Nordic Journal of Political Economy

Volume 36 2010 Article 1

The Survival of the Nordic Welfare State and Social Trust

Urs Steiner Brandt* and Gert Tinggaard Svendsen**

* Urs Steiner Brandt, Associate Professor, PhD, Department of Environmental and Business Economy, University of Southern Denmark, Niels Bohrs Vej 9, DK-6700 Esbjerg, Denmark. E-mail: usb@sam.sdu.dk

** Corresponding author: Gert Tinggaard Svendsen, Professor, PhD, Department of Political Science, Aarhus University, Bartholins Allé 7, DK-8000 Aarhus C, Denmark. E-mail: gts@ps.au.dk, website: www.ps.au.dk/gts

ISSN 0345-8555

Urs Steiner Brandt and Gert Tinggaard Svendsen

The Survival of the Nordic Welfare State and Social Trust[1]

Abstract

Why does free riding not escalate in the universal Nordic welfare state? How is it possible to maintain such a cooperative equilibrium where most people tend to cooperate? Our model suggests that the "missing link" is the accumulated stock of cooperation norms in terms of social trust. Arguably, a sufficient number of norm enforcers facilitate this unique collective insurance system.

JEL classification: D60, H11, P52, Z13

Keywords: Nordic welfare state, social trust, free rider, cooperation

[1] An earlier draft of this paper was presented at the European Public Choice Society Conference in Amsterdam, April 2007. We thank the other participants for helpful comments.

1. Introduction

In the following, we want to focus on the survival of the universal Scandinavian welfare state. How is it possible to maintain such a good equilibrium where most people tend to cooperate rather than free ride? Thus, our main research question is: Why does free riding not escalate and destroy the Scandinavian welfare state?

The backbone of the universal Nordic welfare model is the state as a dominant supplier of social services. Benefits tend to be defined at the individual level but with differences depending on the individual's labor market history. The main financial sources are various forms of taxes (Andersen, 2004). Overall, the universal Nordic welfare model provides adequate benefits for (almost) the entire population such as pension system, child allowances, home help, student allowance, education, unemployment benefits, early retirement benefits etc. As stated by Fong, Bowles and Gintis (2003) the modern welfare state is a remarkable human achievement where a substantial fraction of total income is regularly transferred from the better off to the worse off. Furthermore, publics regularly endorse the governments that preside over these transfers.

When the defining aspect of the universal welfare model is the fact that access to welfare arrangements is an individual right, then there is no direct linkage between entitlements and financing at the individual level: "Any insurance contract has moral hazard problems if the insurance is against contingencies that can be affected by the insured. In this respect, the attitudes and norms associated with receiving social assistance play a key role" (Andersen, 2004:751).

Hence, the insurance system only works if a strong norm prevents free riding. Andersen continues that: "Norms may change over time", which "… can lead to rapid changes in the scope for welfare policies by leading to unsustainable combinations of transfer levels and taxes." In fact, the "Change in norm may be an inherent consequence of the way in which the welfare system is organized" (Andersen, 2004:752).

The likelihood of a change in fundamental welfare norms is increased by the presence of the so-called "Samaritan's Dilemma". Three decades ago, James Buchanan (1975) argued that the altruism and the charity of the rich could induce the poor to act in socially sub-optimal ways to take advantage of transfers from the rich. For example, a poor person will undersave and get a perverse incentive to stay helpless in the current period if he/she anticipates that a rich person will bail him/her out in the future.[2]

If the fundamental welfare norm saying that everyone should contribute when capable is weakened over time, the Nordic welfare state would soon end up in a vicious circle where taxation of income eventually will reach 100 percent. People would stop trusting that other people in general contribute to the welfare state system if free riders are no longer stigmatized and minimized in number. Trusting that most people will contribute to the common pool of resources is a necessary but not sufficient condition for the provision of collective insurance systems and other collective goods.

Concerning related literature, it has been discussed why the Nordic welfare system works (Bjørnskov and Svendsen, 2007). A gap exists. Bjørnskov (2005a), for example, has reviewed the literature and concludes that the direction of causality remains an open question. Rothstein (2003) suggests that the institutional quality of the welfare state is the main reason. The more universal, uncorrupted and impartial the institutions responsible for the implementation of laws and policies are, the better it works. In contrast, Uslaner (2006) emphasizes culture as the main

[2] Amegashie and Kutsoati (2004), Coate (1995), Nannestad (2004), Green-Pedersen, Klitgaard and Nørgaard (2004: 10). Buchanan (1975) considered the increasing economic affluence as one possible explanation for the pervasive importance of what he called the "Samaritan's Dilemma" in twentieth century Western Society. Wealth makes it possible to choose "soft options", such as "kindness for criminals".

explanation and Putnam (1993) civic society, see Ostrom and Ahn (2009) for a comprehensive review of the literature

One way to sustain cooperation and avoid escalating free riding in the universal welfare state is via the possibility of punishment. Even though several motives for cooperation exist, Fehr and Schmidt (1999) argue that the single most effective mechanism to achieve cooperation is through the ability to punish: when players are given the opportunity to punish free riders, stable cooperation is maintained although punishment is costly for those who punish.

Moreover, Fehr and Fischbacher (2004) and Quervain et al. (2004) indicate that the possibility of punishment has a strong disciplining effect on free riders. This is, e.g., seen from comparing experiments on the ultimatum and dictator game. Rejection of positive offers in the ultimatum game is substantially lower when the game is altered so that rejection does not punish the proposer (Abbink et al., 2001). Moreover, offers generated by a computer rather than another person are significantly less likely to be rejected. This suggests that those who reject low offers at a cost to themselves are reacting to violations of fairness norms rather than simply rejecting disadvantageous offers.

When adding a second round to the prisoner's dilemma game where punishment is possible, cooperation also is sustained.[3] Experiments suggest that (some) individuals actually punish even though it is costly for themselves. This type of player is denoted "willing punishers" by Ostrom (2000). When punishment is also costly for the punisher, punishment will, however, only be executed when the punisher receives utility from punishing. Experiments by Quervain et al. (2004) on a trust game show that individuals derive "utility" from punishing norm violators. This enrichment of types of preferences gives a more realistic description of reality when adding the existence of both conditional cooperators and willing punishers. Such insight eventually adds to the understanding of how different societies come around.

Moving from individual preferences to interpersonal norms, a basic definition of a social norm is given as a behavioral regulatory based on socially shared beliefs of how to behave and how and when to punish deviators by informal social sanctions (Fehr and Gächter, 2000).

The emergence of such norms can be analyzed by evolutionary game theory, where the different types of preferences compete against each other. Here the assumption is that a social norm emerges when a sufficient fraction of the total population is of the same type. According to Kolstad (2003), when persons meet in an evolutionary game setting and play e.g. a prisoner's dilemma against each other, the players with highest payoff have a higher reproduction rate. In this setting, and given a population consisting of cooperators (people who always cooperate) and rational egoists (homo economicus), the cooperators will be defeated by rational egoists when different persons are matched randomly.

We show below that in a prisoner's dilemma game, cooperators can only survive (and even expand) if they can identify the type of the other players and then choose who to play against (so-called assortative matching) and if it is not totally possible to identify types then have an effective punishment available. Such a strategy is only possible if the cooperators are also willing to punish.[4] Together, this builds up social norms for cooperation either through an evolutionary process or through learned behavior.

After establishing that a norm of cooperation can be sustained in a group, the last step is to explain the move from this to emergence and sustainability of a welfare state. Here, we argue that social trust is the mechanism that can generate a welfare state. Social trust simply implies that when you meet a stranger, the stranger will reciprocate and thus facilitate the emergence of mutual trust. In other words, trust is defined as the subjective beliefs of an individual when he/she is playing cooperation that a stranger will respond by playing cooperation too. Thus,

[3] Experiments verify that many people are willing to use costly punishment against defectors, see e.g. Fehr and Schmidt (1999) and Abbink et al. (2001).

[4] Since punishment is costly, e.g., Trivers (1971) proposes that the emergence of feelings like guilt and shame gives an evolutionary advantage, which could give another explanation for the emergence of "conditional cooperators".

mutual trust implies that when playing with a stranger, the belief exists that the stranger is also a conditional cooperator.

Thus, our main finding is to suggest that the "missing link" in explaining this empirical puzzle could be the accumulated stock of social trust. This contribution is in line with Fehr and Gächter (2000). They argue that there is reason to believe that social norms are relevant for the amount of tax evasion and the abuse of welfare payments, and for attitudes towards the welfare state in general. Furthermore, they argue that: "In view of the powerful implications of reciprocity, it is also important to know why a sizeable fraction of the people has reciprocal inclinations. Which factors in the evolution of the human species have contributed to this?" (Fehr and Gächter, 2000:21).

The role of social trust in relation to the welfare state has, to our knowledge, not been modeled yet. This paper attempts to fill this gap. Almost all economic models assume that all people are exclusively pursuing their material self-interest and do not care about "social" goals per se. As e.g. Fehr and Schmidt (1999) state, this may be true for some (perhaps for many) people, but it is certainly not true for everybody. Experimental evidence shows that the traditional, selfish model is too limited (Schram, 2000). Crucial to the analysis of the appearance and sustainability of the welfare state is how to explain why players contribute to a welfare state as they essentially face incentives similar to the incentives present in a prisoner's dilemma game.

In order to explain why the Nordic welfare state works, we first introduce the concept of social trust. Second, we present a theoretical model concerning the reciprocity among agents that underlies the remarkable level of cooperation among strangers observed in the Nordic welfare state. Finally, we give a conclusion.

2. Social Trust

Social trust makes it easier to cooperate, for example a person's willingness to cooperate in a one-shot prisoner's dilemma game (Poulsen and Svendsen, 2005). As such, the presence of social trust determines how easily people who do not know each other can work together whereby transaction costs are lowered. Informal self-enforcement of contracts is now possible without third party enforcement.[5] A firm, for example, would lower its transaction costs by having numerous informal, i.e. not formally sanctioned, transactions taking place (see Coase, 1937). It is not necessary to monitor and enforce all transactions. In other words, more social trust in a country will, for example, make it easier for a citizen to do a transaction, do a trade, find a job, get access to resources etc.

Social trust may differ in different countries. For example, an early writer like Adam Smith (1997 [1766]) observed notable differences in trust across nations and found that the Dutch "are most faithful to their word". John Stuart Mill (1848) wrote: "There are countries in Europe … where the most serious impediment to conducting business concerns on a large scale, is the rarity of persons who are supposed fit to be trusted with the receipt and expenditure of large sums of money" (cited from Zak and Knack, 2001). Such differences in the level of social trust across countries survive today.

Fukuyama (1995:153) has defined trust in the following way: "Trust arises when a community shares a set of moral values in such a way as to create regular expectations of regular and honest behavior." Social trust can then be defined as the expectation that a stranger will follow the norm in question. In other words, social trust is the expectation that a stranger will not

[5] These transaction costs will always be positive when the agents do not possess full information. In order to support the exchange of goods and services in a world with incomplete information, the agents need to construct "rules of the game", i.e. institutions (North, 1990). These institutions can both be formal (rules written down) and informal (rules not written down) and both types matter to economic growth (cf. Schjødt and Svendsen, 2002). Oliver Williamson (1975), among others, has emphasized the role of formal institutions in reducing "transaction costs".

break the norm in question and cheat you even if there may be a private net gain from doing so. Social trust basically reflects the likelihood of being cheated.

The norm itself defines what action is right or wrong. Coleman writes that: "… a norm may prescribe certain actions, such as the norm that an athlete on a team should play his best." One could, for example, argue that the "work norm" in the primitive tribe would be to kill a mammoth or to gather many berries, whereas the welfare society prescribes that people, when capable, should get a job and contribute to the common pool of tax revenues from which welfare services are financed. Coleman continues: "It follows that if the person whose action is at issue acts in accord with the norm, he acts in a way that may be less to his immediate interest than if he disregarded the norm. For if this were not so, then his own choice, in the absence of the norm, would lead to the same action and the norm would be superfluous" (Coleman, [1987] 2003:138).

In line with Coleman (1988), Fehr and Gächter (2000) define a social norm as: 1) a behavioral regularity; 2) that is based on a socially shared belief of how one ought to behave; 3) which triggers the enforcement of the prescribed behavior by informal social sanctions. Note that this definition is very broad and does not say that the moral code needs to have a justification in normative ethics.

Ostrom (2000) defines social norms as shared understandings about actions that are obligatory, permitted or forbidden. Which norms are learned, however, varies from one culture to another, across families, and with exposure to diverse social norms expressed within various types of situations. The intrinsic cost that an individual incurs 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 as shame when others know about the failure.

A norm defined as social trust enables us to model why the bumblebee actually flies in spite of strong free rider incentives in the four Nordic welfare states. Social trust is measured as the percentage of a population answering yes to the question "do you think that most people can be trusted, or can't you be too careful?"

This approach was pioneered by the team behind the World Values Survey (Inglehart et al., 2004) and originally developed by Rosenberg (1956). This measure is arguably a good indicator for what it is intended to measure (see Bjørnskov, 2005b for further discussions of the validity of this indicator). Even though social trust does vary over time, it is rather stable across individuals and especially across countries. The mean change in trust is only 0.28 over a decade (Uslaner, 2002:252). Both Volken (2002) and Bjørnskov (2005b) find that the social trust scores are stationary over time too.

The social trust data in 86 countries are drawn from the four waves of the World Values Survey (Inglehart et al., 2004) and our ongoing Social Capital Project.[6] The main result is that the four Nordic welfare states indeed stand out as countries with much more social trust than the rest of the world. Especially Denmark, Norway and Sweden are outstanding with more than 60 percent of respondents trusting most people. Finland is fourth with 56 percent, closely followed by the Netherlands (confirming Adam Smith's observation about the Dutch back in 1766). These high scores presumably make the countries' economies run more smoothly and offer an explanation for the flight of the bumblebee.

Below this leading group, we generally find other Western European countries followed by a mix of Asian, Eastern European, African and South American countries. Brazil, Philippines, Costa Rica and Uganda hit the bottom with scores below 10 percent. The average for the 86 countries as a whole is 28 percent. Most notably, France (23 percent) and Portugal (16 percent) score low compared to their neighboring countries. In summary, the four Nordic welfare states top the list (Svendsen and Svendsen, 2010). Can they stay in this equilibrium? We try to give a theoretical answer to this question in the following section.

[6] Both China and Iran are excluded throughout; they are strong outliers in the trust surveys performed so far, see Uslaner (2002) and Bjørnskov (2004).

3. Model

The theoretical starting point is the classical result from a prisoner's dilemma game that cooperative outcome will never prevail. Therefore, a transformation of the original game is needed to increase the likelihood that players cooperate. In the following, we follow Poulsen and Svendsen (2005) and view cooperators in a one-shot prisoner's dilemma game as a proxy for social trust.

Consequently, we propose two such transformations; both regarding changes in the underlying behavioral assumptions of the players. First we consider that players are willing to punish (norm violators) defectors, even though this is also costly for the punisher, and secondly, we redefine players as being identical to a strategy. In this latter case, the proportion of players (strategies) in a population will be defined as a norm. (If only one strategy survives, this will be the prevailing norm).

As already discussed in the introduction, the possibility to punish is the single most important mechanism to achieve cooperation. We therefore present different ways of incorporating types of players who punish. This could be a situation where a defector is not totally certain about whether he/she might encounter a punishing player, or we could invoke an evolutionary approach where strategies compete and then analyze whether the presence of a punishing type will make the cooperators survive. Both approaches can explain the existence of cooperators and be interpreted to mean that the norm of cooperation survives.

In the following, we set up a model that can explain the existence and persistence of a social norm of cooperation. We consider a type of player, called a cooperator, as a carrier of social trust. The share of such cooperators in a population describes the level of social trust. In order to do this, and in general to analyze the appearance and presence of social norms is by evolutionary game theory, which provides the tools to analyze which strategies, or patterns of behavior, emerge over time through a process of adaptation. 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).

Since social norms are patterns of behavior with certain characteristics, we can use evolutionary game theory to examine the conditions under which these particular patterns called social norms emerge. Most important for our purpose is that it focuses on the process of how a specific equilibrium emerges, through an evolutionary process.

A common feature of the evolutionary models is that players are matched repeatedly to play a game, and a dynamic process describes how players adapt their behavior over time. In an evolutionary setting, players are modeled 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 (ESS). An evolutionary stable strategy (ESS) is based on two conditions: (1) an ESS $\underline{x}$ must earn at least as high payoffs against itself as any mutant strategy $\underline{y}$ against $\underline{x}$; (2) if a mutant strategy $\underline{y}$ does as well against $\underline{x}$ as does $\underline{x}$, then $\underline{x}$ must do strictly better against the mutant $\underline{y}$ than the mutant does against itself. Given a population of cooperators, consider the invasion of a defector. Condition (1) says that in order for the AC population to be stable against invaders, the payoff of AC playing against AC should be higher than AC playing against invaders, which is not true. So a small group of mutant defectors could invade a population of cooperators. Moreover, given that a population of defectors is stable against invasion of cooperators, since a cooperator does strictly worse against a defector than the defector does against itself, mutant cooperators cannot invade a population of defectors. As a consequence, a population that is randomly and repeatedly matched to play a one-shot prisoner's dilemma game thus ends up playing defect.

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. By applying the ESS to a prisoner's dilemma, it is easily shown that defection is the unique ESS (see Figure 1).
Assume as an example that the share x^i of strategy *i* will evolve according to the following replicator function:

$$\frac{\dot{x}_t^i}{x_t^i} = f(\pi_t^i - \pi_t^{Average}) \quad (1)$$

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 game (with restricted strategy spaces) from Table 1.

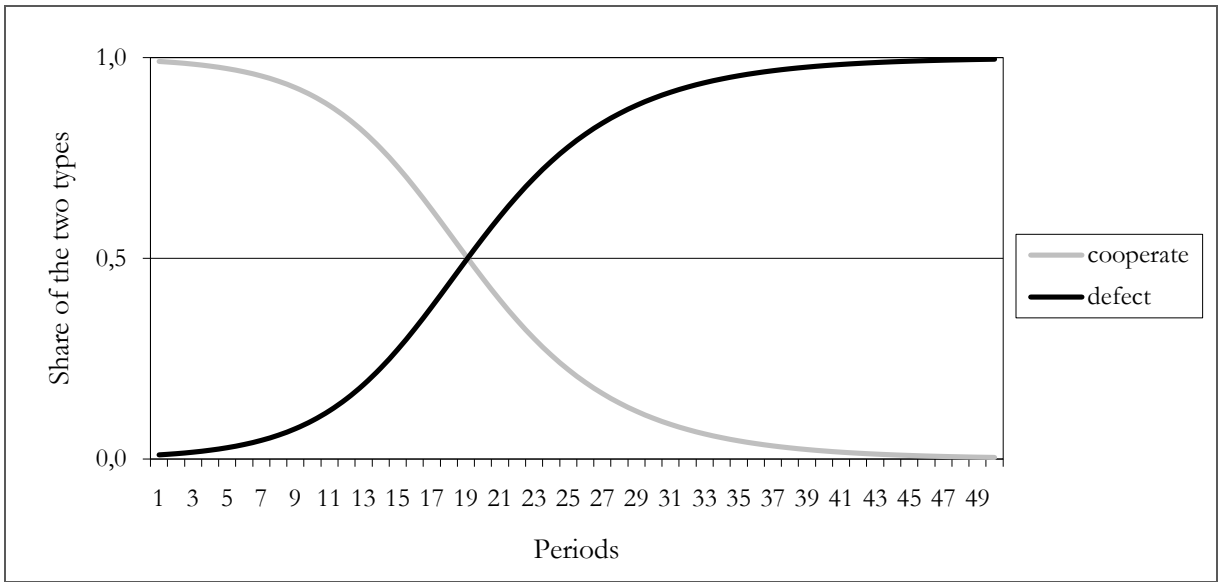
Table 1: The prisoner's dilemma (PD) game

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

For the following examples, let $\dot{x}_t^i = x_t^i \cdot 0.15 \cdot (\pi_t^i - \pi_t^{Average})$.

It is easy to show that if a population of cooperators is invaded by defectors, the defectors will eliminate the cooperators, shown in Figure 1, where a population of AC is invaded by few AD.

Figure 1: Invading defectors outcompete cooperators.



Note: Figure 1 shows how a population of cooperators is outcompeted by a small number of invading defectors when players are randomly matched to play a prisoner's dilemma game.

One key feature of a welfare state is that it reduces income inequality. As stated by Bowles, Fong and Gintis (2006), 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 worse off. Furthermore, the public regularly supports the governments that are in charge of these transfers.[7]

If we want to determine the stability of a welfare state in the framework of evolutionary game theory, that is, determine conditions where the persistence (survival) of cooperators is possible in the presence of defecting types (free riders), we need to be able to propose transformation of the original game such that cooperators can survive invading defectors. Several extensions imply that cooperators can also survive in the long run from the benchmark model of Figure 1.

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, which is called assortative matching. If it is possible for a cooperator to at least partially identify other cooperators, e.g., due to unconscious facial expressions, then cooperators might defeat invading defectors. Essentially, this amounts to saying that the probability that a cooperator meets another cooperator is higher than the share of cooperators in the population. Therefore, consider the following: Assume that the population share of cooperators is s, and let A be the probability that a cooperator meets another cooperator. (Assume further that we only are looking at a small part of the population, such that no matter the size of A, the defectors still face the same probability of meeting a cooperator or a defector). Let the players play the prisoner's dilemma game from Table 1. The expected payoffs to the two types are given by:

$$u(C) = A \cdot 4 + (1 - A)[s \cdot 4 + (1 - s) \cdot 1] = A(3 - 3 \cdot s) + (1 + 3 \cdot s) \quad (2)$$

$$u(D) = s \cdot 6 + (1 - s) \cdot 1 = s \cdot 5 + 1 \quad (3)$$

$$u(C) > u(D) \Rightarrow A > \frac{s \cdot 2}{3 - s \cdot 3} \quad (4)$$

As an example, if $s = 0.5$, then if $A > 2/3$, cooperators receive higher payoff than defectors. Most importantly, in this model, both types will survive also in the long run. Consider that $A = 2/3$, then when $s = 0.5$, both types receive the same expected payoff, and the population is stable.

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, mutant defectors can invade a population of cooperators, but cannot eliminate the cooperators. We can thus substantiate the evolution of a cooperative norm when matching is assortative, and explain the presence of both types in equilibrium.[8] Finally, the better cooperators can identify each other, the larger their share in a stable population.

Bowles, Fong and Gintis (2006) 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 behavior implies adherence to norms. If adherence to such norms implies a larger benefit to the group (like cooperation in a prisoner's dilemma) then the number (share) of strong reciprocal types is likely to increase.

The obvious second extension, therefore, is to include a reciprocal type. To see how the possibility of encountering a "non-strategic" player might change the behavior of a selfish

[7] According to an opinion poll from 2007, a majority of the voters in Denmark prefer more welfare to tax reduction.
[8] As Ostrom (2000: 149) puts it: "If a small core group of users identify each other, then they can begin a process of cooperation without having to devise a full-blown organisation with all the rules that they eventually need to sustain cooperation over time." In small groups with many encounters, a high level of assortative matching is possible, e.g. voluntary dairy movements and voluntary associations (Svendsen and Svendsen, 2004).

rational player, assume that a percentage of the population is what can be called a "reciprocal punisher" (RP).[9] If such types observe defection by other players, they play punish, no matter the personal costs. Moreover, they are fully reciprocal in nature, such that if they observe cooperation, they also cooperate. What does this imply for cooperation? Denote by F_p^e a rational player's expected likelihood of encountering a reciprocal punisher. Assume that encountering such a player is totally random, such that it is not possible to infer the type of the other player in advance. If a player defects against an RP type, then the outcome of this encounter is the payoff p. Define by $EP_i(s_i,\cdot)$ the expected payoff for player *i*, from playing the strategy s_i, when the probability of meeting a reciprocal punisher is F_p^e and the probability of meeting a rational player (= one that always defects in a prisoner's dilemma) is $(1 - F_p^e)$. We have that:

$$EP_i(C,\cdot) = F_p^e \cdot 4 + \left(1 - F_p^e\right) \cdot 0 = 4 \cdot F_p^e \tag{5}$$

$$EP_i(D,\cdot) = F_p^e \cdot p + \left(1 - F_p^e\right) \cdot 1 = 4 \cdot F_p^e \tag{6}$$

$$EP_i(C,\cdot) > EP_i(D,\cdot) \Rightarrow F_p^e > \frac{1}{5-p} \tag{7}$$

The higher the punishment or the greater the likelihood of encountering a reciprocal punisher, the more likely a player is to cooperate.

The implication here is that when the possibility of encountering a punisher exists, then otherwise selfish players will (under certain conditions) find it optimal to cooperate. It is worthwhile mentioning that altruistic punishment is costly, that is, although it must have a positive effect on cooperation, when called upon to punish, a reciprocal punisher will still punish. Therefore, if there were an "automatic understanding" that I should be cooperating, then this would yield an overall larger payoff.

The next step is to introduce the RP type into a population consisting of AC and AD and see if this implies that cooperation (seen as a social norm) can be sustained. Indeed, the introduction of a reciprocal punisher (RP) into the prisoner's dilemma might also change things considerably. Here, the RP-types are able to punish after observing the outcomes of the game. To convey the intuition, it is simply assumed that in the case of defection, this type invokes punishment costs of $p = -1$ (in the sense that both players in total receive –1 of the encounter). The (expected) payoffs to the three types of players are:

$$\pi_t^{AC} = x_t^{AC} \cdot 4 + x_t^{AP} \cdot 4 + x_t^{AD} \cdot 0 \tag{8}$$

$$\pi_t^{AP} = x_t^{AC} \cdot 4 + x_t^{AP} \cdot 4 + x_t^{AD} \cdot (-1) \tag{9}$$

$$\pi_t^{AD} = x_t^{AC} \cdot 6 + x_t^{AP} \cdot (-1) + x_t^{AD} \cdot 1 \tag{10}$$

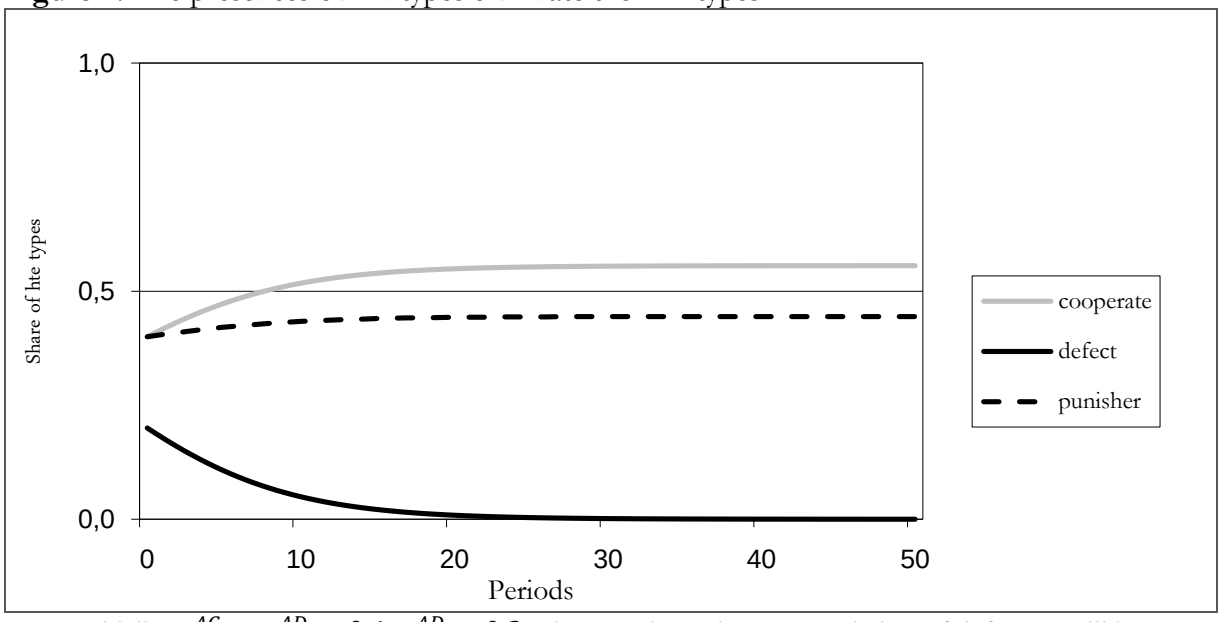
The presence of a sufficiently large share of RP-types implies that the payoff to the AD-types is lower on average than for the AC-types, which implies that the share of the AC-types increases.

[9] This term is found in Ostrom (2000); other authors call this type of player an altrusic punisher. Note that the type reciprocal punisher (RP) resembles the famous Tit-For-Tat (TFT) strategy, which turned out to be very successful in Axelrod (1980, 1984) tournaments. Bendor and Swistak (1995) argue that cooperative behavior supported by these strategies is the most robust evolutionary equilibrium: the easiest to attain, and the hardest to disrupt. The difference is that RP punishes more than TFT and has features of "self-punishment". Another difference is that in noise environments, situations without a one-to-one relationship between a move (strategy profile) and the resulting outcome due to some stochastic process, a strategy will perform better the more forgiving it is (Bendor, Kramer and Stout, 1991). See also Hoffmann (2000) for a thorough discussion of the robustness of Axelrod's results.

The RP-types receive less than the AC-types, but more than the AD-types, eventually eliminating the AD-types and creating an increased share of AC-types. This logic is shown in Figure 2.

The social norm of cooperation can 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.

Figure 2: The presences of PR-types eliminate the AD-types.



Note: Initially: $x^{AC} = x^{AD} = 0.4, x^{AD} = 0.2$. Figure 2 shows how a population of defectors will be outcompeted if the share of reciprocal punishers is initially sufficiently high compared to the share of cooperators.

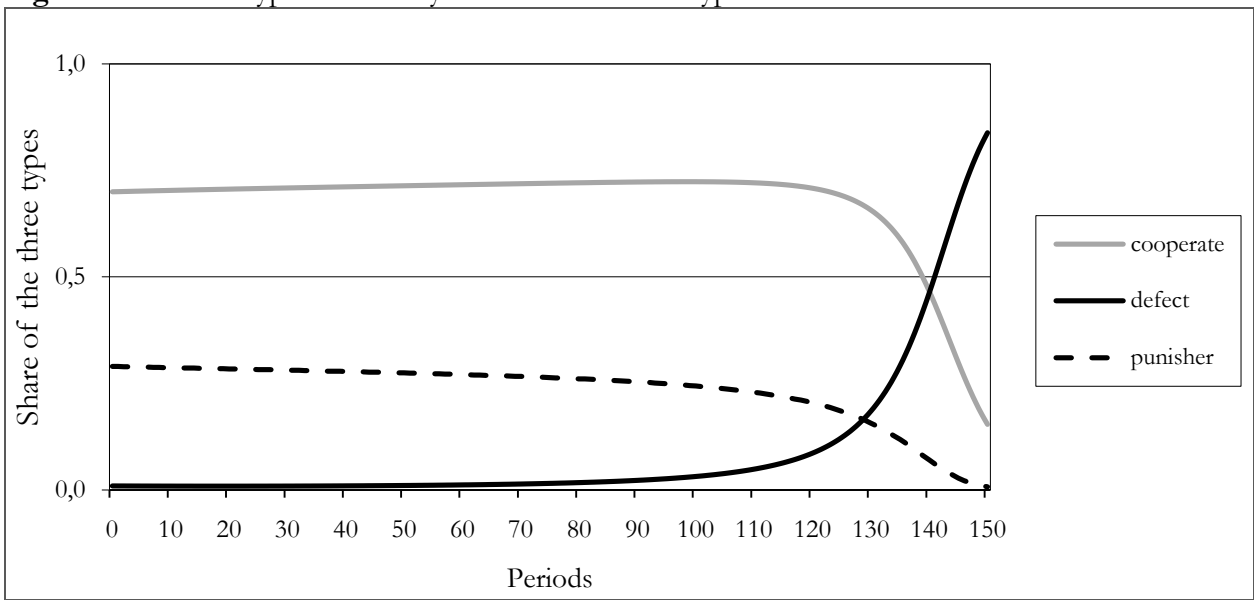
However the evolutionary approach can also be used to show the fragility of the welfare system. Essentially, small changes in the initial distribution of the share of the types can, around certain threshold values, change the outcome totally. This is shown in Figure 3, where the initial shares of types are changed compared with Figure 2. Here, the AD-types eventually can invade and defeat a population consisting of AC and RP-types. As seen, however, it might take a long time. In fact, it takes 112 rounds before the share of the AD-types is above 5 percent suggesting that the population is fairly stable over a long period of time followed by a rapid growth in the share of AD-types. If the initial distribution of the RP-types is reduced only marginally compared to Figure 3, then the AD will not be able to invade this population. The result from comparing Figures 2 and 3 is that it takes a sufficient share of reciprocal punishers to guarantee the welfare state against invading defectors, i.e. norm violators that free ride.

In perspective, certain points deserve to be emphasized. First, it is important to keep in mind that although the norm of cooperation survives, only some of these subjects are "reciprocators" who are prepared to reward friendly behavior and to punish unfriendly behavior even if this is costly to them. Second, whether or not appealing to reciprocal fairness works as enforcement device depends on the fraction of reciprocal types in the population and on the strategic situation in which the subjects interact.

In a broader context, and combining the insight in this section, if the RP cannot identify the type they encounter then they will cooperate if they estimate that the probability of being cheated is sufficiently low (compared to an individual threshold level). Such estimates can be labeled trust. Therefore, if sufficiently many reciprocal players have sufficiently high trust, then the norm of cooperation can be sustained. If on the other hand, when either the number of RP diminishes, and/or the number of AD increases (as seen in Figure 3), the level of trust in the

population diminishes, and at a certain level the population hits a tipping point and reaches a low trust outcome.[10]

Figure 3: The AD-types eventually eliminate the other types.



Note: Initially: $x^{AC} = 0.7, x^{AD} = 0.01,\ x^{AD} = 0.29$. Figure 3 shows how a population of cooperators will be outcompeted by defectors even in the presence of reciprocal punishers when the share of reciprocal punishers is low initially compared to the share of cooperators.

4. Conclusion

Our starting point was the empirical puzzle of why free riding does not escalate in the universal Nordic welfare states. We suggested that one possible explanation is the stock of social trust accumulated in the Nordic welfare states. Such an approach seemed relevant for the Nordic welfare states as they generally rely on the norm that people will work and contribute to the common pool of resources whenever possible rather than free ride and, e.g., stay dependent on social security transfers. Thus, we asked how it is possible to stay in a good equilibrium where most people tend to cooperate.

Our model suggested that cooperation with strangers is possible if a high level of social trust has accumulated. The theoretical results in relation to cooperation based on social trust may be grouped into three main findings. First, different forms of cooperation exist in prisoner's dilemma games, if preferences for fairness are present. Second, social punishments discipline potential defectors. If the share of punishers is high enough, the system can survive in the long run as in the Nordic welfare state. Our model showed that social trust can only survive in the presence of free riders (defectors) if there is sufficient willingness to use costly punishment in a punishment, that is willingness to sacrifice own material well being to punish norm violators (the reciprocal punishers). Overall, the theoretical model suggests that social trust may be one of the "missing links" in explaining the maintenance of the successful Nordic welfare state.

In perspective, if social trust really matters, it is important to understand not only how it is maintained but also how it is accumulated. Further research may try to move one step further

[10] Other types of strategies can be included. Ostrom (2000) assumes the existence of two types of "norm-using" players – "conditional cooperators" and "willing punishers" – in addition to rational egoists. The conditional co-operators as individuals are essential tit-for-tat players. They moreover tend to trust others and be trustworthy (and therefore exposed to exploitation by rational egoists) as long as a sufficient proportion of others who return trust is relatively high.

by attempting to trace the historical origins of social trust in the Nordic welfare state. Here, the direction of causality between the institutional setup and the observed level of social trust needs to be tested further through extensive and rigorous data analysis including time series analyses. Such insight will have wide-ranging implications. It can, for example, tell us how well-functioning welfare institutions from the Nordic welfare states may be successfully exported to developing countries that may wish to copy the universal welfare model.

References

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.

Amegashie, J. A. and E. Kutsoati, 2004. "Welfare Stigma and the Samaritan's Dilemma", Mimeo, University of Guelph.

Andersen, T., 2004. "Challenges to the Scandinavian Welfare Model", *European Journal of Political Economy*, 20:743-754.

Axelrod, R., 1980. "Effective Choice in the Prisoner's Dilemma", *Journal of Conflict Resolution,* 24:3-25.

Axelrod, R., 1984. *The Evolution of Cooperation,* Basic Books, New York.

Bendor, J. and P. Swistak, 1995. "Types of Evolutionary Stability and the Problem of Cooperation", Proceeding of the National Academy of Science, 92:3596-3600.

Bender, J., R. M. Kramer and S. Stout, 1991. "When in Doubt Cooperation in a Noisy Prisoner's Dilemma", *Journal of Conflict Resolution*, 35:691-719.

Bjørnskov, C., 2004. "Determinants of Generalized Trust. A Cross-Country Comparison", unpublished working paper, Department of Economics, Aarhus School of Business.

Bjørnskov, C., 2005a. "How Does Social Trust Affect Economic Growth? Exploring Potential Transmission Channels", unpublished working paper, Department of Economics, Aarhus School of Business.

Bjørnskov, C., 2005b. Investigations in the Economics of Social Capital, PhD Thesis, Aarhus School of Business.

Bjørnskov, C. and G. T. Svendsen, 2007. "Measuring Social Capital: An International Comparison", *Journal of Comparative Policy Analysis,* 9(3):275-292. Best JCPA paper 2007 reward.

Bowles S., C. Fong and H. Gintis, 2006. "Reciprocity and the Welfare State". In Mercier-Ythier, J. (ed.), *Handbook on the Economics of Giving, Altruism, and Reciprocity*, Volume 2, Elsevier, North-Holland.

Buchanan, J. M., 1975. "The Samaritan's Dilemma". In Phelps, E. S. (ed.), *Altruism, Morality, and Economic Theory,* Russell Sage Foundation, New York.

Coase, R. H., 1937. "The Nature of the Firm", *Economica*, 6:386-405.

Coate, S., 1995. "Altruism, the Samaritan's Dilemma, and Government Transfer Policy", *American Economic Review*, 85(1):46-57.

Coleman, J. S., [1987] 2003. "Norms as Social Capital". In Radnitzky, G. and P. Bernholz (eds), *Economic Imperialism: The Economic Approach Applied Outside the Field of Economics*, Paragon House Publishers, New York. Reprinted in Ostrom, E. and T. K. Ahn (eds), *Foundations of Social Capital*, Edward Elgar Publishing, UK.

Coleman, J. S., 1988. "Social Capital in the Creation of Human Capital", *American Journal of Sociology*, 94:95-120.

Fehr, E. and U. Fischbacher, 2004. "Third-party Punishment and Social Norms", *Evolution and Human Behavior*, 25:63-87.

Fehr, E. and S. Gächter, 2000. "Fairness and Retaliation: The Economics of Reciprocity", *The Journal of Economic Perspectives*, 14:159-181.

Fehr, E. and K. Schmidt, 1999. "A Theory of Fairness, Competition and Cooperation", *Quarterly Journal of Economics*, 114:817-868.

Fukuyama, F., 1995. *Trust. The Social Virtues and the Creation of Prosperity*, Hamish Hamilton, London.

Fong, C. M., S. Bowles and H. Gintis, 2003. "Reciprocity, Self-interest and the Welfare State". In Mercier-Ythier, J., S. Kolm and Louis-Andr'e (eds), *Handbook on the Economics of Giving, Reciprocity and Altruism*, Elsevier, Amsterdam.

Green-Pedersen, C., M. B. Klitgaard and A. S. Nørgaard, 2004. *Den danske velfærdsstat: Politiske, sociologiske og institutionelle dynamikker – En rapport til velfærdskommissionen*, Velfærdskommissionen, Copenhagen.

Hoffmann, R., 2000. "Twenty Years on: The evolution of Cooperation Revisited", *Journal of Artificial Societies and Social Simulation*, 3(2).

Inglehart, R., M. Basañez, J. Díez-Medrano, L. Halman and R. Luijkx, 2004. *Human Beliefs and Values*, University of Michigan Press, Ann Arbor.

Kolstad, I., 2003. "The evolution of social norms", CMI working paper 2003:1.

Mill, J. S., 1848. *Principles of Political Economy*, John W. Parker, London.

Nannestad, P., 2004. "Immigration as a challenge to the Danish welfare state?", *European Journal of Political Economy*, 20:755-767.

North, D. C., 1990. *Insitutions, Institutional Change and Economic Performance*, Cambridge University Press, Cambridge.

Ostrom, E., 2000. "Collective Actions and the Evolution of Social Norms", *The Journal of Economic Perspective*, 14:137-158.

Ostrom, E. and T. K. Ahn, 2009. "The Meaning of Social Capital and its Link to Collective Action". In Svendsen, G. T. and G. L. Svendsen (eds), *Handbook of Social Capital. The Troika of Sociology, Political Science and Economics*, Edward Elgar, Cheltenham, UK, Northampton, MA, USA.

Poulsen, A. and G. T. Svendsen, 2005. "Social capital and endogenous preferences", *Public Choice*, 123(1):171-196.

Putnam, R., 1993. *Making Democracy Work. Civic traditions in modern Italy*, Princeton University Press, Princeton.

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.

Rosenberg, M., 1956. "Misanthropy and Political Ideology", *American Sociological Review*, 21:690-695.

Rothstein, B., 2003. "Social Capital, Economic Growth and Quality of Government: The Causal Mechanism", *New Political Economy*, 8:49-71.

Schjødt, E. B. and G. T. Svendsen, 2002. "Transition to Market Economy in Eastern Europe: Interest groups and political institutions in Russia", *Nordic Journal of Political Economy*, 28(2):181-194.

Schram, A., 2000. "Sorting out the Seeking: The Economics of Individual Motivations", *Public Choice*, 103(3/4):231-258.

Smith, A., 1997 [1766]. "Lecture on the influence of commerce on manners". Reprinted in Klein, D. B. (ed.), *Reputation: Studies in the Voluntary Elicitation of good Conduct*, University of Michigan Press, Ann Arbor.

Svendsen, G. L. H. and G. T. Svendsen, 2004. *The Creation and Destruction of Social Capital: Entrepreneurship, Co-operative Movements and Institutions*, Edward Elgar, Cheltenham.

Svendsen, G. T. and G. L. H. Svendsen, 2010. "Social Capital and the Welfare State". In Böss, M. (ed.), *The Nation-State in Transformation: Economic Globalisation, Institutional Mediation and Political Values*, Aarhus University Press, Aarhus.

Trivers, R. L., 1971. "The Evolution of Reciprocal Altruism", *Quarterly Review of Biology,* 46:35-57.

Uslaner, E. M., 2002. *Moral Foundations of Trust*, Cambridge University Press, Cambridge.

Uslaner, E. M., 2006. "Where You Stand Depends upon Where Your Grandparents Sat: The Inheritability of Generalized Trust", working paper, University of Maryland.

Volken, T., 2002. "Generalisiertes Vertrauen: Zur Reliabilität und Validität eines verbreiteten Messinstruments in Wirtschaftssoziologisches Forschungszusammenhängen", Mimeo, University of Zürich.

Williamson, O. E., 1975. *Markets and Hierarchies: Analysis and Antitrust Implications*, Free Press, New York.

Zak, P. J. and S. Knack, 2001. "Trust and growth", *Economic Journal*, 111:295-321.