The Social Dynamics of Learning and Trust

Maurizio Cortesi

January 31, 2011

Organization: University of Luxembourg - CREA[1]

Abstract

Trust is often described as an important variable in social interactions. This has been true not only for disciplines such as sociology, psychology or anthropology. Economics scholars have mainly used trust as a background characteristic, allowing exchanges and potentially influencing economic performance. In the last decades, however, also due to insights coming from other disciplines and to better available measures, economic research on trust consequences has become more and more quantitative. Simultaneously, the understanding of trust antecedents has deepened.

In this paper we present a model integrating some recent insights from experimental research. We investigate how buyer-seller exchanges based on trust and trustworthiness influence both individuals and population performance in terms of learning efficiency and quality of the interactions. We use a definition of trust based on economic primitives, such as beliefs and preferences, to evaluate the ability of agents to learn about their counterparts as to choose the most fruitful relationships.

We show that preferences and learning dynamics play an important role in shaping individuals performance. The results also support the idea that trust generates more trust. Nonetheless, we find interesting evidence that this is not necessarily a desirable outcome.

Keywords: trust, preferences, beliefs, learning, social networks.

1 Introduction

Most of the literature on trust deals with dyadic interactions, both in theoretical and experimental approaches (3, 5, 16). Instead, in this paper we focus on trust dynamics in a social environment.

[1]Email: maurizio.cortesi@uni.lu

This research is made possible under the financial support of the FNR fellowship project AFR PHD-08-017

Interesting network models of trust can be found in computer science. Generally these models deal with trust in terms of reputational systems, with a particular focus on the mechanisms for the evaluation of network nodes, whether they are artificial (computers) or human (6, 25, 28, 19).

On another side, economists are in general more interested in the perfomance effects of trust at various levels (10, 11, 24, 15, 26).

We believe there is a missing level of analysis in the literature, linking theoretical modeling at the dyadic level and empirical research at the social level.

More specifically, we are here interested in investigating how trust dynamics evolve in a social network. On one side trust allows for social interactions to take place in the very first place; on the other side, the outcomes of ongoing interactions allow for an update of trust itself, at least for the participating agents. This is a potentially self-reinforcing mechanism, where trust is needed to generate more trust. We focus therefore our attention on the dynamics of trust learning and on individual and social performance.

The rest of the paper is structured as follows. Section 2 gives an overview on trust related research. Section 3 describes the formalization of the model. Section 4 describes performance measures and discuss the results. Finally section 5 highlights some possible conclusions.

2 Trust literature

Trust is central to all transactions and yet economists rarely discuss the notion. It is treated rather as background environment, present whenever called upon, a sort of ever-ready lubricant that permits voluntary participation in production and exchange (10).

Over than this, some authors correctly claim that most of the research on this topic simply repeats the same results or implications of studies on transaction costs or on Prisoners dilemma-like situations (8).

In fact, game theory approaches have played a dominant role in trust-related research. The so-called Trust game (Figure 1), and closely related games, have received great attention, and even more so in the economic domain. In classical game theory, the game is solved by backward induction. As it is easily seen, the solution is that a buyer would never enter an exchange, since he would know in the first place that a seller would never reciprocate him.

However, plenty of experimental studies developed on the basis of the seminal work of Berg, Dickhaut and McCabe (5) show evidence that in many situations agents cooperate. These theoretically unexpected results can be linked to many causes, which have been and are investigated in a rich stream of research. Confusion about the game setting and players' incentives (18); altruism or inequality aversion (9); kindness (2); and trust-responsiveness (4) are among the possible explanations being investigated.

These results confirm that there is a complex structure of beliefs and goals that are at the basis of agents choices and actions, being these choices and actions socially embedded and strongly influencing agents also in the economic arena.

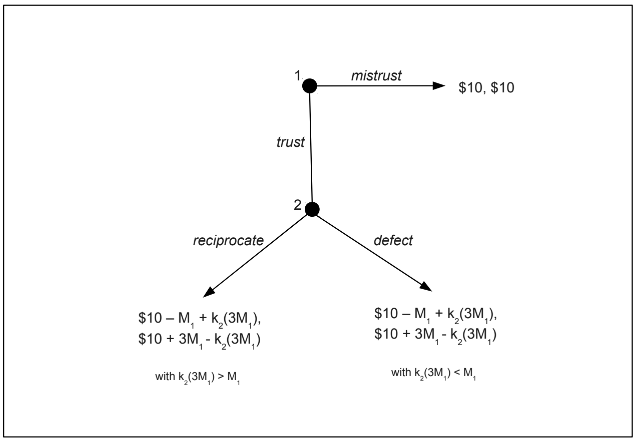


Figure 1: Trust game - Berg, Dickhaut and McCabe, 1995

The social construction of trust is more and more at the center of the research agenda for scholars in many different domains.

In particular, an interesting social aspect of trust is that related to its risk component. In fact, trust appears as a solution to specific problems of risk, that is the risk of being disappointed by others actions and suffering a damage bigger than the advantage one was seeking choosing to trust (23).

Some of these aspects are better understood when one looks to recent studies, linking trust and its mental components in terms of beliefs and preferences. Fehr (12) documents the accumulation of strong evidence that trusting cannot be captured by beliefs about other people's trustworthiness and risk preferences alone, but that social preferences play a key role.

In example, betrayal aversion, that suggests that people are more willing to take risk when facing a given probability of bad luck than to trust when facing an identical probability of being cheated, seems to play a particularly important role in trusting behavior (7).

Trust, for instance, is a possible solution to problems arising from uncertainty, risk, incomplete information, incomplete knowledge, imperfect rationality. As trust diffuses and is practiced, agents get better and better in understanding and predicting the environment in which they act.

As Good (14) suggests, the reaction of another agent to one's own actions is important in confirming prior experience, and this confirmation (predictability) is necessary to make the social world intelligible and seemingly knowable. Though some prerequisites are necessary for these dynamics to be effectively at work. Firstly agents need to interact in an environment rich of trust and trustworthiness, since trust can be depleted through not being used (13). This seems to reinforce the idea that trust may reinforce itself, and so it is for trustworthiness. Also, this seems to be supported by biological studies that claim the evolution of trust as a successful strategy in a population, and its subsequent diffusion and transmission to subsequent generations.

Secondly, agents need some capabilities and rules to discriminate among their potential counterparts and a sufficiently developed information processing ability and flexibility (22) as to make it possible to choose the best partners to place trust in.

Signaling and the ability to decode correctly and profitably the available signals, are therefore two important aspects of social interactions in general, and trust based ones in particular. Signals may come not just from direct perception and experience of the environment, but also from third-party experiences diffused along the network. The sharing of past experiences and informations about others, generates a public record of each of the interacting agents. This record may be freely available to other agents in the network, and is generally defined as agents' reputation.

All the discussed aspects have also a strong economic appeal. Some of these are, in example, central problems in the research areas of market interactions, products' quality and firms' reliability (1, 27). However, very interesting insights and results come from computer science, specifically from referral systems studies and trust management studies. The problem at hand is the possibility to assess the trustworthiness of every node in a network, by means of scoring from other nodes in the same network, as to create a reputation building mechanism.

Researchers discuss important questions related to score generation, discovery and aggregation (28). Some underline that a trust model needs to take into account many sources of information, as to be robust against possibly missing sources or lying from other agents. At the same time, every agent should be able to evaluate and pool all these informations on his own (19).

Trust seems therefore linked to social interactions in many different ways. Not only it is described as a solution to social risks, but also it appears as a very important cognitive device for context framing and predictive ability. Not only it allows social interactions to take place, but also these social interactions are the means for its evolution, through reputational mechanisms and information sharing among agents. Finally, due to these social aspects, not only trust plays a significant role in dyadic interactions, but also it appears as an important variable in groups aggregated performance.

The model we present in the next section develops on these building blocks, with some important specifications. Firstly, it is a model of trust in a social environment, extending over dyadic and triadic contexts. Secondly, it tries to coherently integrate insights from different disciplines, both for agents' characterization and for the definition of interaction rules. Lastly, a simple implementation has been chosen, to allow for future extensions.

3 Model

We consider a population of N agents, indexed with $i = 1, 2, \ldots, N$. For each agent we define a set of characteristics, over which we model interaction

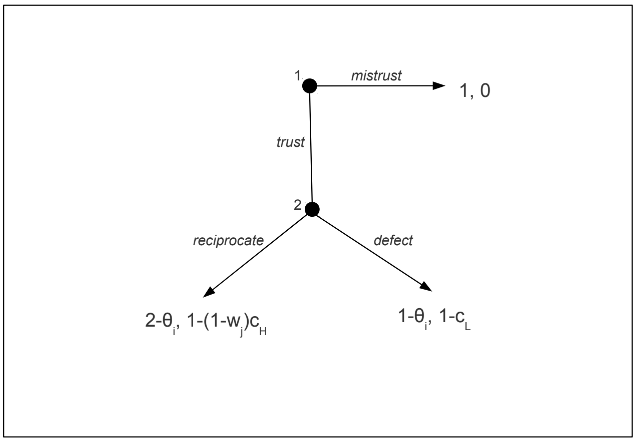


Figure 2: Buyer-seller exchange game

behaviors and learning dynamics. Time is discrete, indexed with $t = 0, 1, \ldots, T$. At each point in time, each agent is randomly assigned the role of buyer or seller, as to create two groups of the same size.

Every time a link between a buyer and a seller is created, the two agents find themselves in Prisoner's Dilemma-like situation. We model this game based on the Trust game (5).

Our interest is to evaluate the effects of trust dynamics on agents' learning and performance. We assume cooperation can arise between a buyer and a seller. We model the exchange on the insights emerged from recent neuro-economics studies on trust (12). This allows us to use some essential construct of economic theory, such as beliefs and preferences. At the same time, we consider utility maximizer agents, that is we do not require agents to be other than purely rational in the homo-oeconomicus sense.

The general game buyers and sellers confront to is shown in Figure 2. The game is different for each buyer and seller's duplet, since agents have heterogeneous characteristic influencing their payoffs. In particular, buyers face different reservation prices, while sellers face different cost structures.

In every period partnerships are formed based on agents' beliefs and preferences. Buyers hold beliefs about sellers' trustworthiness. These beliefs are the result of the informations that each buyer acquired in his past interactions with each seller, and of the information shared by buyers on their past interactions with a common seller. At each point in time t, $p_{ij}^t \in [0, 1]$ defines the belief agent i holds about agent j's trustworthiness. We give a description of the mechanism of formation and update of agents' beliefs in the next paragraph.

Preferences are defined as a threshold value $\theta_i \in [0, 0.75]$, and determine the minimum level of a buyer's belief over which he is willing to propose an interaction to potential sellers. This formalization of preferences resumes two insights from recent research. Firstly, thresholds represent an indicator of agents' pro-social attitude (9, 21). Secondly, they represent an indicator of agents' aversion of socially-embedded risk, such as betrayal aversion (7). For instance, this could be economically described in terms of agents' reservation price.

As we consider exogenous preferences, these are given and stable over time. However, it may well be the case that preferences evolve over time, in response to new information and new experience (14, 20).

Once buyers have selected their counterparts, a network g^t of all the links ij is formed.

Buyers receive a payoff equal to $2-\theta_i$ if reciprocated, that is if they receive the product/service they paid for. On the contrary, if the seller defects, they receive a payoff of $1-\theta_i$. The individual and heterogeneous parameter θ_i defines the maximum price an agent is willing to pay for a product or service (his reservation price).

Interactions fail with probability depending on the sellers' trustworthiness, w_i. This is defined in the interval [0, 1] and is stable over time. As for preferences, this is a strongly simplified assumption. In fact, trustworthiness may evolve due to environmental conditions and/or to each agent's partnering and outcomes history. Moreover, an agent's beliefs and preferences may have a relation with his trustworthiness, and placing trust may influence counterparts' trustworthiness in ongoing interactions. We will investigate these aspects in future developments of the model.

As it is easily seen from the game's payoffs, sellers who reciprocate receive the price of the product/service and incur in a high cost $(1-w_j)c_H$. However, this cost is variable in the level of a seller's trustworthiness. The higher the individual trustworthiness, the lower the cost the seller is facing. On the other hand, if a seller j doesn't reciprocate, he receives the product/service price and incurs in a cost equal to c_L, being the cost of selling a lower quality product/service generally lower.

Once agents have chosen their strategies and actions, the subset $o^t \subseteq g^t$ defines the indicator of positive outcomes for each buyer.

Based on agent i's belief, the condition for a link ij to be created at each period t is thus

$$p_{ij}^t(2-\theta_i) + (1-p_{ij}^t)(1-\theta_i) \geq 1$$

that is

$$p_{ij}^t \geq \theta_i$$

so that i's belief must be higher than his reservation price as to make entering an interaction a convenient and rational choice.

In general, an interaction is added to the network g^t if and only if

$$\pi_i^{g^t} - \pi_i^{g^t - ij} = E\left[p_{ij}^t \mid g^s,\, o^s;\, s < t\right] \geqslant E\left[\theta_i\right]$$

From the perspective of a seller instead, reciprocating is a rational choice if

$$1-(1-w_j)c_H \geq 1-c_L$$

that is

$$w_j \geq 1-\frac{c_L}{c_H}$$

It should be evident by now that agents face a similar game structure, but the actual payoffs are defined in a heterogeneous way, since each agent has idiosyncratic parameters defining the reservation price, for buyers, and the cost, for sellers. Also, the game allows for a cooperative solution to be a rationally viable solution.

3.1 Beliefs formation and update mechanism

As we have said, agents hold beliefs about other agents' trustworthiness. Of course, beliefs (as well as preferences) become relevant only for agents assigned the role of buyer. On the other side, trustworhiness is the only relevant parameter for agents assigned the role of seller.

In our model beliefs guide each buyer's decision to enter an interaction with a seller. After an interaction is made, the seller decides whether to reciprocate or defect and, in turn, this determines the outcome of the exchange. The information relative to each outcome is recorded by the buyer. Information about the outcome with a seller at time t can also be shared with other buyers who interacted in the same time period with the same seller.

We define therefore two mechanisms of formation and update for buyers' beliefs. One comes from direct experience, the other stems from communication along the network of buyers' experiences with common sellers.

Own information is embedded in a β-distributed belief $p_{ij}^t \sim \beta\left(u_{ij}^t,\, v_{ij}^t\right)$, representing the information owned by agent i with respect to his interactions with agents j, from time 0 to time t. The two parameters of the Beta function represent the number of successes and the number of unsuccesses between each couple of players. Based on this function, we can define the expected likelihood of success in a partnership between i and j, where i is the buyer and j the seller, as

$$p_{ij}^t = u_{ij}^t / \left(u_{ij}^t + v_{ij}^t\right)$$

From one period to the next, buyers' beliefs are updated according to Bayes' rule, using the information embedded in o^t. After period t has taken place, positive and negative outcomes are recorded and used to update individual beliefs about sellers. Therefore at time $t+1$

$$p_{ij}^{t+1} \sim \beta\left(u_{ij}^t + o_{ij}^t,\, v_{ij}^t + 1 - o_{ij}^t\right)$$

Given the update mechanism described, learning is only depending on i's interactions with seller j. Although, we want the model to include learning from other buyers' experiences with the same seller. This is where third party information pooling comes in.

To generate learning from others' experiences, we use insights from BMA - Bayesian model averaging (17). This is a method to evaluate the probability of a model being true, given the realizations, and assessing the probability of other models too being true, given the same realizations. In our use, models are the different beliefs that agents hold about someone's trustworthiness.

In the general BMA framework, if Δ is the quantity of interest, such an effect size, a future observable or the utility of a course of action, then its posteriors distribution given data D is an average of the posterior distributions of each considered model weighted by their posterior model probability. All the probabilities are conditional on the set of all models being considered.

In our approach, in particular, a "model" M_k is agent k's belief model about agent j's trustworthiness. When considering also third-party information, thus, the posterior belief at time t is:

$$E\left[p_{ij}^{t} \mid g^{s},\, o^{s};\, s<t\right] = E\left[p_{ij}^{t} \mid g_{ij}^{s},\, o_{ij}^{s};\, s<t\right] Pr\left\{M_i \mid g_{ij}^{s},\, o_{ij}^{s};\, s<t\right\} + \\ + \sum_{k\in g_i, g_j} E\left[p_{kj}^{t} \mid g_{kj}^{s},\, o_{kj}^{s};\, s<t\right] Pr\left\{M_k \mid g_{kj}^{s},\, o_{kj}^{s};\, s<t\right\}$$

where the first term is as defined above, while the second term follows the definition given in the Bayesian model averaging method.

There are two important assumptions regarding information pooling. Firstly, own information is weighted against third-party information, through a parameter $\alpha_i \in [0.2,\, 0.9]$, that is an individual heterogeneous character.

Secondly, third-party's information is assumed to be equally weighted over all the sources. In this paper we assume that agents make no strategic use of information and are reliably communicating their outcomes. Therefore all sources are to be considered equally trustworthy. As a consequence

$$w_k^t = \frac{1}{\#k \mid ij,\, kj \in g^t}$$

so that the weight is proportional on each of the buyers, other than i, who interacted with seller j in time t.

Based on these weights, the formula exposed above is modified assuming

$$Pr\left\{M_k \mid g_{kj}^{s},\, o_{kj}^{s};\, s<t\right\} = \begin{cases} \alpha_i & for\ k=i \\ (1-\alpha_i)w_k^t & for\ ij,\, kj \in g^t \\ 0 & for\ ij \in g^t,\ and\ kj \notin g^t \end{cases}$$

We also assume that, if no other buyer in the network has shared a specific seller with an agent i, then $\alpha_i = 1$, that is only direct experience counts for the described update mechanism.

4 Numerical experiments and results

We run a serie of numerical experiments to investigate the effects of learning and preferences on buyers' individual and aggregated performance at different levels of sellers' average trustworthiness. Specifically, we consider a population of 100 agents that interact over a time span of 60 periods. As previously said, at each point in time half of the agents are assigned the role of buyer, while the other half are assigned that of seller.

Learning weights $\alpha_i \in [0.2, 0.9]$ and social preferences $\theta_i \in [0, 0.75]$ are randomly initialized and uniformly distributed. We assume that every agent starts with idiosyncratic beliefs in each other agent $p_{ij}^t \in [0, 1]$. This makes agents heterogeneous with respect to beliefs too. This allows us to avoid the limit case of $u = 0$ and $v = 0$, which would lead to an impossible value of $0/0$ in the calculation of the beliefs.

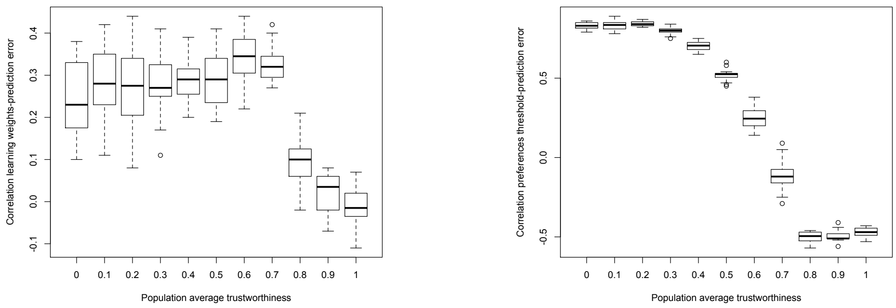
For each experiment agents' trustworthiness is initialized as $N(\mu, \sigma)$, with $\sigma = 0.05$ and $\mu \in [0, 1]$, starting from $\mu = 0$ and going to $\mu = 1$ with increments of 0.1 between experiments.

Coming to the results, on one side we investigate learning dynamics. The faster the learning process, the better the chances for buyers to interact only with reciprocating sellers. On the other side, since social preferences directly influence the number of ongoing interactions, we evaluate how this interacts with learning dynamics and the ability to choose more successful partners.

We use is the average trust prediction error as learning performance indicator, defined as the average of each buyer's beliefs distance to his counterparts' real trustworthiness values after each period. We evaluate the correlation with both individual learning weights and social preferences.

We found some clear patterns in the correlations. In particular, figure 3a shows that relying on indirect experience is in general important for a better prediction of counterparts' trustworthiness. For high values of sellers' average trustworthiness, this evidence disappears. Figure 3b shows instead that for low average sellers' trustworthiness, higher preferences imply bigger prediction errors. On the contrary for high values of sellers' trustworthiness the effect of preferences thresholds on the prediction error is of opposite sign.

The result that when sellers' trustworthiness is on average high, buyers who perform better are the ones who potentially refuse more interactions seems counterintuitive. However, it can well be the case that in this situation buyers tend to over-estimate sellers' trustworthiness, since there is a higher probability of outcomes to be positive. This has a positive feedback on buyers' beliefs, but can undermine a correct learning process about the real levels of sellers' reliability.

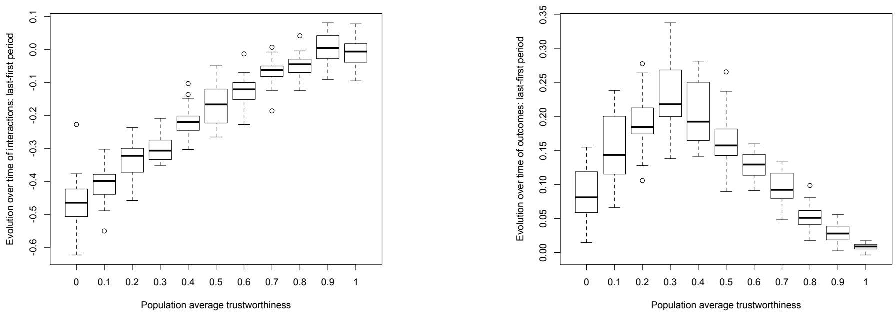


(a) Correlation with learning weights (b) Correlation with preferences

Figure 3: Prediction error

We also evaluate the evolution of interactions along the considered time span, as a percentage over the possible interactions. Moreover, we analyze the dynamics of positive outcomes, as a percentage over the ongoing interactions.

If we observe the interaction evolution in time (Figure 4a) we clearly see that the percentage of interactions is always decreasing, though decreasingly as sellers' trustworthiness increases on average.



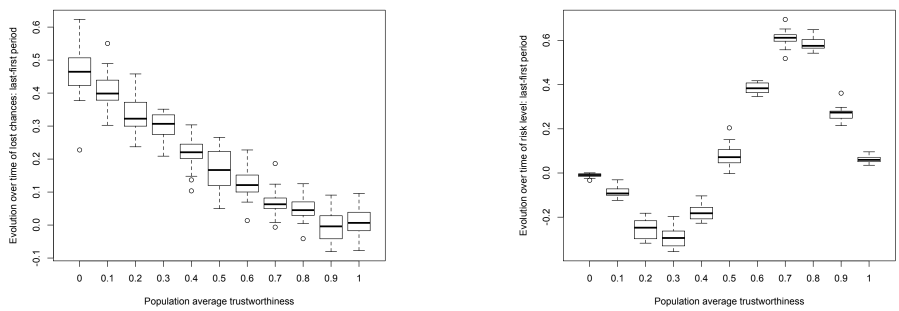
(a) Interactions (b) Outcomes

Figure 4: Evolution over time - Difference last-first period

However, at the same time the percentage of positive outcomes increases (Figure 4b), showing therefore that agents are able to choose and maintain over time generally more trustworthy counterparts, either this is due to preferences thresholds or learning. Nonetheless, such effect disappears when sellers are on average highly trustworthy and the percentage of positive outcomes seems to be stable over time.

Moreover, we define a measure of the social risk and a measure of the loss due to individual thresholds. The latter is defined as the percentage of interactions that are not exploited due to agents' thresholds, over the total of the available interactions. The former is the percentage of accepted and ongoing interactions for which the counterpart's trustworthiness is lower than an agent's belief, with respect to the total of ongoing interactions. In this first case, preferences are too high to allow even potentially fruitful interactions, while in the second case preferences are too low to protect a buyer from misbehaving sellers.

Figure 5a shows that the percentage of interactions loss due to individual preferences thresholds increases over time, but this effect tends to disappear for higher values of sellers' average trustworthiness.



(a) Lost chances

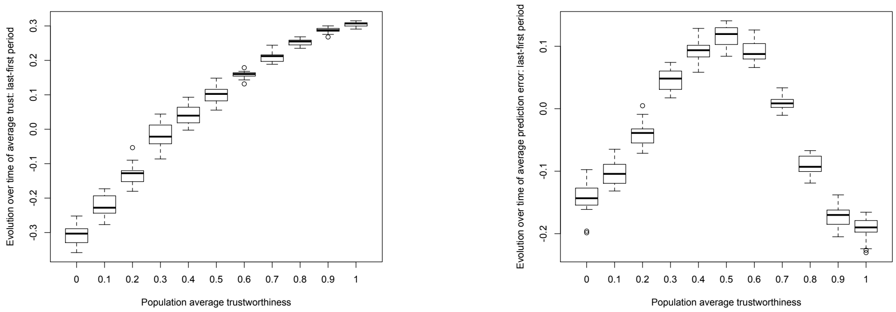
(b) Risk level

Figure 5: Evolution over time - Difference last-first period

On the other side, the percentage of interactions with untrustworthy counterparts over the ongoing interactions at first tends to diminish as sellers' average trustworthiness increases (Figure 5b). However, for values of average trustworthiness higher than 0.5 the relation is inverted and the percentage of risky interactions increases even sensibly, before diminishing again for really high values of average trustworthiness of the population (≥ 0.8).

This seems to sustain what we hypothesized before, that is that for high levels of trustworthiness buyers tend to overestimate their counterparts. The number of interactions is not sensibly changing at these values, and the percentage of positive outcomes remains high, as we have shown before. As beliefs increase over time due to the learning process, we assist to an increased chance of entering potentially riskier interactions with sellers whom levels of trustworthiness are being incorrectly assessed.

If we analyze the evolution over time of buyers' average trust and of the average prediction error made in the estimation of sellers' trustworthiness, we have a confirmation of the dynamics we just depicted. At first sight it might seem desirable that buyers' beliefs tend to diminish for low levels of trustworthiness in sellers' population, and to increase for higher values of sellers' average trustworthiness (Figure 6a).



(a) Average trust (b) Average prediction error

Figure 6: Evolution over time - Difference last-first period

In fact, figure 6b shows that the average prediction error is always decreasing over time, except for average values of sellers' trustworthiness around 0.5. However, we must always confront this with the role of preferences thresholds in the choice of sellers counterparts. Even if learning seems to be efficient and average trust increases as expected when average trustworthiness is high, nonetheless agents are incurring higher levels of risk, as shown in figure 5b.

For high values of sellers' trustworthiness, buyers enter too many interactions and tend to over-estimate their counterparts' trustworthiness, leading to an increased risk of failure. In our model this risk of failure is unrealistic, since the positive outcomes ratio with respect to ongoing interactions seems to be stable over time. However, the potential risk incurred by buyers is higher, due to the incorrect estimate of their counterparts' reliability.

Eventually, also in real markets, when trustworthiness is generally high and widely diffused, buyers are prone to accept higher levels of risk and tend to over-estimate the reliability of potential partners. Trust generates more trust, as it has been said, but this is not necessarily a positive result.

These observations open up interesting question over the fact that in environments like the one just depicted, unexpected falls in the levels of sellers' trustworthiness may lead to strong negative effects on buyers' beliefs. These effects may have a strong negative impact on the system dynamics. This question seems even more interesting if one thinks to the recent events that lead to the collapse of the financial markets and had, and still have, a strong impact on customers' trust levels. We think it'd be interesting to deepen these aspects in future research, and possibly in further developments of the presented model.

5 Conclusions

To understand trust one needs not only to assess its consequences. Research also investigates the link of trust with its antecedents, such as preferences and beliefs. It is exactly this that, in turn, can motivate the impact of trust on agents' decision processes and on their social and economic outcomes.

In this paper we built a social model of trust along these lines. Agents' trust has been defined in terms of beliefs and preferences, allowing for the implementation of recent experimental results.

Our numerical experiments gave some interesting insights on trust dynamics. In particular, we showed that preferences can strongly affect individual/aggregate performance. Higher social preferences thresholds imply a lower attitude towards interactions and this, in turn, can have a double-sided effect.

On one side, for low average level of sellers' trustworthiness, buyers with high preferences tend to interact less and this undermines the process of indirect learning, which has a strong positive impact in low trustworthiness environments. On the other side, when sellers' trustworthiness is high on average, it is exactly buyers with higher preferences who perform better in terms of learning. Interestingly, in these contexts there seems to be a tendency towards over-estimation of sellers' reliability, and buyers with low preferences thresholds tend to enter a high percentage of risky interactions.

In general, more trust, both in terms of higher buyers' beliefs and sellers' trustworthiness, tends to generate more trust. However, this may not always be a desirable outcome. Though in our model risk of failure in interactions is pretty low in highly trustworthy environments, nonetheless one would need to assess what would happen in case of unexpected falls in trustworthiness levels.

We believe this can be interesting evidence, also in the light of recent financial markets' collapse and the consequent fall in customers' trust. It may be possible that when trust beliefs are too high and misaligned with real counterparts' trustworthiness, an unexpected fall in trustworthiness may strongly and negatively impact trust dynamics and the system itself. These points could be worth deeper investigation.

In the end, we aknowledge some simplifications of the present model and the need for future extensions. In particular, we intend to investigate the effects of time-varying preferences and of possible cheating in communication among agents. Also an analysis of the co-evolution of trust and trustworthiness seems to be needed, due to the various ways these variables may interact and influence each other.

References

[1] G. A. Akerlof. The market for "lemons": quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, pages 488–500, 1970.

[2] J. Andreoni. Cooperation in Public-Goods experiments: Kindness or confusion? *American Economic Review*, 85:891–904, 1995.

[3] R. Axelrod and W. D. Hamilton. The evolution of cooperation. *Science*, 211(4489):1390–1396, 1981.

[4] M. Bacharach, G. Guerra, and D. J. Zizzo. *Is Trust Self-fulfilling?: An Experimental Study*. Economics Series Working Papers. University of Oxford, Department of Economics, 2001.

[5] J. Berg, J. Dickhaut, and K. McCabe. Trust, reciprocity, and social history. *Games and Economic Behavior*, 10(1):122–142, 1995.

[6] R. Bhattacharya, T. M. Devinney, and M. M. Pillutla. A formal model of trust based on outcomes. *Academy of Management Review*, 23:459–472, 1998.

[7] I. Bohnet, F. Greig, B. Herrmann, and R. Zeckhauser. Betrayal aversion: Evidence from brazil, china, oman, switzerland, turkey, and the united states. *American Economic Review*, 98(1):294–310, 2008.

[8] C. Castelfranchi and R. Falcone. Trust is much more than subjective probability: Mental components and sources of trust. In *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*, 2000.

[9] J. C. Cox. How to identify trust and reciprocity. *Games and Economic Behavior*, 46(2):260–281, 2004.

[10] P. Dasgupta. Trust as a commodity. In D. Gambetta, editor, *Trust: Making and Breaking Cooperative Relations, electronic edition*, pages 49–72. Department of Sociology, University of Oxford, 2000.

[11] K. T. Dirks and D. L. Ferrin. The role of trust in organizational settings. *Organization Science*, 12(4):450–467, 2001.

[12] E. Fehr. On the economics and biology of trust. *Journal of the European Economic Association*, 7(2-3):235–266, 2009.

[13] D. Gambetta. Can we trust trust. In D. Gambetta, editor, *Trust: Making and Breaking Cooperative Relations, electronic edition*, pages 213–237. Department of Sociology, University of Oxford, 2000.

[14] D. Good. Individuals, interpersonal relations, and trust. In D. Gambetta, editor, *Trust: Making and Breaking Cooperative Relations, electronic edition*, pages 31–48. Department of Sociology, University of Oxford, 2000.

[15] L. Guiso, P. Sapienza, and L. Zingales. Trusting the stock market. *The Journal of Finance*, 63(6):2557–2600, 2008.

[16] R. Hardin. Gaming trust. In E. Ostrom and J. Walker, editors, *Trust and Reciprocity: Interdisciplinary Lessons from Experimental Research*, pages 80–101. Russell Sage Foundation, 2003.

[17] J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky. Bayesian model averaging: a tutorial. *Statistical Science*, 14(4):382–417, 1999.

[18] D. Houser and R. Kurzban. Revisiting kindness and confusion in public goods experiments. *American Economic Review*, 92(4):1062–1069, 2002.

[19] T. D Huynh, N. R. Jennings, and N. R. Shadbolt. An integrated trust and reputation model for open multi-agent systems. *Autonomous Agents and Multi-Agent Systems*, 13(2):119–154, 2006.

[20] B. King-Casas, D. Tomlin, C. Anen, C. F. Camerer, S. R. Quartz, and P. R. Montague. Getting to know you: Reputation and trust in a Two-Person economic exchange. *Science*, 308(5718):78–83, 2005.

[21] M. Kosfeld, M. Heinrichs, P. J Zak, U. Fischbacher, and E. Fehr. Oxytocin increases trust in humans. *Nature*, 435(2):673–676, 2005.

[22] R. Kurzban. Biological foundations of reciprocity. In E. Ostrom and J. Walker, editors, *Trust and Reciprocity: Interdisciplinary Lessons from Experimental Research.* Russell Sage Foundation, 2003.

[23] N. Luhmann. Familiarity, confidence, trust: Problems and alternatives. In D. Gambetta, editor, *Trust: Making and Breaking Cooperative Relations, electronic edition*, pages 94–107. Department of Sociology, University of Oxford, 2000.

[24] B. McEvily, V. Perrone, and A. Zaheer. Trust as an organizing principle. *Organization Science*, pages 91–103, 2003.

[25] L. Mui, M. Mohtashemi, and A. Halberstadt. A computational model of trust and reputation. In *Proceedings of the 35th Annual Hawaii International Conference on System Sciences*, pages 2431–39, 2002.

[26] G. Tabellini. Culture and institutions: economic development in the regions of europe. *Journal of the European Economic Association*, 8(4):677–716, 2010.

[27] J. Tirole. A theory of collective reputations (with applications to the persistence of corruption and to firm quality). *Review of Economic Studies*, 63:1–22, 1996.

[28] B. Yu, M. P. Singh, and K. Sycara. Developing trust in large-scale peer-to-peer systems. In *IEEE First Symposium on Multi-Agent Security and Survivability*, pages 1–10, 2004.