A Simple Model of Conflict

Sebastian Ille

Abstract This article develops a simple dynamic, non-symmetric game between two player populations that can be generalised to a large variety of conflicts. One population attempts to re-write a current (social) contract in its favour, whereas the other prefers to maintain the *status quo*. In the model's initial set up, the free-rider problem obstructs the occurrence of a conflict, leading to a low probability of a successful turn-over. The normative and conventional framework, in which players interact, plays however a vital role in the evolution of conflicts. By relating the individual pay-off perceptions for each strategy to the type and frequency of norm violations, the free-rider effect can be considerably weakened, thus enabling the model to predict the existence of two stable equilibria; one with a high rate of conflict, and another in which no conflict arises. This second equilibrium is caused by a *triggering event*. The model provides an explanation of how and why these events may occur and under which conditions they can be observed more frequently. In addition, it is also shown which factors influence the equilibria's basin of attraction, i.e. the likelihood of a transition and hence the probability of a conflict.

Keywords Social Conflict · Social Change · Evolutionary Game · Stability of Equilibria

JEL Classification C62 · C73 · D72 · D74

1 Introduction

On May 23rd, a group of protestant noblemen enters the Bohemian chancellery by force and defenestrates three protestant senior officials. Yet, all three miraculously survive the seventeen meter fall. Although the three imperial representatives of Bohemia are fortunate, "the Defenestration of Prague" in 1618 marks a fatal day in the European history. This incident triggers a war between Catholics and Protestants that will last for 30 years. It will have severe economic repercussions by drawing six countries into war; bearing the cost of four million lives and depopulating entire areas in the Holy Roman Empire. Some regions will require more than a century to recover economically. The thirty years war is only one of many historical examples that illustrate the impact of conflicts both on the stability of institutions and on economic development, and stresses the importance to understand the evolution of conflicts for the economic theory. The recent *Arab Spring* demonstrates once again that the strong socio-economic implications of large scale conflicts are still valid nowadays.

S. Ille
Laboratory of Economics and Management, Sant'Anna School of Advanced Studies, Piazza Martiri della Libert 33, 56127 Pisa (Italy)
Tel.: +39-050-883343
Fax: +39-050-883344
E-mail: sebastian.ille@sssup.it

Game theory has already been used to explain social phenomena since the middle of the 1950s (for a historical account refer to Swedberg 2001). At the beginning of the 1980's analytical Marxists used Game Theory as the mathematical means for a micro-founded explanation of social problems, structure and change, especially with respect to conflict and cooperation (Elster 1982; Cohen 1982; Roemer 1982b), class struggle (Przeworski and Wallerstein 1982; Holländer 1982; Roemer 1982a, 1985; Eswaran and Kotwal 1985; Mehrling 1986; Cole et al 1998), the stability of social contracts and classes(Starrett 1976; Roemer 1982c; Eswaran and Kotwal 1989; Axtell et al 2000; Bowles 2006), and trade union behaviour, i.e. bargaining models (Ashenfelter and Johnson 1969; Kennan and Wilson 1989; Kiander 1991; Clark 1996).

Yet, this "classical" literature has neglected the role of social conventions, norms and punishment, except for less sophisticated conflict games including variants of Prisoner's Dilemma, Battle of Sexes and Chicken Game (for an overview, see Binmore 1994; Gintis 2000, 2009 and for versions to model the effects of potential threats on cooperation, see Brams and Kilgour 1987a,b, 1988; Brams and Togman 1998). Of special interest in this respect has been the literature on the evolution of conventions following the idea of stochastically stable equilibria and conventional change by Young (Foster and Young 1990; Young 1993, 1998, 2005; Kandori et al 1993; Ellison 1993, 2000; Bergin and Lipman 1996; Morris 2000; Durlauf and Young 2001), and the literature concentrating on strong reciprocity and altruist punishment (Binmore 1998; Bowles et al 2003; Boyd et al 2003; Gintis et al 2005; Brandt et al 2006; Choi and Bowles 2007).

This article bridges the "classical" and "behavioural" literature by illustrating the interdependency between the strategic choice of players during conflicts and the perceived violations of social norms and conventions. This simplified model of social conflict is able to provide analytical proof of some of the intuitions that are observable in real-world conflicts. The first section starts with a game between two player sub-populations, in which a conflict does not occur because of the free-rider effect, though a social conflict and the subsequent change is mutually beneficial for one player sub-population and collective action should result in such a change. The second section develops an analytical representation of the dynamics of emotional violence during a conflict situation that arises from norm violations. The third section then incorporates the "emotional component" into the original model, making social conflict possible under certain conditions. The sixth section constitutes the conclusion and an outlook for future research.

2 The Basic Model

Let there be a game $\Gamma = (S_1, S_2, .., S_{m+n}; \pi_1, \pi_2, ..., \pi_{n+m})$ played in a finite but large population N with players $i = 1, 2, ..., n+m$. Game Γ is thus defined by a set S_i of pure strategies for each player i. Given player i's strategy s_i and for each pure strategy profile $s = (s_1, s_2, ..., s_{n+m})$, in the set of pure strategy profiles $S = \times_i S_i$, the associated individual pay-off to player i's strategy choice is defined by $\pi_i(s) \in R$, implying that $\pi_i : S \to R$ for each player i. Further assume that two distinct sub-populations C_A and C_B, with $C_A \cap C_B = \emptyset$, $C_A \cup C_B = N$, participate in Γ. In addition, suppose that there are n players in C_A and m players in C_B and $n >> m >> 0$.[1]

For simplicity denote each individual player in C_A as A and in C_B as B. Thus, assume that all players in the same sub-population have an identical pure strategy set and pay-off function for each strategy. For an $A \in C_A$ assume that $S_A = \{R, C, S\}$, i.e. each player A has the choice between *(R)*evolting, *(C)*onforming to the current system, or *(S)*upporting it. Revolting implies an active action to change the current (social) contract written between C_A and C_B. Conforming denotes a strategy of "inaction"; a player waits for the other players to act. Supporting is diametric to revolting; a player approves of and actively supports the current social contract.[2] For $B \in C_B$ assume that $S_B = \{P, \neg P\}$, i.e. the player can choose whether or not he provides a bonus payment to the supporters in C_A. (This payment does not necessarily define a direct monetary benefit, but can also be considered as workplace or social amenities, an easier career in a firm controlled by a B, a better reputation or higher social status among the Bs.)

[1] The last assumption is not strictly necessary, but simplifies the model. It could also be assumed that the winning probability is influenced by the relative sub-population sizes. Yet, redefining the parameter values, should have a similar effect.

[2] The Thirty Years' War shall again serve as an example: In the events following the "the Defenestration of Prague" R-players are comparable to the Hussites of Bohemia, S-players are mercenary soldiers supporting Ferdinand of Habsburg, and C-players are mostly peasants.

Suppose that in the current state and with respect to the currently prevalent social contract, there exists an alternative allocation that constitutes a redistribution detrimental for C_B, yet favourable for As playing strategy R or C. This does not necessarily hold for As playing S, especially when paid by the Bs. Consequently, all players in C_B have an interest in maintaining the current (social) contract and in preventing the implementation of the alternative allocation. Strategy R and C players in C_A, on the contrary, benefit from a successful *revolt* that leads to the realisation of the alternative allocation.

To derive the individual pay-off function $\pi_i(s)$ for each strategy and player type, first concentrate on the As by labelling the frequency of revolutionaries in C_A as x, the frequency of supporters as y, and the frequency of conformists as z, for which it must hold $1 \equiv x+y+z$ (with each frequency lying within the unit interval). The additional pay-off derived from the alternative allocation after a successful turn-over is defined by a positive constant δ^r. Assume further that if the attempt of revolution failed, revolutionaries face a negative pay-off defined by a function decreasing in the share of supporters, as those will impose the punishment. For simplicity assume that this is the linear function $\delta^p y$, with δ^p being a negative constant. This negative pay-off can be through death, punishment, imprisonment, social shunning, discrimination, or mobbing etc. This cost is absent if the player has chosen strategy C. (Thus, this model does not take account of collateral damage or second-order punishment. This circumstance can be easily implemented into the model by adding an additional cost to the conformist strategy. As long as it is smaller than $\delta^p y$ the general dynamics of the original model should persist.) Hence, the expected pay-offs of revolutionaries and conformists are equal to

$$\begin{aligned} \pi_i(s_i, s_{-i}|s_i = R) &= P(win|s_i = R)\delta^r + (1 - P(win|s_i = R))\delta^p y \\ \pi_i(s_i, s_{-i}|s_i = C) &= P(win|s_i = C)\delta^r \end{aligned} \tag{1}$$

where $P(win|s_i)$ defines the probability of realising the alternative allocation, if player i chooses s_i, given the strategies s_{-i} of all players other than i.

To derive the functional form of probability P, suppose that in order to engage in a conflict, groups are formed at random from players in C_A, but that conformists never actively join a group. This reflects the general situation prior to a conflict. People congregate to discuss new labour contracts, to meet at a summit, to protest on the streets, to rally forces for battle or covert assaults etc., without exact prior knowledge of who will participate. Supporters, on the contrary, *join* the group to "sabotage a revolutionary attempt", e.g. in the form of police forces, the opposing battalion, the members supporting the counter-faction in the summit.

Since groups are formed at random, a player cannot tell the exact group's composition prior to his choice to participate. He does not know how many players actively engage in the conflict and how many of these will choose strategy R or S. His expected pay-off, however, depends on the frequency, with which each strategy is played in sub-population C_A, since groups are assumed to be defined by an unbiased sample.[3] Consequently, expected group size is determined by the expected number of individuals that join the group, i.e. that are not conformists. In this case, the probability of being in a group of size $s \leq n$ is simply the probability of finding $s-1$ other individuals playing a strategy different from conforming. Since population size n is expected to be large, the hypergeometric distribution is approximated by the easier binomial distribution in order to keep the system tractable. We obtain that the probability of being in a group of size s is:

$$\binom{n-1}{s-1}(1-z)^{s-1}z^{n-s} \tag{2}$$

for both strategies S and R. Group size is thus determined by the frequency of conformists, whereas the composition of a group of size s is only defined by the frequencies of revolutionaries and supporters. Suppose that

[3] Notice that the replicator dynamic, which will be applied later on, does not require a player to know or form expectations about the frequencies, with which each strategy is played, since strategy choice is defined by imitation. Furthermore, the assumption of an unbiased sample excludes that assortment takes place according to a player's strategy, e.g. revolutionaries are not more likely to meet other revolutionaries than supporters or conformists.

each revolutionary adds a marginal unit to the probability that the group revolts successfully, but that this marginal additional unit of probability diminishes in the number of supporters. Assume that this marginal unit is simply $1/n$ minus a constant weighted by the share of supporters.

If a player chooses strategy R, a group of size s can include 1 (only himself) to s (all) revolutionaries. Let τ be the number of revolutionaries in a group. The probability of drawing $\tau-1$ other revolutionaries is $(x/(x+y))^{\tau-1}$ and the probability of drawing the remaining $s-\tau$ supporters is $(y/(x+y))^{s-\tau}$. Furthermore, the share of supporters in a group of size s with τ revolutionaries is equal to $(s-\tau)/s$. If we assume that the marginal negative effect of a supporter in such a group is a, we obtain

$$G_R(s) = \frac{1}{n}\sum_{\tau=1}^{s}\binom{s-1}{\tau-1}\left(\frac{x}{x+y}\right)^{\tau-1}\left(\frac{y}{x+y}\right)^{s-\tau}\tau\left(1-a\left(\frac{s-\tau}{s}\right)\right) \tag{3}$$

The first part defines the expected composition of a group of size s, the second the marginal effect of the revolutionaries minus the marginal effect of supporters on the winning probability of a group of size s. Thus, $G_R(s)$ denotes the probability, with which a group of size s can impose the alternative allocation. It must hold $a \in (0,1)$ for the probability to be restricted to the unit interval. If $a = 0$ supporters have no additional negative effect on the winning probability, except for their inaction (i.e. they do not contribute to the revolutionary attempt). Given $a = 1$, we observe that the marginal effect of a revolutionary is negligible, if the group is composed by a high number of supporters.The former equation 3 can then be placed into equation 2 for the expected group size and determines the probability of imposing the alternative allocation:

$$P(win|s_i = R) = \sum_{s=1}^{n}\binom{n-1}{s-1}(1-z)^{s-1}(z)^{n-s}\left(G_R(s)\right) \tag{4}$$

Since such conflicts oftentimes imply large populations, equation 4 is thus approximated by

$$\lim_{n\to+\infty} P(win|s_i = R) = x - \frac{axy}{x+y} \tag{5}$$

Consequently, for $a = 0$ the probability is simply $P = x$. If $a = 1$ then $P = x^2/(x+y)$, implying increasing returns to scale, i.e. the winning probability increases quadratically in the share of revolutionaries for a given share of conformists.

The probability for a conformist can be derived in the same way by adapting the possible compositions and group sizes to his strategic choice. Since a conformist does not join a group, a group of size s can be composed of 0 to s revolutionaries. For τ revolutionaries in a group of size s, we need to draw τ times a revolutionary, each with probability $x/(x+y)$ and $s-\tau$ supporters, each with probability $y/(x+y)$. This gives

$$G_C(s) = \frac{1}{n}\sum_{\tau=0}^{s}\binom{s}{\tau}\left(\frac{x}{x+y}\right)^{\tau}\left(\frac{y}{x+y}\right)^{s-\tau}\tau\left(1-a\left(\frac{s-\tau}{s}\right)\right) \tag{6}$$

In the presence of a conformist, group size ranges from 0 to $n-1$, and a conformist *draws* s revolutionaries or supporters with probability $(1-z)^s$, and $n-1-s$ other conformists. Consequently, placing the results of equation 6 into the adapted equation 2 gives

$$P(win|s_i = C) = \sum_{s=0}^{n-1}\binom{n-1}{s}(1-z)^{s}(z)^{n-1-s}\left(G_C(s)\right) \tag{7}$$

and for large n

$$\lim_{n\to+\infty} P(win|s_i = C) = x - \frac{axy}{x+y} \tag{8}$$

Comparing equation 5 and 8 shows that for large populations individual strategy choice is irrelevant.

Further let $\pi_i(s_i, s_{-i}|s_i = R) = \pi_R$ and $\pi_i(s_i, s_{-i}|s_i = C) = \pi_C$ for notational simplicity. Putting in the previous results into equation 1 on page 3 yields the expected pay-off function for both strategies:

$$\pi_R = \frac{x(x+y-ay)}{x+y}\delta^r + \left(1 - \frac{x(x+y-ay)}{x+y}\right)\delta^p y$$
$$\text{and} \tag{9}$$
$$\pi_C = \frac{x(x+y-ay)}{x+y}\delta^r$$

Since δ^p is a negative constant, strategy C weakly dominates strategy R for sufficiently large player populations ($n >> 0$) and strictly dominates in the presence of at least one supporter.

Assume that all those players, who choose strategy S are subsidised by a bonus payment from sub-population C_B. They are paid d by the share w of Bs, who are playing strategy P. Since the bonus is assumed to be be shared by the supporters of the group, d increases with the share of revolutionaries in the group. Strategy S, however, incurs an additional cost. Defending the current contract against revolutionaries can be risky and we may assume that this cost is highest if the struggle against revolutionaries is the hardest. This is the case if supporters and revolutionaries meet with equal strength. In addition, we might assume that the conflict escalation is stronger the larger a group. Furthermore, the individual benefit from choosing this strategy can take the form of general acknowledgement, promotions, social honours etc. Hence a supporter has an interest that the population does not consist of too many supporters. It should be hence assumed that expected pay-off is decreasing in the population share of supporters and that the cost function is highest if the group is equally split between revolutionaries and supporters.

The share of revolutionaries for a group of size s is given by τ/s, and the share of supporters in the entire sub-population C_A is $(s-\tau)/n$. Given the former assumptions, assume that expected pay-off of a supporter is given by

$$G_S(s) = \sum_{\tau=0}^{s-1} \binom{s-1}{\tau} \left(\frac{x}{x+y}\right)^{\tau} \left(\frac{y}{x+y}\right)^{s-1-\tau} \left(wd\frac{\tau}{s} - c\left(\frac{\tau}{s}\frac{s-\tau}{n}\right)\right) \tag{10}$$

The value in equation 10 is only indirectly related to the expected probability of a successful revolt, and defines the expected pay-off of an S-player in a group of size s. Setting in into the equation 2, and defining $\pi_i(s_i, s_{-i}|s_i = S) = \pi_S$ gives

$$\pi_S = \sum_{s=1}^{n} \binom{n-1}{s-1} (1-z)^{s-1}(z)^{n-s}(G_C(s)) \tag{11}$$

For $n \to \infty$ the expected pay-off of an individual choosing strategy S is

$$\pi_S = \frac{x(dw-cy)}{x+y} \tag{12}$$

Given that there exists at least one supporter, it must hold that $x = 0$, since strategy C strictly dominates R, and $\pi_C = \pi_S = 0$ in this case. Hence, all points in $y+z=1$ are Lyapunov stable equilibria that cannot be invaded by revolutionaries, but perturbations, i.e. random drift, are not self-correcting. Any point at $y = 0$ (no supporters) provides equal pay-off to revolutionaries and conformists, yet in the presence of at least one supporter the pay-off of the former is strictly smaller than the one of the latter.

To define under which conditions a B will decide to provide the bonus payment to the supporting As, let again w be the share of players in C_B choosing strategy p and, hence $1-w$ be the share of those not paying supporters. Assume that the benefit of paying supporters increases with the share of revolutionaries. We might consider for example a small number of security men that protect a factory owner against his revolting labourers protesting for higher loans. Yet, the higher the numbers of protesters the higher the need for protection. In addition, the conflict between revolutionaries and supporters creates collateral damage to the detriment of the Bs, meaning that if revolutionaries are present in great numbers, it is best to be amongst those that less strongly resisted the attempt. Assume that benefits from protection increase linearly in x, but fall quadratic in x. Hence,

let benefits be described by $bx - bx^2$, with $b > 0$. Let there also be a spillover effect on those not paying the supporters that increases linearly in the share of payers w. Protests are not localised and a probability of spreading violence to other groups exists. Hence, Bs, who choose to *pay*, have also an incentive to "suppress" revolutionaries, when they are only indirectly concerned.

Furthermore, let there be a cost k that also increases with the share of non-payers in a group, since some economies to scale arise, e.g. two factory owners share security personnel, they also share costs. Assume, however, that costs also increase in the number of payers. The underlying idea is that as the number of payers increases, the complexity to set up a security system that satisfy all payers at the same time becomes more complex and thus costly. For example, if a larger number of factory owners invests in security personnel, some amateur security men, who are only recruited in the case of incidents, are substituted by well-trained professional security personnel. It is necessary to set up an infrastructure sufficient to guarantee quick information exchange and access to the various factory sites. Hence, the cost reduction from economies of scale are initially offset by these set-up costs. Non-payers do not bear such costs, but suffer a negative pay-off e by those who chose to pay. e signifies the negative effect of social shunning that increases in the number of payers. The more player choose to pay, the more social pressure is exercised on non-payers. For a player population of size m and ρ being the number of payers, the expected pay-off functions for each strategy are then determined by

$$\begin{aligned} \pi_P &= \sum_{\rho=1}^{m} \binom{m-1}{\rho-1} (w)^{\rho-1} (1-w)^{m-\rho} \left((1-x)xb - k\left(\frac{m-\rho}{m} \frac{\rho}{m} \right) \right) \\ \pi_{\neg P} &= \sum_{\rho=0}^{m-1} \binom{m-1}{\rho} (w)^{\rho} (1-w)^{m-1-\rho} \left((1-x)xb \frac{\rho}{m} - e\frac{\rho}{m} \right) \end{aligned} \tag{13}$$

where the first part of both equations defines the expected composition, and the second part (in brackets) the net pay-off of each strategy. Again for $m >> 0$, approximating by $m \to \infty$, gives

$$\begin{aligned} \pi_P &= bx(1-x) - k(1-w)w \\ \pi_{\neg P} &= w(bx(1-x) - e) \end{aligned} \tag{14}$$

Solving 14 shows that the set of equilibria consists of four components; two interior and two pure equilibria. The interior equilibria are given by the two roots, at which both strategies have equal pay-off, namely

$$w^{*1} = \frac{1}{2k}(k + bx - e - bx^2 - \sqrt{(e-k-b(1-x)x)^2 - 4bk(1-x)x}) \tag{15}$$

and

$$w^{*2} = \frac{1}{2k}(k + bx - e - bx^2 + \sqrt{(e-k-b(1-x)x)^2 - 4bk(1-x)x}) \tag{16}$$

under the constraint that $w^{*1}, w^{*2} \in (0,1)$; the former defines the stable interior equilibrium (henceforth called the low w-equilibrium), the latter provides the unstable equilibrium and frontier between the basin of attraction of the stable pure equilibrium defined by $w^{*3} = 1, \forall x \in (0,1)$ (henceforth called the high w-equilibrium) and the low w-equilibrium.

The lower set of equilibria illustrates that below a certain threshold of w, in the absence of any revolutionaries, no B has an incentive to pay, since the only motivation is provided by the social shunning of *payers*. Similarly if all As choose to revolt, a single B faces a situation, in which he only pays his marginal costs without any benefit, but can evade the costs from social shunning if he did not pay. Hence, also no incentive to pay arises. For an intermediate share of revolutionaries an incentive exists tor protecting his property. If a sufficient number, however, pays supporters the spill over is adequate, i.e. it compensates for the lack of a direct net benefit from paying supporters and the cost of social shunning, and B will have no motivation to play strategy P. On the contrary, in the high w-equilibrium at which all Bs play P, it is also always best response given any x to pay because of cost e.

If the conflict described by the current model is not considered to be defined by a single event but a sequence of repeated events, individuals are able to dynamically re-evaluate their strategy choice between events (as for example, during the Monday demonstrations in the GDR in 1989 or the skirmishes during the Thirty Years' War). Since players meet at random in a group, they might feel unsatisfied with their current strategy and re-assess. In this case, a player compares his strategy choice to a random "model" player. The player adopts the strategy of his "model" with a probability proportional to the positive difference between their model's and their own pay-off, which have been generated in the course of this event (if the difference is negative, i.e. if a player has chosen a better strategy than his model, the player will not switch). Hence, if a player observes another player faring much better with another strategy, he changes strategies less likely. If the model player only received a marginally higher pay-off, the player will less likely to switch. Notice that if an *S*-player observes an *R*-player, who obtains higher pay-off, he will adopt this strategy with a positive probability, though strategy *R* is strictly dominated by *C*. Nevertheless, more *S*-players will switch to strategy *C* than to strategy *R*, since the pay-off difference between players choosing *C* and *S* is greater than between those choosing *R* and *S*.

The dynamics based on this type of assumption are best described by the replicator dynamics (see also Bowles 2006). The replicator dynamics are generally defined by $\dot{\sigma}_i = \sum_j \sigma_i \sigma_j (\pi_i - \pi_j)$, where σ_i denotes the frequency of strategy/trait i (i.e. in this case x, y or z) in the population, see Taylor and Jonker (1978); Taylor (1979); Schuster and Sigmund (1983), and especially Hofbauer and Sigmund (1988); Weibull (1995); Nowak (2006)). Hence, $\sigma_i \sigma_j$ defines the probability, with which a player of trait σ_i (i.e. a player choosing the strategy associated to σ_i) meets another player of trait σ_j. The switching probability is a multiple of the pay-off difference $\pi_i - \pi_j$. For the limited strategy set, the replicator dynamics can be simplified to

$$\begin{aligned} \dot{x} &= x(\pi_R - \phi) \\ \dot{y} &= y(\pi_S - \phi) \text{ and} \\ \dot{w} &= w(1-w)(\pi_P - \pi_{NP}) \end{aligned} \tag{17}$$

where ϕ denotes the average pay-off of the players in C_A being defined by $\phi = x\,\pi_R + y\,\pi_S + z\,\pi_C$. The pay-offs are determined by equations 9, 12, and 14. The dynamics with respect to any distribution of the players in C_A are illustrated in figure 1 on page 7, where $\Delta\sigma_i \equiv \pi_i - \phi$. Consequently, strategy σ_i is stationary at $\Delta\sigma_i = 0$. Solving gives 2 roots for each variable x, y and z, yet the figure only show those within the unit interval, i.e. $x + y \in (0,1)$. The vectors thus indicate the direction of movement relative to these loci. A stationary (equilibrium) state is then defined by those states in which each strategy is either stationary or absent.

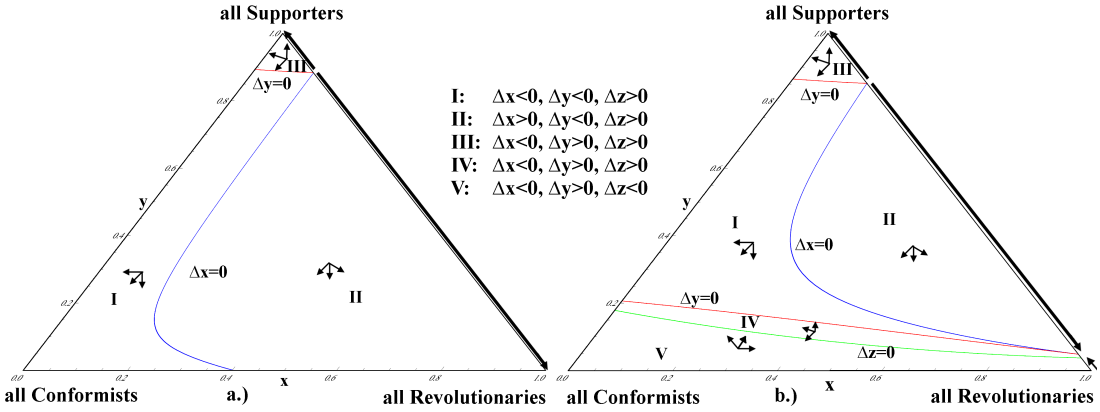

Fig. 1 Projection of the unit simplex and the dynamics for given w: A.) $w = w^{*1}$, B.) $w = 1$, Parameters: $a = 1$, $\delta^r = 4$, $\delta^p = -3$, $d = 5$, $c = 25$, $b = 2$, $k = 5$, $e = 0.5$

In case a.) any state on the x-axis with $y = 0$ is stationary and cannot be invaded by supporters. Notice, however, that *C* weakly dominates *R*. If non-best replies occur infrequently, evolutionary drift will push any population towards the left along the simplex axis, and the number of revolutionaries will converge to zero. Any point on the y-axis with $x = 0$ is also stationary and invasion by revolutionaries is impossible. Small perturbations of non-best response play will, however, push a population back to the equilibrium $z = 1$.[4] The unstable equilibrium for $z = 0$ is defined by point $(0.12, 0.88)$; the former value denotes the share of revolutionaries, the latter the share of supporters.[5]

In case b.) the situation is somewhat different. Any perturbation a point on the x-axis in the direction of *y* will push the population out of region *V* into region *IV*, in which the population converges to the y-axis but remains in this region. In addition to the higher mixed equilibrium at $(0.12, 0.88)$ a second mixed equilibrium exists at $z = 0$ and point $(0.95, 0.05)$. This is also unstable as z-perturbations will again push it into region *IV*. (This implies that the equilibrium can be destabilised by random mutation, but not by imitation.)[6] In the case of infrequent non-best response play the long term equilibria will be defined by the line segment on the *y*-axis of region *IV*. The model is therefore inadequate to efficiently model conflicts. Conflicts would occur with quasi null probability, even if there is little incentive to support the current (social) contract and if a favourable alternative allocation exists. A conflict would require a large number of non-best responses that push the population from the y-axis into region *II*. Further, we observe that all equilibria are only defined by two strategies. A population in a completely mixed equilibrium is unobservable, as the model neglects specific characteristics innate to conflicts. One important aspect is the emotional reaction to the violation of social norms.

3 The Effect of Social Norms

The literature on behavioural and neuro-economics (see for example Frohlich et al 1987b,a; Cooper et al 1992; Rabin 1993, 1998, 2002; Smith 1994; Fehr and Gächter 2000; Camerer and Loewenstein 2002; Camerer et al 2005) has shown that fairness is a complex concept. In general a *fair share* is determined by a reference point. An individual's evaluation of whether an interaction is fair or not is determined by past-interactions, status, expectations etc.. These elements define a reference framework that can be collapsed into a set of social norms and conventions, which govern everyday interactions. Whenever a social norm is violated, individuals feel unjustly treated. Hence, a simple individual pay-off comparison as in Fehr and Schmidt (1999) is only part of the story.[7] Fair does not necessarily mean equal. As Binmore writes: "A person's social standing, as measured by the role assigned to him in the social contract currently serving a society's *status quo*, is therefore highly relevant to how his worthiness is assessed by those around him". (Binmore 1998, p. 459; see also variants of the Ultimatum Game, in which *contest winners* successfully offer lower shares than in traditional version of the game, such as Hoffman and Spitzer 1985; Frey and Bohnet 1995.)

The way in which social norms, and culture in general, are determined is as vague as the concept of fairness. Attempts have been made in the past to model the evolution of social and cultural norms, yet, not only each social/human science has its own definitions of culture, but these definitions are dependent on the current Zeitgeist (Geertz 1987). Therefore no unique normative basis exists. This renders a precise description of a social norm on an analytical basis very difficult and only few papers incorporate norms into their mathematical

[4] Notice that evolutionary drift might push the population to both extremes on the y-axis, namely $y = 1$ or $z = 1$, depending on the parameter values. The dynamics of the low *w*-equilibrium are determined by $\partial\pi_S/\partial x|_{x=0} - \partial\pi_C/\partial x|_{x=0} = -(d\,(e-k+|e-k|) + 2(c-(a-1)\delta^r)ky)/(2ky)$, which is, given the parameters, strictly negative, but the inverse may occur for low values of *c*, and high values of *a* and δ^r. For $w = 1$ the dynamics are determined by $\partial\pi_S/\partial x|_{x=0} - \partial\pi_C/\partial x|_{x=0} = d/y - c + (a-1)\delta^r$, hence the relative marginal effect of a mutant revolutionary on a supporter decreases with respect to the marginal effect on a conformist as *y* increase, leading to the stability in region *IV*.

[5] In addition to the graphical solution using the *zero loci*, the eigenvalues of the system's Jacobian define the stability of the point in the simplex. Both eigenvalues at this equilibrium are positive, defining thus an unstable fixed point.

[6] We observe that the eigenvalues of the system's Jacobian are positive for the former equilibrium and have alternate signs for the second.

[7] Along with the classic solutions provided by Nash (1950) and Raiffa (1953).

models (see e.g. Bernheim 1994; Lindbeck et al 1999, 2002; Huck et al 2001, 2003). It is even unclear whether our actions are determined by global norms or rather by highly local norms (Patterson 2004).

To circumvent this issue, it is sufficient to notice that it is indeed not strictly necessary to explicitly model social norms, since revolutionary behaviour is defined by an aggressive reaction to the *violation* of social norms. Hence, assume that a certain set of social norms, not closer specified, exists and that whenever a norm is violated, it gives rise to an aggressive feeling. This assumption keeps the model general enough, so that it is applicable to a large variety of conflicts, and norms and conventions. There are two aspects that the model should take into account:

Lorenz (1974) has illustrated that, though aggression is immanent to every species and to most interactions, it is nevertheless defined by a high level of ritualisation that minimises the frequency of direct hostile conflicts, and the potential cost they would incur. Furthermore, humans illustrate a general tendency to avoid hostility even in extreme situations. To include this into the model, assume that a general "non-aggression norm" exists that dampens violent responses for all kinds of arising conflicts. On the one hand, the stronger the norm, the less likely an individual will choose a violent response. On the other hand, the more violence an individual observes, the more likely he will also respond aggressively, and the lower the dampening effect of the non-aggression norm.

The formerly mentioned example of *the Defenestration of Prague* illustrates the second important aspect of conflicts. A single event suddenly triggers a conflict, though the actual reasons are much more complex and occur over a longer period of time. In 1555, the Peace of Ausgburg warranted the peaceful coexistence of both confessions. The existential fear created by the small ice age and catholic aggressiveness led to a seething conflict for more than 60 years that intensified from 1600 and suddenly erupted in 1618. Such trigger events can be frequently observed in history: Luther's Ninety-Five Theses on the Power and Efficacy of Indulgences, the storming of the Bastille, the Assassination of Archduke Franz Ferdinand of Austria, the Montgomery Bus Boycott, or the self-immolation of Mohamed Bouazizi. All these events triggered conflicts that had a wide range of economical, political and social reasons. The underlying conflict smouldered for several years, but an open clash was triggered by a single event that taken on its own, was insignificant.

Let there be l different existing social norms and define the violence level, which arises from a violation of norm j, as v_j, with $j = 1,...,l$. Define the total violence level summed over all violations as $v = \sum_j v_j$. Suppose further that for every such norm violation a specific measure is taken that decreases violence. It is assumed for simplicity that those measures appear exogenously and are not subject to strategic choice.[8] Examples of such measures are wage increases, a favourable change in the social security system, work amenities, a greater right of co-determination, but it can also include propaganda that is used to misinform and to shroud the norm violation.[9] Call this measure α_j. In addition, the non-aggression norm being defined by β, reduces the violent reactions to all violations. Assume that its effect is identical for all v_j, thus β does not require an index. Consider the following system of differential equations (heavily inspired by Nowak and May (2000)):

$$\begin{aligned} \dot{v_j} &= v_j\left(r(1-y(1-y)) - s\alpha_j - q\beta - v\varepsilon\right) \\ \dot{\alpha_j} &= hv_j - u\alpha_j(1+v) \\ \dot{\beta} &= (1-x)\kappa - u\beta \end{aligned} \tag{18}$$

where all those variables not previously defined are assumed to be constant. The first equation defines the dynamics of the violence levels. Violence exponentially increases in $r(1-y(1-y))$, where the net growth rate is equal to r in the absence of any supporters. The idea behind the non-monotonic growth of violence is that medium levels of supporters will suppress violent behaviour. Beyond a certain threshold, however, the increasing number of supporters does not scare off revolutionaries, but has the opposite effect, as it ignites violence. Violence also decreases exponentially in $s\alpha_j v_j$ and $q\beta v_j$. Since more violent reactions have the tendency to wear off more quickly, when dampened, the effect of α_j and β increases in v_j. Furthermore,

[8] An additional strategy for players in C_B could be included in the model. Yet, I do not believe that this increases the clarity of the model, since the underlying argument should be similar to the decision of whether or not to pay supporters. It might be still interesting for future research to create a trade-off between paying supporters and paying for anti-aggression measures.

[9] In that sense we might more suitably speak of *perceived* norm violations.

assume that violence naturally "cools down" at the rate $v\varepsilon$, with ε relatively small. The violence level thus has a saturation point in the absence of a non-aggression norm and further counter-measures, due to its decrease in $v_j v\varepsilon$. The last term in $\dot{v_j}$ binds the total value to $v = \frac{r(1-y(1-y))}{\varepsilon}$, since violence levels cannot be infinite.

Assume that a higher level of violence requires a stronger counter-measure, i.e. α_j is increasing in v_j. A high α_j is, however, costly and difficult to maintain and the positive effect of a larger measure wears off more quickly. Furthermore, it is expected that counter-measures absorb resources, and also that their efficiency depends on the acceptance by the individuals concerned. The total violent response v thus decreases α_j. The counter-measure therefore wears off more quickly if it is more efficient and if the total violence level is high.[10]

Finally, the social norm of non-aggression is only an indirect function of the total violence v, since its size is affected by the share of revolutionaries. As discussed above, if a player more frequently meets a revolutionary, who shows a high level of aggression, the player becomes accustomed to this and also shows a higher propensity for violence. If the population consist mostly of revolutionaries, the non-aggression norm can be expected to be much lower than in a population with only few revolutionaries. First consider the equilibrium values, determined by setting the last two equations in 18 to zero.

$$\begin{aligned} \alpha_j^* &= \frac{hv_j}{u+vu} \\ \beta^* &= \frac{\kappa(1-x)}{u} \end{aligned} \tag{19}$$

Setting those into the first equation of 18 defines the equilibrium dynamics of an individual violence level

$$\dot{v}_j = v_j \left(r(1+x) - s\frac{hv_j}{u(1+v)} - q\frac{\kappa(1-x)}{u} - \varepsilon v \right) \tag{20}$$

We are, however, interested in the aggregate value v. Hence, this can be expressed as

$$\dot{v} = \left(\sum_j \frac{v_j}{u(1+v)} \Big((r(1-y(1-y)) - \varepsilon v)u(1+v) - q(1-x)\kappa(1+v) \Big) \right) - \frac{v}{u(1+v)} vsh \sum_j \left(\frac{v_j}{v} \right)^2 \tag{21}$$

Define $D = \sum_j (v_j/v)^2$, with $D \in (\frac{1}{n}; 1)$. If only one violation of a social norm currently occurs, $D = 1$, whereas in the case of n different violations and for identical levels of violent reactions, $D = 1/n$. Substituting and rearranging equation 21 gives

$$\dot{v} = \frac{v}{u(1+v)} \Big(u(r(1-y(1-y)) - v\varepsilon) - q(1-x)\kappa + v\big\{ \mathbf{u}(\mathbf{r}(\mathbf{1}-\mathbf{y}(\mathbf{1}-\mathbf{y})) - \mathbf{v}\varepsilon) - \mathbf{q}(\mathbf{1}-\mathbf{x})\kappa - \mathbf{shD} \big\} \Big) \tag{22}$$

Remember that ε is generally expected to be very small and that $v\varepsilon$ is of the same order as the other variables. It must hold by definition that $r(1-y(1-y)) \geq \varepsilon v$, where equality only holds at the maximum level of v. Consequently, as total violence increases, the dynamics of the system are determined by the second term (in bold) in equation 22, which is multiplied by v. Notice that $\hat{r} = r(1-y(1-y)) - \varepsilon v$ equals the net growth rate of violence. Three cases may occur:

1. Immediate high level of violence: $\hat{r}u > q(1-x)\kappa + shD$.
   In this case, the combined effect of the non-aggression norm and the counter-measures, $q(1-x)\kappa + shD$ is too weak to counter-act the net growth of violence, augmented by the adverse effect of violence on the counter-measures. It describes a society with a general tendency towards violence, little capacity to counter-act violence or with limited resources for programmes that increase social equity.

[10] It might be more intuitive to write $\dot{\alpha_j} = hv_j - u\alpha_j(1+v+\sum_j \alpha)$. Notice, however, that α is approximately proportional to v, thus the addition of another variable should maintain the general dynamics.

2. No violence: $\hat{r}u \leq q(1-x)\kappa$.
   In this case, the non-aggression norm alone is strong enough to stabilise the population. Violence levels exhibit non-positive growth. This is described by a society that has a high degree of democracy and a sense for non-violent conflict solutions (high non-aggression norm) or a pacifist mentality (low net growth rate of violence). We observe that this requires a low population share of revolutionaries. Yet, if a population is defined by a high share of revolutionaries, this inequality is very unlikely to hold.

3. Low levels of violence, followed by a sudden upsurge:
   $q(1-x)\kappa < \hat{r}u < q(1-x)\kappa + shD$.
   This is the most interesting case. If the violence can be counter-balanced only by the joint effect of counter-measures and the non-aggression norm, but not by the norm alone, the stability depends on the frequency of norm violations. An increase in the number of simultaneous or contemporary violations of social norms decreases D. This is, for example, illustrated by a society that suffered over a longer period from a deterioration of working conditions in several sectors, unemployment, corruption and poverty, which only concern specific social strata or classes (again, the Arab Spring might serve as an example). Since the share of revolutionaries x decreases the right-hand side of the inequality, the right part of the condition becomes less likely to be fulfilled, implying that population will observe new and longer conflicts with high probability, once being in a state of conflict.

As a consequence, the threshold, which defines the minimum share of revolutionaries necessary for violence levels to increase, is given by

$$x^* = 1 + \frac{Dhs - u\hat{r}}{\kappa q}, \text{ for } x^* \in (0,1) \tag{23}$$

If violations occur more frequently (i.e. D decreases) or if norm violations affect aggression more strongly (i.e. r increases), fewer revolutionaries are required to trigger the upsurge. If it is assumed that $v = lv_j$ the dynamics are defined by

$$\dot{v}_j = v_j \left(r(1-y(1-y)) - s\frac{hv_j}{u(1+lv_j)} - q\frac{\kappa(1-x)}{u} - \varepsilon l v_j \right) \tag{24}$$

Solving for $\dot{v}_i = 0$ defines the aggregate equilibrium value for the deterministic approximation

$$\hat{v} = \frac{-hs - \varepsilon lu - \kappa^* l + (lr^* u) + \sqrt{-4\varepsilon l^2 u(\kappa^* - r^* u) + (-hs - \kappa^* l + lu(-\varepsilon + r^*))^2}}{2\varepsilon lu} \tag{25}$$

with $r^* \equiv \hat{r} + \varepsilon v = r(1-y(1-y))$ and $\kappa^* \equiv \kappa q(1-x)$. For very small ε, the total violence level is approximated by $\tilde{v} = l\left(r^* u - \kappa^*\right) / \left(hs - l(r^* u - \kappa^*)\right)$ and thus expected to explode after

$$l^* = \lceil \frac{hs}{r^* u - \kappa^*} \rceil \tag{26}$$

violations, with $\lceil \xi \rceil$ denoting the smallest integer not less than ξ. Hence, the number of required violations increases in the effectiveness of the counter-measures, but decreases in the in the joint effect of individual and total violence levels on violence reduced by the impact of the non-aggression norm. Notice that total violence level is bound to $v = r^*/\varepsilon$, and attains its highest level for $x = 1$ and $v = lv_j$. Simplifying equation 25 gives

$$\begin{aligned} v^{max} = \lim_{l \to \infty} \hat{v} &= \frac{r^* u - \varepsilon u - \kappa q(1-x)\sqrt{(u(\varepsilon + r^*) - \kappa q(1-x))^2}}{2\varepsilon u} \\ &= \frac{r^* u - \kappa q(1-x)}{\varepsilon u} \leq \frac{r^*}{e} \end{aligned} \tag{27}$$

4 The Emotional Conflict Model

Obviously, emotions are difficult to quantify. Well-defined and established approaches of how emotions analytically affect best-response play are thus lacking (for a general discussion of the relationship between emotions and rationality, see Kirman et al 2010). If an emotional reaction is considered to "blur" the salience of certain pay-off values (or more general utility values), emotions can be directly incorporated into the pay-off functions. This might take the form of a threshold value, imposed by an emotion, below which an individual is no longer concerned with the potential loss he faces, when playing this strategy. Hence, only if the pay-off difference between best and non-best response is higher than this threshold value, will a player choose his rational best response. Alternatively, giving in to an emotion might directly provide an additional utility, thus influencing the pay-off associated to a certain strategy that channels this emotion.

Assume that in this context, the violence level dampens the salience of the potential punishment that a revolutionary faces in the case, where the conflict does not end in his favour. In general this would change the pay-off of strategy R given by equation 9 to

$$\widetilde{\pi_R} = \frac{x(x+y-ay)}{x+y}\delta^r + \left(1 - \frac{x(x+y-ay)}{x+y}\right)(\delta^p y + \nu(v)) \tag{28}$$

where $\nu(v)$ transforms the value of the aggregate violence level into a pay-off or utility measure. In this configuration, the aggregate violence level directly affects the absolute value of the punishment δ^p and is unaffected by the loss probability. In the given context, assume that a player under-evaluates the negative effect, if he observes uncontested revolutionaries, who do not inter-act with supporters, more frequently. If he perceives, however, too many uncontested rioting revolutionaries on the street, his affinity to them decreases or as Granovetter and Soong explained: "Individuals who would not speak out until some minimum proportion of those expressing opinions were in their camp might no longer feel the need to speak once a more substantial proportion agreed with them and the situation seemed more securely in hand. This seems even more likely when the action in question is more costly than just expressing an opinion."(Granovetter and Soong 1988, p. 86). To take account of this group effect with *decision reversals* assume that the transformation function $\nu(.)$ has the form $\nu(v) = \sigma\, v\, (xz)^2$, where σ is a constant that scales the violence level appropriately into a utility measure. Thus, the compensating effect of v increases in the absence of supporters and has its highest level at the intermediate level of x.

Though the system of equations governing the violence levels is stochastic, equations 25 and 27 provide approximate values for v if the individual v_j's have similar scales. The system's dynamics can thus be adequately modelled, if v is substituted by $\hat{v}$ or v^{max}. Since the former generates values similar to the latter already for relatively few violations, and situations with a high conflict potential – case 1. and especially case 3. – are of greater interest, the analysis can be reduced to the latter value without much loss of generality. Define the probability of winning as $P = x(x+y-ay)/(x+y)$, the dynamics are thus sufficiently approximated by

$$\begin{aligned}
\dot{x} &= x\left(P\delta^r + (1-P)\left(\delta^p y + \sigma\frac{r^* u - \kappa q(1-x)}{\varepsilon u}(xz)^2\right) - \phi\right) \\
\dot{y} &= y\left(\frac{dw^* x - cxy}{x+y} - \phi\right) \\
\dot{z} &= (1-x-y)\,(P\delta^r - \phi)
\end{aligned} \tag{29}$$

where ϕ is defined as the average pay-off as before and $w^{*3} = 1$ or $w^{*1} = \frac{1}{2k}(k + bx - e - bx^2 - \sqrt{(e-k-b(1-x)x)^2 - 4bk(1-x)x})$, depending on the equilibrium, to which sub-population C_B is associated. In contrast to the approach that neglects the emotional component of conflicts, interior equilibria can evolve. These are unlikely to be observed in the low w-equilibrium, since as x approaches 1 and thus w tends to zero, unrealistically high values of d, and low values of c were necessary. In the high w-equilibrium (i.e. $w = 1$), the solution to $\Delta\sigma_i \equiv \pi_i - \phi = 0$ gives 5 roots, each for Δx, Δy and Δz. Some of the roots have values outside the unit interval of $x+y \in (0,1)$. Excluding these, we are left with two roots for both Δx and Δy and one root for Δz.

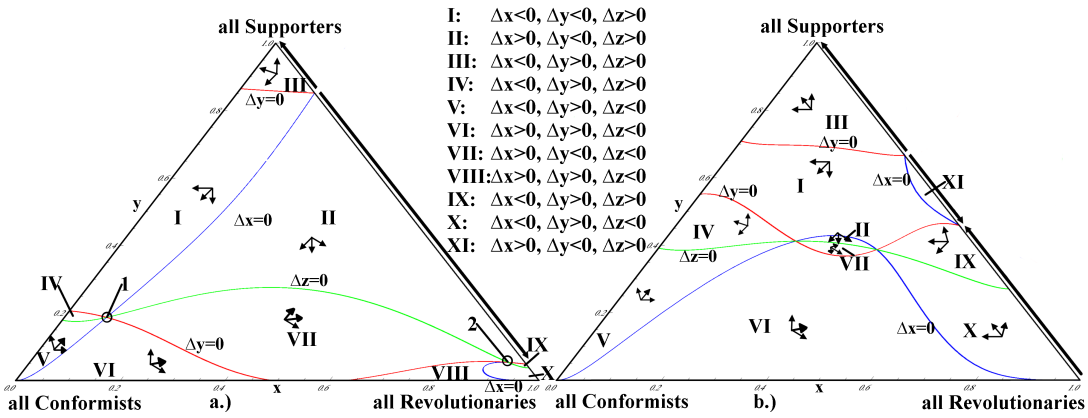

Fig. 2 Projection of the unit simplex and dynamics if norm violations are taken into account, for $w = 1$ and $v = v^{max}$: Parameters: $a = 1$, $\delta^p = -3$, $r = 2$, $u = 2$, $\kappa = 0.2$, $q = 1$, $\varepsilon = 0.05$; A.) $\delta^r = 4$, $d = 5$, $c = 25$,$\sigma = 5$; B.) $\delta^r = 3$, $d = 9$, $c = 20$, $\sigma = 6$

Figure 2 a.) shows a simulation, which uses parameter values identical with those in figure 1 on page 7 for the high w-equilibrium case. As before, equilibria on the y-axis are Lyapunov stable equilibria that cannot be invaded by revolutionaries. Since small perturbations through non-best response play are not self-correcting, evolutionary drift will push the population into region *IV*, representing roughly 20% supporters and 80% conformists (note that in region *III*: $\pi_S < \pi_C$). The x-axis, on the contrary, does not define any weak equilibrium points, since regions *I* and *II* do not touch the x-axis, at which $\Delta y = 0$. As in the simulation shown in figure 1 two unstable equilibria exist at which conformists are entirely absent, given by $(0.15, 0.85)$ and $(0.95, 0.05)$.

In contrast to the previous simulation, two additional and completely mixed (interior) equilibria occur, denoted by 1 and 2. The former equilibrium (situated at $(0.08, 0.19)$) is unstable since the pay-off of revolutionaries increases (decreases) with an additional (absent) revolutionary. Thus, to the right of the equilibrium point, a revolutionary benefits from more players choosing strategy *R* and the equilibrium is not self-correcting for small perturbations to the right or left. The second completely mixed equilibrium (situated at $(0.91, 0.05)$) is stable, since at this point the pay-off of revolutionaries is decreasing in x.[11] The choice of an additional player to play strategy *R* is therefore detrimental to the pay-off of other revolutionaries at this point and the equilibrium is self-correcting. For the given case, a population converges either to the y-axis in region *IV* or to equilibrium 2 in the case, if at least a critical mass of revolutionaries (defined by the blue line, at which $\Delta x = 0$) exists. Intuitively the critical mass increases in the number of supporters. Figure 2 The equilibrium structure of the game is relatively robust to parameter changes. b.) illustrates a simulation with parameter values $\delta^r = 3$, $d = 9$, $c = 20$, $\sigma = 6$, all others being identical with a.). We observe that both equilibria are maintained, but are moving towards the simplex centre as expected.

The model can hence show a further characteristic of conflicts: It is often observed that beneath a threshold people remain bystanders in a conflict situation, although they feel a desire to revolt, but are afraid of being the only one to participate. Only if a sufficient number of other individuals, joining the conflict, is perceived, individuals choose sides and enter the conflict. “The power of the mighty hath no foundation but in the opinion and belief of the people.” (Hobbes 1668)

5 Conclusion and Outlook

In this article, an intuitive approach has been derived to model the general dynamics of conflicts. Though the assumptions highly abstract from the richness of real world conflicts, the model re-creates specific properties that are inherent in conflicts. It can explain why some societies are more prone to conflicts than others caused

[11] Both eigenvalues of the system’s Jacobian have alternate sign at the former equilibrium and are negative at the latter, thus defining it as a stable node.

by the currently prevalent social norms but also the past history and frequency of conflicts and norm violations. The model also illustrates why an open conflict may only be perceived after some triggering event, although reasons for this conflict are to be found in a longer period pre-dating the event. It provides conditions under which these trigger events occur more frequently. In addition, the existence of two stable equilibria, one with high and another with no conflict, provides an explanation for *pluralistic ignorance*.[12] Though the conflict potential in a population may be high, it is still necessary that a certain number of players chooses what is perceived as a *non-best response* in order to trigger a transition out of the basin of attraction of the *no conflict* equilibrium into the basin of attraction of the *high conflict* equilibrium. That may, however, be a large number of individuals, depending on the absolute group size.

The article thus leaves room for further extensions, since it disregards the issue of how transition between both equilibria takes place. The given model generates a set of equilibria defined by simplex edge in region *IV*, in which revolutionaries are absent, and a second interior stable equilibrium. Both are separated by the ($\Delta x = 0$)-locus that indicates the critical mass of players necessary to induce the population to switch to the completely mixed equilibrium. It is often argued that a transition from one equilibrium into the basin of attraction of another can be explained by random idiosyncratic choice (see for example Young 1993; Kandori et al 1993). Yet, a revolution is not caused by a sufficiently large number of players, who idiosyncratically choose a *non-best* response strategy. Revolutionary behaviour is an active choice.

Abandoning the simplifying assumption that players of identical type also have identical pay-off functions provides an approach that can explain the transition behaviour. One option is to integrate the threshold approach of Granovetter and Soong (1988) into the replicator dynamics. First attempts show a number of intuitive results: It explains the bystander effect by showing that it is more likely that a small group is incited than a larger. Yet, if group size is large, a certain number of revolutionaries will be observable with high probability. This approach can be extended into several directions (e.g. networks of groups, in which we observe a domino effect from smaller to larger groups). This is, however, beyond the scope of this article and will be left for future research.

Acknowledgements Financial support by the University of Siena is gratefully acknowledged. I greatly benefited from discussions with Leonardo Boncinelli, Herbert Gintis, Eric Guerci, Nobuyuki Hanaki, Sylvie Thoron, JörgenWeibull, Jean-Benoît Zimmermann, and especially from those wth Alan Kirman.

References

Ashenfelter O, Johnson GE (1969) Bargaining theory, trade unions, and industrial strike activity. The American Economic Review 59(1):35–49

Axtell R, Epstein JM, Young HP (2000) The emergence of classes in a multi-agent bargaining model. Working Paper (9)

Bergin J, Lipman BL (1996) Evolution with state-dependent mutations. Econometrica 64(4):943–956

Bernheim BD (1994) A theory of conformity. Journal of Political Economy 102(5):841–877

Binmore K (1994) Game Theory and the Social Contract, Vol. 1: Playing Fair, illustrated edition edn. The MIT Press

Binmore K (1998) Game Theory and the Social Contract, Vol. 2: Just Playing, illustrated edition edn. The MIT Press

Bowles S (2006) Microeconomics: Behavior, Institutions, and Evolution (The Roundtable Series in Behavioral Economics). Princeton University Press

Bowles S, Choi JK, Hopfensitz A (2003) The co-evolution of individual behaviors and social institutions. Journal of Theoretical Biology 223(2):135 – 147

Boyd R, Gintis H, Bowles S, Richerson P (2003) The evolution of altruistic punishment. Proceedings of the National Academy of Science 100(6):3531–3535

Brams SJ, Kilgour DM (1987a) Optimal threats. Operations Research 35(4):524–536

Brams SJ, Kilgour DM (1987b) Threat escalation and crisis stability: A game-theoretic analysis. The American Political Science Review 81(3):833–850

Brams SJ, Kilgour DM (1988) Deterrence versus defense: A game-theoretic model of star wars. International Studies Quarterly 32(1):3–28

Brams SJ, Togman JM (1998) Cooperation through threats: The northern ireland case. PS: Political Science and Politics 31(1):32–39

Brandt H, Hauert C, Sigmund K (2006) Punishing and abstaining for public goods. Proceedings of the National Academy of Sciences 103(2):495–497

[12] Pluralistic ignorance describes the event in which a norm is individually falsely believed to be appreciated by the majority and thus accepted though it is, in reality, rejected by most players.

Camerer CF, Loewenstein G (2002) Behavioral economics: Past, present, future. draft: 10/25/02
Camerer CF, Loewenstein G, Prelec D (2005) Neuroeconomics: How neuroscience can inform economics. Journal of Economic Literature 43
Choi JK, Bowles S (2007) The coevolution of parochial altruism and war. Science 318(5850):636–640
Clark S (1996) Strike behaviour when market share matters. Oxford Economic Papers 48(4):618–639
Cohen GA (1982) Reply to elster on "marxism, functionalism, and game theory. Theory and Society 11(4):483–495
Cole HL, Mailath GJ, Postlewaite A (1998) Class systems and the enforcement of social norms. Journal of Public Economics 70(1):5 – 35
Cooper CL, Dyck B, Frohlich N (1992) Improving the effectiveness of gainsharing: The role of fairness and participation. Administrative Science Quarterly 37(3):471–490
Durlauf SN, Young HP (2001) Social Dynamics. The MIT Press
Ellison G (1993) Learning, local interaction, and coordination. Econometrica 61(5):1047–1071
Ellison G (2000) Basins of attraction, long-run stochastic stability, and the speed of step- by-step evolution. The Review of Economic Studies 67(1):17–45
Elster J (1982) Marxism, functionalism, and game theory: The case for methodological individualism. geocities URL http://users.auth.gr/kehagiat/GameTheory/05PapersAdvanced/PoliticalScience/001full.pdf
Eswaran M, Kotwal A (1985) A theory of contractual structure in agriculture. The American Economic Review 75(3):352–367
Eswaran M, Kotwal A (1989) Why are capitalists the bosses? The Economic Journal 99(394):162–176
Fehr E, Gächter S (2000) Fairness and retaliation: The economics of reciprocity. The Journal of Economic Perspectives 14(3):159–181
Fehr E, Schmidt KM (1999) A theory of fairness, competition, and cooperation. The Quarterly Journal of Economics 114(3):817–868
Foster D, Young P (1990) Stochastic evolutionary game dynamics. Theoretical Population Biology 38(2):219 – 232
Frey BS, Bohnet I (1995) Institutions affect fairness: Experimental investigations. Journal of Institutional and Theoretical Economics (JITE) / Zeitschrift fr die gesamte Staatswissenschaft 151(2):286–303
Frohlich N, Oppenheimer JA, Eavey CL (1987a) Choices of principles of distributive justice in experimental groups. American Journal of Political Science 31(3):606–636
Frohlich N, Oppenheimer JA, Eavey CL (1987b) Laboratory results on rawls's distributive justice. British Journal of Political Science 17(1):1–21
Geertz C (1987) Dichte Beschreibung. Beiträge zum Verstehen kultureller Systeme, 11th edn. Suhrkamp Verlag
Gintis H (2000) Game Theory Evolving: A Problem-Centered Introduction to Modeling Strategic Interaction (Second Edition), 2nd edn. Princeton University Press
Gintis H (2009) The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences. Princeton University Press
Gintis H, Bowles S, Boyd RT, Fehr E (eds) (2005) Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life. Economic Learning and Social Evolution, The MIT Press
Granovetter M, Soong R (1988) Threshold models of diversity: Chinese restaurants, residential segregation, and the spiral of silence. Sociological Methodology 18:69–104
Hobbes T (1668) Behemoth or The Long Parliament. London, Simpkin, Marshall, 1889, URL http://www.archive.org/download/cu31924028063893/cu31924028063893.epub
Hofbauer J, Sigmund K (1988) The Theory of Evolution and Dynamical Systems: Mathematical Aspects of Selection. Cambridge University Press
Hoffman E, Spitzer ML (1985) Entitlements, rights, and fairness: An experimental examination of subjects' concepts of distributive justice. The Journal of Legal Studies 14(2):259–297
Holländer H (1982) Class antagonism, exploitation and the labour theory of value. The Economic Journal 92(368):868–885
Huck S, Kübler D, Weibull J (2001) Social norms and optimal incentives in firms. Research Institute of Industrial Economics, WP 565
Huck S, Kübler D, Weibull J (2003) Social norms and economic incentives in firms. Revised version of IUI, WP 565
Kandori M, Mailath GJ, Rob R (1993) Learning, mutation, and long run equilibria in games. Econometrica 61(1):29–56
Kennan J, Wilson R (1989) Strategic bargaining models and interpretation of strike data. Journal of Applied Econometrics 4:87–130
Kiander J (1991) Strike threats and the bargaining power of insiders. The Scandinavian Journal of Economics 93(3):349–362
Kirman A, Livet P, Teschl M (2010) Rationality and emotions. Philosophical transactions of the Royal Society of London Series B, Biological sciences (1538):215–219
Lindbeck A, Nyberg S, Weibull JW (1999) Social norms and economic incentives in the welfare state. The Quarterly Journal of Economics 114(1):1–35
Lindbeck A, Nyberg S, Weibull JW (2002) Social norms and welfare state dynamics. The Research Institute of Industrial Economics, Working Paper (585)
Lorenz K (1974) Das sogenannte Bse: Zur Naturgeschichte der Aggression: Zur Naturgeschichte der Aggression., neuaufl. edn. Deutscher Taschenbuch Verlag, 2007
Mehrling PG (1986) A classical model of the class struggle: A game-theoretic approach. Journal of Political Economy 94(6):1280–1303
Morris S (2000) Contagion. The Review of Economic Studies 67(1):57–78
Nash J John F (1950) The bargaining problem. Econometrica 18(2):155–162, URL http://www.jstor.org/stable/1907266
Nowak M, May RM (2000) Virus Dynamics: Mathematical Principles of Immunology and Virology. OUP Oxford

Nowak MA (2006) Evolutionary Dynamics: Exploring the Equations of Life. Belknap Press of Harvard University Press
Patterson O (2004) Kultur ernst nehmen: Rahmenstrukturen und ein afroamerikanisches Beispiel, Goldmann Verlag, pp 309–336
Przeworski A, Wallerstein M (1982) The structure of class conflict in democratic capitalist societies. The American Political Science Review 76(2):215–238
Rabin M (1993) Incorporating fairness into game theory and economics. The American Economic Review 83(5):1281–1302
Rabin M (1998) Psychology and economics. Journal of Economic Literature 36(1):11–46
Rabin M (2002) A perspective on psychology and economics, URL http://repositories.cdlib.org/iber/econ/E02-313, department of Economics, University of California, Berkeley
Raiffa H (1953) Arbitration schemes for generalized two person games, Princeton University Press
Roemer JE (1982a) Exploitation, alternatives and socialism. The Economic Journal 92(365):87–107
Roemer JE (1982b) Methodological individualism and deductive marxism. Theory and Society 11(4):513–520
Roemer JE (1982c) Origins of exploitation and class: Value theory of pre-capitalist economy. Econometrica 50(1):163–192
Roemer JE (1985) Rationalizing revolutionary ideology. Econometrica 53(1):85–108
Schuster P, Sigmund K (1983) Replicator dynamics. Journal of Theoretical Biology 100:533–538
Smith VL (1994) Economics in the laboratory. The Journal of Economic Perspectives 8(1):113–131
Starrett D (1976) Social institutions, imperfect information, and the distribution of income. The Quarterly Journal of Economics 90(2):261–284
Swedberg R (2001) Sociology and game theory: Contemporary and historical perspectives. Theory and Society 30(3):301–335
Taylor PD (1979) Evolutionarily stable strategies with two types of player. Journal of Applied Probability 16(1):76–83
Taylor PD, Jonker LB (1978) Evolutionary stable strategies and game dynamics. Mathematical Biosciences 40:145–56
Weibull JW (1995) Evolutionary Game Theory. The MIT Press
Young HP (1993) The evolution of conventions. Econometrica 61(1):57–84
Young HP (2005)) Diffusion of innovations in social networks, Oxford University Press
Young PH (1998) Individual Strategy and Social Structure. Princeton University Press