From Teleology to Evolution

Bridging the gap between rationality and adaptation in social explanation

Siegfried Berninghaus, Werner Güth and Hartmut Kliemt

Abstract

This paper focuses on the uneasy alliance of rational choice and evolutionary explanations in modern economics. While direct evolutionary explanations of “optimality” rule out "purposeful" rational choice by assuming zero-intelligence and pure rational choice explanations leave no room for "selective" adaptation the indirect evolutionary approach integrates both perspectives. Subsequently we go stepwise "from teleology to evolution" and thereby study the model spectrum ranging from pure rational choice over indirect to direct evolutionary approaches. We believe that knowledge of this spectrum can help to choose more adequate models of economic behavior that incorporate both teleological and evolutionary elements.

1. Introduction

In neo-classical economics all acts are explained by expectations and evaluations of future effects of action as endorsed by the rational actors themselves. Since rational actors typically act in the presence of other rational actors they must also form expectations about what these other rational actors do. As far as this is concerned neo-classical economics assumes a strong form of “theory absorption” as spelled out in full by modern non-co-operative game theory: The theory explaining the actions of rational actors is known to the actors and is in fact applied by them to choose their own actions. The actors are guided by a theory of rational action and conceive all other actors like themselves as being guided by the same theory of rational action (theory absorption is mentioned in a market context for instance by (Morgenstern, Oskar and Gerhard Schwödiauer 1976), and, in a more philosophical vein discussed in (Dacey, Raymond 1976)). In strategic contexts with multiple equilibria this even requires that all players employ the same selection rule for choosing among equilibria. Moreover, everybody knows everybody to know that everybody knows that everybody is rational, follows the theory defining what is rational, knows everybody to do so etc. (see on this (Aumann, Robert J. 1976), (Mertens, Jean-Francois and Shmuel Zamir 1985), and the fundamental study (Fagin, Ronald, Joseph Y. Halpern, Yoram Moses, and Mosche Y. Vardi

1995)). All this is comprised – though not always recognized as such – in the assumption of "commonly known rationality" that is fundamental for a strict rational choice approach to human behavior.

The preceding assumptions are of crucial importance for the philosophically challenging and interesting task of forming a view of a world of completely rational beings endowed with unlimited faculties of reasoning.[1] But as a model of real world social phenomena the neo-classical and game theoretic economic approach to explaining human behavior despite all its formal and analytical charms seems doomed. Therefore some economists have suggested that the traditional rational choice approach to explaining social behavior be substituted by models borrowed from (evolutionary) biology and/or from (learning) psychology. However, the complete elimination of all purposeful choice from the explanation of social behavior amounts to throwing out the baby with the bathwater. There is purposeful forward-looking choice and thus a "teleological" element in real world behavior of higher organisms that must not be neglected. Such organisms are not merely driven by past experience, they are drawn by their expectations. And humans even command the faculty to make forward-looking rational choices.

What is wrong with the traditional rational choice paradigm is not that it takes into account purposeful behavior but rather that it leaves no room for anything else and thereby becomes one-sided. Likewise the traditional evolutionary approach which serves as a paradigm for social theorists from the old Social Darwinists over Hayek to modern evolutionary game theorists tends to become one-sided, too, by eliminating purposeful rational choice altogether. In former work (see in particular (Güth, Werner and Hartmut Kliemt 2000), (Güth, Werner, Hartmut Kliemt, and Bezalel Peleg 1999)) we tried to enter the middle ground between the

---

[1] Relying on a somewhat strange but in this regard instructive Kantian terminology we could say that pure rational choice models deal with the "homo noumenon" while the empirical theory of behavior is dealing with the "homo phaenomenon". But why as in traditional game theory engage the task of analyzing a world populated by fully rational beings? According to Kant we as humans conceive ourselves always as members of both the "noumenal" and the "phaenomenal" world. In doing so we are not only interested in predicting behavior but also in a kind of self-interpretation as purely rational beings. Even though we know that we are not purely rational we are deeply interested in the question of what would emerge if we were; see Kant, Immanuel. 1991. *Political writings. The metaphysics of morals*. Oxford et al.: Oxford University Press..

extremes by integrating purposeful action as well as adaptive roots of human behavior in an “indirect evolutionary approach”. Subsequently we include the extremes and explore for the first time the full spectrum ranging from models based exclusively on “farsighted teleology” to models that explain phenomena in terms of “blind evolution” only. For our exploration we construe a sequence of models each presenting a specific view of the same social interaction of contracting in a large group of potential partners. In the sequence of models the role of teleology decreases “stepwise”. Starting with the most extreme case in which all choices are explained as “purposeful action” of "rational economic men" we gradually substitute rational choice by "blind adaptation" until “purposeful action” and rationality are completely eliminated as explanatory variables of the model.

More specifically, we consider the following cases of interaction models arranged according to the decreasing degree in which teleology or conventional rational choice assumptions are utilized to explain contracting and the underlying problem of (non-)trustworthiness:

case a. all contractual elements including their own trustworthiness are rationally chosen by the players,

case b. the trustworthiness or untrustworthiness of players in contracting evolve whereas everything else is rationally decided,

case c. trustworthiness and the tendency to acquire information about the trustworthiness of others in contracting co-evolve,

case d. contractual choices are not made strategically but rather determined by fixed behavioral programs that evolve.

We hope that after going through the sequence of models a clearer view of the relative merits of rational choice and adaptive modeling of behavior emerges. Section 2 gives a stylized account of the class of social situations that we intend to model with the four cases of game models mentioned before. In section 3 we introduce the basic games that we shall subsequently discuss. In section 4 we embed the basic game in a more comprising one and solve the larger game for the extreme case of purely forward looking deliberation or case a. Sections 5 and 6 characterize solutions of the larger game for the intermediate cases b and c respectively. Section 7 discusses the other extreme case d of direct evolution. In section 8 we shall draw some essential methodological conclusions from our “guided tour” reaching “from teleology to evolution".

2. Pairwise interaction in large societies

Even though most human life takes place within small face to face groups of permanently interacting individuals, interactions involving so many individuals that none of them can keep track of the behavior of all individuals become increasingly important in the modern world. Individuals meet and interact often without knowing much about each other. If not all fruits of co-operation are to be foregone individuals must nevertheless try to co-operate in such situations. They can do so in particular by entering bilateral (quasi-)contractual relationships that aim at mutual gain. To come to grips with this bi-lateral co-operation problem of individuals who are more or less anonymous to each other we make basically three modeling assumptions. Firstly, that the society is large in the sense that conditions akin to anonymity prevail is approximated by the assumption of a pool of infinitely many individuals. Secondly, the presence of both the chance of mutual gain and the risk of default in dyadic (contractual) relationships is approximated by the assumption that the individuals play basic trust games in which trust can be rewarded or be exploited on each round of play. To reduce some of the complexity and to allow for an analytical treatment it is assumed, thirdly, that individuals cannot strategically choose their partners but are randomly matched to form pairs of players engaged in dyadic interaction.

In our model there is no "discipline of continuous dealings" since players substitute each other on each round of play. They do not have a choice with whom to play but they can decide on whether they are going to contract or not with the partner assigned to them randomly and whether or not they are going to fulfill obligations. Since contracts can neither specify all contingencies nor be perfectly monitored an essential trust problem emerges. In contractual relationships there is always a temptation to exploit the trust of others who have already done their part.

Individuals are not only able to seize opportunities of mutually beneficial contracting but also of not fulfilling their contractual promises. Therefore individuals must meet the challenge arising from opportunism. We assume that even though the pool of potential partners is too large to keep track of every prior move and the likely type of others there is often a chance to receive some information about partners at some cost. To incur such costs may be worthwhile since not all individuals behave purely opportunistically. There are in fact (types of)

individuals who are intrinsically motivated to fulfill contracts while others are not and we sometimes can know who they are with some reliability.[2]

3. The basic trust game

The example that we use to illustrate the methodological issues at hand starts from what we call the "simple game of trust":

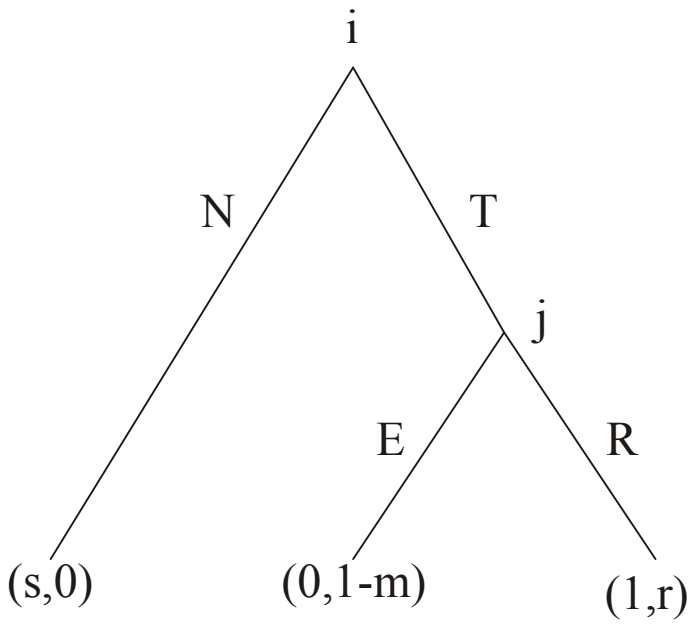

Figure 1: The basic game of trust with payoff parameters 0<r, s<1, m≥0

In the game of Figure 1 player i starts by deciding between N(o-trust) and T(rust). After *N* the game ends with player i earning s and player j earning 0. After *T* the game continues with j's choice between E(xploitation) and R(eward). All payoff parameters, except m represent material reward and, in the context of an evolutionary analysis, may be interpreted as measures of "reproductive" success affecting the relative share of different "types" in the population evolving in the course of time.

---

[2] Conditions of random matching eliminate the folk-theorem logic and it can be studied in isolation how the presence of both intrinsically motivated individuals and some knowledge about them influences the possibility of bi-lateral co-operation in large groups of potential partners.

In an indirect evolutionary approach both, subjective and objective payoffs do play a role. A subjective payoff function and an objective payoff function apply simultaneously. The former is driving choice the latter selection. The subjective and the objective payoffs are represented by the same numerical payoff function. We assume that except for m, actors are motivated exclusively by the material or objective payoffs involved; i.e. they subjectively evaluate the states of the world according to the emergent states' contributions to their individual material or objective success.

Using the same numerical representation for both subjective payoffs representing preferences and objective or material payoffs representing differential "reproductive" success reduces the number of parameters without unduly restricting the substantial content of the dual approach. It also allows for a simple connection of both functions: The objective or material payoff function ("reproductive" success) crucial for evolution emerges by setting m=0 after deriving the solution for all m-types. If m=0 the individual is exclusively motivated by "extrinsic" or "material" rewards as measured by the corresponding objective success function with the same values. With m≠0 a utility function with motives other than factors directly relevant for evolutionary success emerges. If m is positive, we will often speak of "regret" or the presence of a "conscience". If m>1-r her conscience induces the second moving player j to choose *R* rather than *E*. In that case the conscience is sufficiently strong to become "behaviorally effective".

The factor m represents a purely "intrinsic" motive. It is not a measure of objective success but affects that success via potentially influencing behavior. Behaviorally all values of m with m>1-r are equivalent. Subsequently we will therefore assume that whenever m>1-r applies m is fixed at an arbitrary but specific behaviorally effective m=$\overline{m}$>1-r that dictates the choice of *R* in the second-mover role in the simple game of trust.

Likewise if m<1-r individuals in the second-mover role will show the same behavior as individuals who are solely motivated by material payoffs. The motives expressed by m are not strong enough to be behaviorally effective. All m<1-r are behaviorally equivalent. This equivalence class is represented by m=$\underline{m}$.[3]

Since m is *not* directly related to evolutionary success and since all m that fall in the same of the two classes of behaviorally equivalent values of m affect behavior in the same way the

---

[3] If m<0 some malevolence or some negative emotion may be present but will not alter the incentives that are present anyway.

specific value of m – except for falling in one of the classes – is evolutionarily irrelevant. This is brought out also by observing that trustworthy individuals with m>1-r who for some reason or other might make a mistake and regardless of m>1-r choose E in the second-mover role will receive 1 unit of objective or material payoff. Even though the "guilt"-factor m>0 is kicking in – thereby reducing their "subjective satisfaction" below 1 – they receive the same objective payoff as those who lack a conscience altogether or have one that is behaviorally ineffective. The extrinsic reward is the same for all individuals showing the same kind of overt behavior regardless of how they evaluate it subjectively.

The (game) model in Figure 1 describes the archetype of a one-sided trust situation. Though important as an "ideal type" it is as such not of much interest. But basic games of trust are embedded in richer social structures. These richer structures lead to more interesting and more complex interactions. We will analyze a class of such interactions in which

m is assumed to be player j's private information

and

player i can acquire some information about player j's m-type at some cost C.

In the extreme case of purely rational deliberation (the influence of the “shadow of the past” is completely lacking) the m-types will be chosen rationally by the actors themselves. They make these choices of their own dispositions by anticipating the future implications of being endowed with a “conscience” leading to “regret” or not (see for a different, less explicit neo-classical discussion of choosing a conscience (Frank, R. 1987)). In the remaining cases of (in)direct evolution m or rather the behavioral dispositions it represents evolve depending on the past differential success of the various m-types. But let us start with the first extreme case in which the players, in a way, can choose their own type (subjective utility function) operative in the basic game of trust as embedded in the larger interaction.

4. Explanation in terms of forward-looking deliberation only

4.1. The basic game of trust as embedded in fully rational choice

To simplify our illustration of standard rational choice modeling of the game we assume that two players from a large population of players have been chosen to play the game that serves

as a model of bilateral interaction in a large population.[4] When making their decisions on prior stages of the game players do not know whether in the final trust game they will be assigned (with equal probability) to the first- or the second-mover role. The random move of allocating individuals to both roles with equal probability is included since we intend to apply our sequence of models to the same problem. This being said, imagine the following decision process for the two individuals k=i, j with i≠j:

Stage 1: Individuals k decide to become either

trustworthy by "committing" to $m_k>1-r$

or

untrustworthy by "committing" to $m_k<1-r$.

The results of these individual decisions remain private information. But nature provides a signal of the ensuing type distribution. After stage 1 the fraction p of trustworthy $m_k>1-r$ individuals in the population is common knowledge.

Stage 2: Before actually playing the basic trust game of Figure 1 and before knowing whether they shall end up in the first-mover role all individuals k "commit" to become either of "type"

U meaning that such an *u*ninformed player k_U does – not invest in a detection technology or does – not invest in information search about his co-player's type,

or

[4] In subsequent evolutionary analyses it will be assumed that there is an infinite pool of players who are matched randomly to play the game in which the basic game of trust is embedded. For the time being we need to consider strategic considerations after matching only.

I meaning that such an informed player k_I incurs a cost C(≥0) for investing in a detection technology or in information search about his co-player's type.

Since the solution will not depend on it, it may be left open what players might learn about each other's choice of U or I.

Stage 3: On this stage "nature" moves.

An unbiased chance move

decides who becomes first- and who becomes second-mover in the basic trust-game.

Also a stochastic signal revealing the second-mover type with a certain reliability is provided for those individuals in the first mover-role who decided on investing in an information technology on previous stages of the game.

Without loss of generality, assume that of the individuals i, j individual i ends up in the first- and j, j≠i, in the second-mover role. If i has decided to become an informed type i_I at stage 2 of the game then i receives a signal M informing him about the trustworthiness of the second-moving player. M=$\overline{M}$ signals a trustworthy type $\overline{m}$ in the second-mover role. M=$\underline{M}$ signals an untrustworthy type $\underline{m}$. The signal is of reliability $1>\underline{\mu}>1/2$ if originating from an untrustworthy $\underline{m}$-type and of reliability $1>\overline{\mu}>1/2$ if originating from a trustworthy $\overline{m}$-type. Which signal i receives in the first mover-role, if he became an i_I type on stage 2, is decided by a move of chance. If the second-mover is an untrustworthy $\underline{m}$-type then with probability $\underline{\mu}>1/2$ the signal $\underline{M}$ will indicate the co-player type correctly to the informed i_I-type first-mover. With probability $1-\underline{\mu}$ an incorrect signal $\overline{M}$ indicating a trustworthy $\overline{m}$-type will be received by i_I. Likewise, with probability $\overline{\mu}$ the signal $\overline{M}$ will correctly indicate the presence of a trustworthy type $\overline{m}$ while with probability $1-\overline{\mu}$ the signal will be $\underline{M}$, indicating an untrustworthy

co-player of type $\underline{m}$ even though the second-mover is in fact a trustworthy $\overline{m}$-type.

Stage 4: The first-mover chooses between *N* and *T*.

After *N* the game ends. Still assuming that individual i plays in the first-mover role, he receives a payoff of $s-\delta_i C$ and individual j in the role of the second-mover receives a payoff of $0-\delta_j C$;

where $\delta_k = \begin{cases} 0 & in\ case\ of\ U_k \\ 1 & in\ case\ of\ I_k \end{cases} \quad for \quad k = j,i$

After choosing *T* the game continues.

Stage 5: The second-mover decides between *E* and *R*.

Assume still that i is first- and j second-mover. After *E* the game ends with a first-mover payoff of $0-\delta_i C$ and a second-mover payoff of $1-m_j-\delta_j C$. After *R* the game ends with a first-mover payoff of $1-\delta_i C$ and a second-mover payoff of $r-\delta_j C$.

This completes the description of the first model in the sequence. In this initial extreme case "teleology" or purposeful decision-making is extended to the choice of player types. Along with other phenomena the prevalence of player types is "explained" or "predicted" solely in terms of (sequentially) rational choices as derivable from the conventional game theoretic logic of solving a sequential game with incomplete information by means of backward induction.

4.2 Solving the game backwards

Still assuming that individual i moves first and individual j second, we can analyze the game as follows:

On stage 5 the second-mover j's decision depends solely on her m-type:

m_j>1-r leads to the choice of *R* and

m_j<1-r leads to the choice of *E*.

On stage 4 we have to distinguish players i_U who have *not* received a specific signal about the second-mover's type and players i_I who have received a signal $\overline{M}$ or $\underline{M}$ conveying specific type information about the second-mover.

Player i_U

The beliefs of a player i_U are determined by the general signal concerning the population share p of individuals k who have chosen m_k>1-r on the first stage. Since we are assuming an infinite population the player's private information on her own type is irrelevant. The optimal behavior of a player i_U with beliefs determined solely by the general signal is to choose

T if p>s

N if p<s.

Player i_I

A player i_I who has received specific type information about her co-player in form of the signal $\overline{M}$ expects a trustworthy $\overline{m}$-type with probability

$$P(\overline{m}/\overline{M}, p)=\frac{p\,\overline{\mu}}{p\,\overline{\mu}+(1-p)(1-\underline{\mu})}.$$

For p≥0 and 1>$\overline{\mu}$, $\underline{\mu}$>1/2 this probability is always well-defined.[5] A player i_I will choose

T if $P(\overline{m}/\overline{M}, p)$>s

[5] For p=0 the case $\underline{\mu}$=1 can be analyzed as the limit of $P(\overline{m}/\overline{M}, p=0)$ as $\underline{\mu}$ approaches 1.

N if $P(\overline{m}/\overline{M}, p)<s$.

A player i_I who has received specific type information about her co-player in form of the signal $\underline{M}$ expects (nevertheless) a trustworthy $\overline{m}$-type with probability

$$P(\overline{m}/\underline{M}, p)=\frac{p(1-\overline{\mu})}{p(1-\overline{\mu})+(1-p)\underline{\mu}}$$

For $p\geq 0$ and due to $1>\overline{\mu}$, $\underline{\mu}>1/2$ this probability is also always well-defined. Player i_I will choose

T if $P(\overline{m}/\underline{M}, p)>s$

N if $P(\overline{m}/\underline{M}, p)<s$.

On stage 3 an unbiased random move assigns players with equal probability to their roles as first- and second-mover, respectively. Moreover, with a certain reliability "nature" provides a signal $M=\underline{M}, \overline{M}$ to those individuals who happen to end up in the first-mover role and did invest in information technology on previous stages of the game. No rational choices of strategic actors are made on stage 3.

On stage 2 individuals k decide to become of informational type U_k or I_k. When making that decision they do not know whether they play as first- or second-mover on subsequent stages of the game. But as rational decision-makers they anticipate that specific type information about the co-player type becomes relevant only if they are assigned to the first-mover role. Since they know that the latter happens with probability 1/2 the expected payoff differential $\pi(i_I)$-$\pi(i_U)$ of an informed, i_I, and an uninformed, i_U, type in the first-mover role must twice exceed the cost C of acquiring specific type information; i.e. the requirement is $\pi(i_I)$-$\pi(i_U)>2C$.[6]

[6] We chose to model the condition this way in view of the evolutionary analysis to follow since it imposes the more stringent requirement for the emergence of trustworthiness. If stages 2 and 3 would be exchanged and players would know beforehand which role they would be assigned then they would rationally choose to bear the cost of information if C exceeded $\pi(i_I)$-$\pi(i_U)$, yet otherwise the analysis would remain the same.

The difference $\pi(i_I)-\pi(i_U)$ between the payoff expectations of an informed and an uniformed individual in the first-mover role depends on the relation between the probabilities $P(\overline{m}/\underline{M}, p)$, p, $P(\overline{m}/\overline{M}, p)$.

In the limiting cases p=1 and p=0 we have $P(\overline{m}/\underline{M}, p)=p=P(\overline{m}/\overline{M}, p)$. All three probabilities are equal and $\pi(i_I)-\pi(i_U)=0$. Obviously nothing can be gained by investing a positive cost C in a signaling technology.

So let us consider 1>p>0. Note first that $1>\overline{\mu}$, $\underline{\mu}>1/2$ implies $P(\overline{m}/\overline{M}, p)>p$ – equivalent to $\overline{\mu}>(1-\underline{\mu})$ – and $p>P(\overline{m}/\underline{M}, p)$ – equivalent to $\underline{\mu}>(1-\overline{\mu})$, yielding $P(\overline{m}/\overline{M}, p)>p>P(\overline{m}/\underline{M}, p)$. Moreover, cases $s<P(\overline{m}/\underline{M}, p)$ in which T is chosen even after an untrustworthy second-mover has been signaled and $P(\overline{m}/\overline{M}, p)<s$ in which N is chosen even after a trustworthy second-mover has been signaled can be neglected, since choice will not be affected by receiving signals. In view of the preceding only two possibilities remain in case 1>p>0

(i) $$P(\overline{m}/\overline{M}, p)>p>s>P(\overline{m}/\underline{M}, p)$$

(ii) $$P(\overline{m}/\overline{M}, p)>s>p>P(\overline{m}/\underline{M}, p).$$

Since specific information about the co-player type will be acquired only if $\pi(i_I)-\pi(i_U)>2C$ and since players will follow the signal we need to consider

in case (i) $[p\overline{\mu}+(1-p)(1-\underline{\mu})0] +[p(1-\overline{\mu})-(1-p)\underline{\mu}]s – [p1+(1-p)0] >2C$ or

$$p[s+(1-s)\overline{\mu}-\underline{\mu}s-1]>2C+\underline{\mu}s$$

in case (ii) $[p\overline{\mu}+(1-p)(1-\underline{\mu})0]+[p(1-\overline{\mu})-(1-p)\underline{\mu}]s – s >2C$ or

$$p[s+(1-s)\overline{\mu}-\underline{\mu}s]>2C+s(1+\underline{\mu})$$

In sum, whenever some p with 1>p>0 fulfills the condition of case (i) or case (ii) investment in detection technology could conceivably pay, while for p=1 and for p=0 it cannot pay to become an informed player at some positive cost C>0.

On stage 1 decision-makers know that their decisions on this stage affect payoff only if they end up in the second-mover role on the last stages of the game. Bearing this in mind let us distinguish two classes of possible equilibria. On the one hand, there can conceivably be (pure) symmetric equilibria, on the other hand asymmetric equilibria.

4.3. Equilibria in the pure rational choice approach

4.3.1. Pure symmetric equilibria

Conceivably there can be two pure symmetric equilibria characterized either by p=1 or by p=0. Recall first, that we have seen when analyzing stage 2 that nobody would incur the positive cost C>0 of acquiring a technology providing specific type information if p=1 or p=0. Now, if p=1 would characterize equilibrium play then it must be rational for everybody to become a trustworthy $\overline{m}$-type type with the disposition to choose *T* in the final trust-game. Yet, those who would decide on becoming an untrustworthy $\underline{m}$-type would go undetected, since there are no players with the technology to acquire specific type information. Untrustworthy types fare better on the last stage. Therefore the stage 1 decision to become a trustworthy $\overline{m}$-type cannot be optimal. The assumption p=1 is not coherent with the other assumptions of the model since rational individuals making the assumption would choose to become $\underline{m}$-types on stage 1 which amounts to p=0. The only consistent symmetric pure choice behavior on stage 1 is the general choice of becoming an untrustworthy $\underline{m}$-type. Obviously, in a world in which player type remains private information the behavior consistent with the assumption p=0 is in equilibrium. Therefore the single pure symmetric equilibrium is characterized by p=0.

Factoring in occasional mistakes (for a philosophical justification see (Selten, Reinhard 1975) and an evolutionary one (Selten, Reinhard 1988)) of players who in the first-mover role choose *T* will yield an advantage for those who chose to become $\underline{m}$-types over those who chose to become $\overline{m}$-types. Allowing for an occasional mistake in the information decision as well, some I-types who incurred the costs of becoming informed will be around. Then, if somebody mistakenly chose to become a trustworthy $\overline{m}$-type, there will be a potential advantage for such an individual. For, if the individual who became trustworthy by mistake is matched with a first-mover who by mistake became informed and if appropriate conditions apply this will lead to trust where otherwise none would be shown. However, the order of magnitude of the joint occurrence of both mistakes will clearly be smaller than that of a single mistake in choosing *T* on the last round of play. Therefore the potential advantage of becoming a trustworthy type and of being trusted by the occasional informed players is smaller than the potential advantage of exploiting the uninformed who are occasionally choosing *T* by mistake.

4.3.2. Other equilibria

Can there be non-symmetric strict equilibrium outcomes? In such cases we have 1>p>0 and there must be some who choose to become $\underline{m}$-types and some who choose to become $\overline{m}$-types. In an infinite population in which any individual choice of player type cannot affect the population composition as characterized by p asymmetric pure choices can be in equilibrium only if players are indifferent between choosing to become either $\overline{m}$-types or $\underline{m}$-types. The outcome would have to be based on non-strict equilibrium play. The same would obviously apply for the case of mixed symmetric equilibria.

It seems reasonable to be interested only in strict equilibrium play in contexts like those envisioned here. Then in the presence of occasional choices of *T* all individuals should on stage 1 choose to become $\underline{m}$-types. If rational expectations prevail this should be anticipated rendering p*=0 the corresponding first-mover beliefs. The extreme rational choice model of the social situation broadly sketched in section 2 therefore "predicts" that there will be no rational co-operation based on trust in such circumstances.

5. Choice dimensions as evolving

In the strict rational choice analysis the commitment decision to behave trustworthy at the last stage of the game was itself made strategically in view of the payoffs that it might bring about to be so committed. Such explicit commitment decisions may be plausible modeling assumptions in certain circumstances like for example in the presence of explicit formal contracting in social reality (see on legal "hostages" as commitment devices (Raub, Werner and Geroen Keren 1993)). However, the disposition to behave trustworthy or not is often a general trait of character that develops through time (conceivably being even innate). And in many instances it is quite implausible to assume that developing a general disposition to behave trustworthy originates in strategic decisions to that effect. Therefore in our first step towards substituting rational strategic decisions by adaptive processes it is not anymore assumed that the individuals choose their own "commitment-type". In the model emerging after the first step of the modification process the population share p_t, $0 \leq p_t \leq 1$ of trustworthy $\overline{m}$-type individuals develops through time in an evolutionary process that is not driven by strategic type-choices but rather by type-selection. Trustworthiness or the parameter $\overline{m}$=m>1-r become "an endowment" of the individual fixed by "nature" rather than being itself a matter of choice.

5.1. Substituting the choice of trustworthiness by its evolution

Recall that the indirect evolutionary approach relies on a twin payoff function containing both an objective and a subjective dimension. Recall also that it was assumed that except for the parameter "m" the individuals have subjective utility functions coincident with the material or objective payoff functions that measure success as accruing to them when playing the game. Add to this now the assumption of evolutionary population dynamics of p_t that are monotonic in objective or material success: The population share p_t of trustworthy individuals increases if, depending on their (subjectively motivated) individual decisions, the trustworthy are more successful than the untrustworthy. Likewise the population share p_t of trustworthy individuals decreases if the untrustworthy are more successful. And p_t remains constant if trustworthy and untrustworthy individuals are equally successful.

"Subjective preferences" and expectations do not directly affect objective success. But the purely subjective factor m alters behavior in the basic game. The population composition, p_t, represents the prevalence of the subjective factor in the population in general and fixes commonly known priors. The general knowledge of p_t along with specific information about the type, m, of a co-player in the second-mover role will influence rational expectations and rational choices that in turn influence objective success and consequently p_t.

We use a function $R(p_t, m)$ to measure objective success as depending on the determinants of behavior:

$$\dot{p}_t \quad \begin{cases} \text{increases with} \quad t \quad \text{if} \quad R(p_t, m > 1-r) > R(p_t, m < 1-r) \\ \text{decreases with} \quad t \quad \text{if} \quad R(p_t, m > 1-r) < R(p_t, m < 1-r) \end{cases}$$

Success is determined by solving the ("remaining") strategic game G. The "rules of that game" are dependent on the "endowments" m_i, m_j of the two randomly matched players and on p_t such that a class of games each of the form $G(p_t, m_i, m_j)$ emerges. Let $s^*(p_t, m_i, m_j)$ be the solution of $G(p_t, m_i, m_j)$. Assuming that $s^*(p_t, m_i, m_j)$ is unique we can treat the objective success measure R and the subjective evaluation or utility u as functions of (p_t, m_i, m_j). Then the vector $u(m_i, m_j)=u(s^*(p_t, m_i, m_j))$ forms the solution payoff vector in subjective (motivationally relevant) terms and $R(s^*(p_t, m_i, m_j))$ is the solution payoff vector in objective (directly evolutionarily relevant) terms. As in the characterization of the simple monotonic dynamics of p_t we can rely on $R(p_t, m_i, m_j)= R(s^*(p_t, m_i, m_j))$. The values of R are determined in two steps:

Step 1

The solutions of a class of Bayesian games must be derived. The Bayesian games emerge after substituting the strategic stage 1 type-choice of the pure rational choice game of the previous section by fictitious independent type-determining random moves. More specifically, at any time t any player will with probability p_t be of the trustworthy type $\overline{m}$ and with probability $(1-p_t)$ of the untrustworthy $\underline{m}$-type. As far as subjective expectations about the degree of trustworthiness in the population at large are concerned we assume that at each point in time the parameter p_t is common knowledge among the players. Since the population is assumed to be infinite the emerging a priori beliefs of players about the co-player type do not depend on their private knowledge of their own type. Under these conditions for any t the solution $s^*(p_t, m_i, m_j)$ of each of the emerging games G is derived exactly along the lines of argument that have been used in the preceding section. The rational choice solution of the previous game is still (partly) relevant. It proceeds basically from stage 2 of the game as before.

Step 2

The results of step 1 are used to determine whether the share of trustworthy individuals increases, decreases or remains constant. This in turn determines the evolution of p_t through time. Since in step 1 the rational choice solution has been determined for each of the games from the full class of games emerging for each $0 \leq p \leq 1$ we know the objective success corresponding to the solution payoffs for $\overline{m}$- and $\underline{m}$-type players respectively. Therefore we can say which of the types at each state of evolution outperforms the other one.

Steps 1 and 2, respectively, can be somewhat cumbersome but it should be clear in principle how to proceed. In step 1 basically the same backward solution approach as in the last section can be applied after the population composition is fixed and has become known to the players. However, now there is another step that concerns the evolution of the population composition parameter. If the first step describes the motivational factors or the subjective side of the matter in a more or less traditional framework of rational forward-looking decision-making, the second step is of the completely different kind of blind selective adaptation. Combining the two steps in one model we have entered the middle ground between “teleology and adaptation”.

To make this and other basic methodological points it is not necessary to go over all the details of the indirect evolutionary approach here. As indicated in the introductory remarks we have done so elsewhere (again we refer to (Güth, Werner and Hartmut Kliemt 2000)). Therefore, we can confine ourselves to a graphical illustration of the basic intuitions and some intuitive additional remarks that are sufficient to illustrate the basic methodological points at hand.

5.2. Some basic intuitions in a special case

A typical question to be asked in a "comparative statics" evolutionary analysis concerns the so-called evolutionary stability of population compositions. The results of adopting such a method in the circumstances envisioned here differ from the results derived in the limiting case of full rationality. It is in particular not anymore the case that trustworthiness cannot survive. As far as this is concerned all depends on the costs of gaining specific information and the reliability of the available information technology. Consider the following figure:

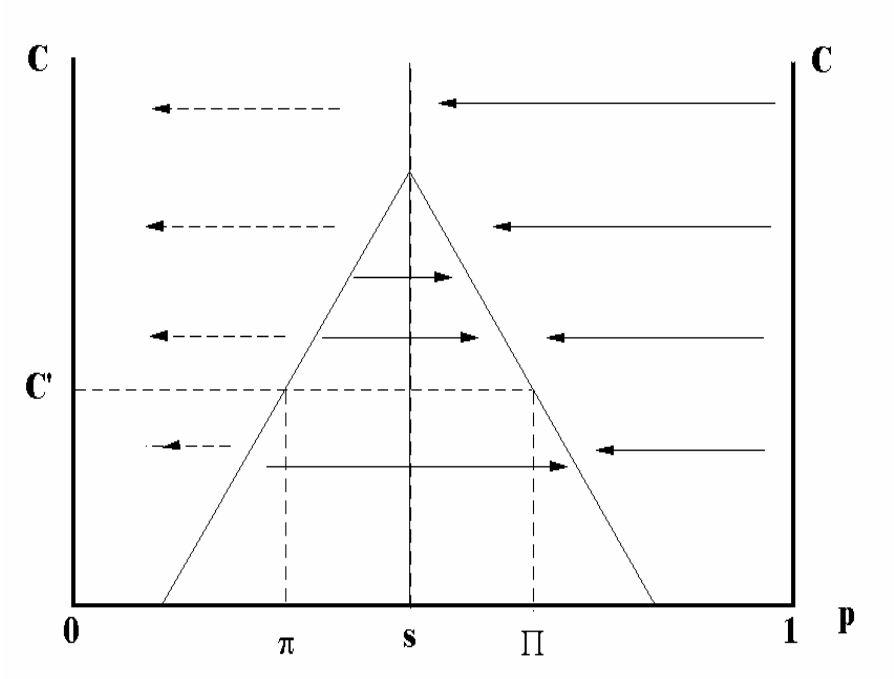

Figure 2: The adaptation of p_t over time

In Figure 2 the horizontal axis represents all possible shares p of trustworthy $\overline{m}$-types in the population. For any initial population composition parameter p_0 the graphic illustrates the development of p_t through time if a detection technology of reliability $\underline{\mu}$, $\overline{\mu}$ to find out $\overline{m}$

and $\underline{m}$ -types, respectively, is available at cost C. For each C solid or dashed arrows show the direction in which p develops if the initial parameter p_0 lies somewhere on the arrow's starting line. Dashed arrows indicate an evolutionary advantage that comes about only if individuals once in a while make mistakes in their choices by deviating from what rationality dictates. Solid arrows show the direction of evolution if under fully rational behavior (in step 2 of the analysis) certain types have the advantage over other types. The shape of the triangle is fixed by the reliability of the available technology and the objective payoff structure. The reliability parameters determine the probability that an individual might gain some additional payoff after investing C (a fixed cost) while the amount to be gained is determined by relations between the payoffs.

The vertical line starting at s indicates the threshold value for the decision to trust or not to trust in the basic game if only the population composition p is known. In that case rational first-movers will show trust if p>s and show no trust if s>p. In the triangle around the s-line we see how the presence of a technology for gathering specific information about the co-player type affects the population composition. Here beyond the parameter p that characterizes the population at large further information about the specific co-player who is assigned to act in the second-mover role is acquired at cost C. For each population composition p_0 that is for some cost level C located in the triangle the population composition parameter p will grow until the right border line of the triangle is hit at cost level C. In this realm individuals have an incentive to reach a positive decision to become informed (and the presence of informed individuals makes it potentially advantageous to be trustworthy).

More generally, for each sufficiently small value of C we get an interval (around s) of population compositions p', p'', … for which it pays to invest in the information technology. Investing in specific type information pays if the probability that this specific information leads to a beneficial alteration of behavior is high enough. This probability depends on the reliability of the information technology and on the population composition.

To start with the latter, assume for instance that p'<s is from the relevant interval for some sufficiently low C (i.e. it lies within the triangle). Then the uninformed individual who did not invest in information technology would play N on the final round. But it is cheap enough to invest in the technology that yields specific information telling with some (sufficient) reliability whether a second-mover deserves to be trusted. If after investment a specific signal of trustworthiness is received a move other than the one dictated by the knowledge of the general population composition will be made. Since making this move selectively leads to

increases in expected gains beyond investment costs, informed individuals even though they incurred C will have the edge over the uniformed.

Intuitively it should be clear also that for population compositions close to p=0 it will in all likelihood not pay to invest in a costly information technology to find out those second-movers who – regardless of the population composition suggesting N – would deserve to be trusted. Close to p=0 there are simply too few trustworthy around to make it worthwhile to seek them. Only if perfectly reliable information is available at no cost it would be worthwhile to acquire specific information for all population compositions p>0. Vice versa, for compositions close to p=1 it will not pay to invest C to find out the individuals who should not be trusted even though the commonly known population composition parameter would suggest the choice of T. If too few untrustworthy are around, it does not pay to bear the cost of finding them.

Summing up this line of argument it is obvious that for suitable values of $\underline{\mu}$, $\overline{\mu}$, C the interval [0, 1] of possible population compositions can be divided into three sub-realms (0,), (, ∏), (∏, 1), with 0< <s<∏<1; where for initial values

$p_0 \in (0,\pi) \Rightarrow [t \to \infty \Rightarrow p_t \to 0]$, it does not pay to invest in information technology since there are not sufficiently many trustworthy to make it worthwhile to find them;

$p_0 \in (\pi,\Pi) \Rightarrow [t \to \infty \Rightarrow p_t \to \Pi]$, it does pay to invest in information technology since in (π, s) there are sufficiently many trustworthy to find them out and in (s, ∏) there are sufficiently many untrustworthy to make it worthwhile to find them out;

$p_0 \in (\Pi,1) \Rightarrow [t \to \infty \Rightarrow p_t \to \Pi]$, it does not pay to invest in information technology and thus to be able to find out the untrustworthy since they are too rare.

Therefore, depending on the initial population composition, we will observe the population dynamics either to eventually decrease to a p=0 population share or to converge – from below or from above – to p=∏. These are the only outcomes that would emerge under plausible monotone dynamics and at the same time the only evolutionarily stable population compositions.

The preceding intuitive illustration of how teleology and evolution are combined in an indirect evolutionary approach demonstrates how a single adaptive dimension can be included in a standard rational choice model. If in the original model all explanations are framed in purposeful individual decision-making only now one dimension of the model is subjective to an adaptive explanation. From this we know in principle how to proceed in case of a model in which several dimensions are treated in terms of rational choice while one is subject to an evolutionary process. It remains to be illustrated how to include several adaptive along with rational choice dimensions in the same model. Rather than going over the analytics of such a model let us turn again to a more or less intuitive illustration of the essential elements.

6. Trustworthiness and informational status as co-evolving

6.1. Elements of a co-evolutionary setting

As opposed to the preceding discussion we now subject two dimensions simultaneously to an evolutionary analysis (see for a detailed analysis of the following (Güth, Werner, Hartmut Kliemt, and Bezalel Peleg 1999)). Individuals do no longer decide on becoming either informed or uninformed. Rather the informed and the uninformed are selected according to their relative success. The population share q of informed I-types (as opposed to the share (1-q) of uninformed U-types) co-evolves together with the population share p of the trustworthy $\overline{m}$-types (as opposed to the untrustworthy $\underline{m}$-types). That is, we get a population share for each of four possible types: I-$\overline{m}$-type, U-$\overline{m}$-type, I-$\underline{m}$-type, U-$\underline{m}$-type.

Let us refer to the U and I type of individuals as their "informational" and to the $\overline{m}$ and $\underline{m}$ type as their "moral" type. As is obvious from the preceding analyses behavior in second-mover roles in the final trust sub-game is determined solely by the moral type and behavior in the first-mover role of that sub-game depends exclusively on the informational type. For instance, a trustworthy uninformed U-$\overline{m}$-type and a trustworthy informed I-$\overline{m}$-type both behave exactly the same way in the second-mover role. Therefore behavior in that role cannot imply differential payoffs that discriminate between U-$\overline{m}$-type and I-$\overline{m}$-type. Likewise in the first-mover role moral type does not matter. Two individuals of the same informational type regardless of their moral type behave equally in the first-mover role and therefore must fare equally in that role.

In a sense the two dimensions of the problem can be separated. However, what happens along one dimension influences what happens along the other. For instance, whether or not the costs

incurred by becoming an informed rather than uninformed individual pay off, crucially depends on the proportion p of individuals who are trustworthy. Likewise, whether or not the trustworthy fare better than the untrustworthy depends on the proportion of informed as opposed to uninformed individuals. Only if first-moving individuals command the faculty to single out trustworthy second-movers with sufficient reliability and thereby discriminate against the untrustworthy can it be a differential advantage to become trustworthy (and thereby to incur the opportunity cost of foregoing the chance of exploiting first-mover trust). In sum, the "evolutionary climate" in which the informed flourish is provided by the presence of trustworthy and untrustworthy types in the "right proportion" – as the preceding analysis shows not too few and not too many trustworthy individuals must be there – while the trustworthy will flourish the better the more informed individuals are around. So what we should expect is that differential "reproductive" success of the trustworthy in comparison to the untrustworthy will depend on the share of informed individuals, q, while the share of the trustworthy, p, determines whether the informed will fare better or not than the uninformed.

In formal terms we move from considering the dynamics of p in the space [0, 1] to considering the dynamics of p,q-constellations in p,q-space or the unit square formed by the Cartesian product [0, 1]x[0, 1]. In the co-evolutionary process individuals of the four possible types are selected by their relative success as measured in objective terms. To determine relative success we must again solve the game. The Bayesian games to be solved in step 1 of the indirect evolutionary approach for each of the p,q-constellations are simplified. But step 2 becomes somewhat more complicated. Fortunately it suffices for our present purposes to convey a more or less intuitive impression of how the (p_t q_t)-dynamics unfolds through time for initial constellations (p_0 q_0). So let us turn to this task for the special case that a perfect information technology providing signals $\overline{M}$ and $\underline{M}$ that are perfectly type-discriminating becomes available at a positive cost; i.e. $1=\overline{\mu}=\underline{\mu}$, C>0.

6.2. A simple illustration of co-evolutionary (p, q) dynamics

As in the one-dimensional case we assume that simple monotonic dynamics that can be represented by linear differential equations apply; i.e. if in one p,q constellation the I-$\overline{m}$-type, U-$\overline{m}$-type, I-$\underline{m}$-type, U-$\underline{m}$-type is relatively more successful than its competitors it will spread and if it is less successful its share will decrease. This will translate to increases and decreases respectively along the two dimensions of p and q. What is going on along these dimensions at each point in time, t, is captured by two differential equations:

$$\dot{q}_t = k\,[R_I(p_t) - R_U(p_t)],$$

$$\dot{p}_t = h\,[R_{m>1-r}(q_t) - R_{m<1-r}(q_t)],$$

where h, k>0 are positive constants.

The first differential equation describes the relative success of the informed as compared with the uninformed the second the relative success of the trustworthy as compared with the non-trustworthy. The equations also clearly illustrate how at each point in time t the change $\dot{p}_t$ of the share p_t depends on q_t and how the change $\dot{q}_t$ depends on the share p_t.

The success function for the informed individuals in the first-mover role is given by

$$R_I(p) = pr + (1-p)s - 2C$$

since they detect all the trustworthy players with whom they are matched with probability p corresponding to the population share of the trustworthy. Trusting precisely the trustworthy they receive pr+(1-p)s at cost C (which must be doubled since they are assigned to the second-mover role only with probability 1/2).

Confining attention to generic cases only and thus excluding p=s the success function of uninformed individuals depends on whether p>s or p<s applies. Since the payoff from being exploited is 0 it is obviously given by

$$R_U(p) = \begin{cases} s & \text{for } p < s \\ pr & \text{for } p > s \end{cases}$$

The difference between the two terms immediately yields

$$R_I(p_t) - R_U(p_t) = \begin{cases} (r-s)p_t - 2C & \text{for } p_t < s \\ s - 2C - sp_t & \text{for } p_t > s \end{cases}$$

Likewise we can derive the payoffs of the two moral types as depending on the prevalence of informed types in the population. All informed individuals are in command of perfect type detection faculties and shall thus choose to trust if and only if the second mover is in fact trustworthy. Since in the second-mover role they will earn 0 if p<s this yields qr as an expectation of the trustworthy in that case. Otherwise, if p>s, everybody will trust the trustworthy who therefore receive r in each and every instance. So depending on q the trustworthy will receive

$$R_{m>1-r}(q_t)=\begin{cases} qr + (1-q)0 & \text{for } p < s \\ r & \text{for } p > s \end{cases}$$

Again, the untrustworthy receive 0 in the second-mover role if p<s. In that case nobody trusts them since the uninformed choose not to trust anyway while the informed will single them out as not worthy of their trust. Analogously in case p>s the untrustworthy shall flourish better due to the trust shown to the undeserving by the uninformed. These considerations amount to

$$R_{m<1-r}(q_t)=\begin{cases} 0 & \text{for } p < s \\ q0 + (1-q)\,1 & \text{for } p > s \end{cases}$$

Finally

$$R_{m>1-r}(q_t) - R_{m<1-r}(q_t)=\begin{cases} rq_t - 0 & \text{for } p_t < s \\ r - (1-q_t) & \text{for } p_t > s \end{cases}$$

The preceding remarks should suffice to get an intuitive grasp of what is going on in the co-evolutionary process. As mentioned already we have determined elsewhere in some detail parameter constellations under which q_t resp. p_t grow or decline. In the present context it is not necessary to go over these formal details. It is more interesting to point out again that the model shows how the two dimensions interact: Looking at the world through the window of a model of the co-evolutionary process we can study how the populations of alternative types along each of the dimensions are providing evolutionary niches, or not, in which the corresponding other type can flourish or not.

As observed the growth and decline of both types interact with each other in an orderly manner describable by differential equations. In such analyses naturally the question of rest-points of the evolutionary dynamics emerges. The process comes to a stop or a rest point (p*, q*) if both $\dot{p}_t=0$ and $\dot{q}_t=0$; i.e.

$$(\dot{p}_t,\ \dot{q}_t) = (k\,[R_I(p^*)-R_U(q^*)],\ h\,[R_{m>1-r}(q^*) - R_{m<1-r}(q^*)])= (0, 0).$$

It can be shown relatively easily that there is a single such rest-point with

$$(p^*, q^*) = \left(\frac{s-2C}{s}, \frac{1}{r}\right) \text{ in the range } E:=\{(p, q)\,) \in [0, 1]^2 \mid p>s\},$$

while for p<s only the line segment [q=0, p≤s) can contain rest points.

But the unique rest-point (p*, q*) is not locally asymptotically stable. There are no small neighborhoods of the point such that for all starting points from the neighborhood the dynamic process will converge towards the rest point. Therefore the rest point is not a likely

candidate for a stable state that might be expected to prevail for any extended period of time. It is rather to be predicted that the dynamics will cycle in a manner illustrated by the following graph:

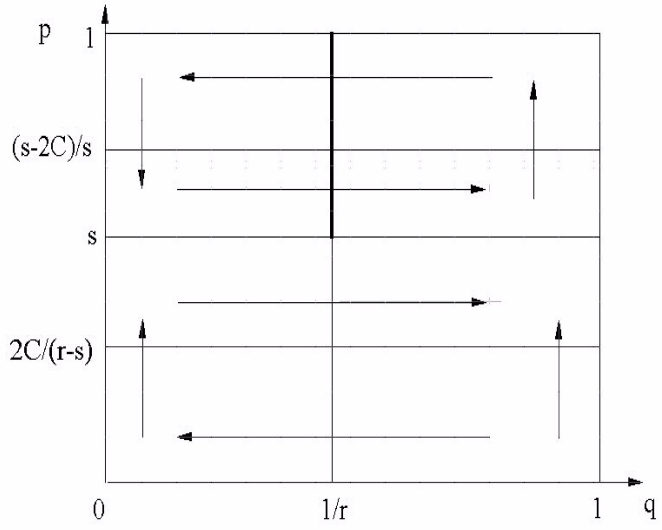

Figure 3: Graph of cycling dynamics

7. Completely direct Evolution

Up to now it has been assumed that the actions on stage 4 and 5 of the game are based on strategic rational choices. Going all the way to a completely non-strategic approach the choices are now modeled as resulting from fixed behavioral inclinations that emerge from an evolutionary process.

On the last stage of the game $m>1-r$ stands for the genetically encoded behavior to use R and $m<1-r$ for the genetically fixed disposition to use E. The actions on the last stage of choice-making are thus explained in terms of the behavioral dispositions of actors in the second-mover role. The population share of each of the types predicts behavior on the last stage of the game according to the behavioral disposition.

Whether the trusting action T is shown on the next to last stage of the game has been determined in the preceding models as the outcome of rational strategic choices in light of the available information. If the action T is to be understood now as the outcome of fixed

behavioral programs then the share of individuals in the population who in the first-mover role would show trust, T, can be subject to an evolutionary process, too.

That the behavioral program is fixed does not imply that behavior needs to be unconditional. In higher organisms behavioral programs are triggered by "information" on states of the world as retrieved by the organisms. In the case at hand the behavior shown depends on whether the individual k is a U_k or an I_k-type. For an I_k-type it is obvious to assume that the signal $\overline{M}$ triggers the choice of T and the signal $\underline{M}$ the choice of N. We thus have to distinguish the population share q of individuals k with I_k and, for the complementary population share 1-q, the sub-share u of those who, as uninformed players, would rely on T.

The dynamics of the shares q_t and u_t over time are, as before, determined by the difference in reproductive success. Again the dynamics of q_t and u_t are only shaped by differences of reproductive success in the role of the first-mover. Since U_k-individuals k receive s if they rely on N and p_t if they use T at time t, the dynamics of q_t are determined as

$$\dot{q}_t = \{p_t\overline{\mu} + [p_t(1-\overline{\mu}) + (1-p_t)\underline{\mu}]s - 2C\} - \{u_t p_t + (1-u_t)s\}$$

by the difference in the reproductive success of I_k- individuals and the weighted reproductive success of U_k-individuals at time t.

For u_t the obvious dynamics are

$$\dot{u}_t = p_t - s.$$

We adjust our preceding discussion of the co-evolutionary process as follows

$$\dot{p}_t = h[R_{m>1-r}(q_t,u_t) - R_{m<1-r}(q_t,u_t)] \qquad \text{with} \quad h>0,$$

where

$$R_{m>1-r}(q_t,u_t) = [q_t\overline{\mu} + (1-q_t)u_t]r$$

and

$$R_{m<1-r}(q_t,u_t) = q_t(1-\underline{\mu}) + (1-q_t)u_t\,.$$

Even in such a complex evolutionary setting results can be derived analytically though it is very demanding to do so in a general way. In particular, we cannot easily characterize the class of rest-points of the evolutionary process completely. Again,to illustrate the principal methodological points at hand it suffices to look at a more specific situation and example here. For instance, imagine that we are interested in whether or not there can be rest-points of

the evolutionary process such that even uninformed individuals who are disposed to choose T can survive. This is of particular interest if s is small in relation to r and thus non-cooperation rather disadvantageous. If even then uninformed trusting individuals could not survive the likelihood that they can survive under different conditions should be even less.

More precisely let (p^*,q^*,u^*) be a rest-point, let $0<u^*<1$, and let s be sufficiently small to fulfill $(\bar{\mu} + \underline{\mu} - 1)(1-s)s < 2C$.

From $0<u^*<1$ we obtain $p^*=s$. Since for $u_t = u^*$

$$\dot{q}_t = (1-s)[\underline{\mu} + \bar{\mu} - 1]s - 2C < 0$$

for s being small, one also gets $q^*=0$ for s being small enough. This, in turn, implies

$$\dot{p}_t = u^*(r-1) < 0$$

and thus $p^*=0$. This proves that there can be no rest point with $0<u^*<1$if s is sufficiently small to satisfy

$(\bar{\mu} + \underline{\mu} - 1)(1-s)s < 2C$.

The analytical treatment of the purely adaptive evolutionary model may become very clumsy and complicated. This may be an obstacle to using the direct or fully evolutionary approach more widely in situations like the one envisioned here. In view of this it might be argued that the complexity of the model stems from starting the discussion from the "wrong end" of an extreme model in which explanations are framed exclusively in terms of purposeful rational choice. Relying on an evolutionary approach from the beginning would perhaps have led to a less complex model. But the adherents of adaptive modeling should be warned that the reduced complexity of their models may not be due to better modeling but rather to a distortion of perspective that emerges if we leave out the dimension of teleology altogether.

8. Discussion

By way of examples we surveyed the full field of approaches reaching from

- pure rational choice based on the assumption of traditional neo-classical economics that there is a shadow of the future but none of the past to
- direct evolution assuming as in traditional (evolutionary) biology that there is only a shadow of the past but none of the future.

Although the game of trust of Fig. 1 is rather simple it represents an archetypical or paradigmatic situation of inter-individual co-operation. The conditions of anonymity according to which the information on m-types remains private, at least initially, are not too far off the mark, too. Neither is it wildly implausible that a detection technology yielding informative but imperfect signals is available. Quite to the contrary, thinking of our experience of large group interactions these modeling assumptions have a certain ring of realism. All the models that were presented here have these modeling assumptions in common. Nevertheless, even though always the same alternatives

- $m_k > 1\text{-}r$ vs. $m_k < 1\text{-}r$,
- U_k vs. I_k,
- N vs. T,
- E vs. R,

are analyzed, the results differ dramatically depending on where on the spectrum between teleology and evolution we locate model:

- Pure forward looking deliberation leads to a social dilemma in which co-operation based on trust cannot be a rational equilibrium choice.
- If only trustworthiness (m-type) evolves while everything else is a matter of rational choice the old dilemma ($p^*=0$) and a bi-morphism in which untrustworthiness coexists with trustworthiness ($1>p^*>0$) can both be evolutionarily stable.
- If the inclination to be trustworthy and the disposition to rely on detection technology co-evolve while choices in the basic trust-sub-game are still made in a forward looking rational manner then stable cycling around a poly-morphic rest-point is the typical result.
- Finally when all inclinations co-evolve, the 3-dimensional dynamics allow for no rest points with specific properties.

Alluding to Nietzsche we can say that we not only need look at the world through different windows but also that what we see strongly depends through which one we choose to look. The insight that the way we construe our models can strongly influence our view of the world is not new. Whether we should even speak of “ways of world making” as Nelson Goodman or

think of different windows on a world that is not of our making may be left open here. In any event, what we see strongly depends on our modelling decisions. More specifically whether we set out to explain a phenomenon as chosen rationally or as the product of evolution and selection will in general make a difference. The view that rational choice behavior and evolutionary stability amount to the same since evolutionary stability implies the best reply-property is rather naïve and unjustified except for special situations. So where on the spectrum between rational teleology and blind evolution we locate our model is an important decision. This decision should be made according to considered prudent judgement of the aspects of a specific situation rather than according to some a priori concept of what an explanation should amount to.

That less can be explained in terms of rational choice than economists tend to admit seems to be clear. In particular behavioral dispositions, which are acquired once and forever, should find their way into economic analysis. We all follow fixed behavioral programs. When the alarm clock rings, we do not always decide anew whether or not to get up. When shopping in a supermarket most of us have once and for all made up their mind to pay and not to steal if the opportunity shows up. To be trustworthy and to fulfill promises is again among our "virtues". We tend to be generally virtuous rather than to behave as if we were as the result of a strategic calculation taking each case on its own merits.

Behavioral dispositions are an important aspect of bounded rationality. Therefore introducing such behavioral dispositions supports the general trend towards theories of bounded rationality in economics. But, of course, not all humans are of the same kind in these regards. Heterogeneity between individuals may prevail if it comes to rational decision-making and self-management. Some may follow routines where others make forward-looking rational choices in a strategic manner. There will also always be situations in which strategic choice in itself is regarded as appropriate and others in which it is deemed inappropriate. Whether we enter the market to exchange goods or the forum for an exchange of opinions will make a difference at least for many individuals. As theorists we have to consider all these factors and see to it that our models follow suit by locating them appropriately somewhere between the extremes of farsighted teleology and blind evolution. We hope that the present contribution may assist researchers in making such modeling choices.

As far as behavior is not genetically fixed but rather the outcome of teleological choice making it cannot be explained in terms of “blind selection”. But contrary to the implicit assumptions of the preceding discussion this does not imply that elements of the traditional perfect rationality assumptions need to be included in our behavioral explanations. Quite to

the contrary, instead of relying on the model of “fully rational or perfect teleology” elements of “workable” or “boundedly rational teleology” should be included (see on this programmatically (Güth, Werner, Hartmut Kliemt, and Yaakov Kareev 2002)). Ideally in an indirect evolutionary model the human ability to anticipate the expected future and thus the limited human capacity to behave in a teleological manner should be represented by the true laws of human cognition along with the true laws representing the evolutionary forces at work.

References

Aumann, Robert J. 1976. "Agreeing to Disagree." *The Annals of Statistics*, 4:6, pp. 1236-39.
Dacey, Raymond. 1976. "Theory Absorption and the Testability of Economic Theory." *Zeitschrift für Nationalökonomie*, 36:3-4, pp. 247-67.
Fagin, Ronald, Joseph Y. Halpern, Yoram Moses, and Mosche Y. Vardi. 1995. *Reasoning about Knowledge*. Cambridge, MA / London: MIT Press.
Frank, R. 1987. "If Homo Economicus Could Choose His Own Utility Function, Would He Want One with a Conscience?" *The American Economic Review*, 77/4, pp. 593-604.
Güth, Werner and Hartmut Kliemt. 2000. "Evolutionarily Stable Co-operative Commitments." *Theory and Decision*, 49, pp. 197-221.
Güth, Werner, Hartmut Kliemt, and Yaakov Kareev. 2002. "How to play randomly without random generator. The case of maximin players." *Homo oeconomicus*, XX:forthcoming.
Güth, Werner, Hartmut Kliemt, and Bezalel Peleg. 1999. "Co-evolution of Preferences and Information in Simple Game of Trust." *German Economic Review*, 1:1, pp. 83-110.
Kant, Immanuel. 1991. *Political writings. The metaphysics of morals*. Oxford et al.: Oxford University Press.
Mertens, Jean-Francois and Shmuel Zamir. 1985. "Formulation of Bayesian Analysis for Games with Incomplete Information." *International Journal of Game Theory*, 14, pp. 1-29.
Morgenstern, Oskar and Gerhard Schwödiauer. 1976. "Competition and Collusion in Bilateral Markets." *Zeitschrift für Nationalökonomie*, 36:3-4, pp. 217-45.
Raub, Werner and Geroen Keren. 1993. "Hostages as a Commitment Device. A Game theoretic Model and an Empirical Test of Some Scenarios." *Journal of Economic Behavior and Organization*, 21, pp. 43-67.
Selten, Reinhard. 1975. "Reexamination of the Perfectness Concept for Equilibrium in Extensive Games." *International Journal of Game Theory*, 4, pp. 25-55.
Selten, Reinhard. 1988. "Evolutionary Stability in Extensive Two-Person Games - Correction and Further Development." *Mathematical Social Science*, 16, pp. 223-66.