The Coevolution of Morality and Legal Institutions – An indirect evolutionary approach –

Werner Güth & Axel Ockenfels

*Max Planck Institute for Research into Economic Systems**

Abstract

Evolutionary game theory is often used to analyze the evolution of moral preferences. A few studies also examine the coevolution of preferences and an institutional aspect of the decision environment. Allowing the adaptation of just one institutional aspect such as litigation or legal insurance to coevolve with morality, however, may be inadequate. If court rulings coevolve with morality the need for legal insurance may vary over time. Applying the indirect evolutionary approach, we therefore analyze the coevolution of morality in the sense of trustworthiness, court rulings (based on rational belief formation), and the population share which is legally insured. If type detection is not possible, the evolutionary interaction of the legal institutions may play a decisive role for the emergence of morality.

---

* Max Planck Institute for Research into Economic Systems, Strategic Interaction Unit, Kahlaische Straße 10, D-07745 Jena, Germany; e-mail: gueth@mpiew-jena.mpg.de or ockenfels@mpiew-jena.mpg.de. Both authors gratefully thank Steffen Huck and Roland Kirstein for helpful comments, and gratefully acknowledge the support of the Deutsche Forschungsgemeinschaft.

I. Introduction

Indirect evolution offers to combine (boundedly) rational decision making, the traditional approach in economics, with purely adaptative ideas, the traditional approach in (evolutionary) biology and partly also in psychology. What can adapt over time, furthermore, must not be behavior but can be any of its determinants. Indirect evolution therefore offers to endogenously derive what usually is assumed as exogenously given, namely the constellation of behavioral determinants or, in game theoretic terminology, the rules of the game.

Only few studies so far have tried to analyze the coevolution of several determinants. All the studies of which we are aware (e.g., Güth and Kliemt, 1994 and forthcoming) are based on a trust game where morality is captured by trustworthiness.[1] Like the current one, two further studies (Brennan, Güth, and Kliemt, 1997, Güth and Ockenfels, 2000) assume the decision framework of the trust game and allow for litigation in case of exploitation. These studies analyze the coevolution of trustworthiness and court rulings since the latter adapts to the degree of morality in society.

We continue this approach by incorporating another institutional aspect which can adapt over time, namely the population share of individuals who are legally insured and therefore as plaintiffs have better chances for achieving a verdict. Clearly, it depends on the court rulings whether or not legal insurance as an institutional aspect will survive. At the same time legal insurance may also affect morality. Thus, we analyze the coevolution of morality (trustworthiness) with court rulings and legal insurance. This for the first time allows to explore not only whether morality will be crowded out or in by institutional aspects of modern societies but also whether one such aspect crowds out or in another such aspect.

That such complex coevolution can be studied analytically is partly due to the simplicity of the basic trust game (see Figure I.1). First player 1 decides between *N* (*n*on-cooperation) or *T* (*t*rust in player 2's reciprocity). After *N* the game ends with payoff s for player 1 and zero for player 2. After *T* player 2 can either reward (*R*) or exploit (*E*). Assume first of all the special case where the payoff parameter m of player 2 after *T* and *E* is zero. Due to $r > 1$ player 2 would exploit and, anticipating this, player 1 rely on *N* since $s > 0$. For $m = 0$ the game of trust is thus a social dilemma since common rationality (plus player 1's awareness of 2's rationality) dictates (*N*, *E*) and thus a payoff dominated payoff vector since both players would gain by playing (*T*, *R*) instead.[2]

---

[1] Bar-Gill and Fershtman (2001) do not really allow for coevolution but only illustrate how economic policy might have to anticipate the evolution of intrinsic motivation.

[2] The payoffs have been standardized without loss of generality in order to reduce the variety of parameters.

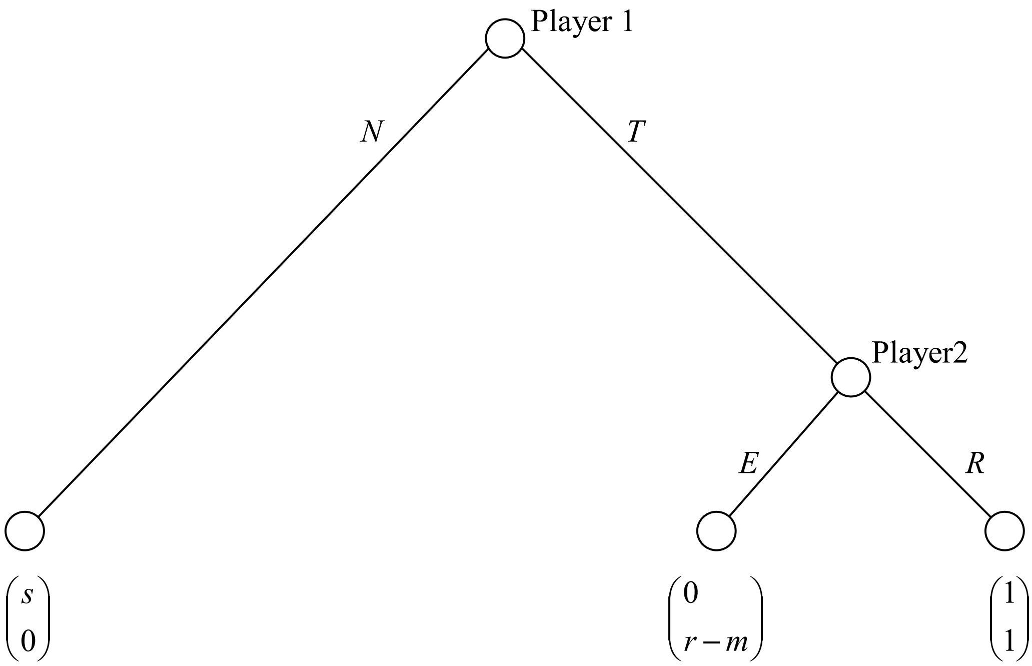

Figure I.1. The basic trust game ($0 < s < 1 < r$)

Morality in the sense of regretting to exploit a trusting player 1 can be captured by $m > r - 1$. The interpretation of m is that of intrinsic motivation. As such the payoff component m is purely immaterial and thus not directly related to (reproductive) success. Although we also will analyze the case where player 1 knows m, the situation where m is player 2's private information is often the more relevant one (see Güth and Ockenfels, forthcoming, for examples, and Ockenfels and Selten, 2000, for experimental evidence). Let p with $0 \leq p \leq 1$ denote the population share of individuals whose personal parameter m satisfies $m > r - 1$. We assume that p is commonly known and thus determines the beliefs of player 1. Thus with private information about m the game of trust is illustrated by Figure I.2 after truncating the decision of player 2 according to their m-types (m with $m > r - 1$ choose T and m with $m \leq r - 1$ choose E). If $p > s$ player 1 will trust (choose T) whereas for $p \leq s$ the dilemma result pertains since 1 chooses N. The evolution of morality in the sense of trustworthiness can then be analyzed via the dynamic adaptation of the population share p of trustworthiness.

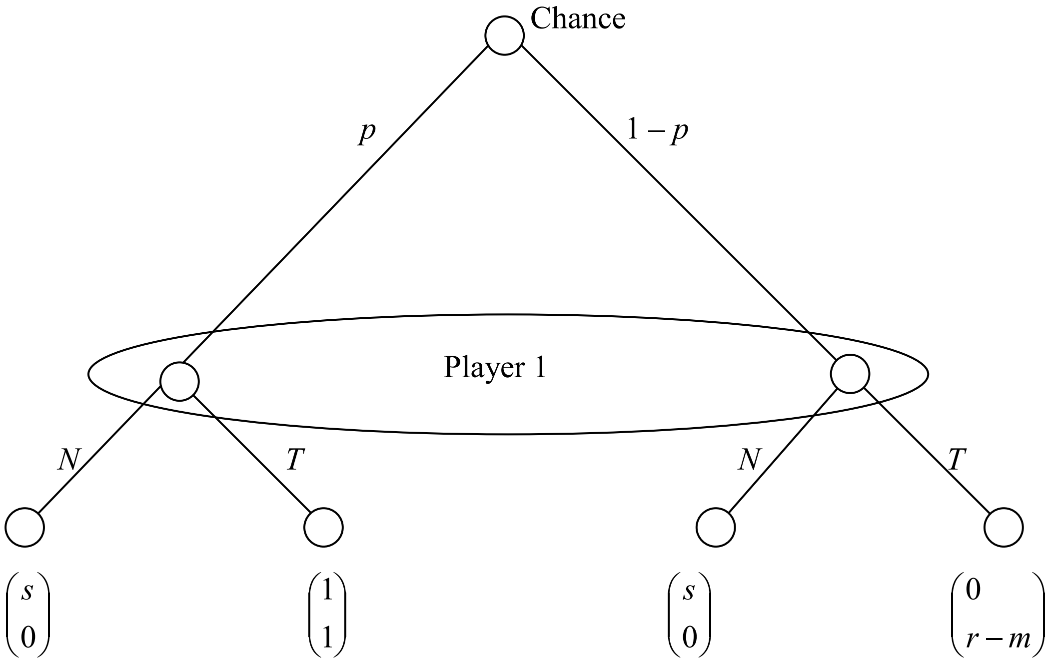

Figure I.2. The (truncated) game of trust when *m* is private information

These two game models will be enriched by including a chance move preventing player 2 from becoming active. Figure I.3 illustrates this point for the game in Figure I.1.[3] After *T* a chance move determines whether player 2 is prevented from moving (by bad luck, or 'force majeure'), which happens with probability *w*. It is crucial that the court can only verify that player 2 did not reward player 1 after *T* but not whether this is due to bad luck or exploitation by player 2.

[3] Figure I.3 omits the fictitious chance move shown in Figure I.2 in order to keep the game tree simpler.

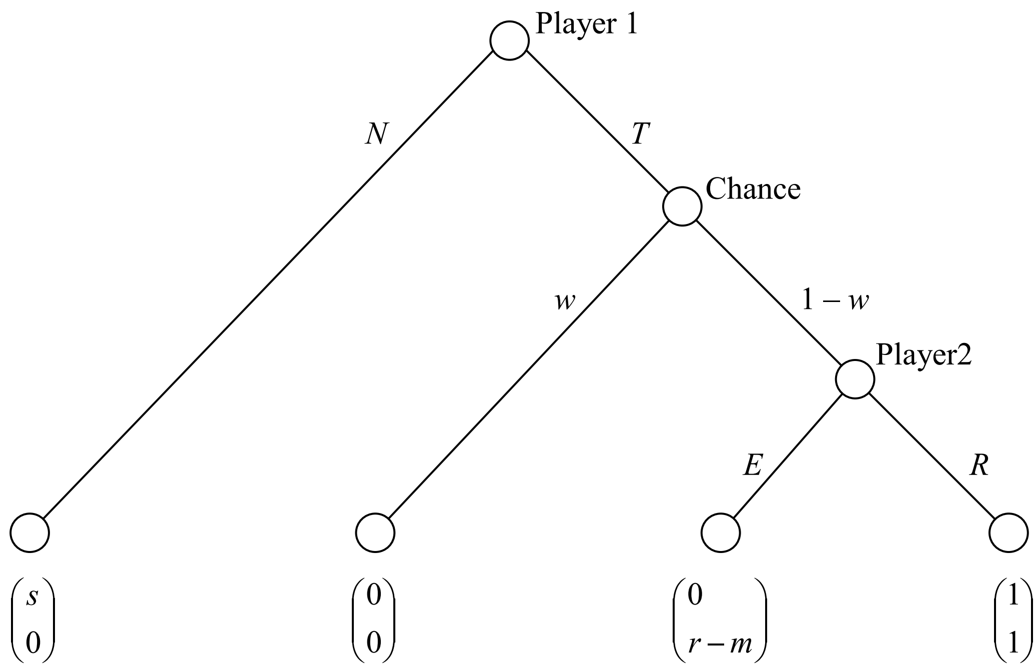

Figure I.3. The game of trust with bad luck ($0 < w < 1/2$)

In the following these situations will be further enriched by including the possibility of litigation in case player 1 is not rewarded. We also allow players 1 to be legally insured or not, but insurance is not rationally chosen but rather an evolving aspect. For this complex model we first solve the games in order to define the evolutionary set up for both the case when m is known to player 1 and the one where m is private information.

II. The trust game with litigation and legal insurance

If m is commonly known among players 1 and 2 (but not by the court),[4] the trust game with litigation and legal insurance in extensive form is described in Figure II.1. In case of non-delivery which can be caused by nature (the move with probability w) or by exploitation (2's move E), player 1 can yield (Y), inducing the same outcome as in the basic trust game in Figure I.1, or appeal (player 1's move A). In the latter case the court becomes active and rules according to its posterior beliefs by conviction (C) or by dismissing (D) the case.

[4] One might argue that it is hard to justify that player 1 can recognize player 2's m-type but that this is impossible for the courts. It may be, however, that among, say, business men m-types are commonly known by past behavior or other signals that are unavailable for the court.

The parameters of the model are:[5]

- w with $0 < w < 1/2$: nature's probability for excluding delivery
- $\delta = \begin{cases} 0 \text{ if player 1 has legal insurance} \\ 1 \text{ if not} \end{cases}$
- c with $1/2 < c < 1$: 2's compensation payment to player 1 in case of conviction [6]
- s with $0 < s < 1/2$: player 1's outside option payoff
- ε with $0 < \varepsilon < s$: (intrinsic) satisfaction due to suing player 2
- C with $0 < C < 1$: legal cost when case is dismissed [7]
- r with $1 + c > r > 1$: exploitation payoff [8]
- m with $m \in \Re$: (intrinsic) regret (for $m > 0$), respectively spite (for $m < 0$) in case of exploiting
- K with $c > K > 0$: the cost of legal insurance (to be subtracted from material success)

---

[5] The parameter restrictions partly avoid too complicated case distinctions.

[6] $c > s$ renders trade as profitable for player 1 even when payment by player 2 requires successful litigation. By $c < 1$ we guarantee that player 1 does not prefer successful litigation over due payment.

[7] The assumption $C < 1$ seems a natural constraint, because unsuccessful litigation should not be more costly than the stakes at hand.

[8] $c > r - 1$ means that the compensation exceeds the exploitation incentive; again a rather natural constraint.

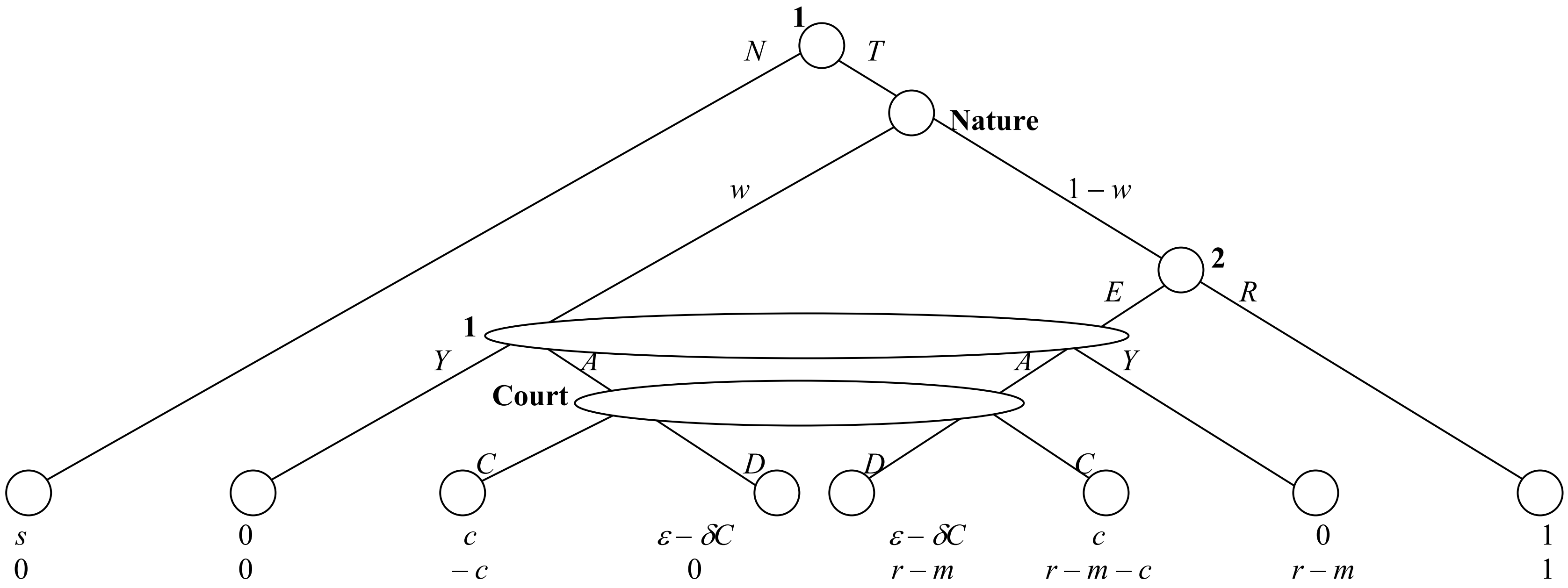

Figure 2.1. The trust game with litigation and legal insurance

Intrinsic payoffs guide behavior but have no direct impact on material or reproductive success. Whereas ε, the satisfaction for not voluntarily yielding to non-delivery, is assumed as a given non-material reward (introduced merely to solve behavioral ties), the other non-material payoff component *m* will be shaped by evolution depending on the relative reproductive success of the possible *m*-values.

If legally insured ($\delta = 0$), player 1 has not to pay the legal cost *C* when the court has dismissed the case. Note, however, that player 1 does not rationally choose for or against legal insurance. Rather being legally insured or not are assumed to be inherited traits determined by evolutionary forces. One possible way to interpret the evolutionary diffusion process of legal insurance is to think of it as chosen by contentious individuals who like going to court whenever there is an opportunity to do so. It has been shown most impressively by Frank (1987 and 1988) that such emotional traits, that may not serve immediate material self-interest, may survive evolutionary competition. The difference here is, however, that such legal insurance also *directly* translates into material payoff consequences caused by the material costs and benefits of the legal institution.

So, we study the coevolution of trustworthiness *m* and legal insurance types $\delta = 0$ and $\delta = 1$. Since the court rules based on rational posterior beliefs, court decision making depends on the population composition what further complicates the analysis. Factually, the three institutions, namely

- legal insurance types ($\delta = 0$ and $\delta = 1$),
- trustworthiness ($m \in \Re$), and
- court ruling (the condition for *C* or *D*)

coevolve together.

If *m* is player 2's private information, the beliefs of player 1 concerning player 2's *m*-type are assumed to be commonly known in the tradition of Harsanyi (1967/68) and have to be captured by an additional, however, fictitious chance move in Figure 2.1 as illustrated by Figure 1.2. We employ the usual assumption (Güth, 1995) that beliefs are determined by the actual frequencies of types and will thus adapt to changes in the population composition. In the same way one can capture private information concerning player 1's δ-type.

As in our earlier study (Güth and Ockenfels, 2000) the court plays an ambiguous role in our game model. On the one hand we do not specify the payoff function of the court (player). Actually, we view it as a considerable advantage that we do not burden our approach with more or less arbitrary assumptions concerning the motives of court decision making. On the other hand, we treat the court player as a rational belief forming institution. The latter is necessary since the predominant principle in civil law is that verdicts hold the party responsible whose guilt is most likely. Thus, courts have to solve the game as any other rational player in order to prepare the ground for applying Bayes-rule when deriving its posterior beliefs (see Güth and Ockenfels, 2000, for a more thorough discussion how this relates to legal practice in civil law). Whatever the solution behavior will be, it is rationally anticipated by the court. This is, furthermore, common knowledge of the other players as is their rationality.

The idea of rationally updating courts essentially denies superior judicial skill when judges try to find out individual responsibilities. In this sense, since courts are not equipped with detection capabilities beyond rational belief forming, our study is a worst case scenario for legal institutions. Nevertheless, as it will turn out, legal institutions can still serve as a substitute for reputation devices and detection capabilities and thus promote trust even in very unfavorable environments.[9]

Let z be player 2's probability for using R, and x or y player 1's probability for T, respectively A. For $y > 0$ the court's posterior probability in Figure 2.1 that 1 has been exploited by player 2 is well-defined and given by

$$e(z) = \frac{xy(1-w)(1-z)}{[w+(1-w)(1-z)]xy} = \frac{(1-w)(1-z)}{w+(1-w)(1-z)}.$$

To justify this formula even in case of $xy = 0$ we can rely on evolutionarily justified strategy trembles (Selten, 1983 and 1988) and the philosophical background of perfect equilibria (Selten, 1975) according to which the unperturbed game should be viewed as a limit of slightly perturbed games for which $xy > 0$ is guaranteed.

Since player 2 without legal help or legal advice is hold responsible if E is more likely than unintentional non-delivery, the court will decide for C (*c*onviction) only when $e(z) > 1/2$ or

---

[9] Other studies, such as Kirstein and Schmidtchen (1997), assume superior judicial detection capabilities. However, one wonders when such skills should develop (when studying or practicing law?) and why such skills are not more widely available. The cost argument (claiming that without superior judicial skills the costly legal system would not exist) is, of course, invalid (large and costly cathedrals or temples do not prove the existence of gods and goddesses).

$$w < \frac{1-z}{2-z} =: f(z).$$

Notice that $f(0) = 1/2$, $f(1) = 0$, and $f'(z) < 0$ for all $z \in [0, 1]$. Thus $f(z)$ can never exceed $w \geq 1/2$ what explains why we restrict the probability for non-delivery without player 2's responsibility to the range $0 < w < 1/2$.

If player 1 is insured and enjoys legal advice, it seems realistic to assume that the odds are improved in his favor.[10] In other words: Player 2 will be convicted even when $e(z) < 1/2$, more specifically, whenever $e(z) > \underline{e}$ with $0 < \underline{e} < 1/2$. With these specific rules governing court decision making for the two types $\delta = 0$ and $\delta = 1$ of the suing party, we have completed the description of the trust game with litigation and legal insurance.

In the following, we first solve the game for all constellations of δ- and m-types. With the help of the solution results we then will define the evolutionary framework in order to analyze the evolution of the population decomposition into $\delta = 0$ and $\delta = 1$ types as well as into the various m-types. Since different m-types of player 2 might rely on different choice probabilities and since the court rules differently when facing $\delta = 0$- and $\delta = 1$-types, the coevolution of δ- and m-types also determines the evolution of court decision making. This justifies our claim that we altogether analyze the coevolution of trustworthiness, legal insurance, and court decision making.

III. The solution for among players commonly known types

Let player 1 recognize player 2's m-value and vice versa let player 2 be aware of player 1's δ-type, whereas the court only learns the δ-type of player 1 when player 1 appeals and remains completely uninformed about 2's m-value. Thus, the court chooses C if $e(\hat{z}) > 1/2$ for $\delta = 1$ and $e(\hat{z}) > \underline{e}$ for $\delta = 0$ where $\hat{z}$ denotes the court's expectation concerning z based on the true composition $\Pi(\cdot)$ of m-types, i.e.

$$\hat{z} = \hat{z}(\Pi) = \int z(m) d\Pi(m).$$

where $z(m)$ is the probability for 2's move R by the m-type of player 2.

Let us first try to intuitively capture what will happen. If player 1 recognizes player 2's m-type, he obviously will choose T with maximal probability and, in case of non-delivery, be absolutely sure that this is due to bad luck and no intentional non-delivery by player 2. If $\delta = 0$, this will not prevent player 1 from appealing to the court due to $\varepsilon > 0$.

Let us therefore consider the court's ruling in case of an appeal. Since an intentional choice of T implies $x > 0$, actually $x = 1$, and $z = 1$ and since for $q > 0$ the probability of appeal is positive, the court's posterior probability is

$$e(z = 1) = 0, \text{ i.e. } e(z = 1) < \underline{e}.$$

Thus, both δ-types would lose the court case when appealing and even $\delta = 0$-types will not gain materially by appealing. Since the (reproductive) costs K of legal insurance are positive this implies $q_t \to 0$ for $t \to \infty$. Furthermore, $m < r - 1$-types will be anticipated by N or $x = 0$ and thus earn zero whereas $m > r - 1$-types are rewarded by T or $x = 1$ and earn 1. Thus, it follows that $p_t \to 1$ for $t \to \infty$. Actually this will be the result of our indirect evolutionary analysis which claims that $(q^*, p^*) = (0, 1)$ is the evolutionarily stable population composition.

What this reasoning neglects is the possibility of out of equilibrium-beliefs $e(z)$ when the overall probability xy for appealing to the court is 0. When not being asked to rule, the court should attribute this to the fact that $x = 0$, i.e. that player 1 has met an $m < r - 1$-type. The court's out of equilibrium beliefs $e(z)$ can thus be reasonably assumed to be determined by $z = 0$ so that $e(z = 0) = 1 - w$ ($> 1/2$). Due to $c > s$ this, however, induces the choice of $x = 1 = y$ by both δ-types of player 1 what contradicts the assumption above that, in case of an appeal, the court should rely on $z = 0$. A purely intuitive reasoning about the likely outcome may thus imply the correct result but in an inconsistent way. This can be avoided by always assuming $xy > 0$ due to trembles so that the court's posteriori beliefs $e(z)$ are always well-defined.[11]

The intuitive reasoning above would be consistent if the court could react to the frequency of appeal. Frequent appeals must rely on $x = 1$ and are thus signaling trustworthiness ($m > r - 1$) if q and w are positive. Rare appeals would result from trembles what signals $x = 0$ and $m < r - 1$. Thus the court could react to the frequency of appeal by ruling D if often appealed to and by

---

[10] Such bias can, for instance, result from the fact that experienced lawyers are more likely to avoid obvious form errors such as missing deadlines or in presenting admissible evidence. A more debatable but nevertheless natural justification could be that judges, as many human beings, yield to pressure.

[11] Applying the Bayes-rule is clearly superior to relying on ad hoc-refinements for signaling games, especially since we have not specified a payoff function for the court player.

convicting if rarely employed. Such more detailed information of the court is, however, not captured by our trust game with court and legal insurance as presented above.[12]

In the following we distinguish three cases according to the range of the posteriori probability of exploitation and thus the court's rulings whose implications for 2's behavior will be obvious for player 2's various *m*-types.

(*a*) If $e(\hat{z}) < \underline{e}$ or

$$w > \frac{(1-\underline{e})(1-\hat{z})}{\underline{e}+(1-\underline{e})(1-\hat{z})} \tag{III.1}$$

both δ-types would lose when appealing. This, however, will not prevent the $\delta = 0$-type from suing player 2. Thus, in case of (III.1), only the $\delta = 0$-type will sue (choose *A*) whose population share is denoted by *q* with $0 \le q \le 1$.

When choosing between *E* and *R* player 2 is supposed to know player 1's δ-type. Since 2's move is, however, independent of $\delta \in \{0, 1\}$, the population share *q* does not matter for 2 in case of (III.1). Thus, player 2's choice only depends on his own *m*-type as expressed by

$$z = \begin{cases} 1 \text{ if } m > r-1 \\ 0 \text{ otherwise} \end{cases}$$

Whenever $m > r - 1$, we will refer to an $\overline{m}$-type of player 2, whereas *m*-types with $m \le r - 1$ we summarize as $\underline{m}$-types. Apparently, all what matters of the *m*-distribution $\Pi(\cdot)$ is the population share *p* with $0 \le p \le 1$ of $\overline{m}$-types in the population. Consequently, rational expectations of the court imply $p = \hat{z}$ so that (III.1) becomes

$$w > \frac{(1-\underline{e})(1-p)}{\underline{e}+(1-\underline{e})(1-p)},$$

or equivalently

$$p > \frac{(1-w)(1-\underline{e})-w\underline{e}}{(1-w)(1-\underline{e})}.$$

[12] One could assume that the conviction threshold depends not only on *e*(*z*) but also on the overall probability of appeal, which would, however, greatly complicate the problem.

The right hand side above is positive (recall that by $0 < \underline{e}$, $w < 1/2$, we have $1 - \underline{e} > w$), and smaller than 1. Thus, Case (a), i.e. $e(\hat{z}) < \underline{e}$, applies for

Case (a): $$1 \geq p > \frac{(1-w)(1-\underline{e}) - w\underline{e}}{(1-w)(1-\underline{e})}$$

(b) If $\underline{e} < e(\hat{z}) \leq 1/2$ holds player 1's decision behavior regarding the choice between A and Y will not change since $\delta = 0$-player types now have even better reasons to choose A since they would win their case. A player 2 confronting a $\delta = 1$-type of player 1 will decide as before. If, however, player 1's type is $\delta = 0$, now z depends on m as follows:

$$z = \begin{cases} 1 \text{ if } m > r - 1 - c \\ 0 \text{ otherwise} \end{cases}$$

Consequently, $\hat{z}$ becomes

$$\hat{z} = \begin{cases} p \text{ if } \delta = 1 \\ \overline{p} \text{ if } \delta = 0 \end{cases}$$

where $\overline{p}$ is the population share with $1 \geq \overline{p} \geq p$ of m -types according to $\Pi(\cdot)$ satisfying $m > r - 1 - c$. Thus Case (b) applies for

Case (b): $$\begin{cases} 0 \leq p < \dfrac{(1-w)(1-\underline{e}) - w\underline{e}}{(1-w)(1-\underline{e})} \text{ for } \delta = 1 \\ 0 \leq \overline{p} < \dfrac{(1-w)(1-\underline{e}) - w\underline{e}}{(1-w)(1-\underline{e})} \text{ for } \delta = 0 \end{cases}$$

(c) The final case $e(\hat{z}) > 1/2$ applies if $z = p < \frac{1-2w}{1-w}$. Here both δ-types appeal and win what implies $z = 1$ and $xy = 1$ regardless of the other m- or δ-type. Since $z = p = 1$ excludes $e(z) > 1/2$, case (c) will be neglected.

Let us finally derive player 1's initial choice between T and N, denoted by $x := \text{Prob}\{T\}$, for the relevant cases solved above. If Case (a) applies, the $\delta = 1$-type in case of T would earn $1 - w$ when $m = \overline{m}$ and 0 otherwise, whereas choosing N yields s, i.e.

$$\delta = 1:\ x = \begin{cases} 1 \text{ if } m = \overline{m} \\ 0 \text{ otherwise} \end{cases}$$

The $\delta = 0$-type would always appeal and then earn ε so that

$$\delta = 0:\ x = \begin{cases} 1 \text{ if } m = \overline{m} \\ 0 \text{ otherwise} \end{cases}$$

due to $\varepsilon < s$.

For Case (*b*) the corresponding results are:

$$\delta = 1:\ x = \begin{cases} 1 \text{ if } m = \overline{m} \\ 0 \text{ otherwise} \end{cases}$$

and

$$\delta = 0:\ x = \begin{cases} 1 \text{ if } m > r - 1 - c \\ 0 \text{ otherwise} \end{cases}$$

For Case (*c*), we have $x = 1$ for $\delta \in \{0,1\}$.

How the solution depends on *m* and the δ-type is graphically summarized by Figure III.1, which, however, does not reflect the frequency of the various *m*-types as, for instance, captured by *p* and $\overline{p}$. For $r - 1 - c < m < r - 1$ and case (*b*) legal insurance ($\delta = 0$) guarantees reward ($z = 1$) whereas $\delta = 1$ leads to exploitation ($z = 0$). Thus, court rulings, when player 1 knows *m* and $r - 1 - c < m < r - 1$, are crowded out by legal insurance, as far as possible, when case (*b*) applies. This obvious competition between court rulings and preventive legal insurance needs, however, to be more thoroughly analyzed, since the relevance of case (*b*) and $r - 1 - c < m < r - 1$ has to be assessed in the form of an (indirect) evolutionary analysis.

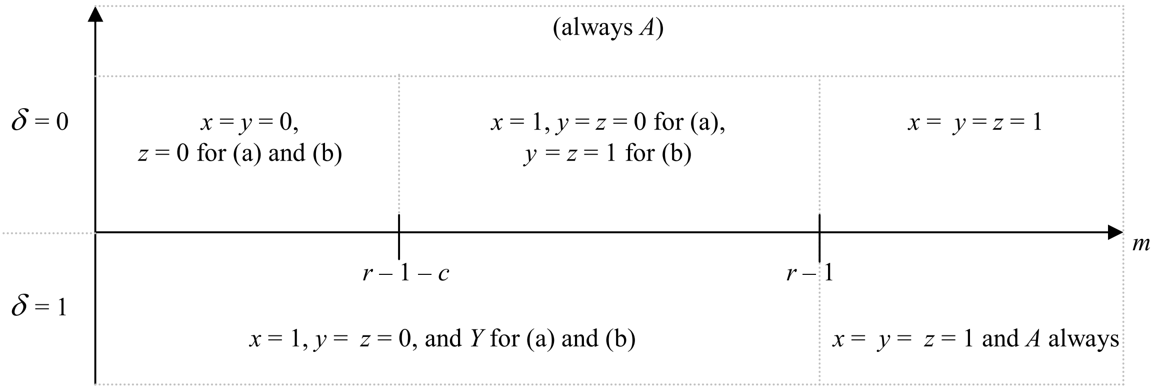

Figure III.1. The solution for among players commonly known types

IV. Expected reproductive success

Assume for the sake of the usual symmetry of evolutionary games that every individual of an infinite population plays two games, one as player 1 and one as player 2, with randomly chosen partners in the other role. Reproductive success is given by the solution payoff when setting ε and m equal to 0, i.e. by neglecting the purely intrinsic payoff components, and after subtracting the positive (reproductive success) cost K of legal insurance in case of $\delta = 0$. Case (c) does not apply. Case (a) rules out any conviction and thus $q > 0$ for any stable constellation due to $K > 0$. So only $\overline{m}$ -types will be trusted. Since an m-type earns 1 if $m = \overline{m}$ and 0 otherwise, this implies that p approaches 1 soon or later. For $q = 0$ and $p = 1$, furthermore, condition of case (a), namely

$$p = 1 > \frac{(1-w)(1-\underline{e}) - w\underline{e}}{(1-w)(1-\underline{e})}$$

is equivalent to $w\underline{e} > 0$ and thus true. This proves that case (a) contains just one evolutionarily stable (q, p)-constellation: $(q^*, p^*) = (0, 1)$.

In case (b), a $\delta = 0$-player earns 1 if his partner in the role of player 2 is of type $m > 1 - r - c$, what happens with probability $\overline{p}$. With complementary probability $1 - \overline{p}$, he chooses N and earns s. A $\delta = 1$-player 1 earns 1 only with probability p and s otherwise. Thus, the expected payoff from being player 1 is

$$U_1(\delta) = \begin{cases} \overline{p} + (1-\overline{p})s \text{ for } \delta = 0 \\ p + (1-p)s \text{ for } \delta = 1 \end{cases}$$

In the role of player 2 payoff expectation depends on the likelihood q of $\delta = 0$ and on the m-parameter as follows:

$$U_2(m) = \begin{cases} 1 \text{ for } m = \overline{m} \\ q \text{ for } r-1-c \leq m < r-1 \\ 0 \text{ for } m < r-1-c \end{cases}$$

Thus, the overall payoff expectation $U(\delta, m)$ of a δ, m-type individual in case (b) is

$$U(\delta, m) = \begin{cases} \overline{p} + (1-\overline{p})s - K \text{ for } m < r-1-c \text{ and } \delta = 0 \\ \overline{p} + (1-\overline{p})s + q - K \text{ for } r-1-c \leq m < r-1 \text{ and } \delta = 0 \\ p + (1-p)s + 1 - K \text{ for } m > r-1 \text{ and } \delta = 0 \\ p + (1-p)s \text{ for } m < r-1-c \text{ and } \delta = 1 \\ p + (1-p)s + q \text{ for } r-1-c \leq m < r-1 \text{ and } \delta = 1 \\ p + (1-p)s + 1 \text{ for } m > r-1 \text{ and } \delta = 1 \end{cases}$$

where K with $0 < K < c$ denotes the reproductive cost of legal insurance. If $\overline{p} - p > K/(1-s)$, legal insurance yields a higher reproductive success, i.e. legal insurance is only superior when enough m-types in the range $r - 1 - c \leq m < r - 1$ exist. Furthermore, for all $q < 1$ the $\overline{m}$-type earns the highest success regardless whether $\delta = 0$ or $\delta = 1$. But then soon or later all intermediate m-types with $r - 1 - c \leq m < r - 1$ disappear so that soon or later $K/(1 - s)$ will become larger than $\overline{p} - p$. With p and $\overline{p}$ approaching 1, the conditions for case (b) are soon or later violated what proves that case (b) contains no stable (q, p)-constellation. We thus have altogether only one stable constellation, namely the one of case (a).

Proposition 1. The only evolutionarily stable type constellation (δ, m) is given by $\delta = 1$ and $m = \overline{m}$ when types are commonly known among players, i.e. ($q = 0$, $p = 1$) is the evolutionarily stable population composition.

As already shown by Güth and Kliemt (1994) for an environment without modern institutions type information crowds in morality. Here we show that even if one allows for courts and legal insurance, type information renders such institutions as useless or crowds them out. (Note that in case of unintentional non-delivery the court is not activated due to $q = 0$.) Let us refer to societies where interacting partners know each others' types well enough as tribal communities. Then Proposition 1 shows

- that tribal communities must not logically exclude modern institutions like courts and legal insurance,
- but that such institutions would be crowded out by type information.

Thus the main aspect of modern life may not be its institutions as such but rather its anonymity questioning that one knows the types of those with whom one is interacting. Whether or not such anonymity crowds in or out modern institutions can be analyzed more rigorously by studying the case of private information.

V. Privately known types

In this section neither player 1 can recognize player 2's m-type nor player 2 player 1's δ-type. It is assumed that the beliefs concerning m, respectively δ, are determined by the true population composition. As before we distinguish the three cases of court ruling.

(*a*) If (III.1) applies the result that only $\delta = 0$ appeals remains valid. As before q denotes the population share of $\delta = 0$-types. For player 2 of type m, R is better than E if $m > r - 1$ in case (*a*) regardless of $\delta \in \{0,1\}$ or $q \in [0,1]$. We denote by p the population share of m-types satisfying this inequality. For player 1 of type $\delta = 1$ the choice of T is better than N if $p > s/(1 - w)$, whereas for the $\delta = 0$-type this requires $p(1 - w) + (pw + 1 - p)\varepsilon > s$, or

$$p > \frac{s-\varepsilon}{1-\varepsilon}$$

One always has

$$0 < \frac{s-\varepsilon}{1-\varepsilon} < \frac{s}{1-w} < 1$$

due to our parameter restrictions. Thus solution behavior depends on parameter p as illustrated by Figure V.1 where one, of course, has to keep in mind the initial assumption (III.1) which can be expressed by p as

(III.1) $$p > 1 - \frac{w\underline{e}}{(1-w)(1-\underline{e})} = \frac{1-w-\underline{e}}{(1-w)(1-\underline{e})}$$

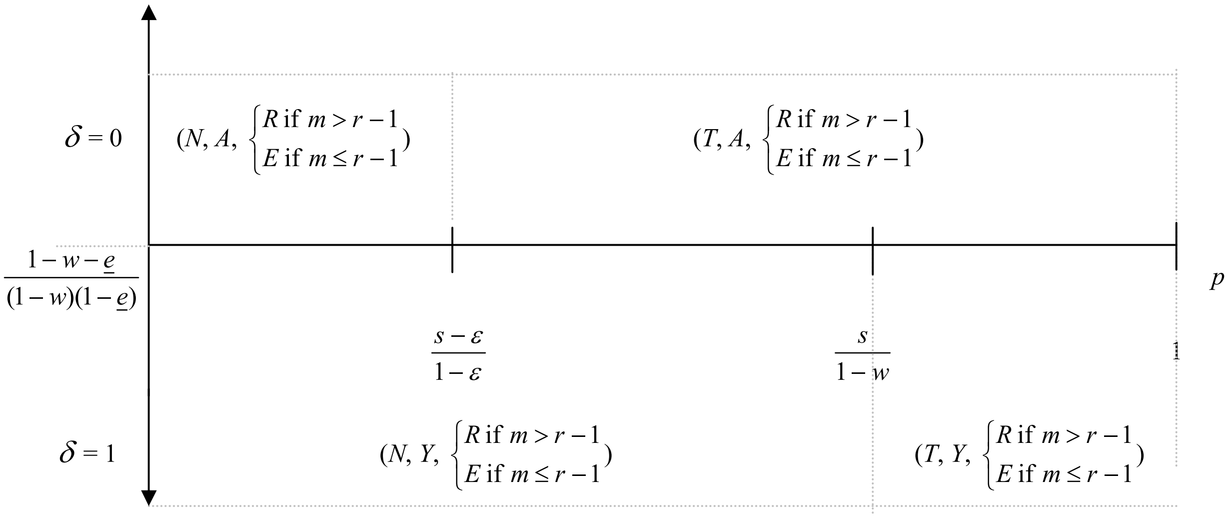

Figure V.1. The solution for Case (*a*)

(*b*) When $\underline{e} < e(\hat{z}) \le 1/2$, δ = 0-types appeal and δ = 1-types not. Thus, $p = p(q)$ is the population share of *m*-types satisfying $m > r - 1 - qc$and δ = 1-types use *T* if $p > s/(1-w)$ whereas for δ = 0-types the corresponding condition is now $p(1-w) + (pw + 1 - p)c > s$, or

$$p > \frac{s-c}{1-c}.$$

Note that

$$0 < \frac{s-c}{1-c} < \frac{s}{1-w} < 1$$

always holds. Since $\hat{z} = p$ case (*b*) thus applies when

$$\underline{e} < \frac{(1-w)(1-p)}{w+(1-w)(1-p)} \le \frac{1}{2}, \text{ or}$$

$$0 < \frac{1-2w}{1-w} < p < \frac{1-w-\underline{e}}{(1-w)(1-\underline{e})}$$

due to 0 < w < 1/2. The result can again be graphically illustrated:

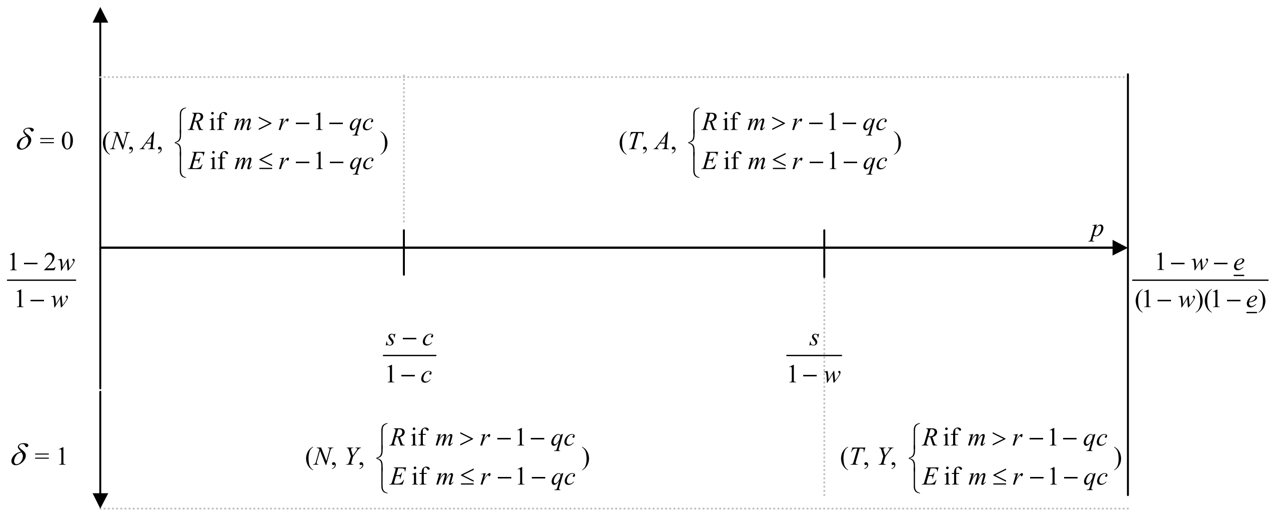

Figure V.2. The solution for Case (*b*)

(*c*) If $e(\hat{z}) > 1/2$ player 1 will always appeal and thus all *m*-types as player 2 reward, i.e. $z = 1$ if one excludes extreme spite in the sense of $m < r - 1 - c$ (< 0). This in turn implies $y = 1$ for $\delta \in \{0,1\}$. Since $z = p = 1$ excludes $p < (1 - 2w)/(1 - w)$, case (*c*) cannot sustain for long as before.

In Case (*a*) or

$$p > 1 - \frac{w\underline{e}}{(1-w)(1-\underline{e})}$$

the result depends on the subcase. If

$$p < \frac{s-\varepsilon}{1-\varepsilon}$$

both δ-types choose *N* so that only $q = 0$ can be stable. But for $q = 0$, trust requires $p > s/(1 - w)$ what excludes the subcase $p < (s-\varepsilon)/(1-\varepsilon)$. On the other hand, if $p > s/(1 - w)$, both δ-types trust (*T*) but only δ = 0-types appeal. Since in case (*a*) the court dismisses the case and since ε is not directly related to (reproductive) success, the success net of the insurance premium is the same for both δ-types, so that only $q = 0$ can be stable. But then m-types with $m \leq r - 1$ earn materially r and m-types with $m > r - 1$ only 1. Thus p will decline till it leaves case (a),[13] i.e.

$$p = p(q \to 0) \to \frac{1 - w - \underline{e}}{(1 - w)(1 - \underline{e})}$$

Let us now consider the coevolutionary process for the two remaining cases (a) and (b) of court ruling. In case (b) δ = 0-types appeal and win their court case whereas δ = 1-types yield. If both δ-types choose would N, only $q = 0$ can be stable. But then p becomes the population share of m with $m > r - 1$. This in turn leads to $p = 0$, when rare (mistaken) choices of T are assumed (see Selten, 1983). Since condition (b) excludes $p = 0$ due to $w < 1/2$, this subcase does not apply.

If both δ-types would choose T, a δ = 0-type earns $p(1 - w) + (pw + 1 - p)c - K$ and a δ = 1-type $p(1 - w)$. For

$$\frac{c - K}{c(1 - w)} < p$$

this leads to $q \to 0$ and, by the same argument as above, to $p = 0$ and a contradiction to condition (b). If, however,

$$\frac{c - K}{c(1 - w)} > p \; (> \frac{1 - 2w}{1 - w}) \qquad (+)$$

the share q would approach 1. But for $q \to 1$, $p = p(q)$ becomes the population share of $m > r - 1 - c$. Any $m > r - 1 - c$ would earn 1 whereas, for $q \to 1$, an m-type with $m < r - 1 - c$ would nearly certainly be sued and lose the court case, i.e. nearly always earn $r - c$ only. Due to $r - c < 1$ we thus would obtain an upward adaptation of $p = p(q)$ till p leaves the region where condition (b) holds,[14] i.e.

$$p = p(q \to 1) \to \frac{1 - w - \underline{e}}{(1 - w)(1 - \underline{e})} \quad \text{for} \quad \frac{c - K}{c(1 - w)} > p\,.$$

---

[13] Since p and q are moving into the same direction, the relative speed by which p, respectively q declines over time does not matter.

[14] Again, p and q are moving into the same direction.

Proposition 2. If condition (+) holds, a restpoint of the adaptive process of q and p is given by

$$p^* = \frac{1-w-\underline{e}}{(1-w)(1-\underline{e})}.$$

For $p < p^*$ this is brought about by $q \to 1$ inducing p to increase and for $p > p^*$ by $q \to 0$ inducing p to decline.

The remaining subcase of (*b*) is

$$\frac{s-c}{1-c} < p < \frac{s}{1-w} \qquad (*)$$

which induces T, A for $\delta = 0$ and N, Y for $\delta = 1$. Success in the role of player 1 therefore depends on δ as follows:

$$U_1(\delta) = \begin{cases} p + (pw + 1 - p)c - K \text{ for } \delta = 0 \\ s \text{ for } \delta = 1 \end{cases}$$

For $\delta = 0$ to be superior one must have

$$p > \frac{K+s-c}{1-c} \left(> \frac{s-c}{1-c}\right),$$

a condition that is always satisfied for sufficiently small costs $K > 0$ in the range of the subcase under consideration. Thus subcase (*) of (*b*) leads to $q = 1$ for small K what makes p the share of m-types with $m > r - 1 - c$. Since by the same argument as above this renders m-types with $m > r - 1 - c$ more successful, $p = p(q \to 0)$ will increase till $p = p(q = 1) = s/(1 - w)$, so that ($q = 1$, $p = s/(1 - w)$) is the stable result. More substantial cost K of legal insurance may either rely on

$$\frac{K+s-c}{1-c} > \frac{s}{1-w}$$

or on

$$\left(\frac{s}{1-w} >\right) \frac{K+s-c}{1-c} > p > \frac{s-c}{1-c}.$$

In both cases one would observe $q \to 0$ and thus a decline of $p = p(q \to 0)$ till $(s - c)(1 - c)$, the lower bound of subcase (*) of case (*b*). But then $p = p(q \to 0) \to 0$ results.

Proposition 3. If m- and δ-types are private information, the situation where everybody is legally insured and where only the share $s/(1 - w)$ for $s + w < 1$ is trustworthy qualifies as a rest point, if compared to the initial value p_0 of trustworthy types, cost K of legal insurance are relatively small in the sense of $p_0 > (K + s - c)/(1 - c)$. Otherwise, only $(q = 0, p = 0)$ qualifies as a rest point.

Even when type information is private there is thus hope for a bimorphism where some trustworthy are being threatened by litigation and others are not. Furthermore, there is path dependence. Once p drops below the threshold $p_0 = (K + s - c)/(1 - c)$, there is no hope for recovery.

VI. Conclusions

In our evolutionary models, the emergence of trust is either due to detection capabilities (Proposition 1) or due to legal institutions such as courts and legal insurance (Proposition 3) depending on the underlying information scenario. In particular, information about m-types crowds out court rulings and the need to be legally insured, but legal insurance can be crowded in as a general behavioral trait if types are private information, provided that legal insurance is not too costly. If so, this stabilizes an interesting m-bimorphism in the form of an $s/(1 - w)$-share of trustworthy individuals in the population. Given our analysis, it seems conceivable that court rulings and legal insurance became crucially important when mankind switched from rural village or tribal life, where commonly known types are rather natural, to modern societies with large metropolitan areas, where information about others' types is rather unlikely. In the latter world institutions like courts and legal insurance may guarantee some reliability when trading or interacting with partners whose reliability is uncertain, even when courts are not equipped with any superior type detection technologies than all other players.

References

Bar-Gill, Oren, and Chaim Fershtman (2001): The Limit of Public Policy: Endogenous Preferences, working paper, Center for Economic Research.

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.

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 (forthcoming): Evolutionarily Stable Co-operative Commitments, *Theory and Decision*.

Güth, Werner, and Axel Ockenfels (2000): Evolutionary Norm Enforcement, *Journal of Institutional and Theoretical Economics*, 156(2), 335-347.

Güth, Werner, and Axel Ockenfels (forthcoming): The Coevolution of Trust and Institutions in Anonymous and Non-anonymous Communities, in: M.J. Holler, H. Kliemt, D. Schmidtchen and M. Streit (eds.), *Jahrbuch für Neue Politische Ökonomie*, 20, Tübingen: Mohr Siebeck.

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

Kirstein, Roland, and Dieter Schmidtchen (1997): Judicial Detection Skill and Contractual Compliance, *International Review of Law and Economics*, 17(4), 509-520.

Ockenfels, Axel, and Reinhard Selten (2000): An Experiment on the Hypothesis of Involuntary Truth-Signaling in Bargaining, *Games and Economic Behavior*, 33(1), 90-116.

Selten, Reinhard (1975): Re-examination of the Perfectness Concept for Equilibrium in Extensive Games, *International Journal of Game* Theory, 4, 25-55.

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

Selten, Reinhard (1988): Evolutionary Stability in Extensive Two-person Games: Corrections and Further Developments, *Mathematical Social Sciences*, 16, 223-266.