Learning to trust strangers: an evolutionary perspective

Pierre Courtois, Tarik Tazdait

L A M E T A

Laboratoire Montpelliérain
d'Economie Théorique et Appliquée

U M R
Unité Mixte de Recherche

DOCUMENT de RECHERCHE

« Learning to trust strangers:
an evolutionary perspective »

Pierre COURTOIS
Tarik TAZDAÏT

DR n°2011-06

Unite de Formation et de Recherche d'Economie
Avenue Raymond DUGRAND C.S. 79606
3 4 9 6 0 MONTPELLIER Cedex 2
Tel : 33 (0) 467158495 Fax : 33(0)467158467
E-mail : lameta@lameta.univ-montp1.fr

Université Montpellier 1
agro Montpellier
Ecole Nationale Supérieure Agronomique
CNRS Centre National de la Recherche Scientifique
INRA

Learning to trust strangers: an evolutionary perspective

Pierre Courtois
INRA LAMETA

Tarik Tazdaït
CNRS CIRED

Abstract

What if living in a relatively trustworthy society was sufficient to blindly trust strangers? In this paper we interpret generalized trust as a learning process and analyse the trust game paradox in light of the replicator dynamics. Given that trust inevitably implies doubts about others, we assume incomplete information and study the dynamics of trust in buyer-supplier purchase transactions. Considering a world made of "good" and "bad" suppliers, we show that the trust game admits a unique evolutionarily stable strategy: buyers may trust strangers if, on the whole, it is not too risky to do so. Examining the situation where some players may play, either as trustor or as trustee, we show that this result is robust.

1 Introduction

Experimental studies report that generalized social trust, i.e. the willingness to trust strangers, is significant (e.g., Berg et al. 1995, Ortmann et al. 2000, Cox 2004) and differ from one country to the next (e.g., Fukuyama 1995, Knack and Keefer 1997). Generalized trust also oscillates between cohorts (e.g., Putnam 2000, Glaeser et al 2000, Alesina and la Ferrara 2000), between cities of different sizes (Yamagishi et al. 1998, Tsuji and Harihara 2002) and between social strata (Yosano and Hayashi 2005). Englemann and Normann (2010), in an experiment on the minimum effort game, have found that recent immigrants behave according to the standards of their home country, but that with time tend to behave according to the standards of the host country. This suggests that, contrary to findings that trust is an anchored cultural inheritance (Uslaner 2004), an individual may learn to trust or to distrust strangers when changing their environment. The literature has paid scant attention to the question of learning to trust, and idiosyncratic subjects are usually described as consistent in their propensity to trust over time (e.g., Katz and Rotter 1969, Uslaner 2002, Bjornskov 2006). An individual changing to a different social environment or a social policy modifying a social environment may however influence the propensity to trust strangers. Our main hypothesis in this paper is that individuals behave according to their general beliefs about the trustworthiness of their partners. We assume trust is not related to a set trait of individual personality (Ulsaner 2002) but rather to a trusting culture grounded on social and political institutions (Putnam 1993, 2000). This presupposes that trust is inextricably linked to generalized trustworthiness in society. If nearly everyone is trustworthy then trusting a stranger is a profitable bet; if nearly everyone is untrustworthy then it is clearly not.[1]

We consider a trust game paradox as defined by Kreps (1990). The only subgame perfect equilibrium of this game is the situation where the trustor withholds trust. This is a social paradox since without trust, social exchange is more costly. Kreps (1990) and Kandori (1992) argue that repetition is the natural solution to lack of trust in this game. If tit-for-tat strategies are followed, and if agents are sufficiently patient, a trusting relationship can be established as a Nash equilibrium. The assumption of different payoff functions, say of the inequity aversion type described in Fehr and Schmidt (1999), may also be conducive to trust.[2] Considering an indirect evolutionary approach, Güth and Kliemt (1994), Güth *et al.* (2000) and Ahn and Esarey (2008) show that trust is established as soon as a detection-type mechanism which reveals trustworthiness is implemented. Bicchieri et al. (2004) and Bravo and Tamburino (2009) offer a direct evolutionary analysis of trust. The former provide a dynamic replicator simulation model showing that within a homogeneous population, trust exchange may emerge spontaneously; the latter study reputation mechanisms and show that generalized trust is present in systems where information regarding the past behaviour of agents can spread sufficiently.

Complementarily to this literature, we are interested in explaining generalized trust under the condition of pure anonymity. We discard both the possibility of repeating the interaction with an identified partner and, more generally, any available detection mechanism. Like Bicchieri et al. (2004), we reconsider two of the standard trust games' strongest assumptions: (i) common knowledge; and (ii) complete information. We suggest that individuals have the ability to learn "good" strategies from observing what worked well in the past, and we consider that they make their choice with information asymmetries, where beliefs are relevant. This translates into considering an evolutionary setting where players learn to trust strangers in case it proves to be a

[1] Note that many contributions study determinants of generalized trust by focusing on the correlation between generalized trust and variables such as democracy (e.g., Bjornskov 2006), political participation (e.g., Montoro and Puchades 2010), professional associations (e.g. Sabatini 2009), economic

[2] Fehr (2009), in a synthesis of leading achievements in neuro-economics and behavioural economics, shows that trust may be explained as the conjunction of three elements: beliefs about other people's trustworthiness, risk preferences, and social preferences (in particular, betrayal aversion).

profitable bet.[3] A key question is then to ask when it is a profitable bet and when it is not. We focus on the behaviour of trustors and consider that trustworthiness is a type that implies a trustee's belief in the value of reciprocating trust (Cook and Cooper 2003, Cochard et al. 2004), Ahn and Esarey 2008). "Good" and "bad" trustees are in a fixed proportion in the population and we analyse how generalized trust comes to pervade a community of trustors.[4] Because a class of agents may have a larger strategy set, acting as both trustees and as trustors, we then enlarge our perspective to the situation where there is also a third type of player deciding alternatively whether to trust or not, and whether to honour or betray the trust.

The paper is organized as follows. Section 2 introduces the main characteristics of the trust game and its outcomes. Section 3 analyses the evolutionary outcome of the game using the replicator dynamics. We consider first that the roles of the players are set, and then turn to a setting where some players have mixed roles. Section 4 concludes. All proof is relegated to the appendix.

2 The trust game

We consider trust exchange between two players involved in a transaction, a buyer and a supplier[5]. Their trust relationship concerns for example the quality or the delivery of a product. Our approach applies however to other transactions and can be illustrative of employment relationship between an employer and an employee as studied by Kreps (1993) or of micro-credit relationship between a lender and a borrower as studied by Karlan (2005). We denote respectively 1 and 2 the buyer and the supplier. The buyer chooses to trust or not the supplier delivering the product and the latter has an incentive to cheat. For example, the supplier could try to sell a good of lower quality than the one expected, or he could not deliver the product on time. Assume there are two types of suppliers: good suppliers (denoted 2^G) and bad suppliers (denoted 2^B). The first are in proportion p, the second in proportion 1-p. When the game starts the buyer does not know what type of supplier she is transacting with, but both players know their own type. It follows that in a prior move, nature (denoted N) determines the supplier's type. Once supplier type is assigned, then the players play the game depicted in figure 1. We assume that after observing the game's outcome, players change behaviour, imitating actions that yield higher benefits.

The game proceeds as follow. The buyer moves first and chooses between two strategies[6]: mistrust (M) or trust (T). If she mistrusts the supplier, the game ends. If the buyer trusts the supplier, it is the supplier's turn to move. The supplier's strategy space encompasses two possible actions: to honour the trust (H), or exploit it (E). Note that in the figure, numbers in brackets are probabilities of nature moves and the terms at the root of the tree represent as usual, the payoffs of the players.

[voir Annexe Figure 1.]

To obtain the trust game we assume: $c_1 > 0 > b_1$, $c_2 > b_2 > 0$. As in Gautschi (1999), the two types of supplier can be distinguished via small differences in their reward c_2. If the supplier honours the trust placed by the buyer, fairness considerations increase his payoff c_2 by ε. More precisely, if the supplier is a good supplier, his additional payoff is $\varepsilon \geq b_2 - c_2$ which may be interpreted as a taste

[3] We examine the game using the replicator dynamics model. The underlying idea is that the frequency of a strategy increases exactly when it yields payoffs above the average. To justify this modelling choice, recall that Schlag (1998) argues that the limiting case of a learning process that depends on imitation yields the replicator dynamics. Also, Erev and Roth (1998) show that some of the learning models in the psychology literature are approximations of the replicator dynamics model.

[4] We are not interested in this paper on the evolution of trustworthiness. Trustworthiness could spread in a population as a result of learning or of evolutionary pressure. This question will be analysed later in this research project.

[5] For readability purpose we assume supplier is a he and buyer is a she.

[6] From an evolutionary perspective, agents are identified with a strategy, and the relative frequency of a strategy within the population is the proportion of agents that adopt it.

for reciprocity. If he is bad, his additional payoff is defined by $\beta \leq b_2 - c_2$ which may be interpreted as a taste for egoism. It results that if the supplier is a good supplier, he honours trust in the transaction; if he is a bad supplier, he betrays trust.
The outcome when the buyer mistrusts the supplier constitutes the reference situation. In this case, the gains are normalized to 0. This is the *status quo* point, *i.e.* the state that exists before trust emerges.
Before considering evolution dynamics, we start focusing on the case where players are rational. Given our setting, player 1 only chooses T if her expected payoff is larger than when choosing M. Therefore, player 1 places trust if the following inequality holds*:* $p > -b_1/(c_1 - b_1)$, and then, when p exceeds the ratio of a potential loss due to unjustified trust $(-b_1)$ and a potential gain due to a justified trust $(c_1 - b_1)$. This ratio corresponds in fact to the risk to a buyer of placing trust (Snijders and Keren, 1999). And she will not place trust if p is smaller than $-b_1/(c_1 - b_1)$. Two Bayesian Nash equilibrium competes: (a) or p > $-b_1/(c_1 - b_1)$ and player 1 chooses T while player 2 chooses H if he is a good supplier, and E if he is a bad supplier; (b) or p < $-b_1/(c_1 - b_1)$ and then, player 1 chooses M and no trust relationship takes place.

3 The evolutionary dynamics

3.1 Trust game with fixed role

We start focusing on the case where buyers and suppliers are fixed in the role of trustor and trustee, they are randomly matched to play the game. Let p_{2G} (respectively p_{2B}) be the proportion of players 2^G (respectively 2^B) choosing H and 1-p_{2G} (respectively 1-p_{2B}) the proportion of players 2^G (respectively 2^B) choosing E. Also, let p_1 (respectively 1-p_1) be the proportion of players 1 choosing T (respectively M). Following Taylor and Jonker (1978) and Zeeman (1980), the distribution of strategies in the population over time can be described by the replicator dynamics:

$$\dot{p}_i = p_i\,[\pi_i - \pi_i^m] \qquad i = 1,\ 2^G,\ 2^B. \tag{1}$$

where π_i is the expected gain of player *i* and π_i^m is the average gain for the i-population. The replicator dynamics describes the variation in the proportion of players of each type *i* within the population in the next generation. It means that when the expected gain of player *i* increases (in relation to the average gain in the *i*-population) the growth rate $\dot{p}_i/p_i$ also increases and more players from the *i*-population will imitate *i*'s action.[7]
Two remarks immediately follow. First, if a pure strategy is absent in the initial population, it will never appear in the game. Second, the deterministic system (1) takes into account that certain strategies may disappear over time. The possibility of extinction accounts for the idea that certain populations are invaded by competing behaviours. However, strategies that disappear can reappear if the environment becomes favourable again (Gintis 2000).
Let us first analyse the behaviour of player 1. We next apply the same reasoning to the other player. Choosing T, player 1 will: (i) with a probability pp_{2G}, interact with a player 2^G playing H, enabling player 1 to get c_1; (ii) with a probability $p(1-p_{2G})$ interact with a player 2^G playing E, enabling player 1 to get b_1; (iii) with a probability $(1-p)p_{2B}$ interact with a player 2^B playing H, enabling

[7] Several authors define alternative learning models, but alternative adaptation rules lead inexorably towards the replicator dynamics (see among others Gale *et al*. 1995; Björnested and Weibull 1996). The axiomatic of the learning rules follows this trend by determining the conditions that induce the replicator dynamics (Börgers and Sarin 1997); the axioms that are introduced define the functional form of a desirable learning rule. Schlag (1998) is the only one that proposes a derivation of the replicator dynamics on the basis of an individual behaviour chosen optimally.

player 1 to get c_1; and (iv) with a probability $(1-p)(1-p_{2B})$, interact with a player 2^B playing E, enabling player 1 to get b_1.
We deduce that the expected gain of player 1 is:

$$\pi_1 = pp_{2G}c_1 + p(1-p_{2G})b_1 + (1-p)p_{2B}c_1 + (1-p)(1-p_{2B})b_1$$

and the average gain of all players 1 is:

$$\pi_1^m = p_1[pp_{2G}c_1 + p(1-p_{2G})b_1 + (1-p)p_{2B}c_1 + (1-p)(1-p_{2B})b_1] + (1-p_1)0 .$$

It follows that the replicator equation describing the rate of growth of p_1 in the population is given by:

$$\dot{p}_1 = p_1(1-p_1)\,[pp_{2G}c_1 + p(1-p_{2G})b_1 + (1-p)p_{2B}c_1 + (1-p)(1-p_{2B})b_1] .$$

By analogy, the growth rate of p_{2G} and of p_{2B} are:

$$\dot{p}_{2G} = p_{2G}(1-p_{2G})p_1(c_2+\varepsilon-b_2) \text{ and } \dot{p}_{2B} = p_{2B}(1-p_{2B})p_1(c_2+\beta-b_2) .$$

In order to study evolutionary dynamics, we examine the evolutionary stable strategies (ESS) of the game. Recall that a strategy $r^* = (p_1^*, p_{2G}^*, p_{2B}^*)$ is said to be ESS if it is a best reply to itself and it is a better reply to any alternative best reply μ than μ is to itself (Maynard Smith and Price 1973). Because the game is asymmetric, if r^* constitutes an ESS, r^* is a pure strategy and there is no better reply to r^* (Selten 1980, 1988). This means in our game that ESS can only be found at the corners of the unit cube, that is the 8 potential ESS: (1,1,1), (1,1,0), (1,0,1), (0,1,1), (1,0,0), (0,1,0), (0,0,1) and (0,0,0).
As shown by Gardner and Morris (1991), if r^* is an ESS[8], then r^* is a dynamic equilibrium which is hyperbolically stable with respect to the dynamics (1). In other words, identifying the ESS translates into identifying dynamic equilibria that are hyperbolically stable.

Definition 1. *Let r^* be a dynamic equilibrium of the system described by (1) and J the Jacobian of the replicator dynamics, r^* is said to be hyperbolically stable if all the eigenvalues of J evaluated at r^* have negative real parts.*

Analysis of all cases let us deduce the following result:

Proposition 1. *In the trust game considered, for $p > \frac{-b_1}{c_1 - b_1}$ the replicator dynamics model admits (1,1,0) as unique ESS.*

This unique ESS depicts the situation where all buyers trust suppliers, good suppliers honour this trust and bad ones abuse it. As soon as $p > -b_1/(c_1 - b_1) > 0$, the risk to place trust is sufficiently low and the probability that a buyer match a good supplier is high enough for systematic trust. Although bad suppliers are encouraged over time systematically to abuse trusting players since this allows them to reap higher benefits, the buyers continue trusting given that the expected gain associated to strategy M is lower than the expected gain associated with strategy T.

[8] Note that for asymmetric games, Selten (1983,1988) proposed the more general concept of limit evolutionarily stable strategy (LESS). The LESS reflects the idea that the equilibrium strategy may be seen as the limit of ESS for close perturbed games. In these perturbed games, admissible strategies may be required to play some actions with arbitrarily small probability. We use to study the LESS of a game when there is no ESS. This is not the case in our model.

We deduce that despite incomplete information, when relations of trust are held, they may spread as if they were new social norms. Trust is behaviourally internalized and players reproduce it naturally. This is consistent with the notion of "*embeddedness*" depicted by Granovetter (1985). The on-going networks of social relations between people discourage malfeasance and trust is viewed as embedded within social relationships. A key difference is however that embeddedness theory focuses on the enhancement of reputation in interpersonal relationships. Trust is portrayed as taking place locally first, within neighbourhood relationships, to extend next to more distant relationships. Instead, in our approach, individual reputation is secondary. A buyer is not interested into a particular seller but to have information (in probability terms) about the set of sellers composing the society. If in average sellers are trustable, a buyer will trust anyone, with an accepted probability to interact with an egoist. According to this view, individuals do not care about the risk of ending in a worse situation than his partner (Ashraf *et al.*, 2003) nor about the risk of being abused (Bohnet et Zeckhauser, 2004). As soon as the risk of ending in a worse situation than when absent of trust (Ben-Ner et Putterman, 2001), the buyer trusts.

3.2 The trust game allowing for mixed role

The assumption of fixed roles can be viewed as a limitation given that in the complexity of real life, some players may sometimes play both roles. We now introduce a third kind of player who can randomly take on either role. We call this player hybrid and denote him with the number 3. These players are typically the ones found buying and selling on Internet sites such as *priceminister* or *eBay* and conform to several situations involving trust transactions. Considering again the extensive form game represented in figure 1, the hybrid player can be in first or second position. When he is in first position, he has the same strategy space as a buyer. Respectively if he is in second position, he has the same strategy space as a supplier and can be of either good or bad type. We call a good type hybrid player 3^G and a bad type one 3^B. We suppose hybrid players are behaviourally consistent and we consider that if they trust as buyers they will honour trust as suppliers, strategies T and H are played by hybrid players proportionally. This assumption is consistent with the experimental studies of Deutsch (1958) or of Dubois and Willinger (2007) according to trusting and trustworthy behaviours tend to be displayed by the same agents.[9]
Introducing this new category of players modifies slightly our game. Now, when it starts, nature makes two random moves. First it determines the role of the hybrid player and second, the type of supplier. We assume the hybrid player to be buyer or supplier with a probability 1/2. Possible random matchings then are: a buyer and a supplier, a hybrid player and a supplier, a buyer and a hybrid player.
In order to study the evolutionary dynamics of this game, we proceed as in the previous case. That is, we start determine the replication equation associated to the behaviour of player 1, the difference being that now she may match with a player 2 or with a player 3. We proceed by analysing expected payoffs of players 2 and 3 which let us, overall, describe the dynamics by a system of five differential equations. Solving this system translates into identifying the set of ESS in pure strategy. We have $2^5 = 32$ potential ESS and studying the eigenvalues of the Jacobian we deduce that only two are ESS.

Proposition 2. *In the trust game with hybrid players:*

- *For* $p > \frac{b_2 - c_2 - \beta - b_1}{c_1 - b_1}$ *the replicator dynamics model admits (1,1,0,1,1) as unique ESS.*

[9] See also the discussion of Fehr (2009) on the self-reinforcing aspect of trust and trustworthiness. According to this author "the empirical evidence suggests that trust can be self-reinforcing" (p. 261).

- *For* $\frac{b_2 - c_2 - \beta - b_1}{c_1 - b_1} > p > \frac{-b_1}{c_1 - b_1}$ *the replicator dynamics model admits (1,1,0,1,0) as unique ESS.*

When the proportions of good suppliers and good hybrids are significant within their respective populations (*i.e.* superior to $(b_2 - c_2 - \beta - b_1)/(c_1 - b_1)$), the only stable state in the population is (1,1,0,1,1). In other words, whether they are hybrids or not, all buyers trust and all hybrids and good suppliers honour this trust. In this stable state, as in proposition 1, all buyers trust. However, the validity condition is now stricter: the proportions of good players must be higher. The principal difference between this setting and the previous one is that now, bad suppliers and notably hybrid ones honour the trust placed by the buyers. Indeed, when buying, a bad hybrid gets a high benefit by trusting because in average, he has a high probability to make a transaction with a supplier that will honour trust. As his behaviour when selling is correlated to his behaviour when buying, the hybrid supplier honours trust. This allows him to reap a high payoff since he matches only with trusting buyers. Given the expected payoff of a hybrid player is the average of what he obtains as a buyer and as a supplier, the overall gain is high. Note that when acting as a supplier, if a bad hybrid player systematically betrays buyers, he also yields high benefits. However, given consistence in behaviour, he never trusts when buying either. It follows that as a buyer he receives a nil benefit which balances the benefit he obtains as a supplier. In the long run, this player ends up in a worse situation than if he always placed and honoured trust. This effect does not exist for bad suppliers that always play second in the game. They never honour trust and this is all the more profitable that they match only with trusting buyers.
If the proportion of good suppliers and of good hybrids within their respective populations is lower than $(b_2 - c_2 - \beta - b_1)/(c_1 - b_1)$ but higher or equal to $-b_1/(c_1 - b_1)$, the game yields another stable state with similar characteristics but in which bad hybrids never trust when buying and always betray trust when supplying. Indeed for a buyer, given that there are fewer good suppliers, trusting them is less profitable. When buying, bad hybrids yield higher benefit on average from mistrusting. Because their behaviour is consistent, they always betray buyers when acting as suppliers. This allows them to increase their benefits and compensate for lack of benefit when playing as buyers. This stable state admits a similarity to proposition 1: all bad players, hybrids or not, always exploit trust. Again, the validity condition is more restrictive given that now the proportion of good players will be between two bounds: high but not too high.
Finally note that under the same conditions as in proposition (1) and for any stable state, trust relationships regularly take place. Proposition (1) is robust to a setting where some agents have a larger strategy space.

4 Conclusion

In the absence of a contract and/or an institution guaranteeing some form of social control, agents with limited knowledge may trust and learn to trust because this is simply a good bet. Focusing on trust in transactions, we show that buyers may trust indiscriminately when the proportion of good suppliers is above a given threshold. Bad suppliers continue betraying trusting buyers but they are few enough not to undermine generalized trust. This leads us to conclude that a trustful world does not necessarily imply a fully trustworthy one. Introducing a third kind of individual, we analysed whether the introduction of agents playing mixed roles threatens trust. The answer is no and our results remain fundamentally similar.
In our opinion, these preliminary results are important in order to estimate the thresholds that may allow for generalized trust to be established in a society. As Nooteboom (2010) shows, trust and institutions are both substitutes and complements. Contracts support reliance but can be destructive to trust, and one can argue that contracts may be used in order to attain a certain threshold ensuring trustworthiness. Complementing the work of Zucker (1986) and Nooteboom (2002), a threshold-

based analysis would allow for a better understanding of the design of contracts and notably the best timing for lasting trust to be established.

References

Ahn T.K., Esarey J. (2008), “A Dynamic Model of generalizad Social Trust”, *Journal of Theoretical Politics*, 20(2), p. 157-180.

Alesina A., Ferrara E. (2000), “Participation in Heterogeneous Communities”, *Quaterly Journal of Economics*, 115, p. 847-904.

Ashraf N., Bohnet I., Piankov N. (2003), "Decomposing Trust and Trusworthiness”, *Experimental Economics*, 9, p. 193-208.

Ben-Ner A., Putterman L. (2001), "Trusting and Trustworthiness", *Boston University Law Review*, 81, p. 523-551.

Berg J., Dickhaut J., McCabe K. (1995), “Trust, Reciprocity, and Social History”, *Games and Economic Behavior*, 10, p. 122-142.

Bicchieri C., Duffy J., Tolle G. (2004), “Trust among Strangers”, *Philosophy of Science*, 71, p. 1-34.

Björnested J., Weibull J. (1996), “Nash Equilibrium and Evolution by Imitation”, In: K. Arrow, E. Colombatto, M. Perlman, C. Schmidt (eds), *The Rational Foundations of Economic Behaviour*, MacMillan, London, p. 155-171.

Bjornskov C. (2006), “Determinants of Generalized Trust: a Cross Vountry Comparison”, *Public Choice*, 130, p. 1-21.

Bohnet I., Zeckhauser R.J. (2004), "Trust, Risk and Betrayal", *Journal of Economic Behavior and Organization*, 55, p. 467-484.

Börgers T., Sarin R. (1997), “Learning through Reinforcement and Replicator Dynamics”, *Journal of Economic Theory*, 77, p. 1-14.

Bravo G., Tamborino L. (2008), “The Evolution of Trust in Non-Simultaneous Exchange Situations”, *Rationality and Society*, 20(1), p. 85-113.

Cochard F., Nguyen Van P., Willinger M. (2004), “Trusting Behavior in a Repeated Investment Galme*”, Journal of Economic Behavior and Organization*, 55, p. 31-44.

Cook K., Cooper R. (2003), “Experimental Studies of Cooperation, Trust and Social Exchange”, in E.Ostrom and J.Walker (eds) *Trust and Reciprocity: Interdisciplinary Lessons from Experimental research*, pp. 209-44, New York: Russell Sage Foundation.

Cox C. (2004), “How to Identify Trust and Reciprocity”, *Games and Economic Behavior*, 46(2), p. 260-281

Deutsch M. (1958), “Trust and Suspicion”, *Journal of Conflict Resolution*, 2, p. 265-279.

Dubois D., Willinger M. (2007), “The Role of Players’ Identification in the Population on the Trusting and the Trustworthy Behaviour: an Experimental Investigation”, Working paper Lameta Montpellier.

Engelmann D., Normann H.T. (2010), “Maximum Effort in the Minimum Effort Game”, *Experimental Economics*, 13(3), p. 249-259.

Erev I., Roth A.E. (1998), “Predicting How People Play Games: Reinforcement Learning in Experimental games with Unique, Mixed Strategy Equilibria”, *American Economic Review*, 88, p. 848-881.

Fehr E. (2009), “On the Economics and Biology of Trust”, *Journal of the European Economic Association*, 7, p. 235-266.

Fehr E., Schmidt M. (1999), “A Theory of Fairness, Competition and Cooperation”, *Quarterly Journal of Economics*, 114, p. 817-868

Fukuyama F. (1995), *Trust: The Social Virtues and the Creation of Prosperity*, New York: Free press.
Gale J., Binmore K., Samuelson L. (1995), "Learning to be Imperfect: The Ultimatum Game", *Games and Economic Behavior*, 8, p. 56-90.
Gardner R., Morris M.R. (1991), "The Evolutionary Stability of Bluffing in a Class of Extensive Form Games", In: R. Selten (Ed.), *Game Equilibrium Models, vol. 1: Evolution and Game Dynamics*, Spinger-Verlag, Berlin.
Gautschi T. (1999), "A Hostage Trust Game with Incomplete Information and Fairness Considerations of the Trustee", Working Paper, Utrecht University, Department of Sociology.
Gintis H. (2000), *Game Theory Evolving*, Princeton University Press, Princeton.
Glaeser E.L., Laibson D.I., Scheinkman J.A., Soutter C.L. (2000), "Measuring Trust", *Quarterly Journal of Economics*, 115(3), p. 811-846.
Granovetter M. (1985), "Economic Action and Social Structure: The Problem of Embeddedness", *American Journal of Sociology*, 91, p. 481-510.
Güth, W., Kliemt, H. (1994), "Competition or Co-Operation: on the Evolutionary Economics of Trust, Exploitation and Moral Attitudes", *Metroeconomica*, 45, p. 155-187.
Güth W., Kliemt H., Peleg, B. (2000), "Co-evolution of Preferences and Information in Simple Games of Trust", *German Economic Review*, 1, p. 83-110.
Holm J.J., Danielson A. (2005), "Tropic Trust versus Nordic Trust: Experimental Evidence from Tanzania and Sweden*", The Economic Journal*, 115, p. 505-532.
Kandori M. (1992), "Social Norms and Community Enforcement", *Review of Economic Studies*, 59, p. 63-80.
Karlan D.S. (2005), "Using Experimental Economics to Measure Social Capital and Predict Financial Decisions", *American Economic Review*, 95, p. 1688-1699.
Katz H.A., Rotter J.B. (1969), "Interpersonal Trust Scores of College Students and their Parents", *Child Development*, 40, p. 657-661.
Knack S., Keefer P. (1997), "Does Social Capital Have an Economic Payo? A Cross-Country Investigation", *Quarterly Journal of Economics*, 112(4), p. 1251-1288.
Knack S., Zak P.J. (2002), "Building Trust: Public Policy, Interpersonal Trust, and Economic Development, *Supreme Court Economic Review*, 10, p. 91-107.
Kreps D. (1990), "Corporate Culture and Economic Theory", In: J.E. Alt and K.A. Shepsle (Eds.), *Perspectives on Positive Political Economy*, Cambridge University Press, Cambridge, p. 90-143.
Kreps D.M. (1993), Game Theory and Economic Modelling, Oxford, Clarendon Press (Fourth edition).
La Porta R., Lopez-de-Silanes F.,, Shleifer A., and Vishny R.W (1997), "Trust in Large Organizations", *American Economic Review*, 87(2), p. 333-338.
Maynard Smith J., Price G.R. (1973), "The Logic of Animal Conflicts", *Nature*, n°246, p. 15-18.
Montoro J.D., Puchades M. (2010), "Social Capital and Political Participation", *Economics Bulletin*, 30(4), p. 3329-3347.
Nooteboom B. (2002), *Trust: Forms, Functions, Foundations, Failures and Figures;* Cheltenham Uk: Edward Elgar.
Nooteboom B. (2010), *The Dynamics of Trust: Communication, Action and Third Parties*, Miméo Tilburg.
Ortmann A., Fitzgerald J., Boeing C. (2000), "Trust, Reciprocity, and Social History: A Re-examination", *Experimental Economics*, 3(1), p. 81-100.
Putnam R.D. (1993), Making Democracy Work: Civic Traditions in Modern Italy, Princeton, Princeton University Press.

Putnam R.D. (2000), Bowling Alone: The Collapse and Revival of American Community, New York, Simon and Schuster.

Sabatini F. (2009), "The Relationship between Trust and Networks. An exploratory empirical analysis", Economics Bulletin, 29(2), p. 661-672.

Schlag K.H. (1998), "Why Imitate, and if so, How? A Bounded Rationality Approach to Multi-armed Bandits", *Journal of Economic Theory*, 78, p. 127-159.

Selten R. (1980), "A Note on Evolutionarily Stable Strategies in Asymmetric Animal Conflicts", *Journal of Theoretical Biology*, 84, p. 93-101.

Selten R. (1983), "Evolutionary Stability in Extensive Two-Agent Games", *Mathematical Social Sciences*, 5, p. 269-363.

Selten R. (1988), "Evolutionary Stability in Extensive Two-Agent Games: Correction and further Development", *Mathematical Social Sciences*, 16, p. 223-266.

Snijders C., Keren G. (1999), "Determinants of Trust", In: D.V. Budescu, I. Erev and R. Zwick (Eds.), *Games and Human Behavior*, Lawrence Erlbaum, Mahwah, p. 355-385.

Taylor P., Jonker L. (1978), "Evolutionarily Stable Strategies and Game Dynamics", *Mathematical Biosciences*, 40, p. 145-156.

Tsuji R., Harihara M. (2002), "Urban and Rural Social Networks and General Trust". Proceedings of the 43rd JSSP Conference (Japan), p. 114-115.

Uslaner, E.M. (2002), *The Moral Foundations of Trust*, Cambridge University Press: New York.

Uslaner E.M. (2004), *Where you Stand Depends on Where your Grand Parents Sat: The Inheritability of Generalized Trust*. Miméo, University of Maryland.

Yamagishi T., Cook, K.S., Watabe M. (1998), "Uncertainty, Trust and Commitment Formation in the United States and Japan", *American Journal of Sociology*, 104, p. 165-194.

Yamamura, E. (2008), "Determinants of Trust in a Racially Homogeneous Society", *Economics Bulletin*, 26(1), p.1-9.

Yosano A., Hayashi N. (2005), "Social Stratification, Intermediary Groups and Creation of Trustfulness", *Sociological Theory and Methods*, 20(1), p. 27-44.

Zeeman E.C. (1980), "Population Dynamics from Game Theory", In: Z. Nitecki and C. Robinson (Eds.), Global Theory of Dynamical Systems, Lecture Notes in Mathematics vol. 819, Springer-Verlag, Berlin, p. 471-497.

Zucker L.G. (1986), "Production of Trust: Institutional Sources of Economic Structure", *Research in Organisationtional Behaviour*, 8, p. 53-111.

Appendix

Proof of Proposition 1

The Jacobian of the replicator dynamics at any point r = (p_1, p_{2G}, p_{2B}) can be expressed as follows:

$$J(r) = \begin{pmatrix} \partial\dot{p}_1/\partial p_1 & \partial\dot{p}_1/\partial p_{2G} & \partial\dot{p}_1/\partial p_{2B} \\ \partial\dot{p}_{2G}/\partial p_1 & \partial\dot{p}_{2G}/\partial p_{2G} & \partial\dot{p}_{2G}/\partial p_{2B} \\ \partial\dot{p}_{2B}/\partial p_1 & \partial\dot{p}_{2B}/\partial p_{2G} & \partial\dot{p}_{2B}/\partial p_{2B} \end{pmatrix},$$

and therefore:

$$J(r)=\begin{pmatrix} (1-2p_1)[pp_{2G}c_1+p(1-p_{2G})b_1+(1-p)p_{2B}c_1+(1-p)(1-p_{2B})b_1] & p_1(1-p_1)p(c_1-b_1) & -p_1(1-p_1)(1-p)(c_1-b_1) \\ p_{2G}(1-p_{2G})(c_2+\varepsilon-b_2) & (1-2p_{2G})p_1(c_2+\varepsilon-b_2) & 0 \\ p_{2B}(1-p_{2B})(c_2+\beta-b_2) & 0 & (1-2p_{2B})p_1(c_2+\beta-b_2) \end{pmatrix}$$

According to definition (1), in order to characterize the ESS we should study the eigenvalues of the matrix J in each of the 8 points that are candidates.

Let $r^* = (1,1,1)$. The Jacobian is:

$$J(1,1,1) = \begin{pmatrix} -c_1 & 0 & 0 \\ 0 & -(c_2+\varepsilon-b_2) & 0 \\ 0 & 0 & -(c_2+\beta-b_2) \end{pmatrix}.$$

The eigenvalues are: $-c_1$, $-(c_2+\varepsilon-b_2)$, $-(c_2+\beta-b_2)$. They are all negative (which shows that (1,1,1) is hyperbolically stable) if:

$c_1 > 0$, $(c_2+\varepsilon-b_2) > 0$ and $(c_2+\beta-b_2) > 0$.

The condition on the gains of 2^B (i.e. $(c_2+\beta-b_2) > 0$) contradicts one of the assumption of the trust game and we deduce that (1,1,1) is not an ESS. In the same manner, we show that (1,0,1) and (1,0,0) are also not ESS.

We focus now on (1,1,0), (0,1,1), (0,1,0), (0,0,1) and (0,0,0). For (1,1,0), we have :

$$J(1,1,0) = \begin{pmatrix} -[pc_1+(1-p)b_1] & 0 & 0 \\ 0 & -(c_2+\varepsilon-b_2) & 0 \\ 0 & 0 & (c_2+\beta-b_2) \end{pmatrix}$$

and this point is an ESS because conditions on the gains of the two types of player 2 do not contradict the assumptions of the model. We have: $c_2+\varepsilon-b_2 > 0$ and $c_2+\beta-b_2 < 0$ and these conditions are verified for: $p > \frac{-b_1}{c_1-b_1} > 0$.

Regarding (0,1,1), (0,1,0), (0,0,1) and (0,0,0), we obtain respectively:

$$J(0,1,1) = \begin{pmatrix} c_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},\ J(0,1,0) = \begin{pmatrix} pc_1+(1-p)b_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$J(0,0,1) = \begin{pmatrix} pb_1 + (1-p)c_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad J(0,0,0) = \begin{pmatrix} b_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

For each of those cases, 0 is an eigenvalue and they are not ESS which concludes the proof.

Proof of proposition 2

We proceed as in proposition 1, the difference being that now there is a third type of player that may both play as buyer or seller. We denote by p_{3G} (respectively p_{3B}) the proportion of players 3^G (respectively 3^B) choosing T when buying and choosing H when selling. The complement 1-p_{3G} (respectively 1-p_{3B}) is the proportion of players 3^G (respectively 3^B) choosing M when buying and E when selling.

We start determine the replication equation associated to the behaviour of player 1. If she matches a player 2, her expected payoff is $pp_{2G}c_1 + p(1-p_{2G})b_1 + (1-p)p_{2B}c_1 + (1-p)(1-p_{2B})b_1$ and if she matches a player 3, her expected payoff is $pp_{3G}c_1 + p(1-p_{3G})b_1 + (1-p)p_{3B}c_1 + (1-p)(1-p_{3B})b_1$. Since player 1 matches a player 2 with probability 1/2 and a player 3 with a probability 1/2, her expected payoffs is $\pi_1 = \frac{1}{2}[p(p_{2G} + p_{3G})c_1 + p(2 - p_{2G} - p_{3G})b_1 + (1-p)(p_{2B} + p_{3B})c_1 + (1-p)(2 - p_{2B} + p_{3B})b_1]$.

The average gain of all 1-players is $\pi_1^m = p_1\pi_1 + (1-p_1)0$ and the growth rate of p_1 is $\dot{p}_1$ = $p_1(1-p_1)\pi_1 = p_1(1-p_1)\frac{1}{2}[p(p_{2G} + p_{3G})c_1 + p(2 - p_{2G} - p_{3G})b_1 + (1-p)(p_{2B} + p_{3B})c_1 + (1-p)(2 - p_{2B} + p_{3B})b_1]$

Consider now player 2^G. If he matches with a player 1, his expected payoff is $p_1(c_2 + \varepsilon)$. If he matches with a player 3, his expected payoff is $pp_{3G}(c_2 + \varepsilon) + (1-p)p_{3B}(c_2 + \varepsilon)$. Again, given player 2^G matches with a player 1 with a probability 1/2 and with a player 3 with a probability 1/2, his expected gain is π_{2G} = $\frac{1}{2}(c_2 + \varepsilon)[p_1 + pp_{3G} + (1-p)p_{3B}]$. The average gain of all 2^G-players is π_{2G}^m = $\frac{1}{2}[p_{2G}(c_2 + \varepsilon) + (1-p_{2G})b_2][p_1 + pp_{3G} + (1-p)p_{3B}]$ and the growth rate of p_{2G} is $\dot{p}_{2G}$ = $p_{2G}(1-p_{2G})\frac{1}{2}(c_2 + \varepsilon - b_2)[p_1 + pp_{3G} + (1-p)p_{3B}]$. Similarly, we deduce that $\dot{p}_{2B}$ = $p_{2B}(1-p_{2B})\frac{1}{2}(c_2 + \beta - b_2)[p_1 + pp_{3G} + (1-p)p_{3B}]$.

Finally, consider a hybrid player 3^G. If acting as a buyer, his expected payoff is $pp_{2G}c_1 + p(1-p_{2G})b_1 + (1-p)p_{2B}c_1 + (1-p)(1-p_{2B})b_1$ and if acting as a supplier, his expected payoff is $p_1(c_2 + \varepsilon)$. Total expected gain is then

$\pi_{3G} = \frac{1}{2}[p_1(c_2+\varepsilon) + pp_{2G}c_1 + p(1-p_{2G})b_1 + (1-p)p_{2B}c_1 + (1-p)(1-p_{2B})b_1]$ and we deduce

$\pi_{3G}^m = p_{3G}\pi_{3G} + \frac{1}{2}(1-p_{3G})p_1 b_2$. The growth rate of p_{3G} is:

$$\dot{p}_{3G} = \frac{1}{2}p_{3G}(1-p_{3G})[p_1(c_2+\varepsilon-b_2) + pp_{2G}c_1 + p(1-p_{2G})b_1 + (1-p)p_{2B}c_1 + (1-p)(1-p_{2B})b_1].$$

Similarly, we obtain:

$$\dot{p}_{3B} = \frac{1}{2}p_{3B}(1-p_{3B})[p_1(c_2+\beta-b_2) + pp_{2G}c_1 + p(1-p_{2G})b_1 + (1-p)p_{2B}c_1 + (1-p)(1-p_{2B})b_1].$$

The Jacobian of the coefficient matrix of the above dynamics is the following:

$$J(r) = \begin{pmatrix} \frac{\partial \dot{p}_1}{\partial p_1} & \frac{\partial \dot{p}_1}{\partial p_{2G}} & \frac{\partial \dot{p}_1}{\partial p_{2B}} & \frac{\partial \dot{p}_1}{\partial p_{3G}} & \frac{\partial \dot{p}_1}{\partial p_{3B}} \\ \frac{\partial \dot{p}_{2G}}{\partial p_1} & \frac{\partial \dot{p}_{2G}}{\partial p_{2G}} & \frac{\partial \dot{p}_{2G}}{\partial p_{2B}} & \frac{\partial \dot{p}_{2G}}{\partial p_{3G}} & \frac{\partial \dot{p}_{2G}}{\partial p_{3B}} \\ \frac{\partial \dot{p}_{2B}}{\partial p_1} & \frac{\partial \dot{p}_{2B}}{\partial p_{2G}} & \frac{\partial \dot{p}_{2B}}{\partial p_{2B}} & \frac{\partial \dot{p}_{2B}}{p_{3G}} & \frac{\partial \dot{p}_{2B}}{p_{3B}} \\ \frac{\partial \dot{p}_{3G}}{\partial p_1} & \frac{\partial \dot{p}_{3G}}{\partial p_{2G}} & \frac{\partial \dot{p}_{3G}}{\partial p_{2B}} & \frac{\partial \dot{p}_{3G}}{\partial p_{3G}} & \frac{\partial \dot{p}_{3G}}{\partial p_{3B}} \\ \frac{\partial \dot{p}_{3B}}{\partial p_1} & \frac{\partial \dot{p}_{3B}}{\partial p_{2G}} & \frac{\partial \dot{p}_{3B}}{\partial p_{2B}} & \frac{\partial \dot{p}_{3B}}{\partial p_{3G}} & \frac{\partial \dot{p}_{3B}}{\partial p_{3B}} \end{pmatrix}$$

For point $r^* = (1,1,0,1,1)$, we obtain :

$$J(1,1,0,1,1) = \begin{pmatrix} -\frac{[(1+p)c_1+(1-p)b_1}{2} & 0 & 0 & 0 & 0 \\ 0 & -(c_2+\varepsilon-b_2) & 0 & 0 & 0 \\ 0 & 0 & (c_2+\beta-b_2) & 0 & 0 \\ 0 & 0 & 0 & -\frac{[c_2+\varepsilon-b_2+pc_1+(1-p)b_1]}{2} & 0 \\ 0 & 0 & 0 & 0 & -\frac{[c_2+\beta-b_2+pc_1+(1-p)b_1]}{2} \end{pmatrix}$$

And we deduce that the eigenvalues are negatives if:

$$(1+p)c_1 + (1-p)b_1 > 0 \qquad (2)$$

$$c_2 + \varepsilon - b_2 > 0 \qquad (3)$$

$$c_2 + \beta - b_2 < 0 \qquad (4)$$

$$c_2 + \varepsilon - b_2 + pc_1 + (1-p)b_1 > 0 \qquad (5)$$

$$c_2 + \beta - b_2 + pc_1 + (1-p)b_1 > 0. \qquad (6)$$

By assumption, (3) and (4) are always true and (2), (5) and (6) lead to the following inequality:

$$p > \operatorname{Max}\left(\frac{-b_1 - c_1}{c_1 - b_1}, \frac{b_2 - c_2 + \beta - b_1}{c_1 - b_1}, \frac{b_2 - c_2 - \varepsilon - b_1}{c_1 - b_1}\right).$$

Given the assumption of the model, we deduce that (1,1,0,1,1) is an ESS when:

$$p > \frac{b_2 - c_2 + \beta - b_1}{c_1 - b_1}.$$

For point r^* = (1,1,0,1,0), the Jacobian is :

$$J(1,1,0,1,0) = \begin{pmatrix} -[pc_1 + (1-p)b_1] & 0 & 0 & 0 & 0 \\ 0 & -\frac{(1+p)(c_2+\varepsilon-b_2)}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{(1+p)(c_2+\beta-b_2)}{2} & 0 & 0 \\ 0 & 0 & 0 & -\frac{[c_2+\varepsilon-b_2+pc_1+(1-p)b_1]}{2} & 0 \\ 0 & 0 & 0 & 0 & \frac{[c_2+\beta-b_2+pc_1+(1-p)b_1]}{2} \end{pmatrix}$$

and we deduce that the eigenvalues are negatives if:

$$pc_1 + (1-p)b_1 > 0 \qquad (7)$$

$$c_2 + \varepsilon - b_2 > 0 \qquad (8)$$

$$c_2 + \beta - b_2 < 0 \qquad (9)$$

$$c_2 + \varepsilon - b_2 + pc_1 + (1-p)b_1 > 0 \qquad (10)$$

$$c_2 + \beta - b_2 + pc_1 + (1-p)b_1 < 0. \qquad (11)$$

Inequalities (8) and (9) are always true and we deduce from (7), (10) and (11) that:

$$\frac{b_2 - c_2 - \beta - b_1}{c_1 - b_1} > p > \operatorname{Max}\left(\frac{-b_1}{c_1 - b_1}, \frac{b_2 - c_2 - \varepsilon - b_1}{c_1 - b_1}\right).$$

Given $-b_1 > b_2 - c_2 - \varepsilon - b_1$, we deduce:

$$\frac{b_2 - c_2 - \beta - b_1}{c_1 - b_1} > p > \frac{-b_1}{c_1 - b_1}$$

which is the condition for (1,1,0,1,0) to be an ESS. End of the proof.

List of Figures

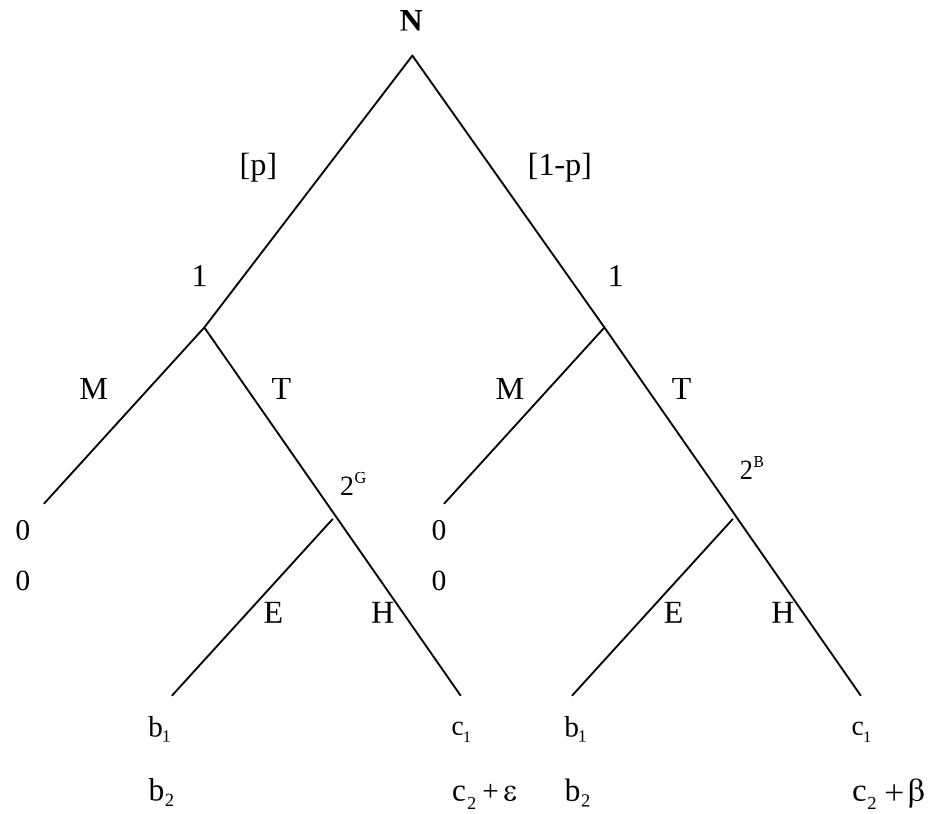

Figure 1. The trust game

Documents de Recherche parus en 2011[1]

DR n°2011 - 01 : Solenn LEPLAY, Sophie THOYER
« Synergy effects of international policy instruments to reduce deforestation: a cross-country panel data analysis »

DR n°2011 - 02 : Solenn LEPLAY, Jonah BUSCH, Philippe DELACOTE, Sophie THOYER
« Implementation of national and international REDD mechanism under alternative payments for environemtal services: theory and illustration from Sumatra »

DR n°2011 - 03 : Stéphanie AULONG, Robert KAST
« A conceptual framework to assess vulnerability. Application to global change stressors on South Indian farmers »

DR n°2011 - 04 : Nicolas QUEROU, Raphael SOUBEYRAN
« Voting Rules in Bargaining with Costly Persistent Recognition »

DR n°2011 - 05 : Pierre COURTOIS, Rabia NESSAH, Tarik TAZDAÏT
« How to play the games? Nash versus Berge behavior rules »

DR n°2011 - 06 : Pierre COURTOIS, Tarik TAZDAÏT
« Learning to trust strangers: an evolutionary perspective »

[1] La liste intégrale des Documents de Travail du LAMETA parus depuis 1997 est disponible sur le site internet : http://www.lameta.univ-montp1.fr

Contact :

Stéphane MUSSARD : mussard@lameta.univ-montp1.fr