_j^2 + p_k^2 + 4(p_h p_i + p_j p_k) - 2(p_h p_j + p_i p_k) - p_i p_j - p_h p_k)$
	$(p_1, p_2, p_3, p_5) (p_1, p_2, p_4, p_6)$ $(p_1, p_3, p_4, p_6) (p_1, p_3, p_5, p_6)$ $(p_2, p_3, p_4, p_5) (p_2, p_4, p_5, p_6)$	$\delta_2 = \frac{1}{2}(p_h^2 + p_i^2 + p_j^2 + p_k^2 + 4p_h p_i - 2(p_h p_j + p_i p_k + p_j p_k) - p_i p_j - p_i p_k)$
	$(p_1, p_2, p_3, p_6) (p_1, p_2, p_4, p_5)$ $(p_1, p_3, p_4, p_5) (p_2, p_3, p_4, p_6)$ $(p_2, p_3, p_5, p_6) (p_1, p_4, p_5, p_6)$	$\delta_3 = \frac{1}{2}(p_h^2 + p_i^2 + p_j^2 + p_k^2 + 4p_h p_i - 2(p_h p_j + p_h p_k) - p_i p_j - p_i p_k - p_j p_k)$
5		$\delta = \frac{1}{2}(p_h^2 + p_i^2 + p_j^2 + p_k^2 + p_l^2 + 4(p_h p_i + p_j p_k) - 2(p_h p_j + p_i p_k + p_i p_l + p_j p_l) - p_i p_j - p_h p_k - p_h p_l - p_k p_l)$

Appendix 2. Equation for the number of possible partitions P of N individuals into different subgroups

Let $N = 1, 2, \dots, \infty$ be the total number of individuals in the model, and the $\{n_1, n_2, n_3, n_4, n_5, n_6\}$ is their partition into 6 subsets which varies as changing the step considered for creating all combinations of probabilities $p_i, i=1..6$. Each subset can be called *cluster*, and the process itself – *clustering*. The size of each cluster can vary from 0 to N , $n_i = 0..6, i = 1..6$, and $\sum_{i=1}^6 n_i = N$. Thus the number of possible outcomes P is a

function of N , and the explicit solution is
$$P(N) = \frac{1}{5!} \prod_{i=1}^5 (N + i).$$

In general, when the number of subsets is $m > 1$, the number of possible partitions P of

N individuals into m subgroups is
$$P(N) = \frac{1}{(m-1)!} \prod_{i=1}^{m-1} (N + i).$$

Some numerical results are presented in the table below.

N	P	N	P	N	P
1	6	11	4,368	21	65,780
2	21	12	6,188	22	80,730
3	56	13	8,568	23	98,280
4	126	14	11,628	24	118,755
5	252	15	15,504	25	142,506
6	462	16	20,349	50	3,478,761
7	792	17	26,334	100	96,560,646
8	1,287	18	33,649	200	2,872,408,791
9	2,002	19	42,504	600	664,353,676,371
10	3,003	20	53,130	1,000	8,459,043,543,951

References

- Axelrod R., Hamilton W.D. (1981), “The evolution of cooperation”, *Science*, 211, 1390–1396
- Boyd R., Richerson, P.J. (1985), *Culture and the Evolutionary Process*, University of Chicago Press, Chicago
- Cavalli-Sforza L., Feldman M.W. (1981), *Cultural Transmission and Evolution*, Princeton University Press, Princeton, NJ
- Fehr E., Fischbacher U. (2004), “Social Norms and Human Cooperation”, *TRENDS in Cognitive Sciences*, 8 (4), 185-190
- Gintis H. (2000), “Strong Reciprocity and Human Sociality”, *Journal of Theoretical Biology*, 206, 169-179
- Gintis H., Bowles S., Boyd R., Fehr E. (2003), “Explaining Altruistic Behavior in Humans”, *Evolution and Human Behavior*, 24, 153-172
- Lumsden C.J., Wilson E. O. (1981), *Genes, Mind, and Culture: The Coevolutionary Process*, Harvard University Press, Cambridge, MA
- Simon H.A. (1993), “Altruism and economics”, *American Economic Review*, 83, 156-161.
- Sober E., Wilson D.S. (1998), *Onto Others: The Evolution and Psychology of Unselfish Behavior*, Harvard University Press, Cambridge, MA
- Sonnessa M. (2004), “JAS: Java Agent-based Simulation Library, an Open Framework for Algorithm-Intensive Simulations”, in Contini B., Leombruni R., Richiardi M. (eds), *Industry and Labor Dynamics: The Agent-Based Computational Economics Approach*, World Scientific, Singapore
- Trivers R. L. (1971), “The evolution of reciprocal altruism”, *Quarterly Review of Biology*, 46, 35–57.
- Wilson D.S., Dugatkin L.A. (1997), “Group selection and assortative interactions”, *American Naturalist*, 149, 336-351.