Analyse & Kritik 26/2004 (© Lucius & Lucius, Stuttgart) p. 203–219

Werner Güth/Hartmut Kliemt

The Evolution of Trust(worthiness) in the Net*

Abstract: The main results of our indirect evolutionary approach to trust in large interactions suggest that trustworthiness must be detectable if good conduct in trust-relationships is to survive. According to theoretical reasoning there is a niche then for an organization offering a (possibly) costly service of keeping track of the conduct of participants on the net. We compare traits of an organizational design as suggested by economic reasoning with those that actually emerged and ask whether institutions like eBay will increasingly have to 'economize on virtue' although so far they could rely on its spontaneous provision.

1. Introduction and Overview

Order can emerge spontaneously if an interaction is repeated and such mechanisms as reputation formation can run their course. Platforms like eBay that organize social cooperation over the Internet are a case in point. They produce 'order without law' (see Ellikson 1991) or without commanding the fundamental coercive powers to raise taxes, to collect information and to punish misbehavior but definitely not without incurring costs. Therefore we must ask why individuals are willing to bear the costs of organizing social cooperation over the Internet, why others are willing to pay them for rendering such organizational services as there are and what the prospects for the survival and development of service providers are.

Addressing such questions in a somewhat indirect way we will first sketch (2.) some results of our former work that show that even in large (almost) anonymous interactions as today prevailing on the Internet trustworthiness can survive if it can be detected with some reliability. This contradicts some common views to the effect that 'large' interactions cannot conceivably be organized without the fundamental coercive power of the state and therefore could in particular not persist on the international level where a common state organization is lacking. It, however, presupposes that individuals have characteristics that as a matter of fact restrict their opportunistic choice making in ways akin to boundedly rational behavior. In the next section (3.) we lay out some 'economists" ideas concerning desirable or likely properties of a net platform—Big Brother or BB—for organizing bilateral exchanges. Comparing this reference model with eBay is instructive since it shows that eBay chose a route relying more strongly on community feel-

[*] Critical comments by Bernd Lahno were very helpful in improving the paper; of course, he is responsible for all remaining errors.

ings and intrinsic motivation than on extrinsic incentives like reputation scores as suggested by 'more economic' approaches to human behavior (4.). Starting from the contrast between eBay and BB we raise additional questions, indicate possible lines of future research, and finally issue a cautionary remark on the scope and limits of eBay as a platform of trade (5.).

2. The Indirect Evolutionary Approach to Trust in Large Transactions' Systems

2.1 The Basic Trust Problem and its Representation as a Simple Game

Imagine a situation in which there are so many potential interaction partners for bi-lateral transactions that in repeated interaction none of them can keep track of the actions of all others. The behavior of each of the actors directly and significantly affects her or his partner but is individually insignificant for the collective result. We refer to such a transactions' system as a *large transactions' system.* Given this definition it is obvious in particular that whether or not there is a 'climate of trust' in a large transactions' system is not a matter to be influenced strategically by individual actions. Individual actions affect the partner of a transaction only. The overall 'trust climate' characterizing the ongoing interaction is the result of all the individual actions but none of the individual acts is *significantly* responsible for the collective result.

Interaction partners are concerned with the results of the particular interactions in which they participate. They are also interested in their own reputation. If there is a reputation effect they will take it into account. But as far as there is no reputation effect what may be called the 'trust predicament' emerges: every participant of interactions taking place in a large transactions' system can act opportunistically behind *the veil of individual insignificance.* In pursuit of their common interest both actors can and in view of their extrinsic profit motives should always deviate from agreements and go for their private interest. When engaging an interaction with a partner both individuals expect to be better off if both act as agreed. However, each knows that showing trust in moving first is risky because the second mover can exploit the first moving actor's trust. The mutually advantageous deals will not be realized if the actor in the first mover role does not trust or the actor in the second mover role does not reward trust. The lack of trust or the presence of the risk of exploitation can thus stand in the way of what is in the best interest of the actors.

The following graph presents a rational choice explication of the trust predicament in unilateral trust problems (the prisoner's dilemma or exchange being the paradigm example of bilateral trust problems):

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).

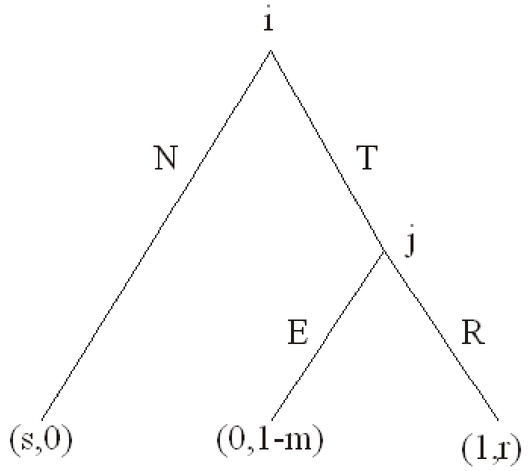

Figure 1: The game of trust with payoff parameters 0 < r, s < 1, m ≤ 0

One could imagine that the game of figure 1 would be represented once with objective payoffs and once with subjective payoffs. Rather than using two game representations we use only one with the understanding that the parameter m is a purely subjective one. There is no objective payoff corresponding to the parameter m while all the other parameters represent objective and subjective payoffs at the same time. Since there are sufficient degrees of freedom the subjective utility functions can be chosen such that the numerical values of the objective and the subjective payoffs coincide.

If we set the parameter m=0 then a game emerges in which only objective payoffs matter as (extrinsic) motivation of individual actions. The second mover in this game is not (intrinsically) motivated by the desire to act in trustworthy ways. Such an actor lacks intrinsic motivation since his actions are guided by nothing but expected objective payoffs or extrinsic motives. Once the parameter m becomes positive or negative it represents some form of intrinsic motivation. We neglect negative values of m representing motives like spite and focus on positive values and thus on the range of parameters that conceivably might further co-operation. The intrinsic motivation to co-operate is sufficient to influence choice only if the parameter m becomes large enough, i.e. $1 - m < r$ or $m > 1 - r(> 0)$ otherwise it affects merely the (subjective) payoff without being behaviorally relevant. If the motivation to act honestly is strong enough then exploitation in the second mover role will be avoided. However, this will further cooperation only if it can be known by the first mover. The first mover will trust only if he knows that the second mover is motivated to act fairly—or at least he must expect with sufficient probability that the second mover is trustworthy. Quite obviously, actors who manage to show trust and to receive the reward for trust in the first mover role will be better off than those who do not trust or are exploited. Likewise the trustworthy individuals who are recognized as such will be trusted and do better than the untrustworthy if otherwise no trust will be shown—this will apply at least under suitable conditions.

2.2 Basic Ideas of Indirect Evolution

We studied the basic game of figure 1 quite extensively within the framework of an indirect evolutionary approach (Güth/Kliemt 1994, 2000; Güth/Kliemt/Brennan 2003; Güth/Kliemt/Peleg 2000). In this approach rational choices that are made by a rational forward looking actor in view of his 'subjective expectations' or 'subjective payoffs' are embedded in an evolutionary process which fixes success of decision determinants in terms of 'objective' payoffs that are brought about by the choices. The duality of subjective payoffs guiding behavior and objective payoffs determining success is at root of the indirect evolutionary approach in general and of the corresponding account of 'trustworthiness' in particular.

As an illustration it may be helpful to think of the evolutionary model of firm competition and innovation that has been suggested by Armen Alchian a long time ago (see Alchian 1950). In this seminal paper it is argued that under suitable competitive conditions the more profitable firms will drive out the less profitable ones regardless of whether or not those who are running the firms are consciously aiming at profit maximization. The subjective aims as represented by the subjective utility function of the staff of a firm may be almost anything. For instance the staff might aim at 'providing the customer with the best quality' available, 'furthering the common weal', 'maximizing market share', 'having an adequate share of the market' or whatever. If a strategy is as a matter of fact objectively profitable then firms that adopt it will survive and otherwise not. Survival is determined by objective success of choices and thus independently of the subjective motives leading to the choices. Firms that pursue for what reasons ever strategies that lead to losses will eventually be eliminated from the game. As Alchian argued, even if initial strategies would be assigned completely randomly to firms, in the end those firms survive that are endowed by chance with a successful strategy. This will hold good as long as there is a growth, birth and, decline as well as death mechanism that will let the objectively more successful flourish and the others not.

Alchian knew Darwinian arguments—that had a long tradition in social theory anyway—very well. But he wrote before more precise concepts of evolutionary game theory like that of an evolutionary stable strategy (ESS) were developed. Later results that rely on such concepts show, that once profit maximizing strategies are established (as a monomorphic population) other strategies cannot invade such a population anymore. The result is evolutionarily stable if interaction forms a competitive market where no seller's profit depends on another seller's behavior (see Güth/Peleg 2001).

Real world examples of objectively successful strategies that are not meant to be profit maximizing might form those (in particular German and Japanese) companies run by engineers that were pursuing a 'quality first strategy'. The strategy was economically doubtful and often seemed foolish in the short perspective but obviously was selected by market forces as the objectively profit maximizing strategy over the long haul. Even though the heads of companies were acting against maximization of subjectively perceived profits, what they did for other reasons was sufficient to be ahead of the average performance of the market as measured in 'true' or objective profits.

2.3 Indirect Evolution in Large Groups

The indirect evolutionary approach to the evolution of trust in large groups (on small interactions see Güth et al. 2002) considers basically two second mover strategies for the game represented in figure 1. For the *first strategy* the parameter m is sufficiently large to affect behavior. The individual actor is endowed with an intrinsic motive that is strong enough to make her behave in trustworthy ways. This makes her a trustworthy actor. She is not only strategically behaving as if she were one but is making strategy choices because she is trustworthy. For the *second strategy* the parameter m is not large enough to affect behavior. The individual actor is not endowed with a sufficiently strong intrinsic motive such as to induce her to reward the trusting initial move of the first mover. She is not trustworthy and therefore will not behave in trustworthy ways unless there is a strategic extrinsic motive to do so.

Social theorists almost across the board have uttered skepticism about survival prospects of such forms of good conduct as 'trustworthiness'. 'The elicitation of good conduct' (see Klein 1997) in large groups seems to be ruled out by the conditions of large scale social interaction. They were even more skeptical when it came to the survival prospects of genuine trustworthiness or the intrinsic motivation to behave in trustworthy ways. People would behave as if trustworthy for strategic reasons for instance in an effort to maintain their good reputation in repeated interactions but there would not be the genuine 'thing' (though even in Kreps et al. 1982 the genuine thing and the 'crazy' belief in its presence are needed to induce others to behave as if trustworthy). Relating this to the preceding discussion in general and in particular to the concept of a large transactions' system as introduced before the issue to be addressed is whether or not trustworthy individuals can survive in large transactions' systems: Are there and if so what are the conditions under which trustworthy individuals will fare well enough to avoid extinction?

2.3.1 A Simple Model of Large Transactions' Systems

To model a large transactions' system we make some assumptions about the scope and kind of the interaction (for formal derivations underlying the following discussion see for instance Güth/Kliemt 1994; 2000): Imagine infinitely many rounds of play of infinitely many individuals (representing a very large group interacting under conditions of individual insignificance). Let the individuals be matched randomly to play the game of trust of figure 1. Half of the time they end up in the first and half of the time they end up in the second mover role. Some of the individuals may be trustworthy in the second mover role. After each round of play objective relative success fixes the population composition on the next round of play. Whether the choice of trust is more advantageous than showing no trust depends on the population composition, i.e. on the presence of trustworthiness. As long as sufficiently many trustworthy individuals are around the payoff expectation of showing trust can be higher than that of showing no trust even without the possibility of screening. Therefore it is worthwhile to incur the risk of showing trust and rational individuals will show trust in the first mover role. If undetected, untrustworthy individuals can exploit first mover trust, however. As long as objective success on each round of play determines how many players of each type will be present on the next round of play relatively higher success of the untrustworthy will drive out the trustworthy until a lower threshold of the population share of the trustworthy is reached from which on it will not be advantageous to trust. This holds good under all plausible dynamics which correlate higher objective payoff with higher population share. Once there are not sufficiently many trustworthy individuals around to make it worthwhile to incur the risk of showing trust no first mover will ever deliberately choose to show trust. However, as long as trust is chosen by mistake occasionally, the untrustworthy will once in a while be presented with the opportunity to exploit a trusting first mover. They will then fare better than the trustworthy and fare equally well as the trustworthy in all other instances when no trust is shown. In this mistake driven process the trustworthy will still be eliminated over the long haul though much more slowly than in the case of rational trust.

Whether the trustworthy or those who are not trustworthy succeed more generally depends on whether or not trustworthiness can be detected by potential first movers. If first movers can trust the trustworthy and avoid trusting those who do not deserve to be trusted the trustworthy individuals will flourish. If the abilities to discriminate between types are perfect then only trustworthy types can conceivably survive. The reasoning is obvious: the trustworthy will be trusted and therefore receive a higher objective payoff than the untrustworthy in each and every transaction in which they take part. The untrustworthy will not be trusted and therefore fare less well than the trustworthy. They never will find a chance to exploit first mover trust. Going to the obvious opposite extreme in which all type discrimination skills are lacking and type information is completely private the untrustworthy will fare better if trusted and fare as well as the trustworthy if nobody ever trusts. In that case the trustworthy will go extinct if only slowly so beyond the threshold from which on trust will be shown only by mistake.

2.3.2 Enriching the Model by a Costly Detection Technology

The model becomes richer if we allow screening to be costly. Consider what may be called a 'costly detection technology of limited reliability' or a 'detection technology' for short. Such a detection technology provides a signal to the actor in the first mover role. Those who pay the price for using the technology can thereby know with higher probability whether the partner in the second mover role is a trustworthy or an untrustworthy type. Whether investing the price or incurring the costs is worthwhile, of course, depends on the reliability of the technology and the type composition of the population. If almost all individuals are trustworthy or if almost all individuals are untrustworthy, it will not pay to invest in screening.

This and some additional insights are, in a way, summed up by figure 2 which shows the population share p of the trustworthy on the horizontal and the cost C for type detection on the vertical axis. If the reliability of the technology is very low then the realm in which it is advantageous to rely on it will be small. In figure 2 the triangle would shrink by lowering its top corner C" while the two other corners on the horizontal axis might be approaching s. If the costs are higher the use of the technology will be advantageous for fewer population compositions. Whenever the population composition parameter p falls for some C' into the interval (π, Π), whose boundaries π and Π vary with C', the population share of trustworthy individuals will grow due to the availability of screening (after costly investment in type detection). For given C' the population share of trustworthy individuals will shrink outside that interval, i.e. for p $<$ π or p $>$ Π where no type information is bought. For other values of C basically the same argument applies. The size of the angles of the triangle depends on the reliability of the detection technology. The sides of the triangle show for each C the lower limiting value $\underline{p}(C)$ from which on up to $\overline{p}(C)$ it pays to invest in detecting the trustworthy and the upper limiting value $\overline{p}(C)$ above which it does not pay to invest in screening.

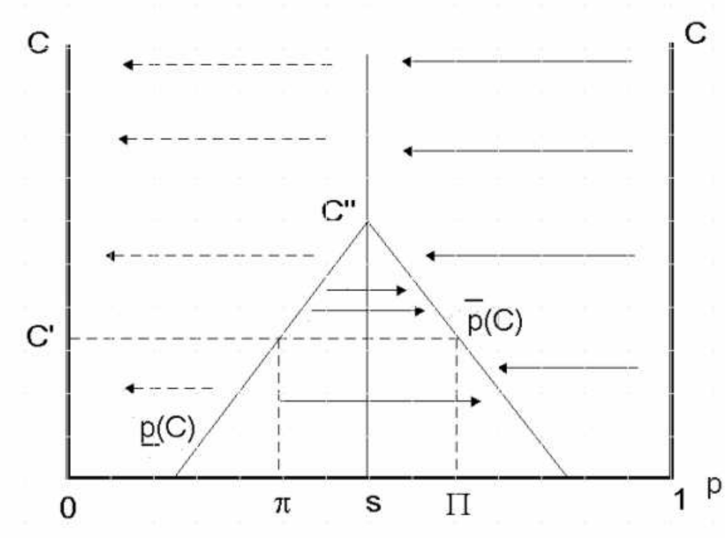

Figure 2: Dynamics of trustworthiness

The more reliable the detection technology the higher its costs can be without preventing its use. If the detection technology works with perfect reliability and its costs are zero then, as previously indicated, the full interval [0, 1] of initial population compositions will lead to a population of trustworthy individuals only. The solid arrows show dynamics that are driven by deliberate choices while the dashed arrows show dynamics that are merely mistake driven. For C > C" there is only one evolutionarily stable type composition, namely p*=0, resp. the m < 1-r-monomorphism, whereas for C=0 and perfect reliability of screening only p*=1, resp. the m > 1-r-monomorphism is stable. In all other cases (i.e. for 0 < C' < C" or C=0 with unreliable detection) there are two evolutionarily stable population compositions p, namely p*=0 and p*=$\overline{p}(C)$. Figure 2 also illustrates the basins of attraction of these two: whenever the initial composition p_o satisfies $p_o < \pi$, the population composition p converges to p*=0 whereas for $p_o > \pi$ it will converge to p*=$\overline{p}(C)$—either by an increase of p if $p_o < \overline{p}(C)$ or a decrease of p if $p_o > \overline{p}(C)$.

2.3.3 Relating Results to eBay

To relate the findings of the indirect evolutionary approach to trust in large interactions' systems to eBay it may be helpful to sum up some general insights of the previous discussion first. If type discrimination is sufficiently inexpensive and sufficiently reliable then:

- there is a niche in which trustworthiness can survive in evolutionarily stable ways;
- there is a niche for those providing at reasonable costs the detection technology or organizing an interaction 'platform' on which trustworthiness may survive;
- if detection is imperfect and/or detection costs are positive then a sufficient degree of trustworthiness must be initially present to stabilize its presence endogenously (there will always be a niche for some untrustworthy individuals but the trustworthy can survive only if there is initially a sufficient number of them).

For successful Internet auction platforms like eBay, this suggests the following conclusions:

- the founders of eBay were lucky in the sense that initially the situation $p_o > \pi$ (see figure 2) prevailed;
- the relative carelessness of behavior when transacting over the net seems to suggest that participants perceive the population share of trustworthy individuals as p > Π and therefore in general do not incur (idiosyncratic) costs of type detection;
- no efforts at screening will be worthwhile as long as the perception of p > Π prevails and therefore the share p of trustworthiness on the Internet will tend to decline and chances for as well as instances of fraud will tend to increase.

Without any doubt eBay is a great success. Its market capitalization is remarkable. It is a company going increasingly international and it might even seem to be a first step towards a global market in which individuals transact increasingly across borders without specialized intermediaries. The Internet is thereby bringing about what may be called spontaneous globalization, not only of information but also of transactions. Nevertheless, in view of the precarious nature of a general climate of trust in large transactions' systems it is by no means a sure thing that eBay will survive.

Interpreted in terms of our general model eBay could work rather well as long as p $> \Pi$ and people were showing trust. Unless the general climate of trust and trustworthiness be eroded the share p of trustworthy individuals must be protected. The safeguard provided by eBay is its reputation mechanism. However, from a rational choice point of view the reputation feedback mechanism of eBay is rather peculiar. The exclusion from interaction that would drive the efficacy of reputation mechanisms in other contexts does not play a central role for eBay. Without possibilities of proper verification of identity a trader in eBay cannot be detected if switching to a new identity. Moreover, the mechanism of building up a 'good' reputation is not cheat proof either. Reputations may be built up strategically for exploitation purposes. What may be called 'eBay brand names' can be traded. So, how to prevent good reputations being bought up by bad guys? True enough, the higher the premium that is carried by positive reputation on the net (but there is not too much evidence for that anyway) the more costly it would become to send a false signal by buying into a good reputation or by willing to 'burn' a good one for the 'big kill'. Still, costs would not preclude all forms of such fraud among strategically rational traders who seek and might find ways to exploit market platforms like eBay and the bigger and the more global transactions become the higher the risks. We submit that eBay will have to change its character and become more 'knave proof' in the process.

3. The Example of Big Brother

With the preceding results on the 'evolutionary niche for trustworthiness' in hand let us imagine an institution that may be called *big brother*. Let us think of 'BB' as providing a net platform and the essential information on the trustworthiness of potential transaction partners that might get into contact through use of that platform. In view of the anarchic character of the net BB is lacking that fundamental coercive power that goes along with the modern state. In particular BB cannot check whether information whispered into his ears is true or not. At least he has no independent coercive means of finding out the truth. BB must rely on whatever he is told and try to distill from this some more reliable information. Can such a powerless BB implement a useful yet costly detection technology and get paid for it if he merely prudently compares information from different sources?

Asking this question is, in a way, asking for something akin to eBay. But let us stay a bit with BB. BB stands more or less for an economist's 'natural way' to set up a large transactions' system that can facilitate exchange and reputation formation over the Internet. Obviously a typical economist should be expected to operate under the assumption that " ... people would act *economically*; when an opportunity of an advantage was presented to them they would take it" (Hicks 1979, 43). But restless maximization and taking of all opportunities is perhaps too far removed from real world behavior to be taken seriously. Some form of merely boundedly rational opportunity taking behavior must, however, clearly be taken into account. Even if not all people take all opportunities all the time some will do. Presumably in an environment like eBay they will do so rather frequently and with relative impunity. Since an economist would anticipate that this might undermine the whole transaction system we would expect BB to be set up in a way that prevents such problems as well as possible.

Obviously it is crucial that we specify what BB would do with the information that he might acquire. The simplest way and a method that comes to one's mind immediately is a check on consistency of reports. Lacking fundamental coercive power and without any backing by such power through state enforced legal institutions this may be almost all BB might be able to do. In forming a reputation for arbitrary individuals i, j that have enlisted as potential transaction partners BB cannot do much more than comparing what the two partners i, j tell him, i.e. by checking whether both parties conceive of a deal as properly completed. Counting the numbers of transactions and to relate them to the number of good, bad, consistent or inconsistent reports seems more or less all BB can do. To see more specifically what may go on here, let us for all i and all times t call the following vector a 'counting type':

$c_t^i = (m_1^i, n_1^i, m_2^i, n_2^i)$; where

m_1^i is the number of participations as seller in successful transactions,

n_1^i is the number of participations as seller in transactions,

m_2^i is the number of participations as buyer in successful transactions,

n_2^i is the number of participations as buyer in transactions.

How a counting type develops through time depends on how the counting is done. It seems reasonable to assume that in all cases in which both participants in a transaction report the same result, counting would just follow the reports. Whenever participants reported differently BB can either count the transaction as unsuccessful, successful or count it not at all. BB must make a strategic decision here. As a most simple yet plausible mechanism imagine the following:

- Initially all counting-types of all individuals i are $c_0^i = (m_1^i, n_1^i, m_2^i, n_2^i) = (0, 0, 0, 0)$.
- BB asks i and j to report simultaneously and
  - reduces the reputation of non-reporters i, by increasing n_1^i *or* n_2^i but neither m_1^i *nor* m_2^i;

- counts according to report if i and j report the same result, i.e. r(i)=r(j);
- increases n_1^i *or* n_2^i but neither m_1^i *nor* m_2^i, for both if r(i)≠r(j), unless one of the self-reports is unfavorable leading to no alteration of count for any party (in that case both were presumably honest and a mistake occurred preventing successful completion of the interaction).

BB does not command the power to tax. But BB cannot provide his services for free either. He must get paid for his services. Therefore to act reasonably, BB should keep all reputation information strictly to himself—or so it seems. He would offer information on the trustworthiness of a potential transaction partner only when asked and he would answer only if a fee for service would be paid. BB as conceived in our thought experiment would continue his dealings with customers only if those customers were willing to identify themselves credibly and would report the results of transactions to him. As far as this is concerned BB would and could require positive proof. (The latter features are clearly different from eBay.)

In 1996 a study group at the University of Duisburg (Stefan Dreckmann, a programmer, Markus Gruner, a simulation expert from physics, and Hartmut Kliemt) simulated a club-like large transactions' system run by BB. As suggested by the former results on the evolutionary niche for trustworthiness we started with quite a high proportion of trustworthy and truthful individuals. These individuals would be trustworthy in second mover roles. Being intrinsically motivated 'to do the right thing' they would truthfully report the results of transactions to BB and thereby give BB access to correct information depending on the number of intrinsically motivated types around. Concerning the other simulated individuals we allowed for arbitrary strategies of cheating and lying that could possibly crowd the trustworthy and truthful out of the population and scrutinized whether or not these strategies would be more successful than those of individuals showing good conduct.

BB relied on counting types trying to extract information from the reports he got and updating the types accordingly then informing potential customers about the trustworthiness of potential transaction partners on condition of a feed back report and a modest fee (counting non-feedback as cheating). According to simulations with more than 50 per cent of trustworthy and truthful individuals in the population good conduct would drive out bad conduct. Since the objective success of strategies on each round of play would fix the proportions of trustworthy and truthful individuals in the next generation the relatively more successful individuals were present in the population in ever higher proportions after every round of play if starting with favorable initial conditions or an exogenous supply of trustworthiness and truthfulness. To put it slightly otherwise, starting out with sufficiently many trustworthy and truthful individuals and sufficiently low costs a BB who could only check on consistency of reports when updating his estimates of the trustworthiness of the transaction partners could make the trustworthy and truthful flourish. And BB could do so by means of a reputation mechanism only.

Insufficient computing power (1996 is a long time back in computing) prevented to check on the long run stability of such results. Still, the simulations did at least indicate that even a powerless BB with very simple means of detecting the trustworthy and truthful might do the trick. This was and presumably still is interesting to some extent. However, BB did not turn into eBay or something akin to that great success. Even though some of us toyed with the idea of trying to exploit its potential, we would certainly have encountered a disappointment. For in fact BB did not have the potential for success while eBay did. It seems to us that eBay succeeded because it was less of an economist vision than BB. It was less concerned with cheating than BB would ever have been. BB appealed to interests, tried to build in checks on truth and hedged his information where eBay was trustful and openly appealing more to community feelings of those transacting over the net on its platform. It could do so since addressing originally only members of a single company it started in a kind of community environment. We speculate that it was triggering thereby a kind of co-operative heuristic in participants. When transcending the borders of the original 'community' additional transaction partners were taken into an already co-operative environment. Again certain heuristics recommending 'communitarian' types of behavior must have thereby been triggered. So the initial success of eBay perhaps was possible only since it was not so much of an economic institution as BB. Nevertheless the long run success may depend on eBay becoming more 'knave proof' or economic.

It seems worthwhile to illustrate such speculative reasoning somewhat further by a short comparison between BB and eBay. We will then be in a position to address again the question of whether what was good for eBay initially will be good for it over the long haul or whether eBay might have to move somewhat closer to BB if it is to survive its own success on the several national and even global levels. In doing so we neglect the problems that might arise for both eBay and BB if these platforms are—as seems to be the case—used to perform transactions in illegally acquired goods.

4. A Comparison of eBay and BB

In particular the fact that reputation information is publicized on the net and thereby provided as a collective good to all members of eBay seems to be significant. Distributing reputation information freely facilitates seeking out partners with certain reputation characteristics. More generally speaking eBay is only in a much weaker sense a club than BB. In particular lying about one's identity is possible quite easily in eBay and has been excluded in the environment created by BB. There is also no incentive provided by eBay at least not such a strong one as in conceivable BB environments to report the results of dealings. Where BB would diminish the reputation of those who are not reporting the results of transactions, eBay does not do any of the kind but relies on an intrinsic motivation to report. People must be motivated to report and to do so truthfully. Here as in other regards eBay seems at least implicitly to appeal to feelings of belonging to a ‘community’ and the like. To put it slightly otherwise eBay seems to rely more on communitarian trust and on retributive emotions than commercial platforms of the BB type as imagined in the corresponding economically motivated thought and simulation experiment.

Emotional trust is fragile in particular in a large transactions’ system under conditions of individual insignificance. It may be that human individuals who cannot really shift their behavioral gears will often act in the new environment of eBay as if participating in a small numbers’ interaction. However, it seems doubtful that we can ‘trust’ in the prevalence of such a systematic error over the long run. Bearing in mind the basic results of the indirect evolutionary approach to trust in large transactions’ systems we must ask whether or not eBay really can create evolutionarily stable co-operation in the following sense: Is type detection in eBay sufficiently reliable and cheap to prevent innovative untrustworthiness from crowding out trustworthy behavior? Will net platforms have to acquire stronger club characteristics and have to impose stricter rules of admission in view of strategic reputation building (e.g. accumulating a reputation for trustworthiness incrementally with small stakes while going for the ‘big kill’ in a fraud eventually)?

BB must make strategic decisions when setting up his reputation formation algorithm as well as other rules of the game that he provides. Likewise eBay must consider modifications of its rules if it intends to be prepared for future challenges to its viability as might emerge rather soon (again increasing trafficking in illegally acquired commodities being a crucial issue). It is fruitful to bear in mind BB when addressing such issues for eBay.

Somebody who would not know anything about eBay would certainly imagine that a reasonable BB would not disclose the full counting type to the participants for free. He would rather insist on giving aggregate information on trustworthiness as seller or buyer respectively and would disclose it only on demand and against payment. If BB discloses all information and perhaps also allows the presentation of qualitative information, e.g. qualitative assessments by the parties to a transaction, then BB gives up control over how the information is used and aggregated by customers. Customers might like that in the first place but it might undermine the viability of the platform since there would not anymore be clear reputation signals applying to all and fixed by the platform providers.

Now, compare that with eBay. eBay partly presents the reputation information in aggregate form. So it seems that it follows the same policy. But quite surprisingly yet also quite in line with its policy of being transparent eBay also allows to check the past transaction records of participants. eBay provides more information than it needs to and than is used most of the time by participants.

Clearly transactions can fail for reasons other than deliberate cheating or deliberate exploitation. Things may go wrong by chance. If we include that possibility then there should be some tolerance in the formation of reputation. The unforgiving approach according to which those who cheated once are ‘out of the game’ so to say might be self-defeating. But forgiving strategies may be subject to exploitation. For instance if people know how close they are to the threshold beyond which they will be trusted or not trusted in a forgiving strategy they could respond to this information strategically. The most dangerous transaction partners should then be those who have never cheated before but might cheat without risking their reputation since the built in tolerance would protect them. On the other hand those who would fall below the threshold of acceptability as a trustworthy partner of interaction should they fail to fulfill their part of a specific bargain might be particularly trustworthy or rather reliable for opportunistic reasons in that deal.

If the full counting type were revealed, actors could themselves draw conclusions. Some actors might be unforgiving, others might be very forgiving still others might act in flexible ways responding to additional information as might be available. As indicated, specialized services judging the trustworthiness of participants might be expected to emerge as secondary organizations on such a platform. This may in one way improve the reputation mechanism but it may also reduce the financial basis of the platform and the willingness of transaction partners to report—in particular if they have special counseling deals with secondary services. In view of this it does not seem to be obvious that the strategy of eBay to disclose more or less all information to all transaction partners is the optimal one or even one that can sustain eBay over the long haul. The issue is not merely whether or not eBay is sufficiently robust against increasingly professional efforts of cheating, it is also whether an informal set up as of eBay might not destabilize its own basis by providing its services too freely and in ways that invite secondary services to enter transactions.

Secondary services already offered by eBay as costly devices, but rarely used might act as 'trustees' who, for instance, guarantee both payment and delivery of items as promised and in a quality as promised. So-called 'pick up points' (typically gas stations and the like) are used nowadays already to facilitate Net-transactions of other providers. The goods are delivered there and after inspection paid typically in cash. Firms offering new goods offer to testify the quality of used products of their own brand that are meant to go for sale on the net etc.—This all may be helpful in containing risks of transactions but it is costly to some extent and widely spread use of such services will render the process much more clumsy. It might in particular also hurt community feelings and as a consequence crowd out the intrinsic motivation to report the results of transactions on which eBay as of now relies.

5. Where Will and Where Should eBay Go from Here?

Bearing in mind the lessons from the indirect evolutionary approach to trust in large transactions' systems it is not at all obvious that eBay in its present form might be a sustainable success. Success may breed new success, though, in that the present platform may develop into a new, different one in ways dependend on the stage already reached on the development path. It seems central to inquire whether or not good conduct can be evolutionarily stable in the environment created by eBay. Our skepticism may be due to our outsider point of view but we would certainly be less skeptical if some of the following issues could be dealt with convincingly:

- Is it possible to go beyond general propositions about reliability and costs and to say something more specific about those reputation mechanisms that lead to evolutionarily stable positive population shares of trustworthy individuals?
- If there is a population of basically trustworthy and truthful individuals, can it be sustained under a counting type characterization of trustworthiness as buyer or seller that is public on the net?
- What are the relative merits of tolerant as compared to intolerant reputation mechanisms and is there an optimal degree of tolerance as well as of information about distance to thresholds of tolerance?
- Would weighing with the monetary value of transactions render counting type mechanisms more stable in some precise sense of evolutionary stability against subversion by strategic reputation building in small deals followed by 'hit and run'-exploitation whenever a big deal comes up?

Some more or less experimental issues:

- Can we design economic experiments to identify how boundedly rational individuals respond to different forms of information in particular to counting type reputation?
- Can we present experimental evidence that and why the inclusion of qualitative ("he is a good eBayer") assessments by the parties into the information vector as practiced by eBay is helpful?
- Would the future availability of personal pictures of transaction partners (and other forms of personalizing relations) on the net affect behavior?
- Is it true that 'thicker' communication will further co-operation in general?

Of course, there has been quite a bit of experimentation on eBay notably by Karen Cook, Alvin Roth and Axel Ockenfels, Chris Snijders and, also mentioned here last but not least, by Toshio Yamagishi. For the present discussion some of Yamagishi's work (see 2002) seems to be particularly relevant. His theoretical argument that positive reputation in an open system may be superior to negative reputation over the long haul seems to be quite convincing. It should however also be factored in that reputation building affords no guarantee that positive reputations can not be strategically manipulated and sold. In particular if operations become more globalized there may be quite some efforts to cheat along that dimension. Whether in the end we will reach a kind of global eBay that would extend auctions to global levels as some might hope seems open to serious doubt. It seems rather likely that there will be countervailing forces that would threaten the evolutionary stability of such a set up.

It should be kept in mind that behaving as if trustworthy and true trustworthiness are two different things. The first will not work without the second over the long haul if something like eBay is to work on as smoothly as it does as of now. As the indirect evolutionary approach suggests the reliable traders must be distinguishable from 'fortune hunters' who behave only as if genuinely trustworthy with some 'reliability'. As far as this is concerned eBay uses its reputation score and provides—for those who are willing to go through the somewhat tedious exercise of retrieving it—additional information on the value and kind of transactions. To reduce information costs weighted measures of performance may be provided such as to prevent strategies of accumulating a good reputation by many small transactions and then to go for the 'big kill'. It may also be true that the untrustworthy suffer from higher discount rates, 'weakness of the will' problems or a lack of the 'indirectly' profit-maximizing intrinsic motivation to realize the long run gains of a good reputation because they just cannot resist the temptation to cheat on small things (see Frank 1988). Still, there will be opportunists who specialize in seemingly good conduct for exploiting trust by big kills. Again there may be remedies by not allowing too many transactions of the same individual of higher value at the same time on eBay. However, going through all the possible remedies, in the end the big deals will depend on other measures to facilitate exchange (guarantees, hostages etc.). A global eBay will be a global flea market but not a global big deal market.

Bibliography

Alchian, A. A. (1950), Uncertainty, Evolution, and Economic Theory, in: *Journal of Political Economy 58*, 211–221

Ellickson, R. C. (1991), *Order without Law: How Neighbors Settle Disputes,* Cambridge/MA-London

Frank, R. (1988), *Passions within Reason. The Strategic Role of the Emotions*, New York

Güth, W./B. Peleg (2001), When Will Payoff Maximization Survive?—An Indirect Evolutionary Analysis, in: *Evolutionary Economics 11*, 479–499

— /C. Schmidt et al. (forthcoming), Fairness in the Mail and Opportunism in the Internet—A Newspaper Experiment on Ultimatum Bargaining, in: *German Economic Review*

— /H. Kliemt (1994), Competition or Co-operation: On the Evolutionary Economics of Trust, Exploitation and Moral Attitudes, in: *Metroeconomica 45*, 155–187

— / — (2000), Evolutionary Stable Co-operative Commitments, in: *Theory and Decision* 49, 197–221

— / — /B. Peleg (2000), Co-evolution of Preferences and Information in a Simple Game of Trust, in: *German Economic Review 1*, 83–100

— / — /G. Brennan (2003), Trust in the Shadow of the Courts, in: *Journal of Institutional and Theoretical Economics (JITE) 159*, 16–36

Güth, S./W. Güth, W./H. Kliemt (2002), The Dynamics of Trustworthiness Among the Few, in: *The Japanese Economic Review 53*, 369–388

Hicks, J. (1979), *Causality in Economics*, Oxford

Klein, D. B. (ed.) (1997), *Reputation. The Elicitation of Good Conduct*, Ann Arbor

Kreps, D./P. Milgrom et al. (1982), Rational Cooperation in the Finitely-Repeated Prisoners' Dilemma, in: *Journal of Economic Theory 27*, 245–252

Yamagishi, Toshio (2002), The Role of Reputation in Open and Closed Societies: An Experimental Study of Internet Auctioning, http://ccs.mit.edu/dell/reputation/YamagishiMIT.pdf