The evolution of social norms

Ivar Kolstad

WP 2003: 1

The evolution of social norms

Ivar Kolstad

WP 2003: 1

Chr. Michelsen Institute *Development Studies and Human Rights*

CMI Reports

This series can be ordered from:
Chr. Michelsen Institute
P.O. Box 6033 Postterminalen,
N-5892 Bergen, Norway
Tel: + 47 55 57 40 00
Fax: + 47 55 57 41 66
E-mail: cmi@cmi.no
www.cmi.no

Price: NOK 50

ISSN 0805-505X
ISBN 82-8062-044-3
This report is also available at:
www.cmi.no/public/public.htm

Indexing terms

Evolutionary game theory
Social norms

JEL: C73

Contents

1. Introduction 1
2. Definition of social norms and evolutionary game theory 1
3. Cooperative norms: The prisoner’s dilemma game 5
4. Fairness norms: Bargaining games 9
5. Conclusion 13

References 14

1. Introduction[1]

A great number of articles and books have been published on the subject of evolutionary game theory and norms (Binmore (1998), Skyrms (1996) and Bicchieri (1997), to name a few). This paper outlines the basic evolutionary approach to social norms, and illustrates the potential and the limits of evolutionary game theory in explaining how social norms are formed and maintained. In more precise terms, evolutionary game theory suggests that social norms emerge due to a process of adaptation. The implications of this idea for the emergence of certain social norms, can be examined by means of a few simple games. However, the basic mechanism by which evolutionary game theory seeks to explain social norms, makes the theory more congenial to analyzing the subset of norms focusing on material payoffs, than norms that cannot be given a reasonable interpretation in these terms.

The paper is structured as follows. In section 2, a suitable definition of the term social norms is suggested, followed by an account of the key features of evolutionary game theory. The next two sections consider two key tenets in the evolutionary literature on norms, represented by two different classes of games. The first of these, the prisoner's dilemma game and related games, highlight the difference between the individually and the collectively optimal. The second, bargaining games such as the ultimatum game, highlight the distributional aspects of interaction. In a final section, the limits of evolutionary game theory in studying social norms is discussed, as well as some general points of criticism against the evolutionary approach.

2. Definition of social norms and evolutionary game theory

There are numerous definitions of social norms, with variations across scientific disciplines. What is striking about the evolutionary literature on social norms, is that everyone uses the term and seems to know what is means, but very few bother with an explicit definition. The implicit definition of a social norm seems to be some sort of behaviour that runs counter to some idea of narrow self-interest.

[1] The author thanks Rashid Sumaila, Arne Wiig, Espen Villanger and Anna Milford for comments to a previous version of this paper. All remaining errors are the responsibility of the author.

For the purposes of this paper, let us use the following definition, employed by Fehr and Gächter (2000):

A social norm is
1) a behavioural regularity
2) that is based on a socially shared belief of how one ought to behave
3) which triggers the enforcement of the prescribed behaviour by informal social sanctions.

A fair interpretation of the first part is that a social norm is a pattern of behaviour, that persists across a significant share of agents and over time. The second part states that the pattern of behaviour reflects some kind of moral code that agents share. The moral code could reflect any kind of basic ethical principle, for instance the utilitarian "one ought to behave so as to maximize the sum of happiness" or the rights-based "one ought to respect the human rights of others". However, the definition does not say that the moral code needs to have a justification in normative ethics, one could also have a socially shared belief that one should steal or lie or cheat, or a belief that one ought always to act in one's narrow self-interest.

On the face of it, the third part of the definition does not preclude self-interested behaviour either. However, Fehr and Gächter add the condition that enforcement entail a net cost to the enforcer, that agents should be willing to punish deviations from the moral code even at a cost to themselves. Since a narrowly self-interested agent would be disinclined to do so, it is reasonable to say that the definition of social norms is inconsistent with narrow self-interest. So in sum, a social norm is a behavioural pattern that reflects something other than narrow self-interest.

Let us turn now to evolutionary game theory. What are the main characteristics of this theory? The term evolutionary game theory now encompasses a large and quite varied set of models. A common feature of these models, is that players are matched repeatedly to play a game, and a dynamic process describes how players adapt their behaviour over time.

A simple example of such a dynamic process, is the replicator dynamics, which states that the population share x_i of strategy i, grows at a rate equal to the difference between the average or expected payoff to strategy i and the average payoff of all strategies in the population:

$$\frac{\dot{x}_i}{x_i} = p_i - p_{average} \quad (1)$$

In other words, strategies that do better than average increase their population shares over time, and strategies that do worse decrease theirs. And the greater is the deviation in payoffs from the average, the faster does the population share increase or decrease. One can thus think of an evolutionary model as a contest between strategies, strategies that do well multiply, strategies that do poorly dwindle.

The distinguishing feature of evolutionary game theory is thus that is sees behavioural patterns as the outcome of a process of adaptation, in which behaviours that do better are selected. The process of adaptation might reflect biological selection, or it might represent learning as agents switch to strategies that are observed to do better. Adaptation is also what sets evolutionary game theory apart from traditional game theory, by focussing on how an equilibrium comes to be played, rather than seeing an equilibrium as something that "springs into being" (Samuelson, 2002).

In general, evolutionary models differ along three dimensions: The way in which players are matched, the dynamic process of adaptation, and the role of mutations. The interaction of these three elements determines which equilibrium is selected, in other words which kind of behaviour survives in the long run. I will not give a full review of these intricacies here, let me instead introduce a few basic tools from the evolutionary tool-kit which allow me to demonstrate how the theory can be applied to the issue of social norms.

A ground-breaking event in evolutionary game theory was the formulation of the principle of evolutionary stability, by Maynard Smith and Price (1973). Assume that from a large population, players are repeatedly and randomly matched to play a game that is symmetric in the sense that players have identical strategy sets and payoffs, and it is impossible to condition strategies on the characteristics of an opponent. Also, assume that payoffs in the game reflect fitness, in other words there is a selection process which favours those earning higher payoffs.

In this setting, we define the concept of evolutionary stability.[2] Imagine that the entire population plays a certain strategy *x*. This strategy *x* is evolutionarily stable, if a small group of players using a different strategy *y* cannot persist in the population. In other words, if we call a small group of players using an alternative strategy mutants, an evolutionarily stable strategy (ESS) is robust to mutations in a certain sense. To be uninvadable in this way, an evolutionarily stable strategy must earn a higher expected payoff than any mutant strategy. In formal terms, this implies that a strategy *x* is evolutionarily stable if these two conditions hold:

$$u(x,x) \geq u(y,x) \tag{2}$$

$$u(x,x) = u(y,x) \Rightarrow u(x,y) > u(y,y) \tag{3}$$

The first conditions says that an ESS *x* must earn at least as high payoffs against itself as does any mutant strategy *y* against *x*. The second conditions says that if a mutant strategy *y* does as well against *x* as does *x*, then *x* must do strictly better against the mutant *y* than the mutant does against itself.

Notice that the two conditions imply that an ESS must be a best reply to itself, and hence a Nash equilibrium (NE):

x is an ESS $\Rightarrow$ (x,x) is a NE

Moreover, the concept of an ESS ties in with the basic dynamic story of evolutionary models in the following way: An evolutionarily stable strategy is asymptotically stable under the replicator dynamics. In other words, if you start in a situation in which almost everyone is playing an ESS, the replicator dynamics will lead you to the state in which everyone plays the ESS.

An ESS does not always exist. Thus sometimes we have to resort to weaker stability concepts. One such concept is a neutrally stable strategy (NSS). A strategy is neutrally stable if a small group of mutants can persist, but not increase their population share. In other words, while an

[2] For a technical introduction to evolutionary game theory, see Weibull (1995).

ESS must earn a strictly higher expected payoff against any mutant, an NSS need only earn as high an expected payoff as any mutant. Against an ESS, no mutant can persist, against an NSS, no mutant can thrive.

Now, how does this help us understand social norms. Evolutionary game theory provides the tools to analyze which strategies, or patterns of behaviour, emerge over time through a process of adaptation. Social norms are patterns of behaviour with certain characteristics. We can therefore use evolutionary game theory to examine the conditions under which these particular patterns called social norms emerge.

3. Cooperative norms: The prisoner's dilemma game

To study the evolution of social norms in this manner, we need some basic setup which contrasts social norms with other patterns of behaviour. A game which highlights one particular contrast is the prisoner's dilemma (PD) game. Indeed, the PD game is often called the prototypical model of social norms. As this game is well known, let me just briefly summarize its characteristics:

		Player 2	
		C	D
Player 1	C	*2,2*	*0,3*
	D	*3,0*	*1,1*

(G1)

There are two players with two strategies, cooperate (C) and defect (D). Defect is a strict best reply to any strategy, and the unique Nash equilibrium is therefore for both players to defect. Two narrowly self-interested players would thus choose to defect, though this leaves both of them worse off than if they both chose to cooperate. In sum therefore, the prisoner's dilemma

game thus highlights the conflict between what is collectively desirable (cooperation) and what is individually desirable (defection).

A series of experiments have been conducted which expose subjects to the prisoner's dilemma game, or public goods games, which are essentially n-person prisoner's dilemma games.[3] The results of these experiments do not entirely confirm the prediction of the Nash equilibrium concept. Subjects do exhibit a positive rate of cooperation in these games. But as the game is repeated many times, the rate of cooperation appears to fall. However, if you introduce the possibility of punishing your opponent after the game is played, a quite high rate of cooperation can be sustained. This happens though self-interested players would not use costly sanctions in this way. Finally, punishment gets even more effective if the game is played repeatedly by a given set of players. Under certain conditions, therefore, a cooperative norm seems to arise when a game of this kind is played.

Can we use evolutionary game theory to understand how a cooperative norm could arise and be sustained? Let us apply the concept of an evolutionarily stable strategy to the PD game. Since I have already mentioned that an ESS must constitute a Nash equilibrium of the underlying game, the conclusion that cooperate is not an ESS follows immediately. In fact, defect is the unique ESS. The intuition for this result is obvious. A defector does strictly better against a cooperator (payoff 3) than a cooperator does against itself (payoff 2), so a small group of mutant defectors could invade a population of cooperators. However, a cooperator does strictly worse against a defector than the defector does against itself, so mutant cooperators cannot invade a population of defectors.

A population that is randomly and repeatedly matched to play a one-shot PD game, thus ends up playing defect. However, by tweaking the assumptions underlying the analysis, we can attain selection of a cooperative norm. I will discuss two ways in which this can be done: Assuming that matching is assortative rather than random, and adding a second stage of punishment to the PD game.

With random matching of players, the probability that you meet a cooperator or defector equals the share of each type in the population. Let us assume now, however, that matching is

[3] See Kagel and Roth (1995) and Fehr and Gächter (2000) for surveys of these experiments.

assortative or viscous, in the sense that players are more likely to be matched with opponents that share their strategy than under random interaction (Hamilton, 1964). A way to formalize this is to say that with probability *r* a cooperator is matched with another cooperator, whereas with probability *1-r* he is matched with a random member of the population.

Assume the population share of cooperators is *s*. Given the specification of the game in (G1), the expected payoff to a cooperator and a defector are now, respectively:

$$u(C) = r \cdot 2 + (1-r)\left[s \cdot 2 + (1-s) \cdot 0\right] = 2r + 2s - 2rs \tag{4}$$

$$u(D) = r \cdot 1 + (1-r)\left[s \cdot 3 + (1-s) \cdot 1\right] = 1 + 2s - 2rs \tag{5}$$

In other words, cooperators do better than defectors if:[4]

$$r > \frac{1}{2} \tag{6}$$

Since matching is not random, we cannot apply the concept of evolutionary stability here. Nevertheless, we can take a view of stability that is analogous. If matching is sufficiently assortative, cooperators on average do strictly better than defectors, which means that a small segment of mutant cooperators can invade a population of defectors. In contrast, mutant defectors cannot invade a population of cooperators. We can thus substantiate the evolution of a cooperative norm when matching is assortative.

Assortative matching in some sense requires that cooperators can recognize each other, that their behavioural dispositions are observable. Frank (1988) suggests that behavioural dispositions of humans are in fact observable through expressions of emotions. A person known to blush, for instance, has an advantage in acquiring the trust of others. To the extent that these emotional expressions are hard to fake, there is a possibility that cooperators can in fact recognize each other in this manner. These ideas suggest an explanation for a greater level of trust in face-to-face interaction and interaction in small groups, as opposed to more

[4] A similar condition is derived for a more general specification of the game in Sethi and Somanathan (2003).

anonymized interaction in large groups. The idea that a predisposition can be signalled has been discussed further in Robson (1990) and Nowak and Sigmund (1998).

Assortative matching does in a sense imply some form of punishment: If you are not a cooperator, you don't get to play with the other cooperators. An alternative way to obtain cooperation in a one-shot PD game, is to explicitly introduce a second sanctioning stage after the PD game (Sethi and Somanathan, 2003). If g is the cost to the punisher, and d is the damage done to the punished, we can write the payoffs at this second stage in the following way:

		Player 2	
		P	NP
Player 1	P	$-g-d$	$-g$
	NP	$-d$	0

(G2)

There are now eight available strategies; cooperation or defection at the first stage, coupled with either of four actions at the punishment stage: no punishment, punish cooperators, punish defectors, or punish both cooperators and defectors. This two-stage game does not have an ESS, so for all strategies there is a mutant that does at least as well. However, both (defect, no punishment) and (cooperate, punish defectors) are neutrally stable strategies. In a population playing (defect, no punishment), a mutant playing (defect, punish cooperators) does as well and can thus persist but not grow. In a population playing (cooperate, punish defectors), a mutant playing (cooperate, no punishment) can persist but not grow. However, neither (defect, no punishment) nor (cooperate, punish defectors) can be eradicated by a mutant, so there is some room here for a cooperative norm to persist.

There are also other ways to get selection of cooperative behaviour in prisoner's dilemma games. In the one-shot case, local interaction provides one such avenue (Eshel et al, 1998). In the case where the PD game is repeated infinitely at each stage, strategies that exhibit cooperative features such as tit-for-tat are neutrally stable but so are strategies that defect

almost always (Axelrod and Hamilton (1981), Fudenberg and Maskin (1990)). These latter strategies are however selected against if one introduces complexity costs or implementation errors (Binmore and Samuelson (1992), Fudenberg and Maskin (1990)).

So, what does the analysis of a PD game really tell us. It tells us that under certain conditions, a cooperative pattern of behaviour can persist under evolutionary pressure, and under certain conditions, evolutionary pressures can eradicate a narrowly self-interest pattern of behaviour. In other words, from the PD game we can deduce the conditions under which a certain type of social norm is sustainable. The norm in question is a cooperative norm, or a norm against opportunism. This is as far as the PD game takes us. Thus for instance, it is hard to relate this analysis to concepts of distributive justice. To do so, we need to study another class of games.

4. Fairness norms: Bargaining games

Bargaining games such as dictatorship games, ultimatum games, Nash demand games and so on, have players split a pie. Strategies available range from an equal split to uneven splits, and games of this sort thus highlight the contrast between behaviour that is narrowly self-interested (more for me) and behaviour that reflects some more egalitarian norm. Games of this type thus provide a perspective on fairness norms.

Take for instance the ultimatum game. In this game, player 1 makes an offer of how to divide the pie (for instance, 70% for me, 30% for you), player 2 chooses whether or not to accept the offer. If the offer is accepted, the pie is divided according to the offer made, if the offer is refused both players get nothing. From the tools of traditional game theory, you would expect the first player to offer the second a pittance, and the second player to accept. This is in fact the subgame-perfect equilibrium of the game.

However, in experiments conducted on the basis of ultimatum games, subjects typically do not accept offers that are low, and players consequently do not make offers that are too low.[5] Behaviour thus seems to be guided by some sort of fairness norm. The question then is, can we use the tools of evolutionary game theory to explain how such norms form and are

[5] See Kagel and Roth (1995) for a survey of bargaining experiments.

maintained. In other words, does being egalitarian in some sense confer an advantage on a player in the presence of evolutionary pressures.

To demonstrate how we could get selection of fairness norms through evolutionary dynamics, let me apply the concepts introduced earlier to a simple bargaining game that makes some interesting distinctions. Consider a large population of agents who are repeatedly and randomly paired to play a Nash demand game with a twist. In each period, two matched players interact twice. In the first interaction, one player is assigned position A and the other position B, and they then split a two-piece pie. In the second interaction, the players swap positions, and split another two-piece pie. Let us assume that each player's demand for the pieces of pie that he wants in positions A and B, is submitted at the start of each period, $\left(p_A^i, p_B^i\right)$, and that the demand in the second interaction cannot be made contingent on play in the first interaction.

The payoffs of the game are assumed to be the following. If the demands of the players in positions A and B do not exceed the total size of the pie, they each get the number of pieces they demanded. Assume, however, that a piece of pie is twice as valuable to a player in position B as to a player in position A. Hence, we can write the payoffs to the player in position A as the number of pieces he gets, and the payoffs to the player in position B as two times the number of pieces he gets. If the demands of the two players exceed the total size of the pie, they each get nothing.

$$\begin{aligned} p^{i}_{A} &= p^{i}_{A} \text{ and } p^{j}_{B} = 2p^{j}_{B} \quad \text{if} \quad p^{i}_{A} + p^{j}_{B} \leq 2 \\ p^{i}_{A} &= p^{j}_{B} = 0 \quad\quad\quad\quad\quad \text{if} \quad p^{i}_{A} + p^{j}_{B} > 2 \end{aligned} \tag{7}$$

Each player plays the game twice in each period, once in position A and once in position B. The total payoffs to a player in a period, is the sum of his payoffs in position A and position B.

$$p^{i} = p^{i}_{A} + p^{i}_{B} \tag{8}$$

The total payoffs of the game are captured by the matrix in table 1. There are nine different strategies in the game, each of which demands 0, 1 or 2 pieces in position A, and 0, 1, or 2

pieces in position B. Now, the number in the matrix are the total payoff to a row strategy that is matched with a column strategy. For instance, if row strategy (1,1) meets column strategy (1,1), the row players gets a payoff of 3. Let me explain why. In the first interaction, the row player occupies position A and demands 1 piece of pie. His opponent occupies position B and also demands 1 piece of pie. The demands of the two players do not exceed the total size of the pie, and they each get 1 piece of pie. For the row player in position A, one piece of pie equals a payoff of 1. The two players then swap positions, the row player in position B demands one piece, his opponent in position A demands on piece, and they each get one piece. However, the row player is now in position B and one piece of pie thus yields a payoff of 2. The total payoffs to the row player (1,1) are thus 1 in the first interaction, 2 in the second interaction, for a total of 3 across both interactions.

	(2,2)	**(2,1)**	**(2,0)**	**(1,2)**	**(1,1)**	**(1,0)**	**(0,2)**	**(0,1)**	**(0,0)**
(2,2)	**0**	0	2	0	0	2	4	4	6
(2,1)	0	**0**	2	2	2	4	2	2	4
(2,0)	0	0	**2**	0	0	2	0	0	2
(1,2)	0	1	1	**0**	1	1	4	5	5
(1,1)	0	1	1	2	**3**	3	2	3	3
(1,0)	0	1	1	0	1	**1**	0	1	1
(0,2)	0	0	0	0	0	0	**4**	4	4
(0,1)	0	0	0	2	2	2	2	**2**	2
(0,0)	0	0	0	0	0	0	0	0	**0**

Table 1. Payoffs to row strategy in the Nash demand game with a twist

Let us take a closer look at the strategies. The topmost strategy (2,2) demands the whole pie in both positions A and B, thus we can dub this strategy selfish. In contrast, the bottommost strategy (0,0) demands nothing in either position, which makes it totally other-regarding. The strategy (1,1) is Rawlsian in each interaction, since an even split of the pie gives the worst off player a payoff of 1, whereas any other split gives the worst off player a payoff of 0. Finally, the strategy (0,2) is utilitarian in each interaction, since it always gives the whole pie to the

player occupying position B, which produces a total payoff of 4 in each interaction, which is higher than that realized by any other division.

We can now use the concept of evolutionary stability to see how these strategies would fare in an evolutionary setting. It turns out that the unique ESS of the game is the Rawlsian (1,1). Let me explain why. Remember the two conditions (2) and (3) of an ESS: An ESS *x* must do at least as well against itself as any mutant *y*, and if the two do equally well against *x*, then *x* must do strictly better against *y*, than the mutant does against itself.

The Rawlsian (1,1) gets payoff 3 when matched with itself, whereas the most any other strategy gets when playing (1,1) is 2. Since the Rawlsian strategy does strictly better against itself than any other strategy, a population of Rawlsians cannot be invaded by a mutant playing another strategy, and the Rawlsian strategy is therefore an ESS.

There is no other ESS. There are three more strategies that meet the first criterion of an ESS. The utilitarian (0,2), the selfish (2,2) and a third strategy (2,0) all do as well against themselves as does any other strategy. However, neither of the three meet the second condition of an ESS. Take the utilitarian (0,2). It gets a payoff of 4 when matched with itself, and so does the strategy (1,2) when matched with a utilitarian. Against (1,2), a utilitarian gets payoff 0, as does (1,2). In other words, a population of utilitarians would get the same expected payoff as a mutant (1,2), and the mutant thus cannot be expelled by the utilitarians. A similar argument explain why the selfish strategy (2,2) and to (2,0) are not ESS.

The Rawlsian (1,1) is thus the unique ESS of the game. The concept of evolutionary stability thus selects a strongly egalitarian pattern of behaviour over selfish ones in this simple model. By the way, note that this strategy equals the Nash bargaining solution in this game. Note also that altering the basic assumptions of the model changes the result. Consider for instance the case of assortative matching. In the extreme case of perfectly assortative matching, where strategies interact only with themselves, the utilitarian strategy (0,2) has a payoff of 4 which is higher than any other strategy. Perfectly assortative matching in this case thus promotes a utilitarian pattern of behaviour.

Several more comprehensive evolutionary models of bargaining have been proposed. Some of these, such as Young (1993), confirm that the Nash bargaining solution is particularly robust

to evolutionary pressures. Remember that in the symmetric case, the Nash bargaining solution is just an even split, so there is a case here for egalitarian norms. Perhaps the most ambitious study of the evolution of fairness norms is that of Binmore (1998), who provides an involved argument for the selection of Rawlsian norms.

5. Conclusion

Evolutionary game theory studies how patterns of behaviour emerge over time through a process of adaptation. Using some simple tools from evolutionary game theory, this paper shows how patterns of behaviour consistent with an idea of social norms can form and persist in evolutionary models. In particular, I have shown how cooperative norms and norms of fairness can be sustained in prisoner's dilemma and bargaining games.

In evolutionary models, selection relates to the material payoffs attained by different types of behaviour. Behaviours that do well in material terms multiply, those that do poorly in material terms dwindle. In the context of social norms, the evolutionary approach thus seems more congenial to analyzing the subset of norms that somehow focus on material payoffs, or that can be given a reasonable interpretation in these terms. It is thus not a given that all types of social norms, such as norms focussing on rights or virtues, can be equally well addressed by evolutionary game theory.

In more general terms, Sugden (2001) has criticized the evolutionary approach for uncritically applying a methodology taken from biology to human decision making. His main point seems to be that the evolutionary approach to human decision making lacks an empirical foundation, that little has been done to empirically substantiate the basic assumptions of evolutionary models. Sugden's point seems a fair one, yet there are signs that these matters are taken more seriously in the evolutionary literature. To cite but one example, Battalio et al (2001) test some basic assumptions of evolutionary game theory in an experimental setting.

References

Axelrod, R. and Hamilton, W. D. (1981); "The evolution of cooperation", *Science*, 211, 1390-1396

Battalio, R., Samuelson, L. and van Huyck, J. (2001), "Optimization incentives and coordination failure in laboratory stag hunt games", *Econometrica*, 69, 3, 749-764

Bicchieri, C., Skyrms, B., Jeffrey, R. (ed.) (1997), *The dynamics of norms*, Cambridge University Press, Cambridge

Binmore, K. (1998), *Game theory and the Social Contract II – Just playing*, The MIT Press, Cambridge Mass.

Binmore, K. and Samuelson, L. (1992), "Evolutionary stability in repeated games played by finite automata", *Journal of economic theory*, 57, 278-305

Eshel, I., Samuelson, L. and Shaked, A. (1998), "Altruists, egoists, and hooligans in a local interaction model", *American Economic Review*, 88, 157-179

Fehr, E. and Gächter, S. (2000), "Fairness and retailiation: The economics of reciprocity", *Journal of economic perspectives*, 14, 3, 159-181

Frank, R.H. (1988), *Passions within reason: The strategic role of the emotions*, W.W. Norton, New York

Hamilton, W. D. (1964), "The genetical evolution of social behavior", *Journal of theoretical biology*, 7, 1-16

Kagel, J. H. and Roth, A. E. (1995), *The handbook of experimental economics*, Princeton University Press, Princeton, New Jersey

Maynard Smith, J. and Price, G.R. (1973), "The logic of animal conflicts", *Nature*, 246, 15-18

Nowak, M.A. and Sigmund, K. (1998), “Evolution of indirect reciprocity by image scoring”, *Nature*, 393, 573-577

Robson, A. (1990), “Efficiency in evolutionary games: Darwin, Nash and the secret handshake”, *Journal of theoretical biology*, 144, 379-396

Samuelson, L. (2002), “Evolution and game theory”, *Journal of economic perspectives*, 16, 2, 47-66

Sethi, R. and Somanathan, E. (2003), “Understanding reciprocity”, *Journal of economic behavior & organization*, 50, 1-27

Skyrms, B. (1996), *Evolution of the social contract*, Cambridge University Press, Cambridge

Sugden, R. (2001), “The evolutionary turn in game theory”, *Journal of economic methodology*, 8, 1, 113-130

Weibull, J. W. (1995), *Evolutionary game theory*, MIT Press, Cambridge, Massachusetts

Young, H.P. (1993), “An evolutionary model of bargaining”, *Journal of economic theory*, 59, 145-68

Summary

Evolutionary game theory provides the tools to analyze which strategies, or patterns of behaviour, emerge over time through a process of adaptation. Social norms can be defined as patterns of behaviour with certain characteristics. Evolutionary game theory thus provides one perspective on how social norms are formed and maintained. Prisoner's dilemma games can be used to study the conditions under which cooperative norms emerge. Bargaining games can be used to address the formation of fairness norms. However, being more congenial to analyzing norms that somehow focus on material payoffs, it is not a given that evolutionary game theory can adequately address norms focusing on rights or virtues.

ISSN 0804-3639
ISBN 82-8062-044-3