CESifo *Working Paper Series*

EVOLUTIONARY NORM ENFORCEMENT

Werner Güth
Axel Ockenfels

Working Paper No. 331

August 2000

CESifo
Poschingerstr. 5
81679 Munich
Germany
Phone: +49 (89) 9224-1410/1425
Fax: +49 (89) 9224-1409
http://www.CESifo.de

CESifo *Working Paper No. 331*
August 2000

EVOLUTIONARY NORM ENFORCEMENT

Abstract

Applying an indirect evolutionary approach with endogenous preference formation, we show that a legal system can induce players to reward trust even if material incentives dictate to exploit trust. By analyzing the crowding out or crowding in of trustworthiness implied by various verdict rules, we can assess how a court influences the share of kept promises of 'truly' trustworthy players who evolutionarily evolved as trustworthy and of opportunistic players who are only trustworthy if inspired by material incentives.

JEL Classification: B4, D8, K0, K4

Werner Güth
Humboldt-University at Berlin
Department of Economics
Spandauer Str. 1
10178 Berlin
Germany
email: gueth@wiwi.hu-berlin.de

Axel Ockenfels
Harvard University
Graduate School of Business Administration
Soldiers Field Road
Baker Library West 188
Boston, MA 02163
USA
email: aockenfels@hbs.edu

1. Introduction

Trustworthiness is a basic prerequisite for both economic and social exchange. In anarchy cooperation based on trust (of keeping one's promises) usually requires small groups where everybody knows who is trustworthy or not or is at least able to detect trustworthiness in a probabilistic sense (see Frank, 1988, 1989, and Güth and Kliemt, 1994, 1999, Bohnet et al., 1999, among many others). One of the characterizing features of modern societies is, however, that due to the prevalent division of labor many economic transactions take place in large, anonymous market environments. This makes it difficult to signal trustworthiness and, by the same token, to trust other members of the same population. Furthermore, recent experimental evidence does not consistently support the hypothesis of a culturally acquired truth detection capability as suggested by Frank (1989) and Gauthier (1978) among others (cf. Ockenfels and Selten, 1999, and the references cited therein). But what mechanism can then lead to the evolution of cooperation based on trust?

Like Brennan, Güth, and Kliemt (1997), who rely on a related but strategically quite different basic game of trust, we examine whether a society's legal system may serve as a substitute for reputation and detection mechanisms that may lead to trustworthiness in small groups. Such a legal system may punish untrustworthy individuals and thereby enforce promises and contracts that could not be enforced otherwise. In the long run, however, there is a price to be paid for the legal system that goes beyond the direct costs of law enforcing. If promises are kept because material incentives dictate so, trustworthy individuals cannot be more successful than untrustworthy ones. This may crowd out trustworthiness and thus endanger cooperation based on trust where the legal system does not or cannot interfere because, for instance, enforcing contracts is prohibitively costly (cf. Williamson, 1993). Successful law enforcement therefore includes norm enforcement, i.e., individual 'internal' commitments to be trustworthy.

How can norm enforcement be captured in an economic model? In our model, preferences are endogenously formed in an evolutionary process. While choices are motivated by the preferences, which may include an internal commitment to norms, objective evolutionary success of preference types depends on 'external incentives'. However, evolutionary success depends on the choice made, which in turn depends on preferences. Thus, preferences *indirectly* affect reproductive success. This is the idea of the so-called "indirect evolutionary approach" (Güth and Yaari, 1992). In particular, the approach captures the idea that individual preferences in social interactions are shaped by evolutionary competition. We will analyze whether internal

commitments to be trustworthy survive such evolutionary competition. Neither individual reputation building nor special detection capabilities are assumed; we just rely on the existence of a court system.

In our two-player court game, player 2, while having an external incentive to exploit trust, may promise to reward a trusting investment of player 1. Player 1, if trusting, either loses his investment or receives a positive net return. Losing the investment can have two reasons. First, player 2 broke his promise, and second, uncontrollable bad luck. Anyway, in case of a loss, player 1 can appeal to a court who does not know the true reason for sure but who can enforce any verdict rule depending on its posterior probability of exploitation. The exploitation probability depends on the average 'reputation' to be trustworthy, which is endogenous in our model, and on the probability of an unintended damage, which is assumed to be exogenous.

As an example, think of an entrepreneur who needs money for a risky project. He offers a contract to a potential investor whom he promises to be trustworthy. While the trustworthiness of the entrepreneur is not known by the investor, both the reputation of the entrepreneur population and the probability of an unintended failure of the project are commonly known. By signing the contract, the investor trusts in the entrepreneur. If the investor then observes that the project fails, he appeals to a court which decides whether and to what degree the entrepreneur is liable for compensation.

In this paper, we investigate how the trustworthiness of the population depends on the actual verdict rule that may in turn depend on the trustworthiness of the population. We also examine the tools of a court that not only wants to enforce the contracts but also wants to inspire internal commitments to trustworthiness. Furthermore, since the verdict rules influence the willingness to sign a contract in the first place and since trustworthiness can only be selected among those who participate in social or economic exchange, the court faces the additional challenge to simultaneously enforce norms and participation.

In section 2, we develop the court game that is the basis of the formal analysis in this paper. In section 3, we examine the influence of a discontinuous verdict rule and in section 4 the influence of a related continuous verdict rule on the degree of trustworthiness in the population. We show that under very weak and reasonable assumptions on the verdict rule and on the selection dynamics there is always a positive and unique stable population share of trustworthy players. In

section 5, we sketch the interaction of verdict rules and the player's willingness to sign the contract before we conclude our analysis in section 6.

2. The contract game and the court game

The contract game is a sequential two-player game that starts after player 2 has offered a contract to player 1. With the contract, player 2 promises to reward trust, i.e., to choose R at his second decision node if this node is reached. Player 1 may trust in the promise and sign the contract (move T) or he may not trust and reject (move N). If the contract is rejected, player 1 receives $s < 1$ and player 2 zero. Then, of course, the Pareto-superior allocation in which both players receive a payoff of 1 is out of reach.

When the contract becomes effective, a chance move determines whether keeping the promise is impossible, i.e., whether an uncontrollable damage occurs (chance move with probability w) or not (chance move with complementary probability $1 - w$). If keeping the promise is impossible, player 1 loses his investment s and both players receive zero payoffs. If compliance is possible, player 2 decides whether to keep the promise, i.e., whether to reward (move R) or to exploit (move E) the trust of player 1. If the contract is fulfilled, the investment is rewarded by a positive net return of $1 - s$ and player 2 receives a payoff of 1. If player 2 exploits player 1's trust, however, player 1 loses his investment, while player 2 receives a *material* payoff of r with $2 \geq r > 1 > s$.

Compared to the trust game analyzed by Güth and Kliemt (1992 and 1994) the contract game is enriched by the chance move determining whether keeping the promise is possible or not. Analogously to Güth and Kliemt, we assume that the payoff of player 2 after having exploited player 1 is the material payoff r minus a parameter $m > 0$. The 'internal' payoff component m represents non-material success that can influence reproductive success only indirectly via behavior, whereas all other payoff parameter represent reproductive success directly. m reflects the degree of internal commitment to be trustworthy.

Note that m is an *individual* preference parameter that is private information of player 2. In the tradition of the incomplete information approach of Harsanyi (1967/1968), however, we assume that the distribution of m in the population is commonly known so that we may apply Bayesian equilibria techniques when solving the game.

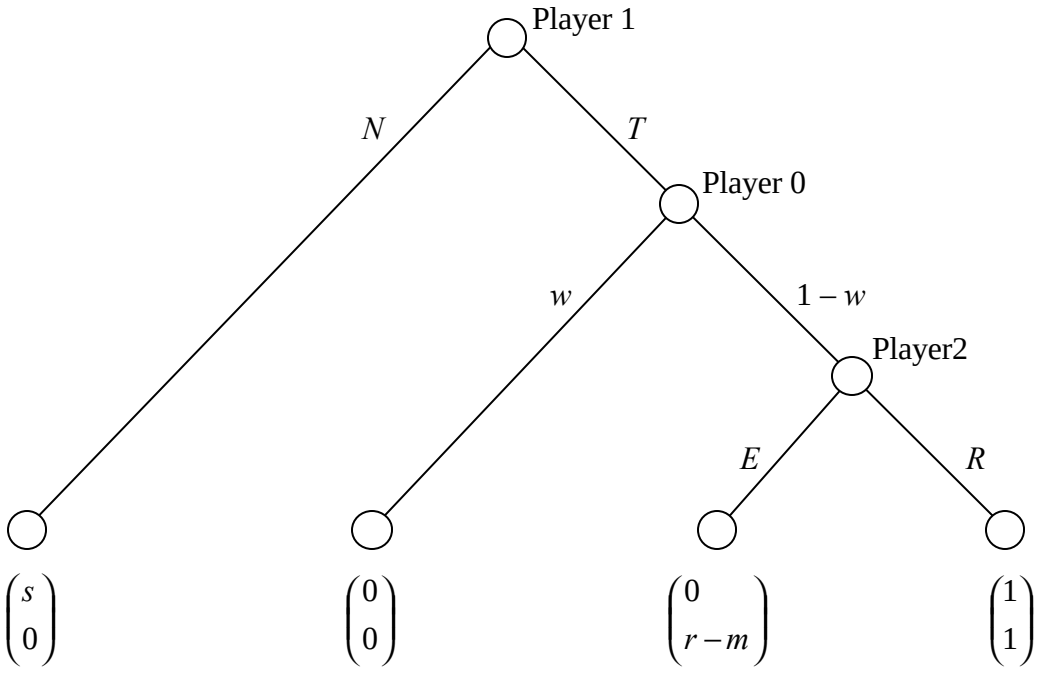

FIGURE II.1. THE CONTRACT GAME

The model parameters are assumed to satisfy conditions

(II.1) $2 \geq r > 1 > s > 0$, $m > 0$, and $0 < w < 1/2$.

How does the court enter the interaction? We assume that whenever player 1 receives zero-payoff, he appeals to the court. The court then either convicts player 1 (the move C in Figure II.2 yielding c for player 1) or dismisses the case (the move D in Figure II.2). Thereby, the court follows a verdict rule $v = (c, q(e))$, $c \geq 0$, $q(e) \in [0,1]$. This rule determines a probability q of conviction and a constant compensation payment c that player 2 has to pay to player 1 in case of conviction.

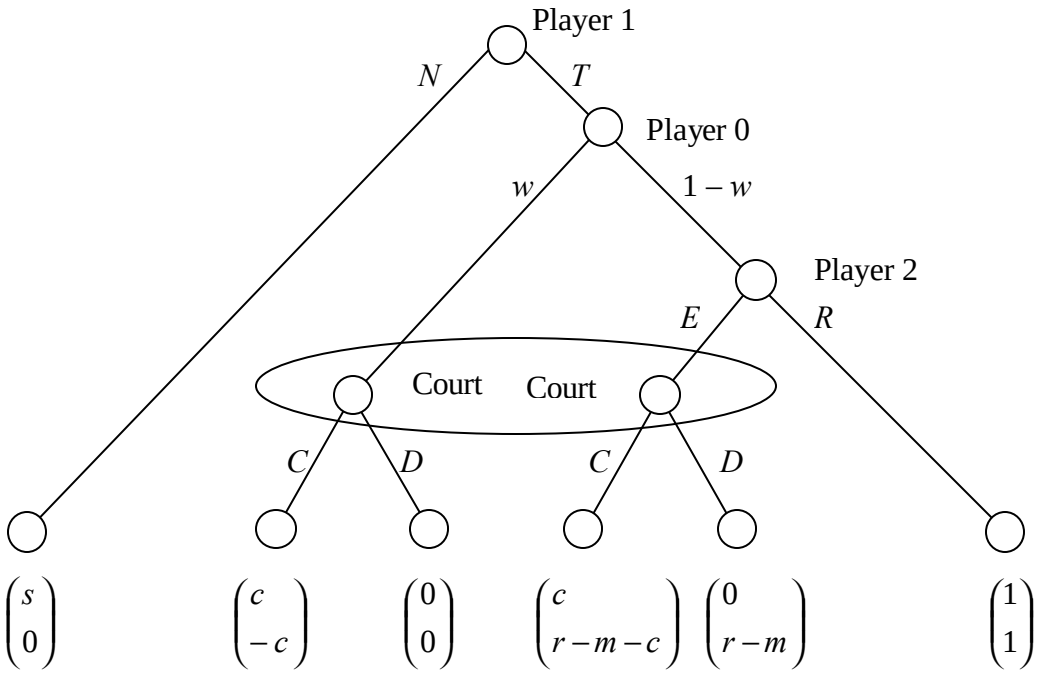

FIGURE II.2. THE COURT GAME

The only indication that can be used by courts in order to come to a verdict is the (conditional) probability e of untrustworthy second movers given that a loss for player 1 occurred. This probability depends on p, the 'reputation' of the corresponding second mover population (for instance, it is a common belief that lawyers have different morals than physicians), and on w, the exogenous probability of an unintended damage. The court may use two instruments in order to enforce contracts and norms, the compensation payment c and the conviction probability q. While most economic studies of law examine the comparative statics of legal institutions, our model allows these institutions to adjust dynamically. (Also, the study of Huck, 1998, does not allow legal institutions to change endogenously but only preferences to be shaped by evolution.) In particular, the probability of conviction q may depend on e which depends on the population's trustworthiness and is therefore endogenously determined. The compensation payment, however, is assumed to be constant. For instance, in case of conviction, player 1 (the plaintiff) gets half of his demand ($c = 1/2$) or player 2 (the defendant) has to fulfill the contract as if he is fully liable for the loss ($c = 1$).

3. The influence of discontinuous verdict rules on trustworthiness

Assume that the contract is signed and that no unintended damage occurred. Thus, player 2 has to decide whether to reward or to exploit the trust of his contract partner. Of course, player 1 may not want to sign the contract in the first place so that player 2 would not reach his decision node. In this case, however, we rely on rare trembles in the sense of a small, but positive minimum probability for the trust move T of player 2 similarly to the coarsening of evolutionarily stable strategies suggested by Selten (1983) and applied by Güth and Kliemt (1994). Then, since play reaches the second decision node of player 2 with positive probability, we need to know the population's inclination to exploit in order to compute how trustworthiness evolves in an evolutionary process. The question whether player 1 actually intends to sign the contract or not is postponed to section 5.

3.1 The case of a 'rational' belief forming court

The exploitation inclination clearly depends on the verdict rule that is implemented by the court. The most natural assumption about the conviction probability is that the court simply decides for the most probable alternative, i.e. if according to its conditional probability the court considers it to be more likely that non-delivery is caused by 2's choice of E, the court decides in favor of player 1 and convicts player 2. Otherwise the court finds player 2 not liable for compensation and dismisses the case. Thus the court enters the interaction as a rational belief forming institution which rules the verdict whose justification is more likely – a rather realistic assumption in civil law. In the following such a verdict rule is denoted by $v = (c, q^{1/2}(e))$.

Let p with $0 \le p \le 1$ denote the proportion of second movers in the population who would not exploit player 1. As mentioned earlier it is assumed that p is commonly known (cf. also Güth, 1995). The verdict rule $v = (c, q^{1/2}(e))$ determines the amount player 1 has to pay to player 2. First, observe that

$$e(w, p) = \frac{(1-w)(1-p)}{w + (1-w)(1-p)}.$$

Then, given our assumptions, the court sets

(III.1) $$v = (c, q^{1/2}(e)) = \left(c, q^{1/2}(e) := \begin{cases} 1 \text{ for } e(w,p) > 1/2 \\ 0 \text{ for } e(w,p) \le 1/2 \end{cases} \right).$$

The highest value of $e(w,p)$ is $e(w,0) = 1 - w > 1/2$ due to $w < 1/2$ whereas p = 1 implies $e(w,1) = 0$. Thus depending on the proportion share p of trustworthy player 2, both verdicts C (conviction) and D (dismissal) by the court are possible. For the verdict C the court's conditional probability $e(w,p)$ must be larger than 1/2 or

(III.2) $$p < \frac{1-2w}{1-w} =: \hat{p}.$$

If (III.2) holds, the choice of E by player 2 of type m would result in the payoff $r - m - c$, and the choice of R would result in the payoff 1, i.e., player 2 prefers R over E if $r - m - c < 1$. Since $m > 0$, we have that if $c \ge r - 1$, all players 2 prefer R. Hence, if $c \ge r - 1$ holds, we have p = 1. Note, however, that an upward movement of p will sooner or later violate (III.2).

If, on the other hand, $c < r - 1$, some players 2 (those with $r - m - c \le 1$ whose population share is p) prefer E in case of $p < \hat{p}$. Let us now consider the case $p > \hat{p}$, i.e. when the court would rule the verdict D. Here, player 2 of type m rewards if and only if $m > r - 1$. Now p is the population share of m-types for which inequality $m > r - 1$ holds. The behavior of the court and player 2 at his second decision node for the different constellations is summarized by Table III.1.

Constellations	$c < r - 1$		$c \ge r - 1$	
	$0 \le p < \hat{p}$	$\hat{p} \le p \le 1$	$0 \le p < \hat{p}$	$\hat{p} \le p \le 1$
Player 2, court	R, C for $m > r - 1 - c$ E, C for $m \le r - 1 - c$	R, D for $m > r - 1$ E, D for $m \le r - 1$	R, C	R, D for $m > r - 1$ E, D for $m \le r - 1$

TABLE III.1. BEHAVIOR OF PLAYER 2 AND THE COURT FOR THE VERDICT RULE $v = (c, q^{1/2}(e))$

3.2 The p-dynamics

In the following sections, we just assume that types m who exhibit the more successful behavior (R or E) increase their population share over time at the cost of the less successful mutants. Obviously, this approach is in line with every reasonable concept of evolutionary dynamics. Furthermore, it will be clear from our analysis that we do not need more demanding assumptions on the dynamics.

When $p < \hat{p}$ and $c \geq r-1$, the trustworthy population share increases till (III.2) is violated. The reproductive success of any m-type is m's solution payoff when setting $m = 0$. Thus, if $p > \hat{p}$ and $c \geq r-1$, the population share p must decrease due to $r > 1$. Similarly, if $c < r-1$, exploitation yields a higher material payoff, so that p decreases.

We conclude that the verdict rule $v = (c, q^{1/2}(e))$ implies the following dynamically stable population composition $p^*(c)$:

$$p^*(c) = \begin{cases} \hat{p} = \dfrac{1-2w}{1-w}, \text{for } c \geq r-1 \\ 0, \text{for } c < r-1 \end{cases}.$$

If $c \geq r-1$, one may interpret the dynamics for $p > \hat{p}$ as adverse selection: since the 'goodies' (those with a high m) are not rewarded by the court they are eliminated. On the other hand, the situation with $p < \hat{p}$ also does not allow to identify the 'badies' (those with a small m) since they are induced to be trustworthy by the material incentives implied by the verdict rule. However, badies and goodies coexist in the form of an evolutionarily stable bimorphism if the compensation is sufficiently high (higher than the net gain $r - 1$ from exploiting).

3.3 Who bears the burden of proof?

Now, assume that the court must be convinced with a probability of more (or less) than 1/2 before it convicts the defendant. For instance, suppose that the burden of proof is on the plaintiff so that the court convicts only if $e(w,p) > y$ for some $y > 1/2$; similarly, one may think of a situation in which the burden of proof is on the defendant so that the court only convicts if $e(w,p) > y$ for some $y < 1/2$. Since $e(w,p)$ is the exploitation probability given the occurrence of a loss for the plaintiff, $y \leq 1-w$. Now, let

$$\text{(III.3)} \qquad v = (c, q^y(e)) = \left(c, q^y(e) := \begin{cases} 1 \text{ for } e(w,p) > y \\ 0 \text{ for } e(w,p) \leq y \end{cases} \right), \; y \in [0, 1-w].$$

Then, the second mover is convicted if

$$\text{(III.4)} \qquad p < \frac{1-w-y}{(1-w)(1-y)} =: \hat{p}(y).$$

It can be shown by the same method as applied above, that $\hat{p}(y)$ is the unique dynamically stable rest point $p^*(c,y)$ for $c \geq r-1$. In particular, $y = 1/2$ is the special case analyzed above. If $y = 1-w$, player 2 is never convicted to pay c. Therefore, in such a situation there is only one dynamically stable rest point $p^*(c,y)$ for all c, namely $p^*(c, 1-w) = 0$. If y = 0, player 2 is always convicted to pay so that $p^*(c,0) = 1$ for $c \geq r-1$.

3.4 Discussion

The following proposition summarizes the results that we derived so far:

Proposition 1. The discontinuous verdict rule $v = (c, q(e; y))$ with

$$q^{y}(e) = \begin{cases} 1 \text{ for } e(w,p) \geq y \\ 0 \text{ for } e(w,p) < y \end{cases}$$

implies the following dynamically stable population composition p^*:

$$p^*(c,y) = \begin{cases} \hat{p}(y) = \dfrac{1-w-y}{(1-w)(1-y)} \text{ for } c \geq r-1 \\ 0 \text{ for } c < r-1 \end{cases}$$

for all $c \geq 0$ and $y \in [0, 1-w]$.

For $w \to 0$ the stable composition $p^* \to 1$. Thus the institution of a court in societies where w is very small produces similar results, namely populations where nearly everybody is trustworthy, as perfect type recognition where player 1 knows the *m*-type of player 2 (see Güth and Kliemt, 1994). For $w \to 1-y$ the results resemble with $p^* \to 0$ those of no type recognition at all.

The proposition reveals that the court can basically use two instruments to enforce trustworthiness in an evolutionarily stable scenario, namely the compensation payment c and the conviction threshold for the exploitation probability y. If the court wanted to maximize p^*, it can set $y = 0$ (so that the court always uses *C*). Then, $p^* = 1$. However, in this case all players are only inspired to be trustworthy by material incentives and norm-guided trustworthiness cannot emerge. Furthermore, $y = 0$ implies the undesirable property of the verdict rule that all players 2 are always liable for compensation, even if they are innocent for sure.

Regardless of whether the goal is to enforce norms (i.e., to induce internal trustworthiness norms) or contracts (i.e., to maximize p^*), the court should always set $c \geq r-1$, because otherwise no trustworthy player would survive. Setting $c \geq r-1$, of course, means to rule out an advantage of player 2 in case of conviction, a quite intuitive requirement.

In sum, for a sufficiently large compensation payment (and disregarding the participation constraints for the moment), the court can implement any proportion of trustworthy players as

evolutionarily stable by choosing an appropriate *y*, provided it rules out an exploitation advantage in case of conviction.

4. Smoothing verdicts

As before, let $q(e) = \Pr\{C\}$. However, instead of the discontinuous decision function $q(e) = 1$ for $e(w, p) > y$ and $q(e) = 0$ for $e(w, p) \leq y$ that is set by the rule $v = (c, q^y(e))$, let us now assume that a court 'smoothes' $q(e)$. In particular, let

$$x(p) := q(e(w, p))$$

where $q(\cdot)$ is a continuous (smoothing) function on $[0,1-w]$ with $q'(e) > 0$, $q(0) = 0$, and $q(1-w) = \bar{x} \leq 1$. Then, $x(\cdot)$ is continuous with $x'(p) < 0$ due to $\partial e(w, p)/\partial p < 0$, $x(1) = 0$ and $x(0) = \bar{x}$ by definition of $e(w, p)$ and $x(p)$.

As an example, think of the verdict rule with $q(e) = e(w, p)$. In this case, we have

(IV.1) $$x(p) = \frac{(1-w)(1-p)}{w + (1-w)(1-p)},$$

where $x(0) = \bar{x} = 1 - w$ and, of course, $x(1) = 0$.

Faced with such a smoothed verdict rule $v = (c, q(e))$, the choice *E* of player 2 at his second decision node would result in the payoff $r - m - x(p)c$ and – as before – the choice of *R* would result in the payoff 1. Thus, player 2 of type *m* prefers *E* over *R* if $r - m - x(p)c \geq 1$.

The *p*-dynamics can be computed in an analogous way as in the discontinuous case. As long as $x(p)c > r - 1$, the share of trustworthy players increases, otherwise it decreases. Since $x(1)c = 0$ and $r > 1$, the dynamically stable rest point must be smaller than one. If in addition the expected compensation payment is not too small when *all* players are untrustworthy, i.e., if $\bar{x}c > r - 1$, there exists a unique positive rest point $p^* > 0$ due to the intermediate value theorem for continuous curves (see Figure IV.1).

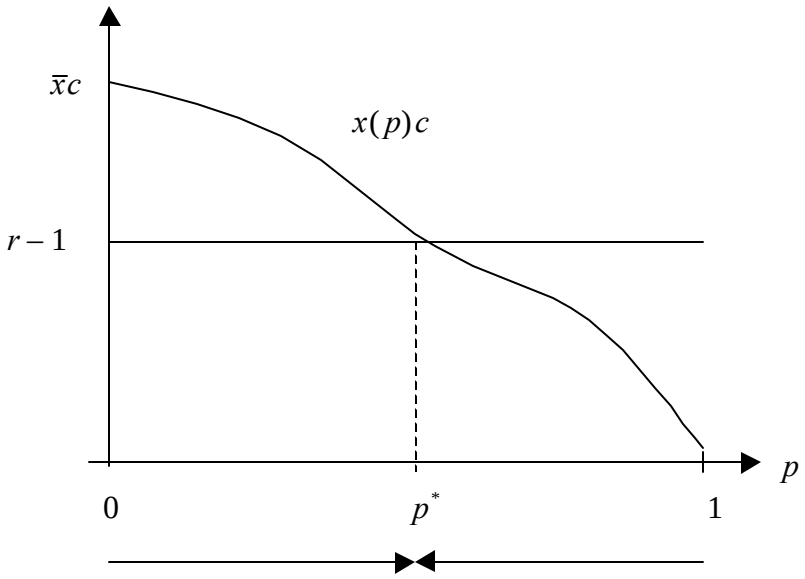

FIGURE IV.1. *P*-DYNAMICS FOR THE SMOOTHED VERDICT RULE

The following proposition summarizes our findings:

Proposition 2. The smoothed verdict rule $v=(c,q(e))$ with $q(\cdot)$ continuous, $q'(e)>0$, $q(0)=0$, and $q(1-w)=\bar{x}\le 1$ implies a unique dynamically stable population composition $p^*<1$ that is for $x(p)=q(e)$ implicitly defined by:

$$x(p^*)c=r-1 \text{ for } \bar{x}c\ge r-1, \text{ and}$$

$$p^*=0 \text{ for } \bar{x}c<r-1.$$

Note that given an interior solution $0 < p^* < 1$, we have $p^{*\prime}(c) > 0$. Hence, given $q(e)$, the court may increase p^* by increasing *c*.

For the example (IV.1), we have

$$p^*(c) = \begin{cases} \dfrac{1-r+(1-w)c}{(1-r+c)(1-w)} \text{ for } c > \dfrac{r-1}{1-w}. \\ 0 \text{ else} \end{cases}$$

5. The participation constraints

A contract and norm enforcing court should consider the participation constraints for at least two reasons. First, if a contract becomes effective only very rarely and unintentionally, then, depending on the exact evolutionary dynamics, the convergence to the target p^* may be very slow, because trustworthiness norms can only evolve among those players who participate in social and economic exchange. Second, since fulfilling of the contract is Pareto superior to any other outcome, a verdict rule should encourage both trustworthiness *and* participation in order to reach Pareto efficiency. Here we shortly sketch the impact of taking participation constraints into account. In particular, we are interested in how the verdict rule influences player 1's and 2's willingness to participate in the discontinuous verdict rule case.

Consider our discontinuous verdict rule in section 3 and assume that $p < \hat{p}$, i.e. whenever a damage occurs the court convicts player 2. Then, independent of the damage probability *w*, players 1 should always participate as long as $s \le c$, i.e. as long as the outside option s is smaller than the compensation payment *c*. If, however, $p \ge \hat{p}$, rational and risk-neutral players 1 participate if and only if $s < (1-w)p$. Hence, the court may encourage players 1 to participate either by a sufficiently high compensation payment in the case ($p < \hat{p}$) that trustworthiness is not very much widespread, or (when $p \ge \hat{p}$) by further enforcing trustworthiness in the case that the defendant is rarely convicted (for instance by shifting the burden of proof to the defendant, i.e. by setting *y* very low).

Analogously, one may also ask under which circumstances player 2 is willing to engage in the social interaction. Since player 2's willingness depends on his *m*-type, however, the analysis is somewhat more complex.

As before, consider the discontinuous case and assume for simplicity that refusing to sign the contract yields zero utility for player 2 (as it would be the case if player 1 refuses to participate). As long as the court intends to dismiss the case (if there is any), it is always better for player 2 to participate than not to participate. If on the other hand the court intends to convict player 2, a sufficient condition for participation for all players 2 regardless of their type is $w(c+1) \leq 1$, or likewise $c \leq (1-w)/w$. One may, of course, be not interested in the participation of player 2 as such, but only in the participation of his *m*-types who would reward trust, i.e. react by *R* to player 1's move *T*. The easiest way to implement this is again to induce $p^* = 1$, e.g. by setting $y = 0$.

In sum, besides enforcing trustworthiness, a court that follows a discontinuous verdict rule $v(c, q^y(e))$ should choose a compensation payment that is neither too small ($s \leq c$) nor too high ($c \leq (1-w)/w$) in order to encourage both parties to sign the contract. Note that since $(1-w)/w > 1$, this is always possible. Do these participation constraints interfere with the norm-enforcing efforts of the legal system? Recall that by proposition 1, for any norm-enforcing verdict rule to be successful, $c \geq r-1$ must hold. Since $r-1<1$, there is always a compensation payment that fulfills $s \leq c$, $c \leq (1-w)/w$ and $c \geq r-1$ simultaneously for all *s*, *w*, and *r*. So for an appropriately chosen compensation payment, there is no basic conflict between norm-enforcing verdict rules and participation.

We just note here (the thorough analysis is more complicated and would not substantially add to the underlying formal mechanisms) that very similar conclusions hold for the continuous case. There also participation of players 1 is encouraged by increasing both, the (expected) compensation payment and the share of trustworthy players while at the same time too high expected compensation payments deter players 2 from signing the contract.

6. Conclusion

Our model shows that in the long run an appropriately designed legal system can induce many players to reward trust even if the purely material, external incentives dictate to exploit trust. By choosing the verdict rule, the court may influence the share of trustworthy players, i.e., the share of 'truly' trustworthy players who are internally committed to norms of trustworthiness and the

share of players who are inspired to be trustworthy by external material incentives. Thereby, in order to boost participation in efficiency-enhancing social interactions as well as to protect trustworthy players from exploitation, the compensation payment should neither be 'too small' nor 'too high'. Most importantly, our analysis suggests that the legal system may serve as a substitute for reputation and type detection capabilities that usually drive evolutionary results in the literature on the emergence of trust. We show that an internal commitment to trust in social interactions may survive evolutionary competition even among rational players and even in large populations where such mechanisms are typically not available.

References

Bohnet, Iris, Bruno S. Frey and Steffen Huck (1999): More Order with Less Law: On Contract Enforcement, Trust and Crowding, Working Paper, Harvard University.

Brennan, Geoffrey, Werner Güth and Hartmut Kliemt (1997): Trust if the Shadow of the Courts are No Better, working paper, Humboldt-University of Berlin.

Frank, Robert H. (1987): If Homo Economicus Could Choose His Own Utility Function, Would He Want One with a Conscience?, *American Economic Review*, 77, 593-604.

Frank, Robert H. (1988): Passions Within Reason: The Strategic Role of Emotions, New York: W.W. Norton.

Gauthier, David (1978): Morals by agreement, Oxford: Clarendon Press.

Güth, Werner, and Hartmut Kliemt (1994): Competition or Co-operation: On the Evolutionary Economics of Trust, Exploitation and Moral Attitudes, *Metroeconomica*, 45(2), 155-187.

Güth, Werner, and Hartmut Kliemt (1999): Evolutionarily Stable Co-operative Commitments, *Theory and Decision*, forthcoming.

Güth, Werner, und M. Yaari (1992): An Evolutionary Approach to Explain Reciprocal Behavior in a Simple Strategic Game, in: U. Witt (ed.), *Explaining Process and Change – Approaches to Evolutionary Economics*, 23-34.

Harsanyi, John C. (1967-8): Games with Incomplete Information Played by Bayesian Players, *Management Science*, 14, 159-82, 320-34, 486-502.

Huck, Steffen (1988): Trust, Treason, and Trials: An Example of How the Evolution of Preferences Can be Driven by Legal Institutions, *The Journal of Law, Economics and Organization*, 14(1), 44-60.

Ockenfels, Axel, and Reinhard Selten (1999): An Experiment on the Hypothesis of Involuntary Truth-Signaling in Bargaining, *Games and Economic Behavior*, forthcoming.

Selten, Reinhard (1983): Evolutionary Stability in Extensive Two-person Games, *Mathematical Social Sciences*, 5, 269-363.

Williamson, O.E. (1993): Calculativeness, Trust, and Economic Organization, *Journal of Law and Economics*, 36, 453-486.