What can facilitate cooperation: Fairness, ineaulity aversion, punishment, norms or trust?

Urs Steiner Brandt

November 2008

Editor: Eva Roth

Department of Environmental and Business Economics
IME WORKING PAPER 80/08

ISSN 1399-3224

Urs Steiner Brandt
Department of Environmental and Business Economics
University of Southern Denmark
Niels Bohrs Vej 9-10
DK-6700 Esbjerg
Tel.: +45 6550 4184
Fax: +45 6550 1091
E-mail: usb@sam.sdu.dk

Abstract

Almost all economic and public choice models assume that *all* people are *exclusively* pursuing their own material self-interests and do not care about "social" goals *per se*. Several (laboratory) experiments address the question of the general validity of this assumption. A consistent conclusion emerges that a significant number of people deviate from the assumption of selfish rational behaviour; this conclusion is robust with respect to the design of the experiments. Therefore, public choice comes with a price: the conclusions are based on the stylized stereotype of economic man, an assumption that is not fully satisfied. The purpose of this paper is to show how to incorporate other-regarding preferences into an otherwise traditional utility approach without losing predicting power or compromising the rationality assumption. On the contrary, since other-regarding preferences are based on observed behaviour, the predicting power increases; this is demonstrated at the end of this paper, where it is shown how other-regarding preferences can explain the existence and persistence of a welfare state and why people might act sustainably.

Acknowledgement
This paper was presented in Munich at the CESifo 2008 confernce and at the Adam Smith seminar at the University of Hamburg. Thanks to all the participants for their comments.

1. Introduction

At the Public Choice Society's 2007 World Meeting in Amsterdam, several keynote speakers praised the introduction of economic man into the realm of political man. An overwhelming number of experiments, however, challenge the general validity of this assumption. A very consistent conclusion emerges that a significant number of people deviate from the assumption of selfish rational behaviour, a conclusion that seems robust with respect to the design of the experiments.[1] Therefore, the advancement of public choice comes with a price: the conclusions are based on a stylized stereotype assumption of economic man, an assumption that has not been fully satisfied.

The main purpose of this paper is to show how to incorporate other-regarding preferences into an otherwise traditional utility approach without losing predicting power or compromising the rationality assumption. This is done by incorporating the results from experiments on simple, mostly static games into information about other-regarding preferences. Obviously, human behaviour is also determined by dynamic considerations, like reputation effects and reciprocal behaviour. Such effects are not easily handled in a static setting. However, the essence of other-regarding behaviour, the focus of this paper, is very likely best extracted in simple settings.

Experiments on games where strictly economic man would never cooperate show a systematic deviation from what non-cooperative game theory suggests, but only for a subset of individuals. Such behaviour is mostly centred on fairness (typically in its relativistic version, i.e., I will be nice to people who are nice to me and unkind to people who are unkind toward me (Rabin, 1993)) and aversion to inequality (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). The experiments show that such behaviour under certain conditions implies that players will cooperate in situations where incentives are against cooperation.

1 Almost all economic models assume that *all* people are *exclusively* pursuing their own material self-interests and do not care about "social" goals, per se. As Fehr and Schmidt (1999) state, this may be true for some people, but it is certainly not true for everybody.

Such behaviour must be incorporated into more realistic modelling of individual preferences, which is what the current paper tries to accomplish.

Another branch of experiments analyses the effects of adding the possibility of punishment. When a second round where punishment is possible is added to a one-shot prisoner's dilemma game, cooperation also is sustained. When the punishment is also costly for the punisher, however, punishment will only be executed when the punisher receives utility from punishing, again suggesting that individual preferences need to deviate from the assumption of "homo economicus". Experiments verify that many people are willing to use costly punishment against defectors (Fehr and Schmidt, 1999; Abbink, Bolton, Sadrieh and Tang, 2001).

Individual behaviour is shaped not only by individualistic preferences, but might also be affected by interpersonal norms. A social norm can be defined as a behavioural regulator based on socially-shared beliefs of how to behave and how and when to punish deviators with informal social sanctions (Fehr and Gächter, 2000). The emergence of such norms can be analyzed by evolutionary game theory, wherein the different types of preferences compete against each other.[2]

2 Here the assumption is that a social norm emerges when a sufficient fraction of the total population is of the same type. To quote Fehr and Fischbacher (2004, page 1): "Cooperation in human societies is mainly based on social norms, including in modern societies, where a considerable amount of cooperation is due to the legal enforcement of rules. Legal enforcement mechanisms cannot function unless they are based on a broad consensus about the normative legitimacy of the rules – in other words, unless the rules are backed by social norms. Moreover, the very existence of legal enforcement institutions is itself a product of prior norms about what constitutes appropriate behaviour. Thus, it is necessary to explain social norms to explain human cooperation." According to Kolstad (2003), in an evolutionary game setting, persons meet each others and play against each other (e.g., a prisoner's dilemma); the players with the highest payoff have a higher reproduction rate. In this setting, and given a population consisting of both cooperators (people that always cooperate) and rational egoists (homo economicus), when different persons are matched randomly, the cooperators will be outcompeted by rational egoists.

The two main classes of games most relevant for the analysis in this paper are voluntary contribution mechanism games, used to determine other-regarding preferences, and proposer-responder games, used to determine the effects of punishment and reward opportunities. These two types of games will be thoroughly defined and discussed in Section 2. Results from a third class of games, evolutionary games, will also be used in this paper. Such games (at least the ones used in this paper) can be characterized as ‘non-strategic’ dynamic games, in which each player’s strategy space is collapsed into only one strategy. This type of game fits into the general framework of this paper; we do not look at strategic dynamic effects. Since evolutionary game theory excludes strategic behaviour and the possibility that each player is equal to a particular strategy, this branch of game theory can and has been used to analyse ‘developments in large groups’, and, in particular, the evolution of social norms. We show below that in a prisoner’s dilemma game, cooperators can only survive if they can identify the type of the other players, then choose who to play against them (so-called assortative matching); if it is not totally possible to identify types, then there must be an effective punishment available. Such a strategy is only possible if the cooperators are also wiling to punish.[3] Together, this creates social norms for cooperation.

All in all, this enrichment of types of preferences gives a better description of reality and therefore, if used properly, also increases the predictive power of the analysis.

The paper is organized as follows. The next section discusses models of individual decision-making and summarizes the main results of relevant experiments. Section 3 provides an overview of other-regarding preferences. Section 4 introduces different types of other-regarding behaviour, while Section 5 discusses the importance of punishment and why costly punishment might be optimal. Section 6 introduces the theory of evolution of social norms. Finally,

---

3 Since punishment is costly, Trivers (1971) proposes that the emergence of feelings like guilt and shame provides an evolutionary advantage, which could offer another explanation for the emergence of “conditional cooperators”.

some applications of the models of other-regarding preferences are made in Section 7. Section 8 concludes the paper.

2. Models of individual decision-making

Both qualitative and quantitative deviations from the assumption that individuals only care about own payoffs have been examined previously in several papers, using different types of underlying games that the individuals participate in. In this section, we briefly describe two main classes of such games in addition to summarizing the main findings.

2.1. Experimental game theory

The two main classes of games most relevant for the analysis in this paper are:

- voluntary contribution mechanism (VCM) games to determine other-regarding preferences, and
- proposer-responder (PR) games to determine the effects of punishment and reward opportunities

Figure 1. A stylized presentation of one round of a voluntary contribution mechanism game

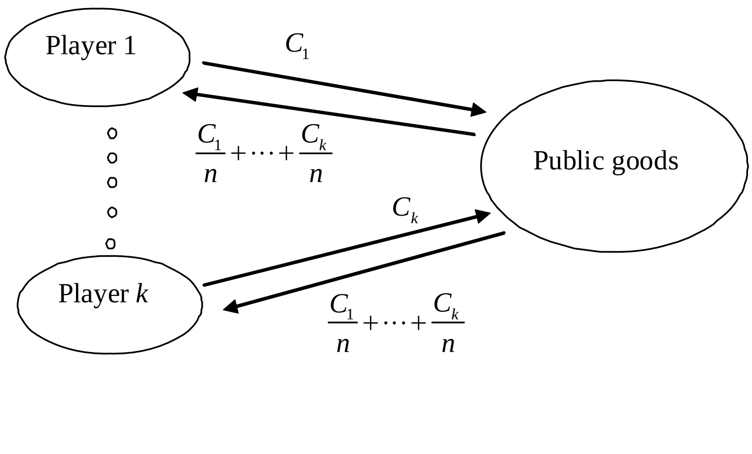
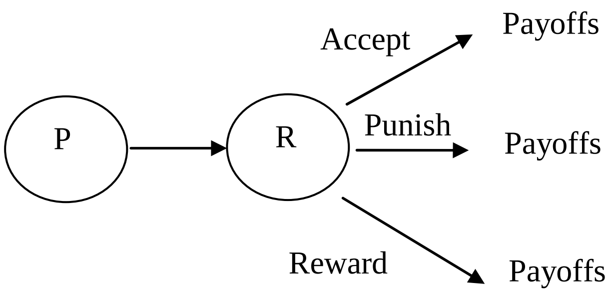


Figure 1 presents a stylized version of one round of a voluntary contribution mechanism game in which k is the number of players, C_i is the individual contribution, and $\frac{C_1}{n}+\cdots+\frac{C_k}{n}-C_i$ is the total net benefits to player i from participating in the game. Finally, to create an interesting condition to study, we set conditions such that $n>1$ and $\sum\frac{C_i}{n}>C_i$. Condition $n>1$ says that the marginal return per capita (MRCP) is smaller than 1. Together, the two conditions indicate that contribution is not individual-rational, but collective-rational. If individuals are rational and care only about their own payoff, no player should contribute anything to this game; free riding is the only behaviour supported by a Nash equilibrium. This type of game is particularly suitable for deriving information about other-regarding preferences. If the game is repeated, learning and reputation effects can also be revealed.

Figure 2 presents a stylized version of one round of a proposer-responder game. Dictator games, ultimatum games, and trust games all belong to this class of games. Such games address concepts of retaliation, positive and negative reciprocity, punishment and rewards and have a sequential feature, although the dictator game is a degenerate version in which the responder has no strategies available. In these types of games, the proposer typically proposes how to split the shares of a pie. This choice is then observed by the responder, who can either reject the choice, leaving both with nothing (ultimatum game) or use the possibility to either punish (often also costly for the responder) or reward the proposer (trust games).

Figure 2. A stylized presentation of one round of a proposer-responder game



In all of these games, a unique subgame perfect equilibrium implies that the proposer keeps everything to himself (or offers only the smallest possible sum to the responder), while the responder accepts this and neither punishes nor rewards the proposer.

2.2. Some results

Mounting evidence from experiments shows that individual behaviour deviates significantly from purely selfish behaviour. A vast number of papers address the question of adequate characterization of individual preferences. This section provides an overview of these results and draws some main lessons relevant for our analysis.

The general results can be summeried into four groups:

- Contribution in VCM-games
- Contribution and acceptance behaviour in PR-games
- The effect of punishment, and
- General results

Contribution in VCM games

A repeated finding is that about half the players contribute in the first round of a VCM game, but if the game is repeated, contribution levels tend to decay, although they remain well above zero; only about 70 percent of subjects contribute nothing in the announced last round of a finitely repeated sequence (Ostrom, 2000; Camerer and Fehr, 2004; and references therein).[4]

Contribution and acceptance behaviour in PR-games

A repeated finding for utimatum games is that proposer offers below 30 percent of the total amount are rejected with a high probability (Fehr and Gächter, 2000 and references herein). The difference between proposal levels in dictator games and proposal levels in ultimatum games is best demonstrated in Table 1, which is built from the reports of Bolton and Ockenfels (2000), and is rather representative of typically reported results.

Table 1. Results from dictator and ultimatum games

	Percent Offered (pie size = $10)						
	0	1	2	3	4	5	6
Dictator game	21	17	13	29	0	21	0
Ultimatum game	0	0	4	4	17	71	4

Source: Bolton and Ockenfels (2000).

Both games indicate that most people offer a non-zero fraction to the other player; only the size of this fraction differs significantly between the two games.

4 The reason for this decay has been heavily debated, but can be attributed either to learning the rules of the game or to the presence of reciprocal behaviour, such that observed non-contribution is heavily punished in later rounds. It has been shown that players will spend personal resources to punish those who make below-average contributions to a collective benefit, including the last period of a finitely repeated game. Regarding the learning effect, Houser and Kurzban (2002) show in a 10-round VCM that all of the decay in the contribution over time can be attributed to confusion, a result also reported in Andreoni (1995).

The effect of punishment
The results of Table 1 also indicate the effect of being able to punish: offers above 30 percent appear only 21 percent of the time in dictator games, but appear 92 percent of the time in ultimatum games. However, even in dictator games, only 21 percent of participants leave nothing to the other player. Similar results are reported by Camerer and Fehr (2004) and Fehr and Fischbacher (2004). Overall, the results indicate that the option of punishment explains much, but not all, of players' concern for the other players.

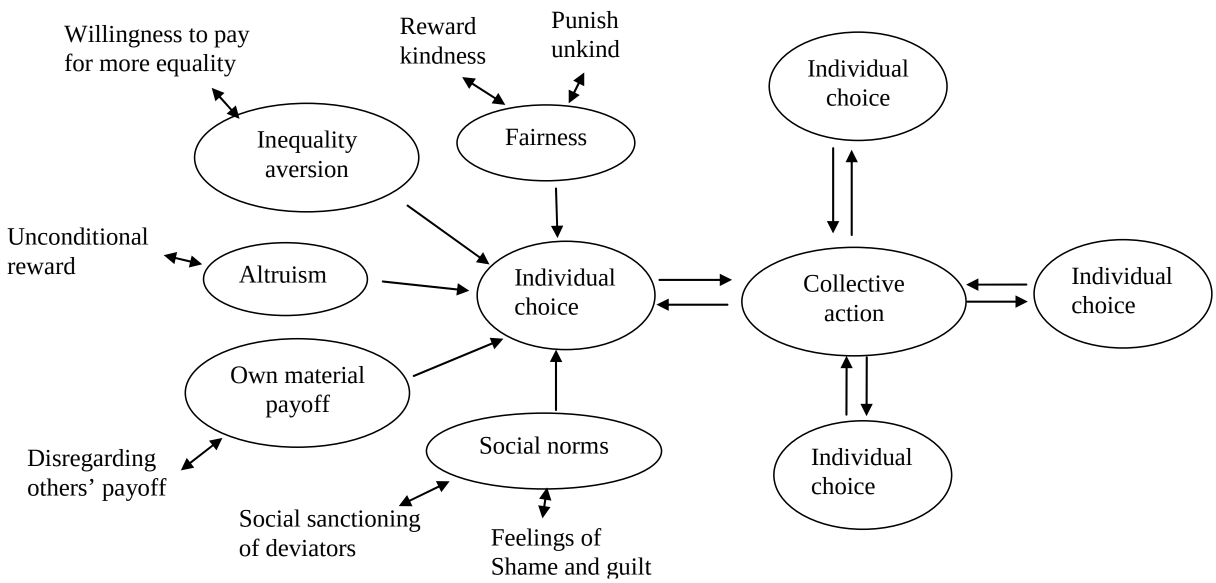
General results
Those who believe others will cooperate in social dilemmas are more likely to cooperate themselves, and face-to-face communication in a public good game produces substantial increases in cooperation. Finally, the rate of contribution to a public good is affected by various contextual factors, including the framing of the situation and the rules used for assigning participants, increasing competition among them, allowing communication, authorizing sanctioning mechanisms, or allocating benefits (Ostrom, 2000).

3. An overview of other-regarding behaviour

The overall result is a systematic deviation from what non-cooperative game theory suggests, but only for a subset of individuals. Such deviation can be grouped into individualistic or group-related behaviour. Individualistic behaviour can be divided into reciprocal or conditional behaviour, altruistic or unconditional behaviour, or behaviour based on the ability to punish. The former can be further divided into behaviour based on either consequences or intentions. Finally, group-related behaviour deals with the emergence and maintenance of social norms, e.g., the question of informal sanctions of norm violators.

Figure 3 illustrates the complexity of individual choice. As described above, many different elements potentially influence our choices. Examples are willingness to sacrifice payoffs to reward good behaviour and punish bad behaviour (individualistic intention-based behaviour) and the temptation to free ride (individualistic self-regarding-based behaviour). These motives are individualistic, but there is also a social (group) component: the presence of social norms and their internalization (which implies feelings of shame and guilt when breaking the norm) and the social sanctioning of free riders.

Figure 3. The main determinants of individual choice



The aim of the next section is to discus the different behavioural pattern in one common model.

4. Modelling other-regarding preferences

By its very definition, cooperation can never be a solution to a prisoner's dilemma.[5] This is because a game consists of both a payoff matrix and a game form defining the order of play, and because of the behaviour of the players, assuming for a prisoner's dilemma game that the players are fully rational and

5 Obviously, if, for example, all the players preferred cooperation, an outcome of mutual cooperation would hardly be a dilemma.

equipped with fully self-regarding preferences. We maintain the assumption that players are fully rational utility-maximizing players, but vary how players evaluate the payoffs and outcome of play.[6]

In this section, we equip the players with a utility function that can incorporate other-regarding behaviours such as fairness and inequality aversion and, by use of simple one-shot prisoner's dilemma payoffs, analyze under what conditions cooperation is possible as an equilibrium strategy. In order to do so, let the players be equipped with the following (linear and additive) utility functions with all the necessary properties of continuity and differentiability.

$$U_i(s) = \pi_i(s) + \sum_{k \neq i} \alpha_{ik}(\pi(s), I_{ik}^e) \cdot \pi_k(s)\,. \tag{1}$$

In this function, *s* represents a strategy profile and $\pi_i(s)$ the material payoff (outcome) of the game for player *i*. Utility to *i* is a function of the material payoff he/she receives; however, the player might also derive utility from the payoff that other players receive in the game The function $\alpha_{ik} = \alpha_{ik}(\pi(s), I_{ik}^e)$ measures the weight player *i* assigns to the payoff of player *k*. The weight in this general setting can be a function of the vector of resulting *outcomes*, $\pi(s)$, or perceived *intentions*, I_{ik}^e, which is the expectation player *i* holds about how player *k* will treat player *i*. The *homo economicus* hypothesis implies that $\alpha_{ik} = 0$ for all *i* and *k*, in which case we would say that the player has *"self-regarding" preferences*. On the other hand, $\alpha_{ik} \neq 0$ implies *"other-regarding" preferences*, such that $\alpha_{ik} > 0$ will be sometimes be labelled as altruistic behaviour and $\alpha_{ik} < 0$ as spiteful behaviour.[7]

---

6 That is, we disregard limited rationality, such as assuming that players are bounded rational or based on procedural rationality.

7 How exactly to interpret these weight functions is not clarified in the literature. Is there (or should there be) a difference between an individual having preferences where $\exists k \in I$ and $\alpha_{ik} \neq 0$ and an individual having preferences where $\forall k \in I$ and $\alpha_{ik} \neq 0$? In our analysis, however, this does not matter, since we focus on a two-player game.

The one-shot game in strategic form presented in Table 2 serves as an illustration of the effects of the different concepts. We denote the payoff structure in Table 2 as prisoner's dilemma-like payoffs if $b > a > d > c$ and $\gamma > \alpha > \delta > \beta$.

Table 2. The generic prisoner's dilemma game[8]

Player 1\Player 2	Cooperate (C)	Defect (D)
Cooperate (C)	(a,α)	(c,γ)
Defect (D)	(b,β)	(d,δ)

4.1. The implication of concern for other players' payoffs

Several papers state that individuals demonstrate varying degrees of concern for how well other players perform in a game. To analyze this case in a two-player setting, focusing on Player 1, rewrite (1): $U_1(s) = \pi_1(s) + \alpha_{12} \cdot \pi_2(s)$. Now apply this to the game in Table 2. Look at the strategy pair (C,C).[9] In order for Player 1 not to defect, the following must be true:

$$U_1(D,C) - U_1(C,C) \le \alpha_{12} \cdot (U_2(C,C) - U_2(D,C)) \Rightarrow \alpha_{12} \ge \frac{U_1(D,C) - U_1(C,C)}{U_2(C,C) - U_2(D,C)}.$$

8 We sometimes refer to the PD with real numbers (referred to as Table 3):

Player 1\Player 2	Cooperate (C)	Defect (D)
Cooperate (C)	(4,4)	(0,6)
Defect (D)	(6,0)	(1,1)

9 Note that α_{12} and α_{21} must be determined simultaneously because when, for example, Player 1 considers playing C, this player cannot determine the value of the strategy profiles without knowing α_{21}. If focus is only on one player, this player will cooperate for certain, given $\min_{s_2 \in S_2} U_1(C, s_2) \ge \max_{s_2 \in S_2} U_1(D, s_s)$, where S_2 is the set of strategies for Player 2. If this condition is met, then C is a dominant strategy for Player 1.

Taking the numbers from the table, for (C,C) to be an equilibrium, $\alpha_{12} \geq \frac{b-a}{\alpha-\beta}$ and $\alpha_{21} \geq \frac{\gamma-\alpha}{a-c}$.[10]

Since $b-a$ measures the gain from defecting, while $\alpha_{12} \cdot (\alpha-\beta)$ measures the loss from doing so, the implication is that when the gain from defecting is large enough, the temptation to deviate is too strong. This essentially highlights the countervailing incentives faced by individuals equipped with the preferences described by Equation 2.

So far, we have not explained why players should have an $\alpha_{ij} \neq 0$. Two different explanations for this are presented below. Players might (to a varying degree) have a sense of fairness, or they might have an aversion towards inequality. These two concepts constitute two different ways of caring about the other players; the first (fairness) is a reaction to how the other players intend to play (intentional farness), while in the other (inequality aversion), players react to the consequences of the play.

4.2. The implication of intentional fairness

For the fairness concept presented here, fairness depends on expectations of in-tentions, with the general notion that a player will be kind to players that are expected to be kind to that player and unkind to a player that is expected to be unkind.[11]

Given Equation (1), insert $\alpha_{ik} = \alpha_{ik}(I_{ik}^{e})$, such that if *k* has good intentions against *i*, player *i* derives utility from an increase in payoff to player *k* such that:

---

10 Given the payoffs in Table 3, cooperation is sustained given that $\alpha_{12} \geq \frac{1}{2}$ and $\alpha_{21} \geq \frac{1}{2}$.

11 Rabin (1993) suggests that the sign of our sentiments is conditional: "[...] the same people who are altruistic to other altruistic people are also motivated to hurt those who hurt them." Moreover, Rabin claims that the direction of our sentiments depends on our beliefs about the others' intentions.

$$\frac{\partial U_i}{\partial \pi_k} > 0 \text{ if } I_{ik}^e = kind \text{ and}$$
$$\frac{\partial U_i}{\partial \pi_k} < 0 \text{ if } I_{ik}^e = unkind \text{ .}$$

These expressions state that if player *i* believes that player *k* is kind (unkind) toward player *i*, then player *i* receives positive (negative) utility from an increase in the payoff to player *k*.

Obviously, it is necessary to more precisely define what is meant by "kind" in order to make the idea of fairness operational. To illustrate the consequence of including fairness, we develop a simplified version of the fairness concept modelled by Rabin (1993): Player *i* will be kind toward player *k*, if player *i*, by playing s_i, expects to receive more than the average payoff of playing s_i:

$$I_{ik}^e(s_i, s_{ik}^e) = \text{kind if } I_{ik}^e(s_i, s_k^e) = \frac{2\pi_i(s_i, s_{ik}^e)}{\pi_i^{\max}(s_i) - \pi_i^{\min}(s_i)} > 1$$
$$I_{ik}^e(s_i, s_k^e) = \text{unkind if } I_{ik}^e(s_i, s_k^e) = \frac{2\pi_i(s_i, s_{ik}^e)}{\pi_i^{\max}(s_i) - \pi_i^{\min}(s_i)} \leq 1$$

where $\pi_i^{\max}(s_i)$ and $\pi_i^{\min}(s_i)$ are the highest and the lowest possible payoffs that player *i* can receive, given that he plays s_i, and s_{ik}^e is player *i*'s belief of what player *k* will play.

That is, if player *i*, by playing s_i, expects to get more than the average of the best and worst payoff given he plays s_i, then he considers player *k* to be kind to him; the reverse is true if he expects to receive less than the average. From this, it follows that if $I_{ik}^e > 1$, then α_{ik} should be positive, and if $I_{ik}^e \leq 1$, then α_{ik} should be non-positive. Therefore, we can include the fairness component as follows: $U_i(s) = \pi_i(s) + (I_{ik}^e - 1) \cdot \pi_k(s)$. In this interpretation of fairness, it follows that the nicer player *i* expects that player *k* is against him as player *i* receives more utility from an increase in the payoff of player *k*, which seems plausible.

To present an example of how to apply this concept, assume a version of the game from Table 1, (presented in Table 4) and look at the following question. Given these payoffs, for which beliefs held by a player (Player 1 in this example) about what Player 2 intends to play will Player 1 find it optimal to cooperate?

Table 4.

Player 1\Player 2	Cooperate (C)	Defect (D)
Cooperate (C)	(a,4)	(0,6)
Defect (D)	(b,0)	(1,1)

Assume that $\rho(s_{12}^{e}=C)=\frac{2}{3}$, which is to say that Player 1 expects that Player 2 will play C with a probability of 2/3. It follows that

$$I_{12}^{e}(C,\tfrac{2}{3})= \frac{2\cdot[\frac{2}{3}\cdot\pi_1(C,C)+\frac{1}{3}\cdot\pi_1(C,D)]}{\pi_1(C,C)+\pi_1(C,D)}=\frac{\frac{4}{3}a+\frac{1}{3}c}{a+c}=\frac{4}{3} \text{ and}$$
$$I_{12}^{e}(D,\tfrac{2}{3})= \frac{2\cdot[\frac{2}{3}\cdot\pi_1(D,C)+\frac{1}{3}\cdot\pi_1(D,D)]}{\pi_1(D,C)+\pi_1(D,D)}=\frac{\frac{4}{3}b+\frac{2}{3}d}{b+d}=\frac{4b+2}{3b+3}.$$

Inserting this into the utility function yields

$$U_1(C,\tfrac{2}{3})=\tfrac{2}{3}\pi_1(C,C)+\tfrac{1}{3}\pi_1(C,D)+(I-1)\cdot[\tfrac{2}{3}\pi_2(C,C)+\tfrac{1}{3}\pi_2(C,D)]=\tfrac{2}{3}a+\tfrac{14}{9}$$
$$U_1(D,\tfrac{2}{3})=\tfrac{2}{3}\pi_1(D,C)+\tfrac{1}{3}\pi_1(D,D)+(I-1)\cdot[\tfrac{2}{3}\pi_2(D,C)+\tfrac{1}{3}\pi_2(D,D)]=\tfrac{2}{3}b+\tfrac{1}{3}+\tfrac{1}{9}\cdot(\frac{b-1}{b+1}).$$

It follows that

$$U_1(C,\tfrac{2}{3})>U_1(D,\tfrac{2}{3})\Rightarrow (b-a)<\tfrac{11}{6}-\tfrac{1}{6}(\tfrac{b-1}{b+1}).$$

since $-\frac{1}{6}(\frac{b-1}{b+1})<0$, $(b-a)$ is smaller than $\frac{11}{6}$, such that cooperation will be optimal if the gain from unilateral deviation is sufficeintly small (smaller than $\frac{11}{6}-\frac{1}{6}(\frac{b-1}{b+1})$).

Table 5. The range of (b-a) for which Player 1 will cooperate is shown for varying beliefs

$\rho(s_{12}^e = C)$	$I_{12}^e(C,\rho)$	$I_{12}^e(D,\rho)$	$U_1(C,\rho)$	$U_1(D,\rho)$	$U_1(C,\rho) > U_1(D,\rho)$
1	2	$\frac{2b}{b+1}$	$a+4$	b	$b-a<4$
2/3	$\frac{4}{3}$	$\frac{4b+2}{3b+3}$	$\frac{2}{3}a+\frac{14}{9}$	$\frac{2}{3}b+\frac{1}{3}+\frac{1}{9}\cdot(\frac{b-1}{b+1})$	$(b-a)<\frac{11}{6}-\frac{1}{6}(\frac{b-1}{b+1})$
½	1	1	$\frac{1}{2}a$	$\frac{1}{2}b+\frac{1}{2}$	$(b-a)<-1$

In Table 5, the range of (*b*-*a*) for which Player 1 will cooperate is shown for varying $\rho(s_{12}^e = C)$. The results are very straightforward. The more Player 1 believes that Player 2 is kind toward him, the more it takes in terms of additional benefits of deviation from cooperation to make the temptation of defection large enough.

This highlights the point made by Rabin (1993) that fairness is relativistic and that there is a trade-off between fairness concerns and material payoffs.[12] It follows that cooperation can be sustained *only* if the gain from deviating ($b-a$) is not too large. Finally, it is important to note that cooperation here is motivated by beliefs. The belief that others might cooperate can form through revision of beliefs after several encounters, if a sufficient fraction of those encounters have been cooperative.

4.3. Inequality aversion

Many individuals perceive egalitarian solutions as focal allocation points such that any deviation from equal division gives some disutility. According to Fehr and Schmidt (1999) and Bolton and Ockenfels (2000), strong evidence exists that relative material payoffs affect people's well being and behaviour. Fehr and Schmidt (1999) also argue that players are heterogeneous with regard to the

12 Rabin's fairness concepts builds on the idea of ”being kind to someone that has the intention to be kind to you and unkind to someone that has the intention to be unkind to you”. Therefore, it is in opposition to other fairness considerations like Kant's categorical imperative.

inequity aversion parameters, and for any individual, the envy parameter is larger than the parameter measuring advantageous inequality aversion.[13]

To analyse the effect of inequality aversion in the set-up of Equation (1), let

$$\alpha_{ik} = -\varepsilon_{ik} \cdot \left| \frac{\pi_i(s) - \pi_k(s)}{\pi_k(s)} \right|$$

where $\varepsilon_{ik} = \varepsilon_{ik}^{+}$ *if* $d_{ik} > 0$, $\varepsilon_{ik} = \varepsilon_{ik}^{-}$ *if* $d_{ik} < 0$ and $\varepsilon_{ik}^{-} > \varepsilon_{ik}^{+}$. $d_{ik} = \pi_i(s) - \pi_k(s)$ is the distance in payoffs between players *i* and *k* as a consequence of the observed payoff vector π. The minus sign in front of ε_{ik} shows the negative impact of payoff inequality on individual utility. Finally, $\varepsilon_{ik}^{-} > \varepsilon_{ik}^{+}$ indicates the above-mentioned heterogeneity such that player *i* experiences a smaller utility loss from payoff inequality if he/she gets more then the other player as compared to the situation where the other player gets more.

All in all, the expression measures the disutility from inequality aversion, and, in accordance with Fehr and Schmidt (1999), the aversion is asymmetric. Inserting the expression for α_{ik} into (1) and reducing the number of players to 2, the utility for Player 1 is given by:

$$U_1(s) = \pi_1(s) - \varepsilon_{12} \cdot \left| \pi_1(s) - \pi_2(s) \right|.$$

Given this, when is (C,C) a unique Nash equilibrium? Take the payoffs from Table 1 and derive for when cooperation is optimal for Player 1:

---

13 Fehr and Schmidt (1999) argue that players are heterogeneous regarding the inequity aversion parameters, and for any individual, the envy parameter is larger than the parameter measuring advantageous inequity aversion. Hill and Neilson (2007) find that inequality aversion exhibits diminishing sensitivity, as the effect of inequality aversion has less and less additional impact the farther the payoffs are moved away from an egalitarian situation. They also identify a so-called Robin Hood effect, i.e., the tendency to take from the high payoff opponent to give to a low-income opponent.

$U_1(C,C) = a$, (since here $d_{12} = 0$) and $U_1(D,C) = b - \varepsilon_{12}^{+} \cdot (b - \beta)$. $U_1(C,C) > U_1(D,C) \Rightarrow \varepsilon_{12}^{+} > \frac{b-a}{b-\beta}$.[14] Given the payoff, $U_2(C,C) > U_2(D,C) \Rightarrow \varepsilon_{21}^{+} > \frac{\gamma-\alpha}{\gamma-c}$.

The obvious conclusion here is that when concern for inequality is sufficiently high (for both players, which due to symmetry yields the same result), then cooperation can be sustained.

We have now used the prisoner's dilemma payoff structure to show that several ways of cooperating behaviour can be explained. This has been done by enriching the players' preferences to include various ways of being concerned about the other players' payoffs and motives by presenting different versions of the benchmark utility function. The specific feature of these functions is that the level of concern about fairness and inequality aversion depends on how each individual perceives such concepts, and is therefore in accordance with experiments showing that most likely, only a fraction of players act non-selfish.

5. The ability to punish defectors

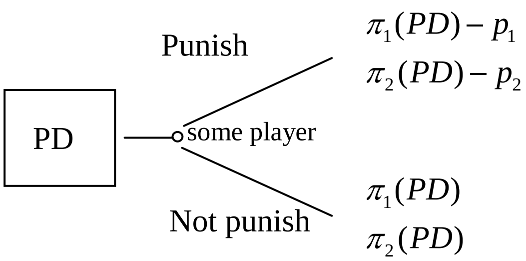
Useful discussions about the importance of punishment are found in Camerer and Fehr (2006), Fehr and Fischbacher (2004) and Fehr and Gächter (2000). The overall impression from these papers is that the possibility of punishing "non-acceptable" behaviour has a strong positive effect on players' willingness to cooperate. Moreover, players are often willing to punish more than seems rational. In this section, we will discuss how punishment works and why players choose to punish, though it does not from the outset seem optimal to do so. In particular, Fehr and Schmidt (1999) show that when players are given the opportunity to punish free riders, stable cooperation is maintained, although punishment is costly for those who punish.

14 Given the numbers in Table 3: $\varepsilon_1 > \frac{6-4}{6-0} = \frac{1}{3}$. Note that given the simple structure of the problem, we need not derive ε_{12}^{-} and ε_{21}^{-}.

5.1. The implication of adding punishment

To analyze the role of punishment, take the following variation (or rather extension) of the prisoner's dilemma game as shown in Figure 4.

Figure 4. A prisoner's dilemma with an additional punishment phase



After playing the ordinary prisoner's dilemma, wherein the payoff to the players is $\pi_i(PD),\ i = 1,2$, another phase is added in which some player(s) can use a punishment strategy. The punishment often takes the form of an option to reduce the payoffs received in the PD game by choosing to either invoke the punishment possibility (if the punishment levels are fixed) or the level of the punishment $p_i \le 0,\ i = 1,2$, such that the payoffs are $\pi_i(PD) - p_i$.

Table 6. Punishment of deviation from cooperation

Player 1\Player 2	Cooperate (C)	Defect (D)
Cooperate (C)	(a,α)	(c,γ-p_2)
Defect (D)	(b-p_1,β)	(d-p_1,δ-p_2)

The normal form in a two-player version of this game (and the one used in the next section) could be as seen in Table 6, where only deviation from cooperation is punished.

It is possible to distinguish between different punishment scenarios. In a situation referred to as self-punishment, one player can, after observing the play, invoke a punishment mechanism that is costly to both ($p_1 < 0$ and $p_2 < 0$). No rational player with only self-regarding preferences would choose this punishment, since it is not subgame perfect.

Another situation is symmetric punishment. One (or both) player(s) can, after observing the play, determine a punishment that is costly only to the other player. If the punishment is credible (and large enough), then this will engender cooperation. This can also be done by letting an exogenous party make the punishment: Both players agree that a third party can punish deviation from cooperation. Both players will cooperate if $p_1 > (b - a)$, $p_2 > (\alpha - \gamma)$ and the arrangement is perceived as credible by both players. In the case of credibility, each player is willing to pay for this service, Player 1 up to $(a - d)$ and Player 2 up to $(\alpha - \delta)$.

Aside from the credibility issue, another problem with the two last situations is that when there are more players involved, punishment itself becomes a public good, such that the issue of free riding must be overcome. However, in both these situations, rational selfish players can achieve cooperation, so we focus in this paper on the first situation and, in particular, situations where self-punishment is needed in order to punish the unwanted behaviour of other players. The next sections present two approaches that go more into depth on the issue of punishment.

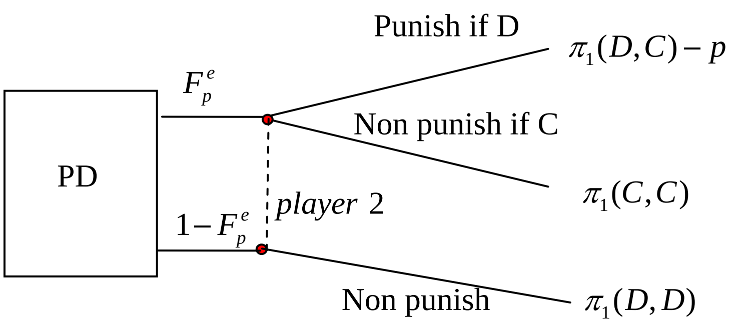
5.2. The presence of reciprocal punishers

The first approach to describing punishment behaviour is a situation where players are modelled as strategies (in the sense that they do not act strategically in the play, but simply follow a pre-determined course of action). The explanation of the appearance of such a strategy is beyond the scope of this paper, but could be explained through either genetic predisposition (see Section 6) or through the updating of beliefs through a history of encounters, such as in Section 4 with regard to fairness concepts, where players, after evaluating the play, might revise their beliefs about their opponents.

To see how the possibilty of encountering a "non-strategic" player might change the behaviour of a sefish rational player, assume that a percentage of the population is what can be called a "reciprocal punisher".[15] If they observe defection by other players, they play punish, no matter the personal costs of doing so. Moreover, they are fully reciprocal in nature, such that if they observe cooperation, they also cooperate.

What does this imply for the possibility for cooperation? Denote by F_p^e the expected likelihood of a rational selfish player encountering a reciprocal punisher. The game is played as shown in Figure 5. First, the players engage in an ordinary one-shot prisoner's dilemma, and afterwards the players can play a punishment strategy. Let Player 1 be the rational selfish type that never punishes, while Player 1 expects that Player 2 will only punish upon defection with probabtilty F_p^e. Assume also that encountering such a player is totally random, and it is not possible to infer the type of the other player in advance.

Figure 5. Introducing the possibility of a reciprocal punisher



Define by $EU_1(s_1,\cdot)$ the expected payoff for Player 1 from playing the strategy s_1, with a probability of F_p^e of meeting a reciprocal punisher and a probability $(1-F_p^e)$ of meeting a rational player (i.e., a player that always defects in a one-shot prisoners dilemma). If we use the payoffs from Table 2 and focus on Player 1, we have:

15 This term is found in Ostrom (2000); other authers call this type of player an altrusic punisher.

$EU_1(D,\cdot) = (1 - F_p^e) \cdot \pi_1(D,D) + F_p^e \cdot [\pi_1(D,C) - p] = F_p^e(b - d - p) + d$ and
$EU_1(C,\cdot) = (1 - F_p^e) \cdot \pi_1(C,D) + F_p^e \cdot \pi_1(C,C) = F_p^e \cdot (a - c) + c$.

It follows that:

$$EU_1(C,\cdot) > EU_1(D,\cdot) \Rightarrow F_P > \frac{d - c}{a - c - b + d + p}.$$

The conclusion is straightforward, since the higher the punishment, the more likely a rational selfish player is to cooperate given the numbers in Table 3, $F_p^e > \frac{1}{-1+p}$. If $p = 0$, no cooperation is possible, no matter the size of F_p^e. For example, if $p = 5$, then Player 1 will cooperate if $F_p^e > \frac{1}{4}$.

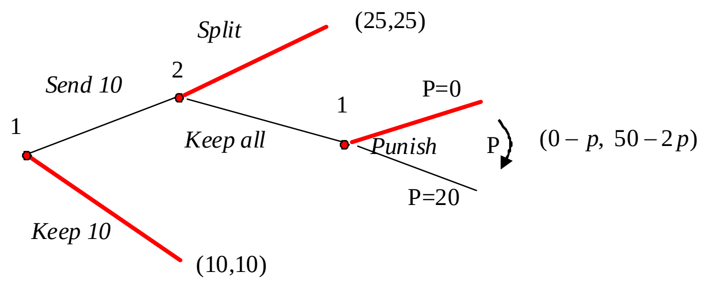
The implication here is that when the possibility of encountering a reciprocal punisher exists, then otherwise selfish players will (under certain conditions) find it optimal to cooperate. It must be noted that the behaviour of the reciprocal punisher is unexplained. In particular, why costly punishment is used has been a puzzle. Moreover, the presence of a “reciprocal punisher” is not explained, but simply assumed. The next section seeks to cast more light on this question of the willingness to punish.

5.3. The neural basis of punishment

Experiments by Quervain et al. (2004) on a trust game show that individuals derive “utility” from punishing norm violators.

A version of the trust game goes like this. For simplicity, assume a two-player game. Both players receive an initial endowment of ten monetary units (MU). Player 1 can either keep his endowment (in which case the game ends) or send it to Player 2, in which case the amount is quadrupled such that Player 2 now has 50 MU. Player 2 can now either keep the 50 MU or send half back. If he/she does not send 25 MU back, Player 1 can use a punishment strategy, choosing from 0 to 20 punishment point, each of which costs the players 1 MU and 2 MU for Player 1 and 2, respectively. In Figure 6, the extended form of this game is depicted. A move for a player (except in the punishment phase) is how much money he/she sends back to the other player. Note that this game is essentially identical to the game shown in Figure 5, but with a particular punishment structure.[16]

Figure 6. The extensive form of the two-player trust game

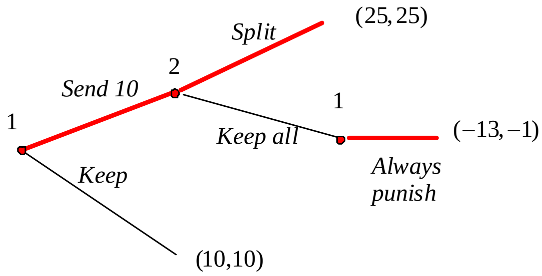


There exist many Nash equilibria, such as if Player 1 chooses more that 12.5 punishment points and Player 2 sends 25 MU back to Player 1, but there is only one subgame perfect equilibrium. The red (bold) moves show the unique subgame perfect Nash equilibrium, where Player 1 sends no money to Player 2, because Player 1's threat of punishment is not credible.

If Player 1 were commited to punishing with at least $p > 12.5$, then (send 10, split, punish on keep all) would be the unique subgame perfect Nash equilibrium, as depicted in Figure 7. That is, if Player 2 has sufficiently high belief that Player 1 is an altruistic reciprocator, Player 2 will cooperate.

16 The only differences are that when Player 1 deviates, then the game ends and both receive the payoff given mutual defection, and that if both players cooperate, no punishment is available.

Figure 7. Trust game with altruistic punishment ($p = 13$)



But why should people punish when it is not in their own self interest to do so? A possible explanation has been provided by Quervain et al. (2004). They give a neural basis for altruistic punishment by measuring brain activity when a player is called upon to actually carry out the punishment. They find that punishment of norm-violators activates reward-related brain circuits, and conclude that "people have a preference for punishing norm-violators". This means that the players' utility functions include more arguments than the payoffs specified in Figure 5 and Figure 6, since people receive reward from punishment.

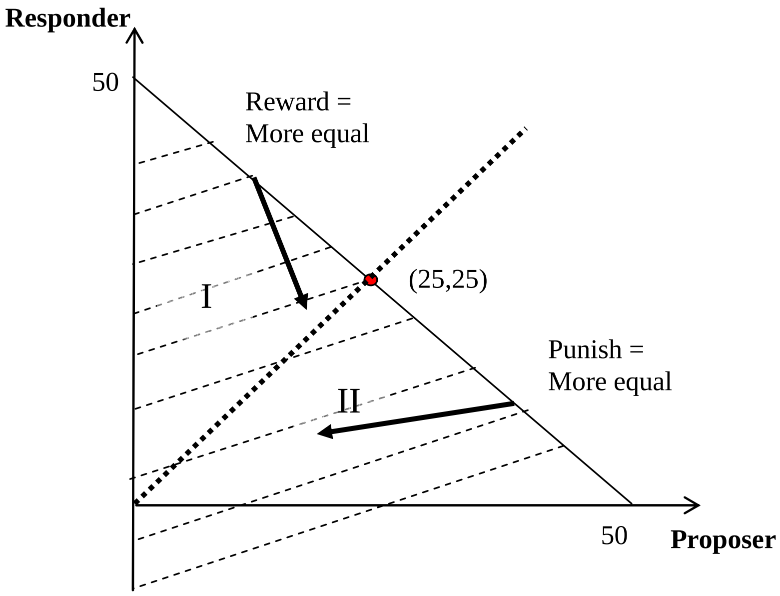
This result can be modelled in the framework of Equation 1 in several ways, by invoking both inequality aversion and fairness considerations. If it is possible to infer that people are willing to punish because they dislike inequality, such that they derive utility not from punishing par se, but from the fact that punishment will make the final outcome more equal, then inequality aversion can explain the above result. Moreover, fairness in the way it was defined in Section 4 can also explain the above behaviour.

The above-mentioned trust game considers only punishment; other experiments have, however, shown a tendency towards rewarding perceived good behaviour (see, e.g., Andreoni et al. (2003)).

Figure 8 gives a more general picture, inspired by Andreoni et al, (2003). Here, the proposer (in the final rounds of the game) can propose any allocation

$(\pi_{proposer}, \pi_{responder}) = (x, 50 - x),\ 0 \le x \le 50$. The punishment structure of the above trust game implies the possibility of moving the final allocation closer to the equal split line at the expense of both receiving less payoffs. (For any allocation $x > 25$, an arrow as shown into Area II in the figure implies such a punishment path). However, if the proposer is generous toward the responder for any allocation $x < 25$, then reward implies a movement into Area I in the figure. The arrow pointing into Area I implies self-sacrifice in order to move towards a more equal split.

Figure 8. A general punish-reward situation



In order to illustrate how inequality aversion can explain the behaviour found by Quervain et al. (2004), assume a more complicated function than in Section 4. Assume that:

$$\alpha_{ik} = -\varepsilon_{ik} \cdot \frac{[\pi_k(s) - \pi_i(s)]^2}{\pi_k(s)}$$

such that the less equal the split turns out to be, the more averse the player. (See footnote 13). Inserting this into Equation 1 yields: $u_i(p) = \pi_i(p) - \varepsilon_{ik}\left[\pi_i(p) - \pi_k(p)\right]^2$. Let player *i* be the responder (R) and player *k* the proposer (P) and consider, for example, the split $(40,10)$. Let us then derive the optimal punishment level.[17]

$$u_R(p) = (10-p) - \varepsilon_{RP}\left[(10-p)-(40-2p)\right]^2 = u_R(p) = (10-p) - \varepsilon_{RP}\cdot(p-30)^2$$

$$\frac{\partial u_R}{\partial p} = -1 - 2\varepsilon_{RP}\cdot(p-30) = 0 .$$

The result is:

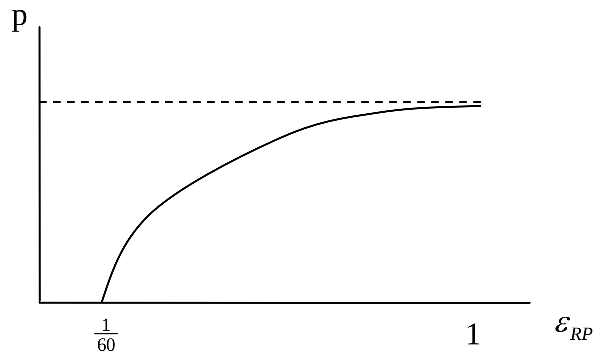
$$p = 30 - \frac{1}{2\varepsilon_{RP}} \text{ and } \frac{\partial p}{\partial \varepsilon_{RP}} > 0 .$$

The results are illustrated in Figure 9, which chows a graph of the connection between inequality aversion and the level of punishment (here it is assumed that $p \geq 0$ and $0 \leq \varepsilon_{RP} \leq 1$). It is seen that for small levels of inequality aversion, no punishment is used, which seems in accordance with the observation that people have a threshold for applying punishment.[18]

17 Note that the derivation here is done for a specific splitting proposal; a more general result could also be interesting.

18 The experiment of Quervain et al. (2004) shows a positive correlation between activity in certain parts of the brain and the willingness to incur costs to punish. Since people differ in these activities, this may explain why only a fraction are willing to punish in a given experiment.

Figure 9. The connection between ineqalitty aversion and the optimal punishment level



Finally, the observation that people are willing to reward also fits into the idea that punishment could be motivated be inequality aversion. Note that in this interpretation, the inequality aversion parameter can be interpreted as an aversion to norm violation.

However, the behaviour could equally well be explained by the notion of fairness. Fairness as presented in Section 4 indicates that players derive utility from punishing unfair behaviour and rewarding fair behaviour. Therefore, a movement from any point on the line of proposals in Figure 8 that results in $x > 25$ and movement into area II as shown in the figure could be motivated by the receiver deriving utility from reducing the payoff to the proposer (which countervails the direct utility loss from smaller payoff to him/herself).

To use the idea of Rabin (1993) that players derive utility from punishing unkind behaviour and rewarding kindness, assume the following utility function (which still fits into the idea of Equation 1, that a player also derives utility from the payoff of the other player):

$$u_R(x) = (50 - x) - p + F(x) \cdot (p)^{½}$$

$$F(x) \begin{array}{l} > 0 \;\; if \;\; x > 25 \\ = 0 \;\; if \;\; x = 25, \\ < 0 \;\; if \;\; x > 25 \end{array}$$

and $\frac{\partial F(x)}{\partial x} > 0$.

The first part of the utility function is the direct payoff to the receiver from the split if he chooses punishment level p. The fucntion $F(x)$ is the fairness function that measures how sensitive the receiver is to kindness and unkindness for each x. Unkindness is here defined as any allocation where $x > 25$. So for $x > 25$, the receiver gets utility from punishment, whereas if $x < 25$, the receiver gets utility from rewarding (assuming that $p < 0$). It is also assumed that there is a decreasing marginal utility of punishment. Finally, the function $F(x)$ is assumed to be monotonically increasing in x. The optimal p is found as:

$$\frac{\partial u_R}{\partial p} = -1 + F(x) \cdot \tfrac{1}{2\sqrt{p}} = 0 \Rightarrow p = \tfrac{1}{4}\left[F(x)\right]^2.$$

The result is that the wilingness to punish depends on *x*, the splitting rule, and how sensitive players are towards kindness / unkindness ($F(x)$). Again, if the function $F(x)$ differs between individuals, we find that people punish differently.

The conclusion is that given the interpretation of punishment in this section, punishment can be due to perfect rational behaviour.

6. The evolution of social norms

Individual behaviour (like the willingness to cooperate) is not only shaped by direct (face-to-face) encounters, but also by norms that define acceptable behaviour for individuals belonging to a group (or who want to be accepted as a group menber). A more formal definition of social norms is given by Coleman (1988) and Fehr and Gächter (2000). They define a social norm as:

1) 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.[19]

Note that this definition is very broad, and does not mean that the moral code needs to have a justification in normative ethics.[20] Ostrom (2000) finds strong support for the assumption that modern humans have inherited a propensity to learn social norms.[21]

An appropriate tool for analysing the development of social norms is evolutionary game theory. The idea that evolutionary game theory provides the tools to analyze which strategies, or patterns of behaviour, emerge over time through a process of adaptation is put forward by Kolstad (2003). He notes that 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.[22]

---

19 The intrinsic cost that an individual suffers from failing to use a social norm, such as telling the truth or keeping a promise, is referred to as guilt (if entirely self-inflicted) or shame (when others know about the failure).

20 Note that the opposite is not true: that a certain pattern of behaviour emerges is not identical to saying that this behaviour establishes a social norm, given the definition of norms. A norm needs to be a shared belief of how one ought to behave and how to behave against defectors. Moreover, norms like stealing or killing are also covered by this definition if a group holds such shared views that group members should do so, and will be punished if they do not. (Such norms can be found among criminal gangs, like Hell's Angels).

21 Which norms are learned, however, varies from one culture to another, across families, and with exposure to diverse social norms expressed within varies types of situations.

22 The human capacity to establish and enforce social norms is perhaps the decisive reason for the uniqueness of human cooperation in the animal world. An evolutionary explanation of the evolvement of social norms is given by Fehr and Fischbacher (2004). The essence is that it needs a high level of cognitive and emotional capacity to enforce social norms. "Punishment must be understand, must be understand as a consequence of certain action, and it must be understand that adjusting to the norms will eliminate the punishment".

A common feature of evolutionary models is that players are matched repeatedly to play a game; a dynamic process describes how players adapt their behaviour over time. The process of adaptation might reflect biological selection or it might represent learning as agents switch to strategies that are observed to do better (Kolstad, 2003). Most important for our purposes is that it focuses on the process of how a specific equilibrium emerges through an evolutionary process. Evolutionary game theory sees behavioural patterns as the outcome of a process of adaptation in which behaviours that do better are selected.

In the context of evolutionary games, different strategies compete against each other. As already noted in the introduction, in an evolutionary setting, players are modelled as non-strategic, following a pre-described rule (strategy). Stability is achieved if, given a population consisting of one strategy, no other strategy can invade the population and achieve a higher payoff than the present strategy. In this case, we have an evolutionarily stable strategy. The strategies that survive can then be interpreted as a norm. If only defectors survive in a society, the prevailing norm in that society is defection.

The formal definition is that if we refer to a small group of players using an alternative strategy (or invading the existing population) as mutants, an evolutionarily stable strategy (ESS) is robust to mutations in a certain sense. To be un-invadable 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:

$$u(x,x) \geq u(y,x) \tag{6.1}$$

$$u(x,x) = u(y,x) \Rightarrow u(x,y) > u(y,y) . \tag{6.2}$$

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

Applying the ESS to a prisoner's dilemma, it is easily shown that defection is the unique ESS.

Given a population of cooperators, consider the invasion of a defector. Condition (6.1) says that in order for the population of cooperators to be stable against invaders, the payoff of cooperators playing against cooperators should be higher than cooperators playing against invaders, which is not true. Therefore, a small group of mutant defectors could invade a population of cooperators.

Moreover, a population of defectors is stable against the invasion of cooperators since a cooperator does worse against a defector than the defector does against itself, so mutant cooperators cannot invade a population of defectors. As a consequence, a population that is randomly and repeatedly matched to play a one-shot prisoners dilemma game thus ends up playing defect. Only the norm of defection survives.

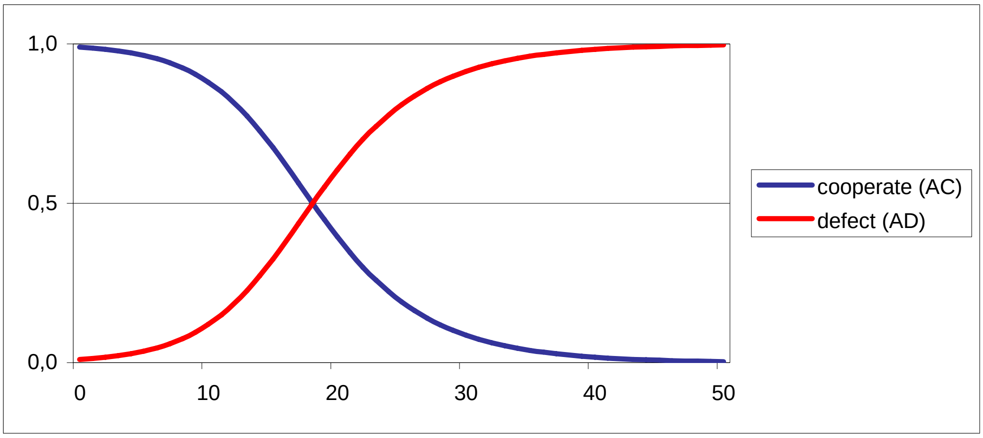
To show an example of this, assume the share *x* of strategy *i* will evolve according to the following replicator function:

$$\frac{\dot{x}_t^i}{x_t^i} = f(\pi_i^t - \pi_t^{Average})$$

where $f_{\pi_t^i}^{'} > 0$ and $f(0) = 0$. $\pi_t^{Average}$ measures the average payoff for the population at time *t*. Let the players consist of "always cooperators" (AC) and "always defectors" (AD) and let the strategies be randomly paired to play the prisoner's dilemma (with restricted strategy spaces) from Table 2. For the following examples, let $\dot{x}_t^i = x_t^i \cdot 0.15 \cdot (\pi_i^t - \pi_t^{Average})$.

It is easy to show that if a population of cooperators is invaded by defectors, then the defectors will eliminate the cooperators, shown in Figure 10, where a population of AC are invaded by few AD.

Figure 10. Invading defectors out compete cooperators



Several extensions that imply that cooperators can also survive in the long run are possible from this benchmark model. First of all, things might change radically if it is somehow possible for players to identify the type of encounter and then select whom to play against.

To see this, let a cooperator be matched with another cooperator with probability *r*, whereas with probability 1-*r*, he is matched with a random member of the population.[23] Assume the population share of cooperators is *s*. Take again the payoff matrix from Table 1:

$$u(C) = r \cdot a + (1-r)\cdot(s\cdot a + (1-s)\cdot c) = r(a-c) - rs(a-c) + s(a-c) + c$$
$$u(D) = r\cdot d + (1-r)\cdot(s\cdot b + (1-s)\cdot d) = s(b-d) - rs(b-d) + d$$
$$u(C) > u(D) \Rightarrow$$
$$r > \frac{d-c-s(a-b-c+d)}{a-c-s(a-b-c+d)}.$$

Given the numbers from Table 3: $r > \frac{1+s}{4+s}$, asan example, if $s = 0.5 =$ then $r > \frac{1}{3}$.

The conclusion is that if matching is sufficiently assortative, cooperators, on average, do better than defectors, which means that a small segment of mutant cooperators can invade a population of defectors. In contrast, in this case,

23 This example is taken from Kolstad (2003).

mutant defectors cannot invade a population of cooperators. We can thus substantiate the evolution of a cooperative norm when matching is assortative.[24]

The social norm of cooperation can be sustained against defectors if group members can be easily identified. One implication could be that such a group, once established, is difficult to become a member of unless aspirants make a significant effort to reveal themselves as cooperators as well since such a group is very vulnerable to free-riding (exploitation) and needs strong adherence to the social norm of cooperation.

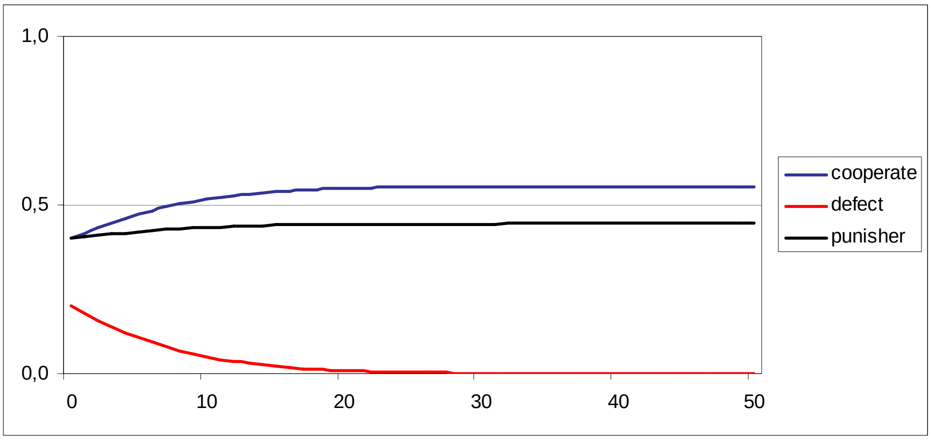
The second extension is to include more types. The introduction of another type, a reciprocal punisher (RP), might change things considerably. Here, the RP-types are able to punish after observing the outcomes of the game in the way described in Figure 2b, such that they, in the case of defection, invoke a punishment that costs $p < 0$ to both. Unless otherwise noted, $p = -1$ (In the sense that both players in total receive -1 of the encounter). The (expected) payoffs to the players are:

$$\pi_t^C = x_t^C \cdot 4 + x_t^P \cdot 4 + x_t^D \cdot 0$$
$$\pi_t^P = x_t^C \cdot 4 + x_t^P \cdot 4 + x_t^D \cdot (-1)$$
$$\pi_t^D = x_t^C \cdot 6 + x_t^P \cdot (-1) + x_t^D \cdot 1 .$$

The presence of a sufficiently large share of RP-types implies that the payoff to the D-types is lower on average than for the C-types, which implies that the share of the C-types increases. The RP-types receive less than the C-types, but more than the D-types, eventually eliminating the D-types and creating an increased share of C-types. The dynamic is shown in Figure 11.

24 If a group of users can determine its own membership – including those who agree to use the resources according to agreed-upon rules and excluding those who do not agree to these rules – the group has made an important first step toward the development of greater trust and reciprocity. (Ostrom, 2000, page 149).

Figure 11. The presence of PR-types eliminates the D-types. Initially: $x^C = x^P = 0.4, x^D = 0.2$



For further examples, see Appendix 1.

The social norm of cooperation can also be sustained against defectors if a sufficiently large number of reciprocal punishers are present, since such types punish the defectors, but reward the cooperators. In this way, these types act as an explicit punishment mechanism for the cooperators.

7. Some examples of the applicability of the enriched preferences

Finally, the enrichment of preferences and the possibility of punishment have been used, and can be used, to explain phenomena like the presence of a welfare state or the propensity of people to act in a sustainable way.

7.1. Explaining the welfare state and its stability

One key feature of a welfare state is that it reduces income inequality. As stated by Bowles et al. (2001), the modern welfare state is a remarkable human achievement wherein a substantial fraction of total income is regularly transferred from the better off to the less well off. Furthermore, the public regularly supports the governments that are in charge of these transfers.[25]

Bowles et al. (2001) provide an evolutionary argument for the emergence of a welfare state and the presence of strong reciprocal types. Their argument is that the presence of strong reciprocal behaviour implies adherence to norms. If adherence to such norms implies a larger benefit to the group (like cooperation in a PD) then the number (share) of strong reciprocal types is likely to increase.

Bowles et al. (2001) argue that it is possible to design an egalitarian institution, a system that is generous toward the poor and rewards those who perform socially valued work and those who seek to improve their chances of engaging in such activities and those who are poor through accidents.

The reason is the observation that people exhibit significant levels of generosity even towards strangers, but beliefs about the causes of the high and low income matter. Some people contribute to public goods and finally, people punish free riders at substantial cost to themselves, even when they cannot reasonably expect future personal gain themselves.

While the emergence of an egalitarian system can be explained, it is more difficult to explain why different countries (that are relatively equal in terms of average per capita income) have significantly different levels of income redistribution (e.g., compare Denmark with the U.S.). However, *reciprocity is context-dependent and path-dependent.*

Again according to Bowles et al. (2001), there is evidence that the larger the social distance (or the level of heterogeneity of the population in general, or even the perceived distance, see below), the less likely people are to accept redistributions (the idea of context dependency). Furthermore, Ostrom (2000) finds strong support for the assumption that modern humans have inherited a

25 In Denmark, a majority of the voters prefer more welfare to tax reduction, according to a recent opinion poll conducted in 2007.

propensity to learn social norms, such that the norms of a group are passed to future generations (the idea of path dependency).

It is possible to model this in the framework of Equation 1. Since the welfare state is an egalitarian institution aimed at reducing inequalities, the underlying behaviour is likely inequality aversion. A player's willingness to accept redistribution can be modelled by:

$$U_i(s) = \pi_i(s) - \varepsilon_{ik}(z_{ik}^e(G_j)) \cdot [\pi_i(s) - \pi_k(s)].$$

Player *i*'s willingness to accept redistribution (assuming that $[\pi_i(s) - \pi_k(s)] > 0$) is dependent on *i's* view of player *k's* characteristics measured by the vector z_{ik}^e, which incorporates individual *i*'s perception of the reason for individual *k* being poor. Furthermore, wilingness to redistribute might also be shaped by prevaling norms in the group that *i* belongs to, measured by G_j.

How can z_{ik}^e be thought of depending on G_j? One group-specific mechanism that can shape individual behaviour is the presences of generalized trust, which can be defined as a measure of the willingness of persons to cooperate with strangers. See, for example, Uslaner (2004) and Brandt and Svendsen (2007); the latter argue for the importance of social capital for the emergence of the welfare state.[26] It has been empirically verified that countries with high levels of redistribution also have high levels of generalized trust (see Appendix 2). The idea is that when you trust strangers, you are more inclined to help the poor.).[27]

---

26 According to Uslaner (2004), in an unequal world, it is difficult to establish bonds between those at the top and those at the bottom. Trust also depends on a foundation of optimism and control. People are optimistic and feel that they have control over their environment when the gap between the rich and the poor is small. In an unequal world, people will be reluctant to take risks in dealing with people who might be different from themselves. They will press for closed markets and work within their own cultures (and will tolerate corruption).

27 Not everybody shares this view, however: "the worship of trust as some magic solution to irreducibly complex collective action dilemmas is in my eyes nearly as incompatible with an academic mindset as is the worship of deities." Ludvigsen (2006), page 1.

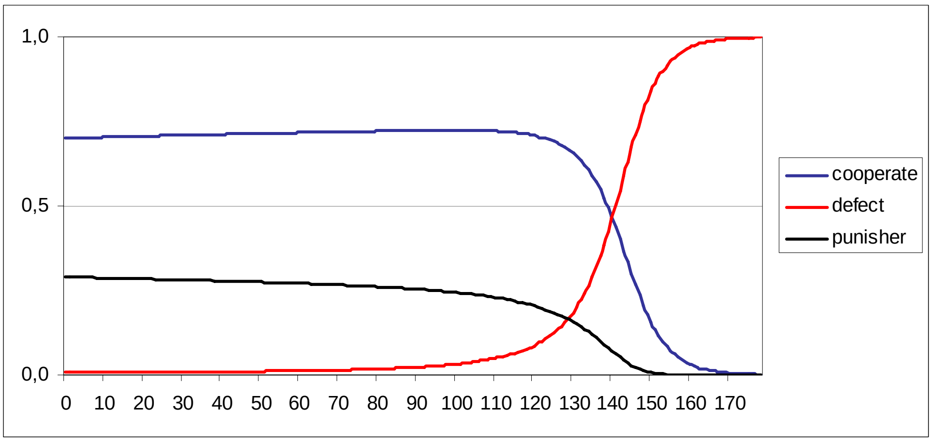
However, the causality between income equality and generalized trust is crucial. It could well be that income equality leads to trust.

To conclude, one explanation for the emergence of a welfare stare is as follows: A group that is able to corporate and hence produce benefits for most of the group is the basis for the evolvement of a large share of reciprocal cooperators. As soon as a sufficiently large proportion of a group is made up of reciprocal cooperators, the level of actual punishment will diminish. Furthermore, Ostrom (2000) provides evidence that humans have a strong propensity for learning social norms. Internalizing punishments for norm violation through sentiments of guilt and shame also means that actual punishment is low. This in turn implies the build-up of trust in a group and the reduction of social distance such that the willingness to redistribute increases.

However the approach in section 6 can also be used to show the fragility of the welfare system. Can AD-types invade a population consisting of both AC and RP types? That is, can they destroy the welfare state. This depends on the relative share of AC and AD types. If sufficiently large share of AC types, then they can. (Compare this with figure 11, where stability is secured).

In figure 12, the AD types eventually can invade and defeat a population consisting of AC and RP types. But as seen in the figure, it might take a long time. It takes 112 rounds before the share of the AD-types is above 5%, such that the population is fairly stable in a long period, but from that point on, the AD-types share growth rapidly.

Figure 12. The AD-types eventually eliminate the other types. Initially: $x^C = 0.7, x^D = 0.01, x^P = 0.29$



The result from comparing figure 11 and 12 is that it needs a sufficiently share of reciprocal punishers to guarantee the welfare state against invading defectors (= norm violators that free ride).

7.2. Explaining sustainability

Acting sustainably can be interpreted as a voluntary contribution to a public good for future generations. Therefore, it seems impossible to explain why a "homo economicus" should ever act sustainably.

However, to recapitulate, according to Fehr and Fischbacher (2004), the need for credible sanctioning threats creates the condition under which the norm of conditional cooperation enables groups to establish high and stable cooperation in the presence of a large minority of selfish individuals who violate cooperation norms. In other words, the punishment opportunities and associated credible punishments generate beliefs that the other group members will provide a high level of cooperation. [28]

28 This leads Fehr and Fischbacher (2004) to the conclusion that in the absence of sanctions for non-cooperation, this pattern of conditional cooperation – in combination with the existence of complete free riders – is likely to cause decreasing contributions over time. High and stable

The future generation cannot, however, reciprocate present behaviour. Therefore, this is a situation where punishment is not possible, since the future generation cannot punish earlier generations, and therefore, observation of sustainable behaviour cannot be a result of behaviour triggered by beliefs about credible sanction threats.

Individual preferences can, however, be shaped by fairness and inequality aversion. Arguments in favour of sustainability are often based on a moral obligation to act sustainably. As Becerman (2007) notes, although advocates of sustainable development may disagree about its precise definition, one thing they all agree on is that society ought to adopt it as a goal.

Fairness as defined in Section 4 could be used as a reason to act sustainably. If a person expects that future generations would act sustainably towards him/her, then he/she should also act in a sustainable way towards future generations. This could also be modelled as a version of Rawlsian ethics: one should act to maximize the welfare (or possibilities) of the worst off generation.

However, since sustainability is explicitly about equating the income of various generations, inequality aversion might be most relevant in connection with explaining the reasons for acting sustainably. Essentially, the sustainablity problem can be expressed as in Table 7. Here, only the present generation has strategic choice, while the second (any future generation) can only accept earlier generations' choices.

---

cooperation levels cannot be maintained if the majority of subjects want to contribute below the average contribution of the other group members.

Table 7. Sustainability Game

Generation 1 (present)	Generation 2 Accept
Cooperate (C)	(6,6)
Defect (D)	(10,1)

The problem that the player in the present generation faces is whether to choose an equal intergenerational equality of income or care only about his/her own income. Sustainable behaviour can therefore be characterized as *intergenerational inequality aversion*. This fit into Equation (1):

$$U_i(s_i) = \pi_i(s_i) - \sum_{k=2}^{\infty} \varepsilon_{ik}(G_j) \cdot [\pi_i(s_i) - \pi_k(s_i)].$$

First, only the strategy of player i matters; secondly, *k* is an index of the future generation; and finally, it is reasonable to expects that $\varepsilon_{ik} = 0$ for some $k \geq 2$ and onwards. $\varepsilon_{ik} = \varepsilon_{ik}(G_j)$ includes the idea that group norms might shape individual preferences towards future generation.

Given the logic of inequality aversion, (for any individual, the envy parameter is larger than the parameter measuring advantageous inequity aversion), and given the observation of decreasing sensitivity of inequality aversion, this suggests that that type of behaviour would suggest a decreasing consumption path over time. Deceasing sensitivity of inequality aversion also implies that even if individuals care about inequality, this aversion against inequality diminishes over time.

This can be expressed as follows in a two-player version:

$$\frac{\partial U_i}{\partial d_{ij}}\Big|_{\pi_i+\pi_2=\bar{\pi}} < 0 \text{ for } d_{ij} > 0 \qquad (7.1)$$

$$\frac{\partial^2 U_i}{\partial (d_{ij})^2}|_{\pi_i+\pi_2=\bar{\pi}} > 0 \text{ for } d_{ij} > 0 \tag{7.2}$$

$$\frac{\partial^2 U_i}{\partial d_{ij}\partial t}|_{\pi_i+\pi_2=\bar{\pi}} < 0 \text{ for } d_{ij} > 0 . \tag{7.3}$$

(7.1) expresses inequality aversion, which simply says that for the same total payoff, Player 1 derives less utility as the distance between the two payoffs increases. (7.2) implies decreaing sensitivty of ineaulity aversion. In (7.3), the variable t measures the distance in time between i and j. Therefore, (7.3) is the time analog to (7.2), where the same absolute payoff and distance between payoffs implies less loss of utility as distance in time increases.

The time effect can be tested, but common wisdom indicates that this effect is rather strong. Furthermore, facts seem to support the view that people in general do not act sustainably. A new report from the UN seems to support this hypothesis: Two decades ago, our common future emphasized the urgency of sustainable development, yet environmental degradation continues to threaten human wellbeing, endangering health, physical security, social cohesion and the ability to meet material needs. Analysis throughout GEO-4 also highlights rapid disappearing forests, deteriorating landscapes, polluted waters and urban sprawl. The objective is not to present a dark and gloomy scenario, but an urgent call for action. (GEO-4, 2007, Page 34).

As a final remark, if there is a desire to follow a sustainable path but incentives are not in favour of such a path, then according to Sandler (1997), one should seek solutions that are beneficial for the present generation and at the same time also imply benefits for future generations (see Sandler for examples).

8. Conclusion

This paper shows that by using experimental findings to enrich the preferences of people, it is possible to make more and better predictions about individual decision making. Results are no longer clearcut, but this is in accordance with real life, where decisions typically hinge on several conditions and circumstances. The enrichment of preferences can deliver a characterisation of such conditions. The inclusion of other-regarding preferences makes it possible to provide explanations not possible when sticking to the hypothesis of homo economicus; for example, the appearance and stability of, the welfare states in the Nordic countries, or the limited (but not absent) incentives to act sustainably.

As stated in the introduction, although the introduction of economic man was thought to be an advance, this comes with a price. It was mentioned that public choice has rather uncritically introduced economic man into the political arena. This has been criticised by some. Kelman (1987, p. 81) states rather ironically:

> *Do you want to understand why government officials behave the ways they do? All you need to know is that they are trying to maximize the budgets of their agencies. Do you want to understand what drives politicians? All you need to know is that they want to be re-elected. Do you want to understand legislation? Just see it as a sale of the coercive power of government to the highest bidder, like a cattle auction.*

Essentially, Kelman lacks motives other than the purely selfish in such analysis, such as the presence of norms, in particular, norms of ‘public spirit’.

Applying the main lessons from decades of experiments on the subject of economic behaviour, we should not model the political arena as a marketplace; on the other hand, it is easy to be convinced rather than blindly accept that politicians and bureaucrats are simply norm-abiding agents devoted to serving a ‘public spirit’.

9. References

[1] Abbink K., G.E. Bolton, A. Sadrieh and F-F. Tang (2001). 'Adaptive Learning versus Punishment in Ultimatum Bargaining', *Games and Economic Behavior*, 37, 1-25.

[2] Andreoni, J. (1995). 'Cooperation in Public Goods Experiments: Kindness or Confusion?', *American Ecnomic Review*, **85**, 891-904.

[3] Andreoni, J., W. Harbaugh and L. Vesterlund (2003). 'The Carrot or the Stick: Rewards, Punishment and Cooperation', *American Economic Review*, **93**, 893-902.

[4] Beckerman, W (2007). 'The Chimera of Sustainable development', *Electronic Journal of Sustainable Development*, vol 1, issue 1.

[5] Bolton, G.E. and A. Ockenfels (2000). 'ERC: A Theory of Equity, Reciprocity and Competition', *American Economic Review*, **90**, 166-193.

[6] Bowles, S. C. Fong and H. Gintis (2001). 'Reciprocity and the Welfare State', Working paper.

[7] Brandt and Svendsen (2007). 'Money for Nothing?The Nordic Welfare State, Social Capital and Reciprocity', Presented at PC2007, Amsterdam (april 2007).

[8] Camerer, C.F. and E. Fehr (2006). 'When Does ''Economic Man'' Dominate Social Behaviour?', *Science*, **311**, 47-52.

[9] Coleman, J.S. ([1987] 2003). 'Norms as Social Capital', in Gerard Radnitzky and Peter Bernholz (eds), *Economic Imperialism: The Economic Approach Applied Outside the Field of Economics*, New York: Paragon

House Publishers, 133-55. Reprinted in Ostrom, E. and Ahn, T.K.(eds.) *Foundations of Social Capital*. Edward Elgar Publishing, UK.

[10] Fehr, E. and S. Gächter (2000). ‘Cooperation and Punishment’, *American Economic Review*, **90**, 980-994.

[11] Fehr, E. and S.Gächter (2000). ‘Fairness and Retaliation: The economics of Reciprocity’, *The Journal of Economic Perspectives*, 14, 159-181.

[12] Fehr, E. and K. Schmidt (1999). ‘A Theory of Fairness, Competition and Cooperation’, *Quarterly Journal of Economics*, 114, 817-68.

[13] Fehr, E and U. Fischbacher (2004). ‘Third-party punishment and social norms’, *Evolution and Human Behavior*, **25**, 63–87.

[14] Fehr, E and U. Fischbacher (2004). ‘Social Norms and Human Cooperation’, TRENDS in Cognitive Sciences, 8, 185-190.

[15] GEO-4 (2007). UNEP's Global Environment Outlook: environment for development, http://www.unep.org/Documents.Multilingual/Default.asp?DocumentID=519&ArticleID=5688&l=en.

[16] Houser, D. and R. Kurzban (2002). ‘Revisiting Kindness and Confusion in Public Goods Experiments’, *American Economic Review*, **92**, 1062-1069.

[17] Hill, S.A. and W. Nielson (2007). ‘Inequality aversion and Diminishing Sensitivity’, *Journal of Economic Psychology*, **28**, 143-153.

[18] Ludvigsen, S.S. (2006). ‘The Trust Delusion – or why cooperation is not irreducibly complex’, paper presented at the First World Meeting of the Public Choice Society, 2007.

[19] Kelman, S. (1987). ‘Public Choice and Public Spirit’, Public Interest, 87, 80–94.

[20] Kolstad, I. (2003). ‘The evolution of social norms’, *CMI working paper* 2003:1.

[21] Ostrom, E. (2000). ‘Collective Actions and the Evolution of Social Norms’, *The Journal of Economic Perspective*, 14, 137-158.

[22] Quervain, D. J-F, U. Fischbacher, V. Treyer, M. Schellhammer, U. Schnyder, A. Buck and E. Fehr (2004). ‘The Neural Basis of Altruistic Punishment’, *Science*, 305, 1254-1258.

[23] Rabin, M. (1993). ‘Incorporating Fairness into Game Theory and Economics’, *American Economic Review*, 83, 1281-1302.

[24] Trivers, R. L. (1971). ‘The Evolution of Reciprocal Altruism’, *Quarterly Review of Biology*, 46, 35–57.

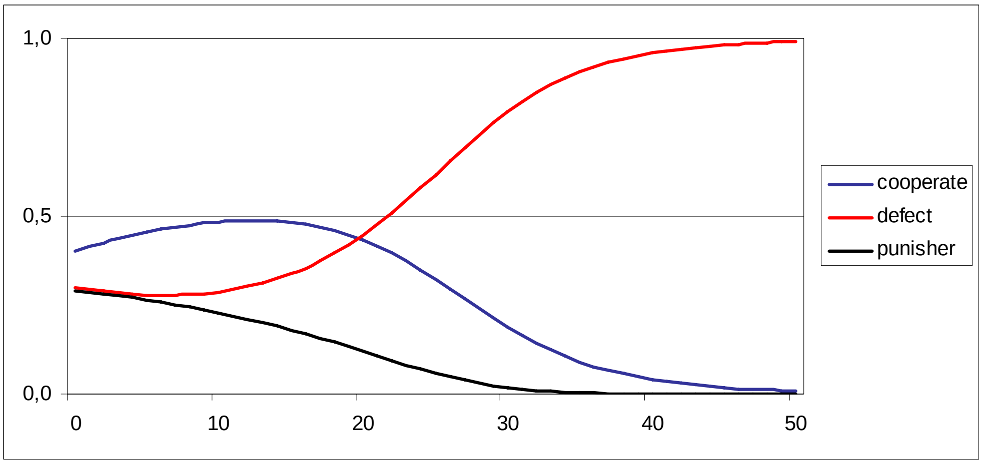
[25] Uslaner, E.M. (2004). ‘Where You Stand Depends Upon Where Your Grandparents Sat: The Inheritability of Generalized Trust.’ Working paper, University of Maryland, US.

Appendix 1.

Can AD-types invade a population consisting of both AC and AP types? It depends on the relative share of AC and AD types. In Figure 11, it was not possible, while in Figure 12 it was. Figure A1 presents another example: Although initially the share of AC-types increased, since the share of P-types is falling, AD-types will eventually earn more than AC-types.

Figure A1. The D-types eliminate also all the PR-types. Initially:

$$x^C = 0.4, x^D = x^P = 0.3$$

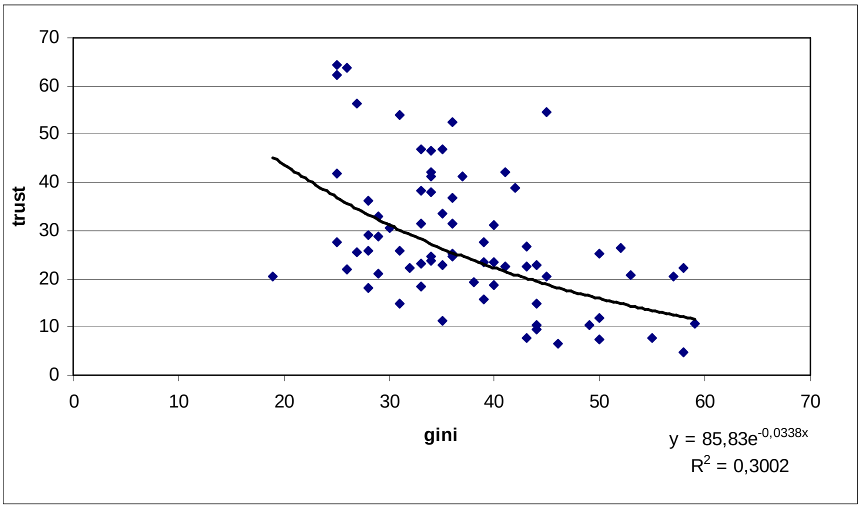


The result is that the initial distribution of types is important for the development of the population dynamics.

Appendix 2

Empirically, there is a connection between generalized trust and income equality

Figure A6. Empirical relationship between trust and income equality (UN Survey 1992 to 2003)[29]



29 For most countries, the numbers are provided as averages from 200-2003; for some countries, only one year is available (e.g., Namibia only 1993).

Department of Environmental and Business Economics

Institut for Miljø- og Erhvervsøkonomi (IME)

IME WORKING PAPERS

ISSN: 1399-3224

Issued working papers from IME

Udgivne arbejdspapirer fra IME

No.		
1/99	Frank Jensen Niels Vestergaard Hans Frost	*Asymmetrisk information og regulering af forurening*
2/99	Finn Olesen	*Monetær integration i EU*
3/99	Frank Jensen Niels Vestergaard	*Regulation of Renewable Resources in Federal Systems: The Case of Fishery in the EU*
4/99	Villy Søgaard	*The Development of Organic Farming in Europe*
5/99	Teit Lüthje Finn Olesen	*EU som handelsskabende faktor?*
6/99	Carsten Lynge Jensen	*A Critical Review of the Common Fisheries Policy*
7/00	Carsten Lynge Jensen	*Output Substitution in a Regulated Fishery*
8/00	Finn Olesen	*Jørgen Henrik Gelting – En betydende dansk keynesianer*
9/00	Frank Jensen Niels Vestergaard	*Moral Hazard Problems in Fisheries Regulation: The Case of Illegal Landings*
10/00	Finn Olesen	*Moral, etik og økonomi*

11/00	Birgit Nahrstedt	*Legal Aspect of Border Commuting in the Danish-German Border Region*
12/00	Finn Olesen	*Om Økonomi, matematik og videnskabelighed - et bud på provokation*
13/00	Finn Olesen Jørgen Drud Hansen	*European Integration: Some stylised facts*
14/01	Lone Grønbæk	*Fishery Economics and Game Theory*
15/01	Finn Olesen	*Jørgen Pedersen on fiscal policy - A note*
16/01	Frank Jensen	*A Critical Review of the Fisheries Policy: Total Allowable Catches and Rations for Cod in the North Sea*
17/01	Urs Steiner Brandt	*Are uniform solutions focal? The case of international environmental agreements*
18/01	Urs Steiner Brandt	*Group Uniform Solutions*
19/01	Frank Jensen	*Prices versus Quantities for Common Pool Resources*
20/01	Urs Steiner Brandt	*Uniform Reductions are not that Bad*
21/01	Finn Olesen Frank Jensen	*A note on Marx*
22/01	Urs Steiner Brandt Gert Tinggaard Svendsen	*Hot air in Kyoto, cold air in The Hague*
23/01	Finn Olesen	*Den marginalistiske revolution: En dansk spire der ikke slog rod?*
24/01	Tommy Poulsen	*Skattekonkurrence og EU's skattestruktur*
25/01	Knud Sinding	*Environmental Management Systems as Sources of Competitive Advantage*
26/01	Finn Olesen	*On Machinery. Tog Ricardo fejl?*
27/01	Finn Olesen	*Ernst Brandes: Samfundsspørgsmaal - en kritik af Malthus og Ricardo*
28/01	Henrik Herlau Helge Tetzschner	*Securing Knowledge Assets in the Early Phase of Innovation*
29/02	Finn Olesen	*Økonomisk teorihistorie* *Overflødig information eller brugbar ballast?*

30/02	Finn Olesen	*Om god økonomisk metode – beskrivelse af et lukket eller et åbent socialt system?*
31/02	Lone Grønbæk Kronbak	*The Dynamics of an Open Access: The case of the Baltic Sea Cod Fishery – A Strategic Approach -*
32/02	Niels Vestergaard Dale Squires Frank Jensen Jesper Levring Andersen	*Technical Efficiency of the Danish Trawl fleet: Are the Industrial Vessels Better Than Others?*
33/02	Birgit Nahrstedt Henning P. Jørgensen Ayoe Hoff	*Estimation of Production Functions on Fishery: A Danish Survey*
34/02	Hans Jørgen Skriver	*Organisationskulturens betydning for vidensdelingen mellem daginstitutionsledere i Varde Kommune*
35/02	Urs Steiner Brandt Gert Tinggaard Svendsen	*Rent-seeking and grandfathering: The case of GHG trade in the EU*
36/02	Philip Peck Knud Sinding	*Environmental and Social Disclosure and Data-Richness in the Mining Industry*
37/03	Urs Steiner Brandt Gert Tinggaard Svendsen	*Fighting windmills? EU industrial interests and global climate negotiations*
38/03	Finn Olesen	*Ivar Jantzen – ingeniøren, som beskæftigede sig med økonomi*
39/03	Finn Olesen	*Jens Warming: den miskendte økonom*
40/03	Urs Steiner Brandt	*Unilateral actions, the case of international environmental problems*
41/03	Finn Olesen	*Isi Grünbaum: den politiske økonom*
42/03	Urs Steiner Brandt Gert Tinggaard Svendsen	*Hot Air as an Implicit Side Payment Arrangement: Could a Hot Air Provision have Saved the Kyoto-Agreement?*

43/03	Frank Jensen Max Nielsen Eva Roth	*Application of the Inverse Almost Ideal Demand System to Welfare Analysis*
44/03	Finn Olesen	*Rudolf Christiani – en interessant rigsdagsmand?*
45/03	Finn Olesen	*Kjeld Philip – en økonom som også blev politiker*
46/03	Urs Steiner Brandt Gert Tinggaard Svendsen	*Bureaucratic Rent-Seeking in the European Union*
47/03	Bodil Stilling Blichfeldt	*Unmanageable Tourism Destination Brands?*
48/03	Eva Roth Susanne Jensen	*Impact of recreational fishery on the formal Danish economy*
49/03	Helge Tetzschner Henrik Herlau	*Innovation and social entrepreneurship in tourism - A potential for local business development?*
50/03	Lone Grønbæk Kronbak Marko Lindroos	*An Enforcement-Coalition Model: Fishermen and Authorities forming Coalitions*
51/03	Urs Steiner Brandt Gert Tinggaard Svendsen	*The Political Economy of Climate Change Policy in the EU: Auction and Grandfathering*
52/03	Tipparat Pongthanapanich	*Review of Mathematical Programming for Coastal Land Use Optimization*
53/04	Max Nielsen Frank Jensen Eva Roth	*A Cost-Benefit Analysis of a Public Labelling Scheme of Fish Quality*
54/04	Frank Jensen Niels Vestergaard	*Fisheries Management with Multiple Market Failures*
55/04	Lone Grønbæk Kronbak	*A Coalition Game of the Baltic Sea Cod Fishery*

56/04	Bodil Stilling Blichfeldt	*Approaches of Fast Moving Consumer Good Brand Manufacturers Product Development "Safe players" versus "Productors": Implications for Retailers' Management of Manufacturer Relations*
57/04	Svend Ole Madsen Ole Stegmann Mikkelsen	*Interactions between HQ and divisions in a MNC - Some consequences of IT implementation on organizing supply activities*
58/04	Urs Steiner Brandt Frank Jensen Lars Gårn Hansen Niels Vestergaard	*Ratcheting in Renewable Resources Contracting*
59/04	Pernille Eskerod Anna Lund Jepsen	*Voluntary Enrolment – A Viable Way of Staffing Projects?*
60/04	Finn Olesen	*Den prækeynesianske Malthus*
61/05	Ragnar Arnason Leif K. Sandal Stein Ivar Steinshamn Niels Vestergaard	*Actual versus Optimal Fisheries Policies: An Evaluation of the Cod Fishing Policies of Denmark, Iceland and Norway*
62/05	Bodil Stilling Blichfeldt Jesper Rank Andersen	*On Research in Action and Action in Research*
63/05	Urs Steiner Brandt	*Lobbyism and Climate Change in Fisheries: A Political Support Function Approach*
64/05	Tipparat Pongthanapanich	*An Optimal Corrective Tax for Thai Shrimp Farming*
65/05	Henning P. Jørgensen Kurt Hjort-Gregersen	*Socio-economic impact in a region in the southern part of Jutland by the establishment of a plant for processing of bio ethanol*
66/05	Tipparat Pongthanapanich	*Options and Tradeoffs in Krabi's Coastal Land Use*
67/06	Tipparat Pongthanapanich	*Optimal Coastal Land Use and Management in Krabi, Thailand: Compromise Programming Approach*

68/06	Anna Lund Jepsen Svend Ole Madsen	*Developing competences designed to create customer value*
69/06	Finn Olesen	*Værdifri samfundsvidenskab? - nogle reflek-sioner om økonomi*
70/06	Tipparat Pongthanapanich	*Toward Environmental Responsibility of Thai Shrimp Farming through a Voluntary Man-agement Scheme*
71/06	Finn Olesen	*Rational Economic Man og Bounded Ratio-nality – Nogle betragtninger over rationali-tetsbegrebet i økonomisk teori*
72/06	Urs Steiner Brandt	*The Effect of Climate Change on the Proba-bility of Conservation: Fisheries Regulation as a Policy Contest*
73/06	Urs Steiner Brandt Lone Grønbæk Kronbak	*Robustness of Sharing Rules under Climate Change. The Case of International Fisheries Agreements*
74/06	Finn Olesen	*Lange and his 1938-contribution – An early Keynesian*
75/07	Finn Olesen	*Kritisk realisme og post keynesianisme.*
76/07	Finn Olesen	*Aggregate Supply and Demand Analysis – A note on a 1963 Post Keynesian Macroeco-nomic textbook*
77/07	Finn Olesen	*Betydningen af Keynes' metodologi for aktuel makroøkonomisk forskning – En Ph.D. fore-læsning*
78/08	Urs Steiner Brandt	*Håndtering af usikkerhed og betydningen af innovationer i klimaproblematikken: Med udgangspunkt i Stern rapporten*
79/08	Lone Grønbæk Kronbak Marko Lindroos	*On Species Preservation and Non-Cooperative Exploiters*
80/08	Urs Steiner Brandt	*What can facilitate cooperation: Fairness, ineaulity aversion, punishment, norms or trust?*