Maurizio Cortesi

ESSAYS ON TRUST

Supervisor: Prof. Gianmaria Martini

PhD Economics and Management of Technology

XXII Cycle

1

"The imagination of nature is far, far greater than the imagination of man."

R. Feynman

Acknowledgments

I am grateful to the members of the Luxembourg Business Academy and of the Luxembourg School of Finance, University of Luxembourg, where I spent 20 months of my PhD period for their hospitality, useful comments and suggestions in developing my research.

Special thanks to Prof. Nicolas Jonard for every single chance he has given and still is giving to me. Above all for his support and the invaluable human presence.

I also wish to thank Robin Cowan, from UNU-Merit Maastricht, for help in the definition of some research questions, as well as for making me approach and discover R. Even more thankful for the chance to be around for some months at UNU-Merit, participating to very interesting classes and getting to know wonderful persons, over than inspiring teachers, in a town that I got to love so much.

Many thanks also go to Tristan Boyer, for the funny and nice conversations we had during our cigarette breaks. I quit smoking in the meanwhile, but the fruit of those discussions is hopefully not gone in ashes. Thanks to Gwaenel Piaser for attention to a newbie's questions on some game theory matters. Finally, Jean Luc Arregle and Tibor Neugebauer for their interesting comments.

And of course I recall here also my Italian professors and PhD committee, first of all my supervisor in the committee Prof. Gianmaria Martini and Prof. Riccardo Leoni.

I also would like to specially thank Prof. Luigi Buzzacchi for his attention and interesting comments.

Important moments of discussions always took place during our WIP meetings, both in Rome and in Turin. Great chances to get useful comments from both professors and PhD fellows.

More than this, you all, my mates, Ana, Anna, Cristina, Giordano and Michele, have made this experience great and so humanly important. The

first year with you all during classes was an intense moment of exchange and happiness. And so have been and are all the chances to spend some time with you in the last years.

I really believe a PhD is a profound human experience I have been gifted with, before being an adventure in research on whatever topic we may choose for a doctoral dissertation.

Finally, but since the first moment and above all, in my heart and in my life, my great family and wonderful friends. You are farrer in space now, but as always so close to me.

Contents

6

List of Figures

1 Introduction

Trust is central to all transactions and yet economists rarely discuss the notion. It is treated rather as background characteristic, present whenever called upon, a sort of ever-ready lubricant that permits voluntary participation in production and exchange (Dasgupta, 2000). Moreover trust permits increased efficiency, through the possibility of a fair degree of reliance on others people words (Arrow, 1974). Luhmann (2000) highlights that a system where trust lacks may even shrink below a critical threshold level necessary for its own reproduction at a certain level of development.

The concept was already widely studied in other domains, such as biology, sociology or psychology (Trivers, 1971, Axelrod and Hamilton, 1981, Smith, 1982, Coleman, 1988, 1990, Dasgupta, 2000, Bruni and Sugden, 2000). However, in the last decades a strong interest has emerged also in economics, a domain in which trust had always been called upon from a qualitative perspective till then (Akerlof, 1970, Mishra et al., 1996, Zaheer et al., 1998, Dasgupta, 2000, Dirks and Ferrin, 2001, McEvily et al., 2003), but never that much quantitatively approached it.

In recent years the relevance of trust has been often recognized and also deeply challenged, and scholars have adopted a more quantitative approach to its analysis. Many studies investigate the role of trust in shaping macroeconomic results, as well as the persistence of differences among countries with respect to macroeconomic performance indicators.

Nowadays, research shows, for instance, growing evidence of the importance of trust for economic outcomes, both at the individual and aggregate level (La Slemrod and Katuscak, 2002, Carter and Castillo, 2002, Hoff and Pandey, 2004, Tabellini, 2005, Guiso et al., 2006, Hoff and Pandey, 2006, Fehr, 2009). The contrasting evidence in empirical research requires however a clearer definition of the measures of trust and a better clarification of the causality link between trust, institutional setting and macro-economic performance (Fehr, 2009).

Trust emerges also as an important aspect of inter-organizational cooperation (Mishra et al., 1996, Zaheer et al., 1998, Nooteboom, 1999). The uncertainty of a strongly innovative and competitive market generate the incentives for research partnerships and long-term relationships, and this in turn requires and shapes trust dynamics among firms. All this appears to be true also for organizational performance (Zaheer et al., 1998, Davis et al., 2000, Dirks and Ferrin, 2001, McEvily et al., 2003, Dur and Sol, 2008).

More recent disciplines too, such as computer science, investigate and underline the importance of trust in the constitution of social networks in the world wide web or in the functioning of e-commerce platforms (Bhattacharya et al., 1998, Mui et al., 2001, Ba and Pavlou, 2002, Mui et al., 2002, Buchegger and Boudec, 2003, Yu et al., 2004, Guha et al., 2004, Huynh et al., 2006, Pavlou and Dimoka, 2006, Lu et al., 2009).

However, it has to be noted that not all about trust is good news. In fact, it may account also for increased rigidity and lock-in effects in ongoing interactions, therefore preventing individuals and organizations to enter new potentially fruitful interaction. Interactions require generally a horizon of stability to be based on trust, and to foster trust. At the same time, trust is more valuable in uncertain environments, where flexibility and rapid change are important assets (Nooteboom, 1999, Carter and Castillo, 2002). Moreover, trust may also be a strong glue for cartels, oligopolies, and even criminal organizations (Gambetta, 2000).

To complement this macro approach to trust, many other studies are conducted at the micro-level, and the focus is typically cooperation. However, as Good (2000) puts it, while cooperation and trust are intimately related in that the former is a central manifestation of the latter, the former cannot provide, for either the actor or the analyst, a simple redefinition of trust. Broadly speaking, the scope of trust extends in fact to all the situations characterized by information asymmetries, impossible or imperfect monitoring, uncertainty, and risk.

Recently research has moved into the investigation of the primitives of trust. An important point, at least for the understanding of the economic value of trust, is to study how the concept relates to economic theory concepts such as preferences and beliefs (Fehr, 2009). Results from various disciplines (Cox, 2004, King-Casas et al., 2005, Kosfeld et al., 2005) show evidence that trust is linked to beliefs about others trustworthiness, but that social preferences also play an important role in shaping trust decisions and relations.

This Doctoral dissertation is built around these insights and the main objective is to define a theoretical framework for the analysis of trust dynamics. It is constituted of three papers. The first one is a broad and deep overview of the research on the topic of trust. The objective is to demonstrate how pervasive the concept is in academic research, independent of disciplines. Moreover, main insights and gaps are also highlighted.

In the other two papers, two models are presented, one dealing with a dyadic interaction setting, and another dealing with a social setting. Both models investigate trust dynamics with a particular focus on its primitives. The aim is to demonstrate how beliefs and preferences play a fundamental role in shaping trust and its consequences. Attention is also given to the social and cultural dimensions of trust.

2 An overview on trust: different perspectives and a missing paradigm

Abstract

Over the last decades an increasing number of scholars have moved their attention to the relevance of cultural variables in shaping individual and social performance, also in the economic domain. But research on trust has been going on since much longer, first of all in disciplines such as sociology and psychology. More than this, biologists, life and evolution scholars, computer scientist have also produced interesting contributions on the topic of trust.

Due both to the broad nature of the concept, and to the wide nature of the contributing disciplines, trust still lacks a commonly held definition in academic research. On the contrary, many different perspectives have shed in time more light on both the antecedents and consequences of trust.

The aim of this overview is to bring together the different disciplines with their useful insights and to discuss the contributions and the gaps in research. This will also be the basis for the development of the models presented in the other two sections.

Introduction

In the last decades, a growing body of literature has started to deal with trust and its economic consequences. The concept was already widely studied in other domains, such as biology, sociology or psychology (Trivers, 1971, Axelrod and Hamilton, 1981, Smith, 1982, Coleman, 1988, 1990, Dasgupta, 2000, Bruni and Sugden, 2000). However, a strong interest emerged also in economics, a domain which had till then called upon trust mostly from a qualitative perspective (Akerlof, 1970, Mishra et al., 1996, Zaheer et al.,

1998, Dasgupta, 2000, Dirks and Ferrin, 2001, McEvily et al., 2003), but never that much quantitatively approached it.

Economics scholars have observed how trust is this ever-ready lubricant that permits voluntary participation in production and exchange (Dasgupta, 2000). Moreover trust permits increased efficiency, through the possibility of a fair degree of reliance on others people words (Arrow, 1974). Luhmann (2000) highlights that a system where trust lacks may even shrink below a critical threshold level necessary for its own reproduction at a certain level of development.

More recently there has been a surge of empirical research on trust. The development of experimental tools for trust measuring, as well as the availability of survey measures at national and international level, greatly facilitated research on trust. This enabled both the analysis of the impact of institutions on trust and cross national comparisons of trust effects on performance and other variables.

Aggregate measures of trust at the country level have been related to important economic variables such as GDP growth, inflation, or the volume of trade between countries; several papers suggest that trust may be an important determinant of these variables. La Porta et al. (1997) show that a larger share of trusting people is negatively correlated with inflation rates and positively correlated with GDP growth across countries. Others (Knack and Keefer, 1997, Knack, 2001) report positive correlations between a measure of trust and a country's average annual GDP growth rate.

More recently, Guiso, Sapienza and Zingales (2004) show that higher bilateral trust between two countries is associated with more trade between the countries. The same authors (2007) also provide micro-economic evidence on the role of trust in financial markets. They suggest that lack of individual trust in the stock market could partly explain the "participation puzzle", that is, why so few people take advantage of the existence of a stock market.

Nowadays, research shows, for instance, growing evidence of the impor-

tance of trust for economic outcomes, both at the individual and aggregate level (La Slemrod and Katuscak, 2002, Carter and Castillo, 2002, Hoff and Pandey, 2004, Tabellini, 2005, Guiso et al., 2006, Hoff and Pandey, 2006, Fehr, 2009).

More generally, economists link social environments, trust and economic outcomes along three main different perspectives. The first perspective refers to the social capital stream (Granovetter, 1985, Dasgupta, 2000, Berggren and Jordahl, 2006). Secondly, managerial literature focuses on the link between trust dynamics in groups and organizational performance (Mayer et al., 1995, Mishra et al., 1996, Zaheer et al., 1998, Dirks, 1999, Davis et al., 2000, Dirks and Ferrin, 2001, McEvily et al., 2003, Dur and Sol, 2008). Finally, an increasing number of studies focuses on the link between macro-economic performance and cultural aspects of the relevant social group, typically a country (Knack and Keefer, 1997, Porta et al., 1997, 1999, Raiser et al., 1999, Glaeser et al., 2000, Knack, 2001, Radaev, 2002, Guiso et al., 2004, Tabellini, 2005, Guiso et al., 2006, 2007, Bohnet et al., 2008, Fehr, 2009, Rainer and Siedler, 2009).

Moreover, many authors have recently focused on the understanding of the link between trust, beliefs and preferences (McCabe et al., 2003, Cox, 2004, Fehr et al., 2005, King-Casas et al., 2005, Fehr, 2009). As these are also economic primitives, it becomes even more interesting to analyse the effects of trust in economic interactions.

More recent disciplines too, such as computer science, investigate and underline the importance of trust in the constitution of social networks in the world wide web or in the functioning of e-commerce platforms (Bhattacharya et al., 1998, Mui et al., 2001, Ba and Pavlou, 2002, Mui et al., 2002, Buchegger and Boudec, 2003, Yu et al., 2004, Guha et al., 2004, Huynh et al., 2006, Pavlou and Dimoka, 2006, Lu et al., 2009).

Such a broad attention has led to different conceptualizations, and this has implied the absence of a common paradigm. Some commonalities emerge

among the definitions used in the different approaches, but still the concept is so broad, even in its street level perception and use, that it can be difficult to find a shared meaning across all these approaches.

The relevance of trust for everyday life is though broadly recognized. Despite this, many studies don't go beyond a broad conceptualization, and many times the concept remains in the background, while related concepts, such as trustworthiness, cooperation, reputation or others are the real object of the research. This paper aims at giving an overview of such broad research production.

2.1 Key concepts

The majority of studies on trust have been carried out in domains far away from economics. Some useful insights come from domains such as biology, computer science, and the relevant definitions are mainly brought from sociology or psychology.

Moreover, studies on trust are most of the times about cooperation. While cooperation and trust are intimately related in that the former is a central manifestation of the latter, the former cannot provide, for either the actor or the analyst, a simple redefinition of trust (Good, 2000). In effect trust is essential not only for cooperation, but even for competition (Gambetta, 2000).

To our view, however, trust is relevant for competition as far as it allows agents to agree on the definition and respect of the rules of the competitive arena, which brings us back to trust relation with cooperation. In other words, we believe trust allows agents to cooperate on the rules of competition, and fair competition is a consequence of this antecedent cooperation.

Generally trust is a more valuable asset in situations where information asymmetries and uncertainty are strongly present. Mainstream economics gives a primary importance to contracts or pre-commitment practices, as means to solve these problems.

But, as it has been frequently observed, a contract cannot always be complete, and the more it is so, the less flexible the agreement and the lower the possibility for the involved agents to face the complexity of uncertain environments. And signals, such as pre-commitment practices, need some predictability of the environment and some common information to be held as valuable by partners.

Therefore, given the incompleteness of these instruments, something else is needed to make both parties believe that a pre-commitment practice or a contract sufficiently reduce uncertainty and asymmetries (Dasgupta, 2000). Trust specifically allows parties to agree on the relative value of a signaling practice, as well as on the reliability of an incomplete contract.

This is so, we believe, for two main reasons. In fact, firstly trust is needed towards the institutions that generate and eventually enforce those instruments. And, secondly, agents must trust each other on the common reliance on these institutions and on the same set of rules.

For instance, as it may appear from the above observations, the concept of trust has been adopted both in micro and macro analysis. However, even if its relevance may go beyond the social interaction among two persons, it is in this micro environment that trust is generated and sustained, as to potentially diffuse in a wider community, generating consequences at the macro-level. We will come to macro aspects in the next sections. In the first step we are interested in focusing our attention on the micro aspects of trust.

At this micro-level trust has been defined as the "correct expectation about the actions of other people that have a bearing on one's own choice of action when that action must be chosen before one can monitor the actions of those others" (Dasgupta, 2000), so that the presence or absence of trust constrains the set of chosable actions (*ibidem*; Luhmann, 2000).

It has to be underlined that, for some scholars, whilst trust allows actions otherwise unpredictable from a rational theory of choice, trust itself is not action (Hardin, 2002). Trust opens the way to a set of actions wider than

the one assumed as available to a perfectly rational agent, but it remains a predisposition to action, while the choice itself may be based on trust or on other respects.

Others describe instead trust in a behavioral perspective. It is therefore the actual action of giving own resources to another agent, without any commitment from the counterpart, that depicts trust (Coleman, 1990, Fehr, 2009).

In effects, if trust has to be just an attitude, then all its consequences are mostly potential, while the concept is economically relevant only when it is expressed by choices and actions, leading therefore to concrete consequences that one may possibly be able to estimate and evaluate. It is also to observe that some of the problems related to trust are such only if we consider trust for its behavioral consequences.

A crucial matter in trust dynamics is, i.e., the impossibility to monitor others actions, where trustee's choices and actions, as well as the consequent outcomes, are imperfectly known or uncertain for the trustor. From a theoretical perspective, this implies the necessity to abandon the assumptions of complete information and perfect rationality of standard game theory.

We refer here to game theory, because of its relevance in the studies on trust, both as a widely adopted tool, and as an epistemological counterpart. Indeed, research on trust has often used games and theoretical insights from game theory to analyze situations involving trust. Moreover, this research has also demonstrated the need to depart from some assumption, seeking to explain experimental results mostly contrary to theoretical predictions.

As it has been stressed, trust is mostly valuable in situations characterized by uncertainty, risk, incomplete and/or asymmetric information. Over than this, plenty of possible limitations to perfect rationality are at work in the real world (Williams, 2000), such as imperfect understanding of limits and limitations, problems in the acquisition of knowledge, and recursive complexity of calculations. In particular, extensive processing work can only

be done under particular events and circumstances, while often individuals strongly follow habits and routinized behaviors. This in turn may explain the strong role of some cognitive biases, i.e. confirmation bias or cognitive inertia, that have been shown to preserve trust more over one should expect (Good, 2000).

It appears then with little surprise that standard game theory cannot predict as such the natural emergence of trust, and of cooperation. Some constraints or an infinite horizon of the game are necessary for reciprocating and cooperative strategies to emerge. But, as it has been previously noted, many experimental studies show, on the contrary, strong evidence of the presence of trust even in one-shot, anonymous interactions (Berg et al., 1995). Along this review this evidence will be presented and discussed with more detail.

Some other points need to be highlighted as a necessary introduction to the key aspects of trust. In particular, side by side with trust, one needs to take into account the concept of trustworthiness. Agents generally get an opinion about others trustworthiness, that is a belief inferred from background elements, culture, social environment, previous direct and indirect contacts.

In other words, agents get some idea about others reputation. Reputation is described as a public record about an agent trustworthiness and behaviour. However, this is of some value only if it is credible, built on past interactions (historical data) and behaviors or choices made under well understandable circumstances (Dasgupta, 2000).

For some scholars, both trust and trustworthiness come from incentives, norms or common shared interests (Hardin, 2003). This reflects the importance of interest (personal or common) in the choice of being a trustworthy partner, or to trust others. If cooperation is about at least one partner being in a dependent position on the other, then people must be motivated to enter this dependent position (Williams, 2000).

Trust is for instance required as a minimum level of assurance that the non-dependent party won't defect; and this assurance is well based if agents are also in general motivated not to defect if they are in a non-dependent position. Interestingly experimental studies show that those who are more willing to trust others, are also likely to be equally trustworthy in that they are less likely to lie, cheat, or steal (Good, 2000).

The concept of risk is also frequently used in the studies on trust. In fact, trust appears as a solution for specific problems of risk, that is the risk of being disappointed by the others actions and suffering a damage bigger than the advantage one was seeking choosing to trust (Luhmann, 2000). In recent neuro-economic studies this socially embedded risk is defined as betrayal aversion (Bohnet et al., 2008).

Interest and risk have an intimate relationship in trust dynamics. In effect, there appears to be a threshold of the interest/risk rate, over which trust is placed, and under which no trust is possible. This level is both objective, i.e. context dependent, and subjective (Gambetta, 2000). In other words, disposition to cooperate is cost sensitive, so there is a cost threshold, possibly changing over time, that defines when cooperation is feasible for an agent (Williams, 2000).

These considerations may probably reconcile the view of trust as an attitude or as a behaviour. If risk is seen as a component of action and decision, trust is the attitude that allows for risk-taking decisions. But, we may add, trust becomes evident only when these risky decisions are taken. Interestingly, we may observe and evaluate the consequences of trust only when this attitude becomes an actual decision and transforms therefore into action.

Luhmann (2000) observes that where trust lacks the set of possible decisions and actions is reduced, and a system where trust lacks may even shrink below a critical threshold level necessary for its own reproduction at a certain level of development.

Trust is also sometimes conceptualized as a public good. But, it is ob-

served, opposite to a public good it is not depleted through use, but through not being used (Gambetta, 2000). For some others, trust can be described as a commodity, such as knowledge or information (Dasgupta, 2000).

It is curious to note that not only humans, but even animals appear to be sensitive to others behavior, and this observation may suggest an explanation for the evolution of the mental state that we recognize as trust in ourselves (Bateson, 2000). Studies on evolution and in biology (Smith, 1982; Trivers, 1971) have shown that trust may appear as a reciprocating successful strategy. Once this strategy spreads sufficiently in the population and demonstrates itself as a successful one, biologists say that it is codified in genes and transferred to future generations, as a mean to undergo specific situations (Bowles, 2006).

The way trust emerges may be depicted as a game theory tit-for-tat mechanism (Axelrod and Hamilton, 1981). However some prerequisites in agents endowment are requested, as some capabilities in discriminating among agents and a sufficiently developed information processing ability and flexibility (Kurzban, 2003).

Predictability is a strong and possible solution to problems arising from uncertainty, risk, incomplete information, incomplete knowledge, imperfect rationality. And, this is also confirmed at human level, as in Good (2000), where he highlights that the reaction of another agent to one's own action is important in confirming prior experience and evidence and, as observed in preceding literature, this confirmation (predictability) is necessary to make the social world intelligible and seemingly knowable.

To conclude, we may stress that trust is not always welcomed and not always leads to better social results or higher social welfare. Cartels and oligopolies are in example also sustained and sustainable on the basis of trust among the participants. On the opposite, trust may fail to emerge in situations where it would lead to better social outcomes, due not to a lack of motivation, i.e. interests and incentives, but to the lack of belief (Gambetta,

2000).

As we have shortly depicted, trust has led to many different conceptualizations and approaches. Even though we may reconstruct a very broad common meaning of the concept, it still lacks a stringent definition. As highlighted from the interchangeability of words in previous paragraphs, trust is most of the time used side by side with concepts as trustworthiness, cooperation, reputation, and others.

For these reasons, the next section tries to give a broad picture of the definitions of trust. Section three present some of the theoretical and empirical evidence on trust. The section is built around the links among trust and the concepts of cooperation, reputation, performance, networks and preferences. In section four, the available measures of trust are described, as used in some empirical studies. The last section concludes the review with some critical remarks, underlining the gaps in the existing studies, and suggesting possible future lines of research.

2.2 Definitions of trust

In the previous section we underlined some key aspect of trust, as emerged from previous studies. Hereafter we move a step further into the understanding of the concept, analyzing how it has been defined in different disciplines and approaches. Clearly, there is not better place to start than the definition a dictionary gives.

It is interesting however to note that in different languages and/or in different countries and cultures, the word "trust" may have different implications and shadings. In this overview, though, we remain close to the conceptualization of trust as it has emerged in the academic research, that is stick to the anglosaxon idiom.

The Oxford dictionary reports three definitions of trust:

> "1 firm belief in the reliability, truth, ability, or strength of someone or something. 2 acceptance of the truth of a statement

without evidence or investigation. 3 the state of being responsible
for someone or something."

The concept is related to an appreciation of someone else's abilities or to
the acceptance of an information without further proof requirements. In
Italian, i.e., these concepts are expressed with two different terms, the first
"fiducia", and the second "fede", that is more linked to a strong belief in some
religious credo, at the same time not requiring and not allowing for further
investigation, a sort of ungrounded trust (Hardin, 2002).

In academic literature on trust, the concept got a broader definition. It
is linked to other concepts such as cooperation, reciprocity, trustworthiness,
reputation, vulnerability, betrayal, and it is sometimes confused and blurred,
in a way that makes it difficult to identify if trust is at study, or those other
related concepts. Moreover, this confusion doesn't always allow to assess
whether the concept of trust has a peculiar and specific relevance to the
domain under investigation.

It is also interesting to note that definitions of trust used in the literature
have frequently originated in domains far away from the ones where they are
adopted. In particular, sociology and psychology had an important impact
in the definition and study of trust, and therefore their influence goes well
beyond the domains boundaries.

It comes with no surprise then, that an important starting point in many
studies is Coleman and his Foundations of social theory (1990), where he
states:

> "Placement of trust allows an action on the part of the trustee
> that would have not been possible otherwise; if the trustee is
> trustworthy, the person who places trust is better off than if trust
> were not placed [and the opposite]; the action of placing trust
> involves the trustor's voluntary placing resources at the disposal
> of another party, without any real commitment from that party;
> a time lag is involved."

Translating this definition into the economic domain, some observations seem relevant. First, trust allows for actions otherwise impossible, i.e. in a game theoretical approach trust may lead to solve games such as the prisoner's dilemma or cooperative games, reaching the socially optimal equilibrium of cooperation.

Moreover, this equilibrium is reached even in one-shot games, whilst game theory recurs to repeated interactions and infinite game horizons to solve the suboptimal outcomes problem. On this point, one should question the stability of the solutions reached via trust, and this question has probably no clear answer.

Secondly, trust in this definition involves a commitment of resources, without any prior commitment from the other party, which means that resources of one agent, the trustor, become more or less freely available to whatever use the other agent, the trustee, may decide to make of them.

The important distinction with a mainstream economic approach is that usually pre-commitment practices and agreements are essential part of contract theory, whilst here trust allows optimal solutions even without such instruments, but simply via a social interactive coordination among agents. On this point one may also note that resources may be not only material, but possibly also psychological ones.

Finally Coleman is defining the "placement of trust", and the "action of placing trust", which means that the conceptualization is about trust in action and in its deployment. Trust in itself appears therefore of little value, unless it gives place to an action (Hardin, 2002). In effect, an interaction is the only context in which one may understand trust dynamics and its implications for the interaction itself and the subsequent outcomes.

Nonetheless, some other authors move from this behavioral definition. Dasgupta (2000) sees trust as the "correct expectations about the actions of other people that have a bearing on one's own choice of action when that action must be chosen before one can monitor the actions of those others".

Trust appears as a cognitive device, purely calculative, used when interests are at stake in an uncertain interaction among different agents. There is no reference to the action of trusting, but to trust as an expectation. There is no reference to the outcomes of the interaction.

This account is perhaps more general, but may be confused with a mere calculative process, which is nothing more that the concept of bounded rationality. Whilst uncertainty is surely a matter in every situation where trust may play a role, this definition seems to miss some peculiarities of the concept, if we are to use it as a specific factor in an account of economic theory.

A similar definition comes from Gambetta (2000), who depicts trust "(or symmetrically, distrust), [as] a particular level of the subjective probability with which an agent assesses that another agent or group of agents will perform a particular action, both before he can monitor and in a context in which it affects his own action.". Here too there in no clear cut evidence about the way the interaction may affect the trustor's outcomes.

This is anyhow an important operational definition, as it allows for an easy mathematical approach to trust modeling in terms of probabilities; on the other side, many authors have highlighted that trust is not only about a correct calculation of expectations about others actions, because it may well be the case that trust is also motivated by some personal well-being gain in trusting. In this sense, trust becomes part of the preferences structure of the agent (Castelfranchi and Falcone, 2000).

Differently from Coleman's definition (1990), and in line with the two previously cited (Dasgupta, 2000, Gambetta, 2000), Mayer, Davis and Schoorman (1995) highlight the importance not of a time lag, but of the impossibility of monitoring. They define trust as "the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other party will perform a particular action important to the trustor, irrespective of the ability to monitor or control the party.". An important and

new aspect of this definition is the concept of vulnerability. While one may agree that trusting involves being in a vulnerable position, at the same time this concept is to be taken in a broad sense, where the vulnerability is not a necessary prerequisite of trusting.

An interesting definition is that of "encapsulated interest" (Hardin, 2002), characterized as the situation where "I trust you because I think it is in your interest to attend to my interests in the relevant matter." Even though this conceptualization may fail a direct applicability to research, it highlights the relevance of interest and moves the focus from personal intentions to trust, to expectations about others interest in reciprocation.

We may observe that the definition in itself lacks some further specifications that Hardin makes clear in his book, but we may infer that trust is context-dependent and a three-part relation. The context-dependence is not only related to the interaction in itself or the abilities/competencies required for the task, but more as the broad context where the interaction goes on and the potential interrelations with other actions or choices going on at the time of the interaction. The three-part relation claim is about the relevance of the specificity of trust to a particular individual(s) and to particular matters.

Lastly it is interesting to report an operationalised definition that Bhattacharya, Devinney and Pillutla (1998) use in their outcomes-based model of trust: "trust is an expectancy of positive (or non-negative) outcomes that one can receive based on the expected action of another party in an interaction characterized by uncertainty".

While one cannot expect a definition to clarify all the mechanisms relative to the concept it defines, still it is to underline that this definition isn't clear with respect to the character of uncertainty, whether it is linked to the environment, the interaction, or the party's choice of action. Moreover, as in Gambetta (2000), trust appears as reduced to a mere process of probability calculation, in which sense it may not be different from bounded rationality.

We may therefore resume the principal elements that have been high-

lighted as relevant and characterizing trust. All of these aspects have a bearing in trust dynamics, and may be depicted as components of trust. A belief component, a risk component and a social component are hereafter resumed.

Firstly there is the belief component, since trust is an expectation and this expectation is both subjective and contextualized to a precise interaction with another agent in a specific context or to particular matters. It has to be stressed that, while the street-level conceptualization of trust may lack the clear and precise indication of this context, so that one can say "I trust you" without putting the matter or the context of trust explicitly, it is necessary, to the operationalization of the concept, to underline its clear relevance and dependence on the context and the specific agents involved.

Secondly there is a risk component, that is trust has to do with uncertainty or risk. We may stress here that the trustor is in an open position, which means that trusting someone doesn't implicate an action, but more a predisposition, a higher or lower preference to the interaction with another agent in respect to some objectives, possibly depending on the joint involvement of trustor and trustee. A trustor, indeed, may always choose not to enter the relation with the trustee, so the risk is in a sense controlled. However, as it has been noted before, trust dynamics cannot be understood and are not relevant if trust is a mere expectation, and no action is involved.

On the other side, that of the trustee, it is to highlight that, apart from the necessary abilities to perform the required expected action, a trustee may also be able to influence the outcomes of the interaction with his choices and actions. If not, trusting someone who has no control over his actions and the subsequent outcomes, is like playing a lottery against nature, that means, one cannot use the concept of trust where the uncertainty is about nature itself and alone.

Thirdly there is a social component. In fact, while trust in itself may well be an individual predisposition, acting on trust and evaluating its importance

and relevance requires an interaction among at least two agents. Acting on trust requires then mutuality, and it's embedded in a social structure of relations, institutions and norms.

It has to be noted that there is an overlap between the social component and the risk component, as trust involves a particular kind of risk dependent on social interactions. Researchers name this risk betrayal aversion, that suggests that people are more willing to take a risk when facing a given probability of bad luck than to trust when facing an identical probability of being cheated.

Finally, another component may be added, relative to the outcomes of a trust relationship. Trust generally accounts for some positive outcome for the involved parties. While theory has operationalised this concept giving, as in the prisoners dilemma, a positive outcome to both of the agents, it is not always the case, to our view, that the trustee has some gain from reciprocating trust in the ongoing interaction. It may well be the case that his gain is delayed in time, possibly in future interactions among the same agents.

Over than this, we believe that since trust is about reducing risk, agents may choose to enter such a relationship to lower negative outcomes. This in turns implies that outcomes may not always be positive, but, more generally, as in Coleman (1990), trustor needs to be better off trusting than not doing it.

2.3 Studies on trust

Research on trust is characterized by many different approaches. This is due not only to the differences among the disciplines interested in trust, but also to those within every discipline, as many perspectives have been adopted over time. This is also evident in the choice of the key concepts used to analyze trust, such as cooperation, reputation, trustworthiness, vulnerability, beliefs and preferences, and so on.

Hereafter, some of these concepts are used to aggregate and categorize previous research. This in turn allows to underline the differences among the various research streams, and to highlight the potentialities and gaps in the different approaches.

The aggregating concepts used hereafter are: cooperation, reputation, performance, networks and preferences. The performance category is itself divided into three other sub-categories: organizational, inter-organizational and macro-economic.

We believe this categorization may be useful to understand the main streams of research on trust. At the same time, however, we recognize that for some studies the categories are not unique. Therefore some overlapping is surely possible, without anyway loosing any clarity in the exposition of the main results.

2.3.1 Trust and cooperation

The stream of literature relating trust to reciprocity and cooperation is strongly influenced by the settings of game theory, most importantly by the Prisoner's dilemma model. The main contributions go back to Axelrod and Hamilton (1981) and to Berg, Dickhaut and McCabe (1995). Much of the research developed to test this conceptualization of trust relate to the latters, and to the setting they used to test their hypothesis, the so-called Trust game. Other studies use different game settings to test different hypothesis, while remaining strongly linked to this stream. This approach has developed in the last decades and the ongoing research tries to replicate some of the results and the implications of this approach. For a broad overview on reciprocity see also Sethi and Somanathan (2003).

The main limit of this perspective resides in the game theoretical setting it takes as instrument to test the relevant hypothesis on trust. Most of the times these studies consider just two agents, and remain to a really simple description of trust, mainly depicting situations where the theoretical

27

outcomes of game theory are violated, more than assuring that the outcomes are determined by trust alone and itself.

Some authors try to give a picture of a more historical and social embeddedness of trust, but the theoretical instruments limits the study to the relation among two person, not focusing on the social dynamics of trust and its evolution, but primarily on its appearance.

Moreover, as far as most of the studies involve monetary rewards and gains, one cannot fully stress the concept of trust. As it has been highlighted above, the decision to trust may be related not only to monetary rewards, but also to social and psychological gains. The monetary nature of the game may therefore affect the perception of the game and participants choices.

On the other side, it is important to underline that these studies had a strong role in bringing trust into economic research agenda and discussion, as a variable that plays its own role in the exchanges among agents.

From the modeling perspective, there are other interesting aspects in this approach. Importantly, in most of the cases anonymity is one of the game hypothesis. This, together with the one-shot game setting, brings light on the evidence that some different process than a rational calculation is going on in the interaction among agents.

None of the classical economic theory instruments, such as contracts, pre-commitment practices, or repetition of the interaction, are necessarily in place in the experiments. This leads to the consideration that agents behaviour is not necessarily rational in the homo-oeconomicus sense, but mainly agents use cognitive and decision making devices that may be strongly related to norms, habits, routines or preferences.

Whether these "rules" are to be called trust alone, or may possibly be described by or together with others concepts, this stream of research gives no clear cut on the point. However some of the studies pose this problem and try to analyze and disentangle the different explanations at hand. We will turn to these questions after giving an account of the reference model.

In their fundamental paper, Berg, Dickhaut and McCabe (1995) present the results of an experiment to study trust and reciprocity in an investment setting. The investment game is played as follows (see Figure 1). A number of participants is divided in two rooms, A and B. All of the participants receive $10 as show-up fee, but while participants in room B pocket this fee, participants in room A have to decide how much of their $10 to send to an anonymous counterpart in the other room. Subjects are informed that this amount will be tripled by the experimenter by the time it reached room B. Subjects in room B have then to decide how much of the received money to send back to room A counterparts.

The perfect and unique Nash equilibrium for the game, with perfect information, is to send zero money. Being only motivated by self-interest, a rational agent in room B will send back zero to his counterpart. Therefore a rational agent in room A should anticipate this decision and keep all of his show-up fee. The outcomes of the experiment show on the contrary a different picture.

Most of the agents in room A sent money to their anonymous counterparts, indicating, to the authors, that they trust their counterparts to reciprocate them in the second stage of the game. In fact, some agents in room B reciprocated sending back some of the tripled amount they received. This indicates, in turn, that they probably interpreted receiving money as a decision to trust from their counterparts. However only one third of room B subjects decided to reciprocate.

Figure 1: The trust game (Berg, Dickhaut and McCabe, 1995)

The results of this experiment showed a propensity to trust for most of the agents in room A. However, as the authors state, this behavior can be grounded in the so-called socialization of norms (Coleman, 1990), since all the participant were chosen from a specific population. Therefore, it can be difficult to assess whether agents decisions were taken depending on the actual game structure or mostly influenced by externally acquired norms.

To account for some of these questions, Berg, Dickhaut and McCabe (*ibidem*) also explore similar outcomes in a so-called social history treatment, where, before the game takes place, all subjects are given informations about previous results of the experiment. In this second setting, in face of the

evidence of only one third of the trusted agents reciprocating in previous experiments, most of room A agents still enter the game sending money to their counterparts. On the other side, the response of room B agents is more in line with a reciprocity hypothesis, since the amount sent back is higher on average and the number of reciprocators is higher.

To our view, some points can be questioned on these experiments. First of all, it is not clear what the authors mean by the fact that room B subjects may interpret receiving money as initiating trust. One cannot say if the room A agents are really placing trust unless intentions or beliefs are fully clear and analyzable. It may well be the case, at the extreme, that agents in room A are just playing some of their money since the game setting is not clear; they don't understand or misunderstand the rules; they are not really playing with their own money; the amount at stake is low.

We believe that some doubts on these points are clearer if we take a look at the second experiment, where room A agents still play the game, even without any evidence of positive payoff from the investment. The social history treatment gives evidence, although not precise, that the invested amount is higher on average, as well as the payoff, and this may be in line with the fact that in the previous experiment those who received better paybacks where those who invested more. This doesn't necessarily mean that room A players are placing more trust, but possibly it can also mean that they are trying to increase the probability of a positive payback, by stimulating a better response from their counterparts. As we would put it, it is somehow as if they were playing a lottery, more than choosing to trust. In fact the authors themselves claim that room B agents were not probably always interpreting as trust placement the fact of receiving money.

Moreover, we may add, the setting can no longer be strictly anonymous if all agents come from a specific population. As externally acquired norms are typically shared across individuals in the same population, agents are probably not considering to play against really anonymous counterparts, but

31

they average their beliefs about their counterparts based on the shared norms and experiences acquired in the day-by-day life in the social group (Tanis and Postmes, 2005, Tirole, 1996). Agents decisions may be therefore influenced by routinized behavior and some cognitive biases may be at work in this setting (Good, 2000).

Finally, one may question whether the history treatment is really historical. The subject are given the results of a previous experiment where they were not involved, which implies that the history is not grounded in the game itself and in the ongoing interactions among agents, but the information is taken from an external source. While one may surely define his choices on the basis of external evidence from others, the definition of history given to this treatment doesn't seem clear and appropriate.

Moreover, we may stress that if results in the first game setting, the no-history treatment, may have been influenced by socialization of norms and group-belonging, this effect cannot probably be override only giving to subjects an information about results in a previous game. As we said, since norms are culturally defined and socially shared, it may require time to change them, and therefore also for an individual to choose on completely different grounds than those norms themselves.

To address part of the underlined problems, Ortmann et al. (2000) try to replicate the results from previous research while modifying those parts of the experiment or of the experiment presentation that may have caused confusion and/or lack of understanding, lack of salience, and the framing of the situation. In other words, they try to investigate if the agents are clearly reasoning about the game setting and they try to stimulate a strategic thinking. Their results replicate pretty clearly the outcomes already highlighted by Berg and his colleagues. This is also true for the fact that expectations of the participants are most of the times in contradiction to the available evidence. However, this point is not investigated and remains without explanation.

Andreoni (1995) tries to disentangle effects of kindness, altruism or warm-

glow in cooperation, from the effect of errors and confusion. The paper stresses two interesting points. The first one is that confusion plays an important role in determining the observed cooperative outcomes, at least as long as learning has not reached a critical level. Secondly, kindness is even more important, since it allows cooperative outcomes to remain stable over time, even in face of contrary evidence, as far as this preference for cooperation has not been frustrated to a critical point.

This stream of research gives evidence that more than trust may actually be at work when we observe cooperation. Confusion has an important role in experiments, as long as the free-riding problem is maybe not so well understood from participants on a theoretical ground, but repetition needs to take place so that learning may lead the participants to a direct understanding of the problem. Evidence of this confusion problem is also found in another similar experiment by Houser and Kurzban (2002).

Other scholars try to discriminate trust from altruism or inequality aversion (Cox, 2004). They stress that the broad development of literature around experimental designs, such as the investment game, fails to address the possible noise in the data resulting from other-regarding preferences different from trust. These preferences involve the idea of fairness of the outcomes, which leads to an unconditional kindness, in contrast with conditional kindness, that is in turn a response specifically motivated by a generous action by another.

On one side, therefore, is one agent trusting his counterpart when he chooses an action that benefits the latter, or is he motivated by other-regarding preferences, such as altruism? On the other side, when an agent receives some beneficial action from his counterpart, is he responding positively because of a reciprocation choice, or because of his other-regarding preferences, such as inequality aversion?

Without entering in the details of the experiments, this paper is important since it underlines the relevance of the question whether trust is the only

possible explanation for the outcomes highlighted in this stream of literature. And again, as in the previously reported papers, there seems to be evidence that different possible variables may be at work to raise the level of cooperation and to give place to results inconsistent with a pure self-regarding preferences hypothesis, generally applied in the classical game theoretic and economic approach.

On the other hand, Ashraf, Bohnet and Piankov (2006) base trust both on beliefs of trustworthiness and on unconditional kindness, as well as they base trustworthiness both on reciprocity and unconditional kindness. Opposite to the previous study they don't disentangle these preferences as different explanations of the cooperative behavior, but they stress the idea that unconditional kindness is itself a sub-component of trust, as if the agents derive personal satisfaction from trusting and being trustworthy.

Glaeser et al. (2000) conducted a study to investigate possible measures of trust and trustworthiness in the stream of social capital literature. The objective is to evaluate whether survey measures of trust correlate with behavioral measures obtained through experiments. A survey is submitted to the participants, and two experiments are conducted, as to evaluate trust behaviour.

Trust choices appear significantly linked to past trusting behaviour, more than to survey questions about trust. Moreover trust is correlated with attitudinal survey questions about trusting strangers, implying a stable individual component that goes well beyond group belonging. These results are also relevant in terms of the historical and social building of trust.

In addition, as the authors link their results to social capital, it appears clear that social connections are an important variable in the prediction of trustworthiness and reciprocation. Evidence shows that race and nationality matter in terms of the levels of reciprocity, as well as background characteristics capturing the status or organizational memberships of the trustors.

Similar results hold in other studies showing that social identities, such

as caste in India, once revealed may affect trust behaviour and the economic performance of individuals, even when such social identities are no longer binding and discriminating from a legal perspective (Hoff and Pandey, 2004, 2006).

The important point in these experiments is that there is evidence of the social construction of trusting and reciprocating behaviors. Although the experimental settings may share the confusion or other-regarding preferences problem with the study conducted by Berg et al. (1995), it highlights the importance of a repeated social interaction for the emergence of cooperative outcomes. Glaeser and his coauthors (2000) underline that trust is best predicted by previous trust, which indicates, in turn, that an important component of trust has to be studied outside the single interaction observed in the experimental setting, if one wants to address the problem of how trust emerges and is established.

Other motives may be at work when cooperation among individuals arise. Trust responsiveness (or the self-fulfilling prophecy of trust) (Bacharach et al., 2001) describes the situation where one fulfills trust because he believes trust has been places in him. From this perspective, where the trustee has the perception that trust has been placed, he fulfills it for a sense of aversion to letting down the trustor, this aversion possibly coming from sentiments of sympathy or respect. An important consideration is that there is more place for a cooperative outcome based on trust, the more there are clear signals that trust has been placed.

Somehow, this recalls the idea that reciprocity comes from the perception that trust has been places by counterparts sending money in the Trust game setting. For instance, this is the same interpretation made by Berg, Dickhaut and McCabe (1995).

In some situations it is even the case that cooperation is fostered not by means of individual trustworthiness or direct knowledge among the interacting agents, but by means of group membership. This point was already

made in the no-history treatment in Berg, Dickhaut and McCabe experiments (*ibidem*), as we have noted before. When individual agents are not identifiable, then trustworthiness may be inferred from group membership (Tanis and Postmes, 2005, Tirole, 1996). However, the inference through some social identity indicator, as associational or group membership, may have paradoxical or trust lowering effects (Hoff and Pandey, 2006).

Finally, it is interesting to stress one more point about cooperation based on trust and trustworthiness. Following Chaudhuri et al. (2002), trusting can also be described as a perfectly rational choice, once we move our focus to the relative weight that the trustor puts on a sure gain today or a possible gain tomorrow. In fact, if the weight on and the subjective probability assigned to future probable outcomes is high enough, the agent is better off if he places his trust, participating to an interaction. Where, on the other hand, a sure gain today is preferred, then the agent chooses rationally not to participate.

On the opposite, trustworthiness is justifiable in an economic sense only where the interaction is going to be repeated in time, with an unknown horizon. Otherwise, reciprocating is never increasing the economic welfare of the trustee, unless, as reported before stating results from other research, we consider in this welfare also psychological or social motives.

The problem is, to our view, that these insights cannot clearly identify trust from any other rational choice mechanisms. Therefore, in this view trust looses its potential economic role and meaning, making it irrelevant to study it as a specific aspect of agents decision processes. Nonetheless, we believe this is not the case, since this account misses one important aspect of trust, that is its social component, both residing in shared norms and in individual social preferences.

These conclusions allow us to enter the discussion on the relevance of reputation, seen as a mechanism to signal trustworthiness and to enhance the chances of being trusted, simultaneously increasing the chances of trust to be deployed and to develop. The next paragraph deals with these aspects.

2.3.2 Trust and reputation

Reputation is a matter of great relevance in the context of trust studies and it encompasses the studies on both trust and trustworthiness. Most of the quantitative and theoretical research on this topic is linked to computer science and the study of networks as means to assess the trustworthiness of the nodes and to transfer reputation scores along the network itself.

The problem has a strong economic appeal, and it is mostly analyzed in the research areas of markets, products quality and firm reliability. Economics scholars have tried to model mathematically the incentives for reputation in many different models. Moreover, the problem to which reputation may be a possible solution, is a well-known and well-defined economic subject, that is the market for lemons. Failures similar to the one depicted by Akerlof (1970) may obstacle the development of a market unless there is an efficient structure of incentives or well developed contract and pre-commitment practices. On the other side, whereas these instruments may not be available, trust and trustworthiness are pre-conditions to the development of the market itself, and reputation building on the seller's side is considered strongly important.

Some studies deal with the problem from a theoretical perspective. Usually market modeling or exchange modeling is the focus of the study, and matters of trustworthiness, information diffusion among agents and reputation building practices are the means to solve market failures. This approach is related to problems such as the persistence of low quality products; the reliability of firms on a competitive market to sell quality products at the right price (Tirole, 1996); the information possibly gathered from consumers from prices; the information gathered from agents in ongoing and past interactions; the diffusion of the information along the network among different agents; the persistence of corruption and the problem of partner selection and reliability (Tirole, 1996); efficiency gains from trustworthiness and reputation in competitive markets (Marimon et al., 1999, Braynov and Sandholm,

2002).

Reputation is also strongly conditioning trust behaviour, expected trustworthiness and observed trustworthiness. In an interesting experiment (Fehr, 2009), a gift exchange game conducted with and without a reputation building mechanism shows interesting results. In fact, allowing for reputation building increases not only trust behaviour expressed by employers wage offer levels, but also reciprocity from the counterparts, expressed by workers effort levels. Therefore, higher effort levels that at first sight might appear as caused by higher wages, are for instance a consequence of higher reciprocity inducted by an informal system of reputation building.

Some of these points are addressed also in non-economic studies, but even these studies mostly rely on economic exchanges (such e-commerce) to motivate their interest on the topic or to test the hypothesis. Some scholars have analyzed the problem from a biological perspective and found out that reputational mechanism even work at the level of our brain, shortening the timing of our decisions in an interaction game, and demonstrating a reputation building effect (King-Casas et al., 2005).

A growing research related to trust and reputation is going on in computer science. Most of the times this literature is linked to Bayesian learning and goes under the broad categorization of reputation/referral systems studies or trust management (Josang et al., 2005). In peer to peer networks, notably in e-commerce websites, but even in the structure of informatic communication among servers and nodes in the world wide web or in mobile networks, trust and reputation are topics of strong interest for many scholars.

The problem at hand is the possibility to assess the trustworthiness of every node in the network, by means of scoring from other nodes in the same network, as to create a reputation building mechanism. To do so, one needs to structure an evaluation scheme/mechanism, and the necessary rules and incentives to foster its reliability in the correct evaluation of the individual nodes.

There is indeed a secondary problem that is linked to the reliability of the evaluating nodes, that is to their own reputation, and to the possibility for subsequent agents to assess not only the reputation of the agent with whom they may have an opportunity for valuable interaction, but also the reputation and reliability of the previous partners that gave that agent an evaluation.

Some authors underline that a trust model needs to take into account many sources of information, as to be robust against some possibly missing ones or lying from other agents. At the same time, every agent should be able to evaluate and pool all these informations on his own. (Huynh et al., 2006). This seems to be an important prerequisite for a reputation system to work properly: on one side, the more the communication and the information sources, the more efficient the system in isolating misbehaving agents; on the other side, the sooner misbehaving agents are recognized and isolated, the better the information obtained and the more stable the system.

An extensive communication network provides a reputation system with stronger incentives for agents to behave in the expected way under the rules of the system. Moreover, and to this objective, protection against unfair ratings is a basic requirement for the robustness of the system. In these contexts, misbehavior is not only intended in the sense of not respecting the rules of the interaction, but, more interestingly, not respecting the rules of a fair scoring mechanism (Whitby et al., 2005).

There seems to be a number of important questions on the mechanisms underling scores generation, discovery and aggregation (Yu et al., 2004). Agents may use their own information, as well as that gained from other sources, to guide their own decision making, but the possible sources of information may not be known from previous interaction, and there is no direct way to assess their identity and motivations.

Certificates (Mass and Shehory, 2001) or context-dependent and subjective scoring systems (Mui et al., 2001) are potentially means to gather some

information on the trustworthiness of the source one is referring to, rely-ing on third-party certifiers, or on statistically efficient systems to purge the evaluations.

There are many different solutions to the problem of generation, aggre-gation and post-evaluation of the scores, and some of them are investigated also from an empirical perspective, to evaluate their effectiveness in the real world. In example, Guha et al. (2004) suggest a model of diffusion of both trust and distrust, and test its efficiency based on some available data from Epinions, a web site where users can give evaluations on many different sub-jects.

To conclude, we underline again that more than trust itself, the main focuses of these studies are trustworthiness and the possibility to assess it at the individual level, aggregating different sources. The mechanism of ag-gregation allows agents to build a reputation on the basis of their previous interactions, and to advertise it at the network level. Given the reliability of the evaluation and assessment mechanism, this in turn allows easier in-teractions, and gives place to a higher level of trust in the network. But this higher trust is a mere result, and this literature is most of the times not directly focusing on it.

2.3.3   Trust and performance

While a wide research has gone into the understanding of the antecedents of trust, many other scholars have focused their attention onto the effects of trust. For the concept to be relevant from an economic perspective one needs in fact to assess whether trust has a causal role in shaping and determining economic results. And more than this, it has to be understood whether trust has a direct or indirect causal role in determining economic outcomes.

In this stream of research, trust role and effects are evaluated at differ-ent levels, from micro to macro. In particular research may be divided for its focus on organizational, inter-organizational and macro-economic perfor-

mance.

As it will clearly appear from the next sections, research has since now been unable to identify a clear cut evidence about a causal role of trust in shaping economic results. This has been due to two main reasons. Firstly, measures of trust have yet not been clearly defined, and the existing ones are often under investigation as effective measure of trust. Secondly, the causality link is sometimes difficult to identify. Since trust is strongly culturally and socially embedded, then research deals with potential endogeneity problems in the analyzed variables.

Some observe, however, that institutionally similar realities may present strongly different economic results. Therefore, it is said, these different performance has to be motivated by some other variables, such as trust differences among analyzed individuals and groups. Nonetheless, the opposite can also be claimed, that is trust cannot develop or properly work where the institutional context is unable to provide sufficient stability and predictability.

The section is structured following the main research focuses: organizations, inter-organizational cooperation, macro-economic outcomes.

Organizations   The management literature has been one of the most active streams on the topic of trust. Most of the studies have been qualitative, but more and more investigation assumes a quantitative perspective. One of the main challenges of this stream is the translation of an individual level concept to an organizational outcome in terms of performance (Zaheer et al., 1998). Many authors agree that trust has positive effects on organizational outcomes, but no agreement is present on the way trust influences these outcomes.

Some authors suggest that trust may be treated as an organizing principle (McEvily et al., 2003), that is it affects several important organizational attributes through different causal pathways. In particular, two are the main means in which trust influences organizational structure and responsiveness.

The first is structuring and it is related to the development, maintenance and modification of a system of relative positions and links among actors in a social space. These relations and links are either formal and informal. Trust then operates on these variables via many different channels, affecting the social structure of an organization.

The second is mobilizing, that has implications with the process of converting resources into finalized activities by interdependent actors. Specifically, trust influences information and knowledge sharing, commitment, and monitoring, motivating actors to contribute and coordinate their resources, and directing them toward the achievement of organizational goals.

Notwithstanding these considerations, still trust has potentially negative effects. Unconditional trust may lead to even great fraud, and uncertain situations or rapidly changing environment may put a strong challenge on trust. This is why reconstruction of trust is also an important matter for organizations, beside trust building and maintenance. On another side, as trust is a useful tool for analyzing environment, allowing for a better predictability, it has also a possible downturn producing systematic biases of interpretation.

Two different approaches are suggested for the analysis of trust effects on organizational performance (Dirks and Ferrin, 2001). On one side, some describe trust as having main effects on outcomes, that is trust enables directly more positive attitudes, higher levels of cooperation, and superior levels of performance. This is called the main effect model.

Others, on the contrary, suggest that trust has indirect effects via other determinants of economic performance. In this sense, trust provides the conditions under which certain outcomes, such as cooperation and higher performance, are likely to occur. This perspective is called the moderation model.

The basic idea underlying the main effect model is that trust influences the behaviour of agents in their interactions, allowing for higher risk taking propensity in cooperation and information sharing. This, in turn, is expected

to lead to better individual performance. The focus is then the evaluation of the effects of this assumption on many different aspects: communication and information sharing, organizational citizenship behaviour, effort, conflict, negotiation, individual performance, group performance. For many of these outcomes, the estimated effects are generally not statistically robust, and the evidence seems to be inconsistent across studies. The most robust result appear to be the effects of trust on organizational citizenship and individual performance (*ibidem*).

Some studies concentrate on the effects of trust on workplace attitudes. In this sense, enhanced trust in the management seems to affect general satisfaction and organizational commitment. Most of the literature find positive effects of trust on variables related to satisfaction, such as satisfaction with decisions, supervisor, relationships, job. Job satisfaction is strongly linked to social interactions workers build in the organization (Davis et al., 2000). Support from colleagues and good interpersonal relationships at work are positively associated with general job satisfaction, and negatively with stress, absenteeism and turnover (Dur and Sol, 2008). Incentive schemes designed to create externalities among workers, such as team incentives or relative incentives, foster the engagement of workers in cooperation. This increases the general level of job satisfaction, that, in turn, has been found to be negatively related to wages level (*ibidem*).

However, the evidence on direct effects of trust on performance doesn't seem to be robust nor straightforward. Dirks and Ferrin (2001) suggest an analysis of the moderation approach. As trust may represent how individuals understand and approach their relationships, it may foster two different aspect of these interactions. Firstly there may be an effect on how one assesses the future behaviour of the counterpart. Secondly, trust also affects the interpretation of past and present behaviour. These effects reduce uncertainty and ambiguity inherent to the relationship and, in turn, this affects outcomes, that is trust has an indirect effect on performance, facilitating the

conditions under which a better performance may occur.

Two ways in which these indirect positive effects may arise are suggested. On one side, trust has an indirect effect on workplace behaviour by providing an assessment of the potential behaviour of one's work partners, i.e. to the extent an employee trusts his supervisors, he is more likely to devote his resources to role performance, norms conformance, rule compliance, supervisor's requests, because he believes he will receive appropriate rewards. On the other side, as trust provides a way to interpret partner's actions, it may well influence the reaction to these actions, i.e. if an employee receives a negative feedback from an untrusted supervisor, he is less likely to react putting more effort in his workplace behaviour to improve his own performance.

It has been suggested that there is a tendency for trustees to fulfill trust because they believe they are trusted, what is called trust responsiveness (Bacharach et al., 2001). The propensity to fulfill trust is found to be sensitive to the structure of payoffs. Work payment schemes based on trust rather than monitoring are described as having positive effects on the rate of fulfillment, via the transmission of credible signals of trustor's confidence. This is in line with the concept of moderation effects of trust on performance.

In line with these results is also the gift exchange presented in the previous section (Fehr, 2009). In this type of game higher performance derives from higher effort level of the workers, that in turn is motivated by a reputational mechanism inducing higher trust levels and reciprocating propensity.

Another study analyzes the effects of trust level within a group on trust performance (Dirks, 1999). In the study an experimental approach is applied. Both the main effect model and the moderation model are used. Dirks presents positive effects of trust on group performance and tests whether these effects are directly correlated with performance. The evidence suggests that higher performance is not necessarily related to higher trust groups, but that trust influences how motivation is translated into agents behaviors. In particular, in high trust groups, motivation translates into cooperative action

and joint efforts, while in low trust level groups motivation translates into individual efforts.

Trust also relates to the responsiveness of an organization to environmental contingencies, such as a crisis (Mishra et al., 1996). Literature has found that in these occurrences organizational response usually implies more rigidity and centralization. However, research has also tried to investigate why performance is not always at risk during crisis, in face of the evidence that some organizations fail to respond actively and efficiently to crisis, while others doesn't seem to be affected so much.

The authors adopt a moderation effect approach. Trust is conceptualized as a mediator of three organizational behaviors: decentralized decision making, undistorted communication, and collaboration within and across organizations.

Trust within the top management group, trust between the top management and lower levels of the organization, and trust between organization member and its suppliers and customers, seem to have a direct effect on the behavioral responses adopted in the organization facing a crisis. In turn, these behavioral responses affect the general level of performance and the responsiveness of the organization to crisis challenges.

Practically, trust allows the implementation of solutions that may affect the ability and timing of resources allocation and use in the whole organization. Decentralization frees resources and allows for higher flexibility and efficiency; higher quality communication, good information flows and better accountability of the information itself are expected to foster the ability of response and the efficiency of resources allocation; internal and external collaboration also affects the efficiency and timing of resources reallocation responses.

Some evidence is also given of the potentially negative effects of trust in crisis contingencies, mainly if trust is to foster security feelings, therefore lowering employees motivation.

Lastly, another stream of literature links trust to knowledge transfer and learning. Knowledge production and its diffusion within organization is now well investigated and appears as one of the most important challenges for managers. Levin and Cross (2004) surveyed employees of three different organizations, an American pharmaceutical company, a British bank and a Canadian oil and gas company. The authors focused their attention on perceived trustworthiness, as a mediator of positive effects of social ties on learning. They use two different categories of trust, that is competence based trust and benevolence based trust.

Their findings suggest that trust mediates the link between strong ties and knowledge exchanges. Moreover, they show that competence based trust is especially valuable for tacit knowledge exchanges. Different constructs of trust and knowledge may interact in different ways in the knowledge management process. An extensive discussion on the topic is also provided in Ford (2002).

Inter-organizational cooperation Trust is considered to have performance implications also for inter-firm relationships, to the extent it may affect the efficiency of these interactions and increase their innovative results. Moreover, trust may also facilitate the creation of inter-firm collaborations, increasing the predisposition towards alliances and allowing for easier contractual practices.

As seen in the previous paragraph, trust may influence cooperation predisposition among organizations even during crisis contingencies, leading to a better deployment of resources and increasing the speed of adaptation in face of a critical environment (Mishra et al., 1996).

Many scholars have highlighted the importance of understanding trust in cooperative inter-firm relationships. However, also at this level there is no clear cut evidence about the precise role of trust and its influence on performance. Zaheer, McEvily and Perrone (1998) investigate how trust operates

at the individual and organizational levels, how these two are related, and how the individual level phenomenon translates into an organizational outcome, that is performance. The conceptualization of trust refers to the extent to which members of a focal organization commonly held a trust orientation toward the partner firm.

The focus of the research is on boundary spanning individuals. Interviews are administrated to a number of purchasing managers from different firms in the electronic equipment sector, to reach a better understanding of trust perceptions and relationships, and to allow the construction of measures of trust both at the individual and at the organizational level.

The findings show that organizational trust and individual trust are related, although they operate separately. The more the boundary spanning individual in the buying organization trusts his counterpart, the more his organization trusts the supplier organization. This relationship works both ways, suggesting mutually reinforcing effects of trust.

Other effects are described in the paper as relevant and significant. Primarily organizational trust is related to lowered costs of negotiations and conflict resolution, while individual trust strikingly positively relates to negotiation costs. Moreover, it appears that organizational trust has a direct effect on performance.

Institutionalized practices and routines designed to deal with partner organizations seem to have major effects, transcending the importance of the individual boundary spanning employee. This institutionalized environment creates a stable and normative context in which the partnership develops, well beyond the individual agents that operate at the boundaries of the two organizations.

Trust seems also to be relevant for innovative joint efforts or for research partnerships. Because collaborations among firms bring risk and are strongly uncertain, and because contracts, monitoring and control cannot solve all the inherent problems, trust is a necessary requisite to positive results of joint

ventures (Nooteboom, 1999). Of course trust has limits also in this context. Specific forms of trust may come to dominate a specific environment due to the evolution of communication channels and of information costs (Lorenzen, 1998).

It has been noted that trust might lack more exactly where it would be more needed for cooperation, i.e. in strongly uncertain and highly competitive environments. At the same time, trust may lead to lock-in effects into ongoing partnerships, making organizations blind to new opportunities and profitable alternatives.

To resume, we observe that this duality of trust is clear in the contrast between stability and flexibility of organizational settings and response. In fact, trust is typically involved in long-lasting relationships, as it requires a minimum horizon of stability to foster positive outcomes. This necessary prerequisite of stability, nonetheless, may cope with the flexibility requested to deal with uncertain environments and rapid change. On the other side, as it has been observed, it can precisely be trust to foster this flexibility, allowing for easier interactions and more rapid communications and response timing.

Macroeconomic indicators and culture This stream of research goes well beyond the boundaries of organizational or inter-organizational performance. During the last decades research on trust has broaden its limits to the analysis of the impact of trust on institutions and on macro-economic indicators. At the same time, research has addressed the impact of institutions and economic development on trust, trying to analyze the opposite causality link.

Most of this research is broadly categorisable under the studies on how culture may affect economic outcomes. Whether or not the effects analyzed are clear and the measures used are really showing evidence of a causal effect going from cultural variables, among which trust, to economic outcomes, this remains questionable. This is clearly shown by the contrasting evidence of

the findings and by some researchers accounts.

Notwithstanding these opened questions, lot of research tries to link various macro-economic indicators to cultural aspects, among which trust. Guiso, Sapienza and Zingales (2006) give evidence that culture has an effect on economic outcomes, and this effect may persist over time despite of its lower profitability. Other studies relate trust and financial participation (Guiso et al., 2007), where evidence emerges that financial participation in the stock market may be influenced by cultural matters, among which individual trust has a particular incidence.

Tabellini (2005) measures four values in a principal component analysis: trust, beliefs in the importance of individual effort, generalized morality, and obedience, inferred from questions on the World Values Survey. Using a measure of education and historical political institutions in the European countries, he analyzes the effect of these institutions on today's social values and finds a positive effect. He also finds that countries with a higher level of "good values" such as trust, have both higher GDP per capita and growth rates.

Cultural similarities or dissimilarities among pairs of countries may affect trust between the countries themselves (Guiso et al., 2004). The formation of cultural priors defined by inherited cultural aspects, is weakened for more educated people, suggesting that education lowers the idiosyncratic component of trust. Government performance, civic engagement, importance of large firms, are found to be variables relating trust to performance.

Social identity differences may have impacts on individual performance (Hoff and Pandey, 2004). This cultural effect may indeed be the cause for differences among countries in entrepreneurship and economic development. Once a social identity difference, as the one posed by the caste discrimination, exist or persist in a society, then the group belonging to that discriminated social identity may well be out of the economic arena, or under-perform, lowering the social welfare of the country, that looses part of its potentialities.

Trust in itself is also influenced by cultural differences. Guiso, Sapienza and Zingales (2003) investigate how trust is linked to religion or ethnic origins. They use data from the World Values Survey and find strong evidence in favor of their claim. They also replicate the experiment using data from the GSS in the U.S.. Similarly, they find positive and persistent effects of ethnic origins, indeed of religious background, on trust levels as declared in the answers to the survey. Furthermore, trust seems lower in countries with dominant hierarchical religions, which may deter the formation of horizontal form of cooperation among people (La Porta et al., 1997).

It has to be stressed that the use of surveys questions to analyze trust has been put in doubt by many scholars. The broad generality of the questions, the de-contextualization of the concept of trust, the possibly arising confusion, have been underlined as some of the possible reasons that undermine the validity of survey measures.

At the same time, it has to be noted that while respondents may understand and answer correctly, they are most probably not giving a general answer, but thinking to their previous experiences. Therefore, the answer is not depicting today's level of trust in people, but an historical data about past trusting experience. Even more, some studies find that general survey attitudinal question about trust more properly account for trustworthiness (Glaeser et al., 2000, Holm and Danielson, 2005, Sapienza et al., 2007).

Among others, Knack and Keefer (1997) analyze the relationship between trust, measured at the country level as the percentage of positive respondents to the World Value Survey, norms of civic cooperation, and economic performance. On the evidence coming from an investigation based on 29 countries, they conclude that indeed trust matters and has a positive impact on economic performance and on aggregate economic activity.

This seminal work has been lately revisited and the results tested for robustness (Beugelsdijk et al., 2004). Different robustness checks for the regression analysis conducted by Knack and Keefer lead to the conclusion that

the results were of limited robustness, strongly depending on the conditioning on the set of variables taken into account. The predictions stated in the conclusions of the study were then far-reaching.

Knack (2001) highlights that when studying the relationship between trust and economic welfare, the choice of units of analysis is crucial. Cooperation generated by trust has costs and benefits, the benefits being stronger for the in-group, whilst the costs being imposed more on the non members. It has been already highlighted that belonging to a group is beneficial to an individual when the group welfare is high, cause member benefit more from the other group members.

Knack claims, among others, that a "radius" of trust must be defined before investigating the welfare effects of the concept. When the radius has no coincidence with the population for which welfare is measured, then we may expect, at best, ambiguous effects. In his study he focuses his attention on the effects of wide radius trust on economic performance measured at the national level. This choice is made because in highly developed societies a sizable proportion of ongoing exchanges are at the level of strangers, parties without prior personal ties. In high-trust societies of this type, individuals may contract without extended written agreements; run activities without too much monitoring of employees, partners, suppliers; support efficient economic policies independent of the personal benefit. In general the author presents evidence in favor of a positive impact of trust on economic growth.

Slemrod and Katuscak (2002) investigate whether more trusting people are better off, and whether being more trustworthy influences individuals prosperity. They use data from the World Value Survey and data on real household income, adding some other socio-demographic variables as controls in their regressions. They suggest evidence that the personal return to trustworthiness is negative in most countries, but increasing with the average level of trust. On the other hand, the personal return to trust is positive in most of the countries, and increasing in the average level of trustworthiness

in the society. However, the robustness checks confirm the former result, but not the latter, due to the survey questions not really measuring trust.

Other scholars more broadly consider trust as part of social capital. Carter and Castillo (2002) study fourteen separate South African communities, seven urban and seven rural, previously selected for a study on national living standards in South Africa. The researchers implemented three different game settings to test for altruism, trust and trustworthiness, and used a multi-stage experimental approach to disentangle the effects of trust and reciprocity from the effects of altruistic caring of others. They find that these norms show large positive payoffs in urban areas, where probably opportunities are greater, or trust permits people to broker opportunities linking to other people. On the other side, norms of trust and altruism negatively affect livelihood in rural areas, negatively impacting on incentives and incomes.

We believe this is an interesting example of how stability and flexibility mix in trust dynamics, with sometimes contrasting effects. Most probably, in rural areas trust creates lock-in effects, making new opportunities far reachable, even though potentially valuable. On the contrary, trust facilitates the flexibility needed to cope with a new fast moving environment, such that of urban areas. Here more trust increases the propensity of individuals to contact and interact with other agents, therefore increasing chances and, potentially, individual performance.

Berggren and Jordhal (2006) investigate the other side of the coin, that is how institutional settings may affect the emergence and diffusion of trust. They run cross-country regression for almost 50 different countries to explain trust levels registered in 1995 and 2000. To account for social capital they use a measure of generalized trust taken from the World Value Survey. They use a specification of economic freedom of a country, that resumes in five different areas the level of freedom and development of institutions and market in that country. The areas of interest and under investigation are: size of the government; legal structure and security of property rights; access to sound

money; freedom to exchange with foreigners; regulation of credit, labor and business. Results show that trust is affected positively by the presence of a legal system and a well developed system of property rights. In some of the specifications they also find evidence of a positive effect of access to sound money and regulation of credit, labor and business.

We conclude this section presenting two other studies going along the same line (Raiser et al., 1999, Radaev, 2002). Both studies analyze transition economies, that is the case of Russia and of eastern Europe countries. Findings show that these countries face heavy legacies of distrust in the state and closed social and business networks, which reduce competitive pressures and mechanisms of adjustment. Formal rules are typically contradictory and unstable, and there is a lack of legal enforcement, which fosters the level of uncertainty. The unpredictability of the state and of regulatory policies is not compensated by an increased reciprocal trust on the market, due to the frequent infringement of business contracts and non-transparency of business transactions.

Both studies conclude that a linked mechanism of trust building at the generalized level, that is trust in the institutions, and trust at the relationships level, that is trust at the business and network of economic actors level, needs to be simultaneously operating as to develop a higher level of trust and economic development. For instance, the enforceability of laws and a stable system of norms are at the basis of the process of trust building and diffusion.

2.3.4 Trust and networks

In the previous section we have seen how formal and informal networks of interaction may foster trust and, eventually, performance. It has also been underlined that trust allows for higher propensity towards interactions and new opportunities arising from the social and economic context in which agents and/or organizations operate.

This is not only true from an economic perspective. The importance

of social interactions for trust building and diffusion appears clear and well recognized in many studies and in many different domains. Processes of learning and information exchange go on in social networks among all agents involved in relationships with other agents, allowing for the creation and diffusion of reputation, whether good or bad it may be. In turn, this process allows for a restructuring of the network based on the available information and on the active exchange channels.

Studies of trust in network literature are strongly related, indeed, to studies of information diffusion. Different question are at hand in this research, that is how information spreads, how the underlying network structure influences the information diffusion process and the economic outcomes, and how efficient is the resulting network structure after information diffusion has taken place.

Literature on reputation systems also focuses on networks and information. Most of the findings presented in the section about reputation with respect to referral systems show the relevance of network studies for the understanding of trust and its dynamics. Moreover, social or artificial networks are the contexts in which models of trust are experimented and tested.

In this section we want to add some insights deriving from research not directly relating to trust, but focusing primarily on networks themselves. Some of the results are, to our view, of some interest also in the modeling of trust and in the understanding of some of the involved mechanisms influencing its dynamics.

Bala and Goyal (1997) study the problem of communication networks asking which structure is more reasonable, and which is the relationship between socially efficient networks and the networks derived from choices made by self-interested individuals. The problem of efficiency is a relevant matter for the timing and the speed at which information flows in the network, allowing agents to gather the relevant data about the individual agents they may interact with, and about the opportunities present in their community.

In their model agents weight the cost of link formation with other agents and the potential benefit deriving from the information obtained through that link. Cost and benefits structure accounts for the public good aspect of information, that is well-connected people collect information from many different sources and generate a positive externality. As a consequence, agents try to link directly to these ones, rather than linking to all their partners.

The results they present are striking, as they demonstrate that self-interested agents choices lead to a well defined network structure with probability one, that is a wheel network. While many different structures may be sustained in equilibrium, the wheel network is the only efficient architecture. Even robustness tests, allowing for some hypothesis relaxation, give similar results, demonstrating that wheel-similar networks always emerge at the equilibrium. In the case of potential decay of the information quality along the transmission, they find that the resulting network is constituted of local neighborhoods.

Learning processes are another interesting aspect considered in this research stream linking trust and networks. Bala and Goyal (1998) also discuss the process of learning from neighbors. In particular they consider a network of agents who have to choose an action with uncertain outcome. After the action is chosen, a random reward is assigned. Agents try then to infer information about the most profitable outcomes, updating their beliefs based on their own experience, and on information they gather from neighbors in the network. They show that the structure of these neighborhoods has important implications for agents choices, and for the temporal and spatial patterns of diffusion in a society.

As it has been said, this research has not been conducted with a specific focus on trust. However, the emerging results are appealing also for the development of trust research and some of the insights on network properties, individual network attributes, learning, and communication are easily shared between the studies just presented and those on trust. In fact, the

implications of networks attributes and mechanisms for trust emergence and diffusion have been studied by many scholars. We give hereafter some evidence about this research.

Buskens (1998) study the way in which the level of trust in cooperative relations depends on network structure. The idea is based on the findings that a network may affect trust diffusion in at least two ways: a buyer obtaining information about a seller's earlier transactions, and a buyer informing potential buyers that a seller is abusing trust. These reputation effects are studied with a focus on their origins as the result of network position and specific network structure. The dependent variable is the level of trust a buyer can have in a seller, and it is defined at the individual level.

Previous research stresses many points related to this topic. In particular, the level of trust that can be placed is higher if the network has a higher density. A second finding is that the more an actor is central, the faster he can receive and spread informations. Finally, if both in-degree and out-degree are high for many agents, then the information diffusion is more efficient, opposite to the situation where there is considerable variation in out-degrees and in-degrees.

Buskens suggests some possible conclusions. In particular, the density of the network becomes more important as the trust situation becomes more difficult, that is the seller has a higher incentive to abuse trust, or the punishment potential for the buyer is low or, again, the seller has a lower weight on future interactions. At the same time, buyers with a higher out-degree show more trust, even after controlling for density. Level of trust increases if buyers direct their ties to other buyers with high out-degrees. Finally, individual centralization has a positive effect for the individual agent involved, but the level of trust across all buyers decreases in the centralization of the whole network.

Similar results hold in a network model dealing with cooperation problems (Buskens and Weesie, 2000). Among the other results in accordance with

previous literature, the main finding of the research is that when agents have heterogeneous out-degrees, then the network that is centralized around the trustors with the higher out-degree produces the higher level of trust. However, the model used in this paper is strictly bounded in the assumptions on the timing of actions and communication among agents. Secondly, it is not analyzed whether the preferences structure of agents and the payoffs may have an effect on the levels of trust.

To conclude, we present evidence from another interesting perspective on networks and trust, that is the effects of control and learning. Social embeddedness may affect trust among agents in two different ways. Firstly there is learning, that defines the possibility to base trust on previous experiences with a partner, or, at best, to even gather information from other agents in the network about their previous experiences. Secondly there is control, that describes the situation where trust is built upon the possibility to sanction untrustworthy behaviour through own or third-party sanctions.

We may note, the similarities with the gift exchange game in Fehr (2009), however there no network consideration is done, and both control and learning work only at the dyadic level, without any communication among other actors external to the interactions.

Buskens and Raub (2002) give instead evidence on the simultaneous working of learning and control mechanisms at the dyadic and at the network level. They analyze trust situations where the main problem is the incentive of the trustee to abuse trust. The authors present some empirical results from previous research on the topic, claiming that the distinction of the two different mechanisms remained unclear, and their effects have not been disentangled. They investigate the results of two experiments, using hypothetical bargaining situations.

They find evidence of learning and control mechanisms both at the dyadic level and at the network level. These mechanisms facilitate trust building and diffusion, highlighting the relevance of the embeddedness effect. Not only

trust prevails in embedded settings, but these settings allow for a better development of the underlying mechanisms at the base of trust, that is learning and control at every considered level level. They finally claim that learning becomes more important relative to control in situations where uncertainty is higher.

Finally, Buskens (2003) compares three different trust settings, to evaluate the effects of exit, control and learning on trust building. In particular he presents mathematical modeling of the different settings under investigation: a model of a dyadic interaction, used as a baseline model to evaluate the results; a model where two trustees are involved in an interaction with a trustor, that is the trustor has an exit option; and finally a model where two trustors play with one trustee, allowing for communication among the trustors about the trustee's behaviour.

The author find that exit options have no implication for trust, while adding a second trustor, that is allowing for control, only has an impact if the communication rate among the trustors is high enough relative to the trustee incentive to abuse trust. About this communication, it is found that control has only an effect on trust when agents share among themselves their own experiences, and learn from this information. Where this process is not combined, the effects on trust are negligible.

2.3.5 Trust and preferences

As we have seen in previous sections, trust is claimed to shape economic outcomes and performance, both at the individual and the macro levels. It has been analyzed for its strategic implications in agents choices. These streams try to develop a theory of trust typically using well developed analysis tools, making at the same time trust a treatable and simple concept, while recognizing its complexity.

However, to understand trust one needs not only to assess its consequences, but also to analyze its antecedents. In particular, for a better

understanding of trust dynamics, research has to clarify whether trust is a well-defined attribute and eventually disentangle it from other attributes. More than this, research tries to link trust to micro-economic theory concepts, such as preferences and/or beliefs. It is exactly these steps that, in turn, may motivate the impact of trust on agents decision processes and on economic outcomes.

In fact, some authors claim that most of trust research is not conducting to any other result or implication than studies on transaction costs or on Prisoners dilemma-like economic situations (Castelfranchi and Falcone, 2000). It is underlined how a richer view and description of trust is needed to make it a valuable concept for economic analysis, and not just a mere word to make a more user-friendly theory or a redundant and useless concept. Both the opposite approaches of reduction of trust to subjective probability, such as in Gambetta (2000), and of trust elimination, are criticized.

The major claim is that a cognitive theory of trust is needed as part of an enriched description of the 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. Castelfranchi and his colleagues, therefore, try to put explicitly what is assumed to be described by the single number given in the subjective probability approach.

But some points need to be stressed, to our view. In fact, one may not deny that all these considerations are necessary to understand what trust is and how it may work through different channels, but it seems too complex to define an operational model of trust with such a degree of specificity.

Moreover, while a subjective probability number may not describe all the mental processes that give place to its definition in the trustor's mind, at the same time it seems particularly useful, after a discussion of how this number is formed, and what it may mean, to be able to synthesize the concept and the cognitive function of subjects, in a treatable and simple instrument. While

the authors claim a more complex approach to trust, in the article they end up with the same single number of a subjective probability, and the difference to other studies lies not in the operationalization, as they seem to claim, but in the more detailed picture of the process that leads to that number.

The authors also suggest a decomposition of trust, as partly defined by the trust in the trustee, and partly by a generalized trust in the context of the interaction. This view seems redundant, as the context is better described in terms of uncertainty or risk, and one doesn't necessarily need to trust the context of the interaction, while more is needed to contextualize the trust level in the trustee. To our view, this decomposition is better understood, on the contrary, in terms of agents preferences, such as risk aversion.

Authors also claim another distinction on competencies, as they define a variable of trust based on the perceived capabilities of the partner. However, this aspect is normally defined as confidence. We observe, in addition, that competencies are not easily detected and perceived, unless direct experience has already taken place. Otherwise, to our view, trust in not in perceived competencies, driven by, to say, the professional title, but more in the value of the title itself, or in the reliability of our counterpart's words about his competencies.

In this stream of literature, linked to the mental components of trust, to beliefs, interests, and preferences, some interesting insights come from studies conducted by Fehr and some colleagues (Fehr et al., 2005, Kosfeld et al., 2005). These studies use an approach referred to as neuro-economics, that brings together methods from neuroscience and economics, to investigate how human decisions in social and economic context are generated in the brain. In particular the authors analyze some evidence on other-regarding behaviors. The setting is similar to the one used in game theoretical approaches, and usually a trust game or an investment game are the context under which the study is conducted.

In a recent paper, Fehr (2009) documents the recent accumulation of

strong evidence – neuro-biological (Kosfeld et al., 2005), genetic, and behavioral (Cox, 2004) – 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 (Bohnet et al., 2008).

Kosfeld et al. (2005) study on oxytocin, an hormone that has been studied as an important enhancer of pro-social and affiliation behavior in many different non-human mammals, demonstrate that this substance may increase pro-social behavior of humans, acting on their preferences towards social behaviors. This doesn't affect in any way reciprocity, hence trustworthiness, but only the component of preferences that is linked to trusting others. It is also to note that beliefs about others trustworthiness remain unchanged with or without oxytocin.

The study also pose the question whether trust is the same as risk taking, and the evidence suggest that more specifically trust requires a risk that is generated in a social environment and interaction, where, indeed, the specific risk is linked to the agents choices. Lastly, Kosfeld and his colleagues exclude every possible effect of oxytocin on the outcomes (such as a possible psychotropic effect). Humans (but not only) biologically reply to the substance increasing their preferences towards trusting.

Fehr (2009) reports more research evidence that the risk present in trusting is of a different kind than the general case of risk taking. Trusting is a risk taking involving social interactions, and studies in the biological and evolutionary domain are demonstrating that this behavior is codified and appears in specific ways.

Risk and social preferences affect measures of trust such as the ones derived from the surveys. This finding suggests that survey measures do not just capture beliefs about people's trustworthiness but are also influenced by

their preferences. It is shown that preference measures do not affect beliefs about others' trustworthiness in a trust game.

The author also argues that although it seems possible that trust has a casual role in affecting long-term outcomes, evidence is yet not fully convincing. In his view, economists still lack instrumental variables for trust that support causality claims beyond doubt.

Other scholars demonstrate, using an experiment based on the trust game and functional magnetic resonance, that a neural process develops and changes as far as trust and interactions are developed (King-Casas et al., 2005). Decision making processes of subjects entering a trust relation are strongly influenced in magnitude and timing by the signals of trustworthiness coming from their counterpart.

This seems to demonstrate that the structure of preferences may change over time, as soon as useful information is gained from the interaction, so that the preference towards trusting becomes strongly predominant.

Apart from this seemingly increasing literature using non conventional tools of analysis, there is also a broad and growing literature using more traditional tools. A similar result is reported in McCabe, Rigdon and Smith (2003), where the authors investigate the difference between outcomes-based and intention-based approaches to trust.

Outcomes-based approaches would predict that only the pay-offs of a game matter to guide agents choices and behaviors. On the other side, intention-based approaches, explore the possibility that, depending on the available alternatives, identical outcomes may be interpreted differently.

Authors suggest a trust game, with two different settings. The first is a traditional game, where the first agent has an exit strategy, that is he can mistrust. The second setting, on the contrary, has no exit strategy, therefore the first agent is forced to enter the game and to send some money.

It is suggested that when the set of available choices differs, the same behavior by an agent may be interpreted differently by his counterparts.

The authors find that a different structure of the game gives players different signals about the motives underlying the observed choices, whether those are based on trusting or not. This in turn enhances or lowers the preference towards reciprocating. We may note that, however, the structure of the game, that is the tree of available choices, is not always common knowledge in trust interactions.

Andreoni (1995) stresses that subjects have a clear view of the free-riding problem in cooperation and trust, but choose anyway to enter the game out of forms of kindness, but for specific preferences for cooperation. He also points out that the experiments are most probably already eliminating a large amount of the subjects natural preference for cooperation, being the laboratory conditions designed to control and minimize social effects like kindness.

Diekmann and Lindenberg (2000) also highlight the presence in sociology of different streams of research that rely on rationality, but add specific assumptions on preferences. These studies are linked to what they call social rationality. The preference structure of individuals and the different typologies are brought together to analyze how cooperation may spread over time, taking over other strategies. And trust is a major concept in the analysis of cooperation in the models considered.

As the topics of preferences and decision making are strictly linked, it is also interesting to observe that psychologists have more and more deeply analyzed the problem of framing, that is how a specific context, with its social values and connections, may influence selectively the attention of individuals on specific aspects of a situation/event. It has been previously stated and observed, i.e., that the presentation of the rules of the game in the experiments may influence the results, as long as the participants perception of the objectives and of the structure of the game is influenced.

Lindenberg (2000) observes that the theory of framing has important implications also for a theory of cooperation and of trust. In fact, framing

allows a different perspective on rational choice, that is it is rational to stick to rules, even though this implies opportunity costs, where agents act in a stable normative frame. In this situation, actors may expect their counterparts to react cooperatively, being the frame and the context in which the interaction takes place normatively stable.

Where the frame becomes unstable, group memberships, associations and other informal means to create a common frame are lacking or failing their objectives, implying a default in the stability and reliability of ongoing and future exchanges. The ability to read the signals coming from the partners becomes highly important, but even more it is fundamental the direct observation of the social environment where the interactions take place.

The problem of social preferences is also linked to another important aspect of the research on trust, that is the strong attention to social norms. And this is not surprising since it was already Coleman (1990) to stress the importance of norms in trust emergence and building. Once one goes a bit further, investigation takes place to explore whether these norms reside at an individual or at a more aggregated level, what is called social identity.

Butler (2008) uses the typical setting of a Trust game to analyze the problem. He gives evidence that categorizations, that is social identity belonging, significantly affects subjects behavior. In particular, he report results that subjects hold member of their own social identity to a higher standard; moreover, people with a higher status, hold all the others to a higher standard, as for a sense of benevolence. Subjects norms and preferences seems then to be linked to the particular social identity they think themselves into. A change in this identity framing process, implies a change in the relative weight of norms in the individual preference structure, opening the possibility for different behaviors.

These effects may go even further. A clear example and proof, although not strictly linked to the topic of trust, comes from a study on social identities and inequalities in the Indian caste context (Hoff and Pandey, 2006).

The public disclosure of information about the social identities of subjects involved in experiments, seem to change their behavior, even if the information has no relevance for the pay-offs. It's like being in front of a social identity lock-in effect. The persistence over time of these effects possibly indicate that, in example, a bad reputation, whether correctly or wrongly assigned to a subject, may have long-lasting effects in the dynamics of trust building and in the cooperative outcomes.

All this discussion shows how active is the research on trust antecedents, and how this is strictly linked to the understanding of trust consequences on individual and aggregate performance. One of the biggest problems remains therefore the identification of reliable measures of trust, disentangling the different components. Findings and opened questions on this topic are the focus of the next section.

2.4 Measures of trust

As it should be more than clear at this point of this overview, recent literature contains multiple meanings and measures of trust. At an empirical level, research on trust has mostly relied on associational density measures (e.g., number and strength of civic associations), or on survey questions asking respondents to self-report trust (e.g., how much they trust family, neighbors, and government; how much they contribute to charities; how often they lend money to neighbors; etc.).

Neither approach is entirely satisfactory. Associational density may measure perfectly incentive compatible information sharing through networks and may at best be weakly related to trust (Carter and Castillo, 2002). Self-reported trust measures have been criticized as suspect by many authors (Putnam, 1995).

The empirical literature on trust has typically focused on responses to the question: ''Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?'' This ques-

tion, similar to the one in the World Value Survey, is taken from the National Opinion Research Center's General Social Survey (GSS). The survey is the primary source for U. S. evidence on trust and social capital, administrated since 1972.

Subject responses to the GSS trust question are difficult to interpret. Variation in responses might arise for numerous reasons: e.g., differences in beliefs about the trustworthiness of a common set of people; differences in interpretation of who comprises ''most people''; differences in interpretation of what it means to be able to trust someone; or differences in the ability to elicit trustworthy behavior from other people. Variation may also arise because some respondents are not willing to answer truthfully.

Different respondents might understand such questions differently, or they may respond differently according to the identity of the interviewer. More importantly, even if these questions do reveal information about the subject, it is difficult to understand what is exactly uncovered.

On the contrary, other authors (Knack, 2001) claim the reliability of survey measures of trust and show strong evidence of the validity of these measures. In particular they present a strong correlation with external assessments, trustworthy attitudes and trustworthy behavior. Moreover, these correlation are significant due to the independence of the possible sources of errors of these external assessments from the possible sources of errors in the trust surveys.

However, even on this point there is no clear cut if, as it may appear from other research, measures of trust as presented in surveys better predict trustworthy behavior, more than trust in itself, while trust would be better predicted by past trusting behavior in previous social interactions (Glaeser et al., 2000).

Finally, most of the empirical approaches offer no clear separation of the effects of different norms (e.g., altruism versus reciprocity), despite the fact that these may have radically different economic impact. This is partly due to

the lack of survey measures able to differentiate among the different concepts.

Experimental economic methods offer a potentially more interesting way to measure such effects. Many scholars (Henrich et al., 2001, Camerer and Fehr, 2002) advocate the use of economic experiments to measure the relative importance of social norms and preferences, while others (Carpenter, 2002, Carter and Castillo, 2002) suggest the use economic experiments to measure social capital and trust.

On the other side, many underline that experiments may present problems similar to the ones described for surveys. In particular they highlight possible confusion of the participants on rules and game structure and payoffs; lack of understanding; noise in the results and confusion with other norms or other-regarding preferences (Ortmann et al., 2000, Houser and Kurzban, 2002, Cox, 2004).

Experimental economists have used dictator games to measure the strength of other-regarding, or altruistic norms. Glaeser et al. (2000) propose an experimental measure of trust and trustworthiness using a trust game. Since without trust a selfish trustor would be better off by keeping all the money for himself, Glaeser and his colleagues consider the amount of money sent to the trustee as a measure of trust. However, amounts invested in trust games do not necessarily isolate trust. Trust games reveal only how much purely selfish trustors trust. Therefore, measuring trust with this trust game assumes that no other motives explain acts of giving.

This hypothesis, and its experimental implementation, is somehow at odds with the implicit notion that people trust because they are immersed in a social normative universe. People can return money in the trust game out of fairness or inequality aversion rather than out of reciprocity. In the same manner, people can send money out of altruism as well as trust; it may be the case, at worst, that agents are just playing and cheating on rules of the experiment; or there may even be misunderstanding about the rules, or confusion about the real structure of the game and of the payoffs.

Many (Andreoni, 1995, Cox, 2004) have attempted to disentangle norms of altruism, trust and reciprocity. This decomposition of norms is potentially useful because different norms may have distinct economic impacts. Research tests whether or not these different norms really matter (and matter differently) in terms of influencing people's capacity to get ahead economically, as the work on social capital has suggested.

Therefore, as we have shown, the debate is open in academic research not just on the role of trust and of its dynamics, but also on the proper ways to measure it. Not surprisingly, the wide spectrum of definitions and approaches, corresponds to an equally wide number of measures of trust. Nonetheless, it is to be stressed that some of the presented measures are indirect measures, such as in the case of associational habits or other social capital related measures. On the contrary, recent experimental evidence seems to show promising results in terms of a correct evaluation of trust components. This, moreover, appears to disentangle trust from other concepts and attributes, therefore allowing for better measures.

2.5   Concluding remarks and critics

Trust is widely studied for its potential effects on performance, both at the individual and the aggregated level of analysis. For this reason, trust is potentially relevant also from an economic perspective, since it may directly and indirectly foster economic activities, exchanges, communication, partnerships, innovation.

While in a first step trust has been used more as a background concept, such as a facilitating element not suitable for a quantitative analysis, nowadays research has moved its focus into empirical investigation.

As we have noted, this had been done mainly in two directions. The first one is the study of trust consequences; the second, instead, is the investigation of its antecedents. In between, some scholars, mostly in biology, also studied the emergence of trust and its establishment as a successful strategy in the

evolutionary path.

However, either the studies focus on dyadic interactions, or they investigate possible links among cultural variables and macro-economic indicators.

In the first case, we believe research should try to better integrate insights coming from studies on both trust antecedents and trust consequences in dyadic models. Most frequently, instead, studies focus on one aspect or the other, loosing the necessary integration aspect.

Recent studies, though, as we have stressed in this overview, account for empirical evidence on the role of economic primitives such as beliefs and preferences in decision processes involving trust. And there seems to be some empirical evidence of the potential link between trust and increased individual and social performance, whether this is a direct or indirect link.

Therefore, the insights from such deep research should now start to be integrated as tools to solve typical economic problems, finding a way to model trust in general economic settings and to demonstrate its usefulness for a theoretical approach and possible solution to well-known economic problems.

A second step relates more specifically to the literature on the effect of cultural variables on economic performance. We believe here an intermediate step is possibly missing, that is the analysis of how trust dynamics may affect social networks, both in their efficiency and in their composition. This should ideally be the linking element moving the analysis from the dyadic level to the macro level.

It is therefore the objective of the next two essays on trust following in this dissertation, to investigate a theoretical approach to modeling these insights and to suggest some possible analysis to fill these gaps.

References

G. A. Akerlof. The market for" lemons": quality uncertainty and the market mechanism. *The quarterly journal of economics*, pages 488–500, 1970.

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

K. J. Arrow. *The limits of organization*. Norton New York, 1974.

N. Ashraf, I. Bohnet, and N. Piankov. Decomposing trust and trustworthiness. *Experimental Economics*, 9(3):193–208, 2006.

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

S. Ba and P. A Pavlou. Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior. *MIS quarterly*, 26(3):243–268, 2002.

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

V. Bala and S. Goyal. Learning from neighbours. *The Review of Economic Studies*, 65(3):595–621, 1998.

V. Bala and S. Goyal. Self-Organization in communication networks. 1997.

V. Bala and S. Goyal. A theory of learning with heterogeneous agents. *International Economic Review*, 36(2):303–23, 1995.

P. Bateson. The biological evolution of cooperation and trust. In D. Gambetta, editor, *Trust: Making and Breaking Cooperative Relations, electronic edition. Department of Sociology, University of Oxford*, pages 14–30. 2000.

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

N. Berggren and H. Jordahl. Free to trust: Economic freedom and social capital. *Kyklos*, 59(2):141–169, 2006.

S. Beugelsdijk, H. L. F. de Groot, and ABTM van Schaik. *Trust and economic growth: a robustness analysis*, volume 56. Oxford Univ Press, 2004.

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.

S. Birkhchandani and D. Hirshleifer. Learning from the behaviour of others: Conformity, fads, and informational cascades. *Journal of Economic Perspectives*, 12(3):151–170, 1998.

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.

S. Bowles. Group competition, reproductive leveling, and the evolution of human altruism. *Science*, 314(5805):1569–1572, 2006.

S. Braynov and T. Sandholm. Contracting with uncertain level of trust. *Computational Intelligence*, 18(4):501–514, 2002.

L. Bruni and R. Sugden. Moral canals: trust and social capital in the work of hume, smith and genovesi. *Economics and Philosophy*, 16(01):21–45, 2000.

S. Buchegger and J. Y. Le Boudec. The effect of rumor spreading in reputation systems for mobile ad-hoc networks. volume 4, page 4.1, 2003.

V. Buskens. The social structure of trust. *Social Networks*, 20(3):265–289, 1998.

V. Buskens. Trust in triads: effects of exit, control, and learning. *Games and Economic Behavior*, 42(2):235–252, 2003.

V. Buskens and W. Raub. Embedded trust: Control and learning. *Group Cohesion, Trust and Solidarity*, 19:167–202, 2002.

V. Buskens and J. Weesie. Cooperation via social networks. *Analyse & Kritik*, 22(1):44–74, 2000.

J. V. Butler. Trust, truth, status and identity an experimental inquiry. *Not published*, 2008.

C. Camerer and E. Fehr. Measuring social norms and preferences using experimental games: A guide for social scientists. *Forthcoming in: Foundations of Human Sociality - Experimental and Ethnographic Evidence from 15 Small-Scale Societies, edited by Henrich, Boyd, Bowles, Camerer, Fehr, Gintis and McElreath*, 2002.

J. Carpenter. Measuring social capital: Adding field experimental methods to the analytical toolbox. *Social capital and economic development: Well-being in developing countries*, pages 119–137, 2002.

M. R Carter and M. Castillo. The economic impacts of altruism, trust and reciprocity: An experimental approach to social capital. *Wisconsin-Madison Agricultural and Applied Economics Staff Papers*, 448, 2002.

C. Castelfranchi and R. Falcone. Trust is much more than subjective probability: Mental components and sources of trust. volume 6, pages 10–12, 2000.

A. Chaudhuri, B. Sopher, and P. Strand. Cooperation in social dilemmas, trust and reciprocity. *Journal of Economic Psychology*, 23(2):231–249, 2002.

J. S Coleman. *Foundations of Social Theory*. Harvard University Press, 1990.

J. S Coleman. Social capital in the creation of human capital. *American Journal of Sociology*, 94(S1):95, 1988.

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

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

J. H. Davis, F. D. Schoormann, R. C. Mayer, and H. H. Tan. The trusted general manager and business unit performance: empirical evidence of a competitive advantage. *Strategic Management Journal*, page 563, 2000.

A. Diekmann and S. Lindenberg. Cooperation: Sociological aspects. *International Encyclopedia of the Social and Behavioral Sciences (4). Oxford: Pergamon-Elsevier*, 2000.

K. T Dirks. The effects of interpersonal trust on work group performance. *Journal of applyed psychology*, 84:445–455, 1999.

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

R. Dur and J. Sol. Social interaction, Co-Worker altruism, and incentives. *CESIFO Working Paper*, 2476, November 2008.

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

E. Fehr, U. Fischbacher, and M. Kosfeld. Neuroeconomic foundations of trust and social preferences: Initial evidence. *American Economic Review*, 95 (2):346–351, 2005.

D. P. Ford. Trust and knowledge management: The seeds of success. *Handbook on Knowledge Management*, 2002.

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

J. Geanakoplos, D. Pearce, and E. Stacchetti. Psychological games and sequential rationality. *Games and Economic Behavior*, 1(1):60–79, 1989.

E. L Glaeser, D. I Laibson, J. A Scheinkman, and C. L Soutter. Measuring trust. *Quarterly Journal of Economics*, 115(3):811–846, 2000.

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

M. Granovetter. Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, 91(3):481, 1985.

M. Granovetter. The impact of social structure on economic outcomes. *Journal of Economic Perspectives*, 19(1):33–50, 2005.

R. Guha, R. Kumar, P. Raghavan, and A. Tomkins. Propagation of trust and distrust. pages 403–412, 2004.

L. Guiso, P. Sapienza, and L. Zingales. People's opium? religion and economic attitudes. *Journal of Monetary Economics*, 50(1):225–282, 2003.

L. Guiso, P. Sapienza, L. Zingales, and V. D. Macelli. Cultural biases in economic exchange. *NBER working paper*, 2004.

L. Guiso, P. Sapienza, and L. Zingales. Does culture affect economic outcomes? *The Journal of Economic Perspectives*, pages 23–48, 2006.

L. Guiso, P. Sapienza, and L. Zingales. Trusting the Stock-Market. Technical report, January. Mimeo, 2007.

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.

R. Hardin. *Trust and trustworthiness*. Russell Sage Foundation Publications, 2002.

S. H Heap and Y. Varoufakis. *Game theory: a critical text*. Routledge, 2004.

J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, and R. McElreath. In search of homo economicus: Behavioral experiments in 15 small-scale societies. *American Economic Review*, pages 73–78, 2001.

J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky. BAYESIAN MODEL AVERAGING: a TUTORIAL. *STATISTICAL SCIENCE*, 14(4): 382–417, 1999.

K. Hoff and P. Pandey. Belief systems and durable inequalities: An experimental investigation of indian caste. *World*, 2004.

K. Hoff and P. Pandey. Discrimination, social identity, and durable inequalities. *American Economic Review*, 96(2):206–211, 2006.

H. J. Holm and A. Danielson. Tropic trust versus nordic trust: Experimental evidence from tanzania and sweden. *The Economic Journal*, 115(503):505–532, 2005.

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

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.

H. S. Jr James. The trust paradox: a survey of economic inquiries into the nature of trust and trustworthiness. *Journal of Economic Behavior and Organization*, 47(3):291–307, 2002.

A. Josang, C. Keser, and T. Dimitrakos. Can we manage trust. 2005.

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.

A. P Kirman and N. J Vriend. Learning to be loyal. a study of the marseille fish market. *Interaction and market structure: essays on heterogeneity in economics*, page 33, 2000.

S. Knack. Trust, associational life and economic performance. volume The Contribution of Human and Social Capital to Sustained Economic Growth and Well-being: International Symposium Report, Human Resources Development Canada and OECD, pages 172–202, 2001.

S. Knack and P. Keefer. Does social capital have an economic payoff? a Cross-Country investigation*. *Quarterly Journal of Economics*, 112(4): 1251–1288, 1997.

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

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.

D. Z. Levin and R. Cross. The strength of weak ties you can trust: The mediating role of trust in effective knowledge transfer. *Management Science*, pages 1477–1490, 2004.

S. Lindenberg. It takes both trust and lack of mistrust: The workings of cooperation and relational signaling in contractual relationships. *Journal of Management and Governance*, 4(1):11–33, 2000.

M. Lorenzen. Information cost, learning, and trust: lessons from co-operation and higher-order capabilities amongst geographically proximate firms. *DRUID Working Papers*, 98-21, 1998.

G. Lu, J. Lu, S. Yao, and Y. J Yip. A review on computational trust models for multi-agent systems. *The Open Information Science Journal*, 2:18–25, 2009.

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

R. Marimon, J. P Nicolini, and P. Teles. *Competition and reputation.* European University Institute, 1999.

Y. Mass and O. Shehory. Distributed trust in open multi-agent systems. *Lecture notes in computer science*, pages 159–174, 2001.

R. C. Mayer, J. H. Davis, and F. D. Schoorman. An integrative model of organizational trust. *Academy of management review*, pages 709–734, 1995.

K. A McCabe, M. L Rigdon, and V. L Smith. Positive reciprocity and intentions in trust games. *Journal of Economic Behavior and Organization*, 52(2):267–275, 2003.

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

A. K Mishra, R. Kramer, and T. Tyler. Organizational responses to crisis: The centrality of trust. In *Trust in organizations*, pages 261–287. London: Sage Publications, 1996.

L. Mui, M. Mohtashemi, C. Ang, P. Szolovits, and A. Halberstadt. Ratings in distributed systems: A bayesian approach. 2001.

L. Mui, M. Mohtashemi, and A. Halberstadt. A computational model of trust and reputation. pages 188–196, 2002.

B. Nooteboom. Innovation and inter-firm linkages: new implications for policy. *Research policy*, 28(8):793–805, 1999.

A. Ortmann, J. Fitzgerald, and C. Boeing. Trust, reciprocity, and social history: A re-examination. *Experimental Economics*, 3(1):81–100, 2000.

M. J. Osborne and A. Rubinstein. A course in game theory. *Cambridge/MA*, 1994.

P. A Pavlou and A. Dimoka. The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation. *Information Systems Research*, 17(4):392, 2006.

R. La Porta, F. Lopez-de-Silanes, A. Shleifer, and R. W Vishny. Trust in large organizations. *The American Economic Review*, 87(2):333–338, 1997.

R. La Porta, F. Lopez-de-Silanes, A. Shleifer, and R. W Vishny. Trust in large organizations. *Social Capital: A Multifaceted Perspective*, 1999.

R. D. Putnam. Tuning in, tuning out: the strange disappearance of social capital in america. *PS: Political Science and Politics*, pages 664–683, 1995.

M. Rabin. Incorporating fairness into game theory and economics. *The American Economic Review*, pages 1281–1302, 1993.

V. Radaev. How trust is established in economic relationships when institutions and individuals are not trustworthy (The case of russia). *Collegium*, 22:23, 2002.

H. Rainer and T. Siedler. Does democracy foster trust? *Journal of Comparative Economics*, 37(2):251–269, 2009.

M. Raiser, European Bank for Reconstruction, and Development. *Trust in Transition*. European Bank for Reconstruction and Development, 1999.

F. Roth. The effect of the financial crisis on systemic trust. *Intereconomics*, 44(4):203–208, 2009.

P. Sapienza, A. Toldra, and L. Zingales. Understanding trust. *NBER working paper*, 2007.

R. Sethi and E. Somanathan. Understanding reciprocity. *Journal of Economic Behavior and Organization*, 50(1):1–27, 2003.

J. B. Slemrod and P. Katuscak. Do trust and trustworthiness pay off? *NBER Working Paper*, 2002.

J. M Smith. *Evolution and the Theory of Games*. Cambridge University Press: Cambridge, 1982.

G. Tabellini. Culture and institutions: Economic development in the regions of europe. *Working paper*, 2005.

M. Tanis and T. Postmes. A social identity approach to trust: Interpersonal perception, group membership and trusting behaviour. *European Journal of Social Psychology*, 35(3), 2005.

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

F. Tonkiss. Trust, confidence and economic crisis. *Intereconomics*, 44(4): 196–202, 2009.

R. L Trivers. The evolution of reciprocal altruism. *The Quarterly Review of Biology*, 46(1):35, 1971.

G. Weisbuch, A. Kirman, and D. Herreiner. Market organisation and trading relationships. *The economic journal*, 110(463):411–436, 2000.

A. Whitby, A. J\textbackslashtextbackslashosang, and J. Indulska. Filtering out unfair ratings in bayesian reputation systems. *The Icfain Journal of Management Research*, 4(2):48–64, 2005.

B. Williams. Formal structures and social reality. In D. Gambetta, editor, *Trust: Making and Breaking Cooperative Relations, electronic edition, Department of Sociology, University of Oxford*. 2000.

B. Yu, M. P. Singh, and K. Sycara. Developing trust in large-scale peer-to-peer systems. pages 1–10, 2004.

A. Zaheer, B. McEvily, and V. Perrone. Does trust matter? exploring the effects of interorganizational and interpersonal trust on performance. *Organization Science*, pages 141–159, 1998.

3   A psychological game for economic exchange

Abstract

Market exchanges are often characterised by uncertainty on the reliability interacting parties, as well as by information asymmetries. This is true not just for second hand markets, as in the so-called 'market for lemons', but for most of the economic exchanges. Generally, these problems may arise on both sides of an exchange on matters as product quality, counterparts reliability, reputation, quality of the information about the product or the parties, and so on.

In these contexts, trust has been defined as an important variable to reduce uncertainty and information asymmetries, allowing for economic interactions and exchanges. However, trust has been generally treated, at least since more recent studies, as a rather background element.

In this paper we investigate the effects of trust on agents decisions and interactions. We characterise agents trust in terms of beliefs and evaluate how beliefs may be linked to preferences and the context of interaction. Available information on the counterparts, as well as other elements, may influence agents beliefs about their partners, therefore modifying their perceptions and preferences.

We demonstrate how modeling trust in terms of beliefs and preferences allows to define cooperative equilibria even in one shot settings.

3.1   Introduction

In the last decades, a growing body of literature has started to deal with trust and its economic consequences. The concept was already widely studied in other domains, such as biology, sociology or psychology (Axelrod and Hamilton, 1981, Smith, 1982, Coleman, 1988, 1990, Dasgupta, 2000, Bruni and Sugden, 2000). However, a strong interest emerged also in economics, a domain which had always called upon trust from a qualitative perspective

since then (Akerlof, 1970, Mishra et al., 1996, Zaheer et al., 1998, Dasgupta, 2000, Dirks and Ferrin, 2001, McEvily et al., 2003), but never that much quantitatively approached it.

Nowadays, research shows growing evidence of the importance of trust for economic outcomes, both at the individual and aggregate level (La Porta et al., 1997, Slemrod and Katuscak, 2002, Carter and Castillo, 2002, Guiso et al., 2004, Hoff and Pandey, 2004, Tabellini, 2005, Guiso et al., 2006, Hoff and Pandey, 2006, Guiso et al., 2007, Fehr, 2009).

Moreover, many authors have recently focused on the understanding of the links between trust and beliefs and preferences (McCabe et al., 2003, Cox, 2004, Fehr et al., 2005, King-Casas et al., 2005, Fehr, 2009). As these are also economic primitives, it becomes even more interesting to analyse the effects of trust in economic interactions.

In this paper we address the relevance of trust as a mean to solve market failures arising from Prisoner's dilemma like situations. In particular, we give evidence of how trust dynamics may influence agents decision processes, as to allow exchanges. Our focus is primarily on the belief component of trust. At the same time, we consider endogenous preferences determined by the beliefs structure itself.

3.2 Trust in market exchanges

Trust is the ever-ready lubricant of economic activity, as it allows for participation to production and exchange (Dasgupta, 2000). Trust is indeed particularly relevant in exchanges characterised by uncertainty, asymmetric information, risk. Luhmann (2000) highlights that a system where trust lacks may even shrink below a critical threshold level necessary for its own reproduction at a certain level of development.

Market exchanges are often characterised by uncertainty on the reliability of both parties, as well as by information asymmetries. This is true not just for second hand markets, as in the so-called 'market for lemons' (Akerlof,

1970), but for most of the economic exchanges. Generally, these problems may arise on both sides of an exchange on matters as product quality, reliability of the counterpart, reputation, quality of the information about the product or the parties, and so on.

In the 'market for lemons' information asymmetry leads to market failure. Reproduction of the system is therefore impossible and economic activity defaults. However, in real life many economic exchanges go on even among perfect strangers or anonymous counterparts, like most frequently happens for e-commerce sites, but more generally when one enters a shop for the first time and buys something.

It has to be observed that most of the uncertainty resides on the buyer's side of the market. The biggest problem on the seller side is about the reliability of the buyer for the payment. However, this is often solved by credit card payments, or a proof of the feasibility of the payment via some account check. Internet exchanges for instance are solved this way; room reservations at hotels typically require a credit card number for confirmation, and so on.

On the buyer side, on the contrary, most of the uncertainty still remains. Firstly one needs to assess the quality of the seller, in his reliability and in the information quality about the products. Secondly, the quality of the product itself needs to be assessed, and this is not easy, mostly for complex or technologically advanced products. Therefore, a buyer needs to rely on some means to solve this uncertainty, or to reduce it.

Many researchers, starting from Akerlof (1970), stress solutions related to trust and reputational mechanisms. Most of the times these solutions are simply approached and presented in informal ways. In the last decades, however, the diffusion of internet exchanges has caused an increased interest from scholars and more quantitative and formalised models have emerged (Marimon et al., 1999, Braynov and Sandholm, 2002).

For instance, computer scientists have focused on information asymme-

tries on internet markets and analysed the systems that allow for reputational mechanisms and trust in e-commerce platforms (Mui et al., 2002, Yu et al., 2004, Whitby et al., 2005, Huynh et al., 2006). Referral systems, as the ones provided for the evaluation and scoring of sellers on e-commerce websites, are of great importance for the creation of incentive schemes allowing for the exclusion of bad sellers and for higher returns to reputation for the best ones (Ba and Pavlou, 2002, Pavlou and Dimoka, 2006).

These studies cannot account for real market transactions where evaluations or scores are more frequently not available. Nonetheless, they give evidence of how some belief or evaluation of a counterpart's trustworthiness may be obtained, through third party comments and scores. This also poses a problem on the reliability of the system in detecting untrustworthy evaluating behaviours (Yu et al., 2004, Whitby et al., 2005).

It has to be underlined that reputation and trust are not the only possible solutions to market failures deriving from asymmetries. Trial and error and learning processes based on past experiences may lead to stable exchange relationships based on product availability (Kirman and Vriend, 2000, Weisbuch et al., 2000). In these accounts, however, the focus is more on the simultaneous emergence of stable and erratic exchange habits, than on the reasons underlying the emergence of the exchanges themselves.

In the proposed analysis we present a model where the belief component of trust is the primary focus of the exchange. This is in line with recent research in many different domains, highlighting the relevance of beliefs and preferences for trust dynamics and consequent cooperative behaviours (Cox, 2004, Fehr et al., 2005, Kosfeld et al., 2005, Fehr, 2009).

Moreover, we are interested in the effect of beliefs on the preferences structure of agents. Different informations and different expectations may lead agents to hold different preferences towards similar exchanges. Preferences may therefore vary depending on beliefs and available information (King-Casas et al., 2005).

Research also suggests that agents derive psychological satisfaction from trusting and being trustworthy (Ashraf et al., 2006). It appears therefore plausible to consider a psychological component in agents payoff, while modeling an exchange game based on trust dynamics.

Based on these research results we build a simple model of buyer-seller interaction, where beliefs about trust and trustworthiness play a conditioning role in determining the outcomes for each player and the interaction itself.

The rest of the article is structured as follows. In the next section we introduce the model and give some examples of the context to which it may adapt. In section 4 we present the setting of psychological games, and discuss the reasons leading to the choice of this setting for our modeling purposes. We also briefly highlight the differences towards more classical approaches. In section 5 we give a formalisation of the model and analyse the possible solutions in the simultaneous and sequential settings. Finally, we conclude in section 6.

3.3 A model of economic exchange based on beliefs

We model a simple game depicting an exchange among a buyer and a seller . The latter owns private information about the product(s) he is selling, while the former has no evidence on the seller's reliability and product quality. The buyer needs to infer somehow both these hidden characteristics.

The game we design is inspired to the so-called Trust game (Berg et al., 1995), that has strongly influenced research on trust and cooperation. Generally speaking, we consider a Prisoner's dilemma like situation. However we modify it to analyse the effects of agents beliefs on the equilibrium outcomes of the game, following the literature on psychological games (Geanakoplos et al., 1989, Rabin, 1993).

Agents beliefs are considered in a way that enhances or reduces players utilities. Instead of considering pure monetary payoffs, agents receive utility also from their trust and trustworthiness. This picture is in line with the

idea that agents preferences may include more than material satisfaction (Andreoni, 1995, Fehr, 2009).

In the game we present, a buyer has to decide whether to trust a seller, that is if to interact and buy some products from him. This happen if he perceives the seller as reliable. The seller, on the other hand, has to decide whether to be trustworthy or not. We define a seller trustworthy, i.e., if he is selling the product at a price reflecting its real quality and characteristics and he is giving a good sale service. He decides to be trustworthy based on his perception of buyer's intentions in terms of placed trust.

In this game setting, beliefs play a major role in determining agents choices. Moreover, the choice of a psychological game as tool for modeling these dynamics, allows for endogenised preferences and outcomes. Therefore, beliefs determine each agent's outcomes and, consequently, his preferences and best strategies. We will return in more detail on these points in the next sections.

We consider both a simultaneous and a sequential version of the game. In the first, the buyer and the seller have to decide simultaneously whether to trust (the former), and whether to cooperate trustworthily (the latter). As an example, we may imagine that the seller is a private owning a car and willing to sell it on the second hand market. This is the typical situation of the so-called 'market for lemons' (Akerlof, 1970).

In the second setting, decisions are on the contrary sequential. This is probably the most typical setting in trust interactions. The buyer decides whether to interact or not, and the seller has the strategic advantage of knowing buyer's decision before choosing his strategy. To give an example, in this setting the seller has more pieces of the same (at least apparently) product to put on the market. He knows that some are good, while others are not (i.e. some are new pieces, and some used ones; or, some are original products and some fake ones). Another example can be the quality of the assistance, when a customer care service is sold. The buyer has not enough

information to discriminate among the different pieces on sale or cannot be sure in advance of the quality of the service. Where the seller receives buyer's trust, he decides if to conclude the exchange trustworthily or not.

If the seller is untrustworthy, he receives the payment and ships a bad product/service. The buyer will have less than his initial endowment, due to the payment he made. On the contrary, if the seller is trustworthy the buyer receives the good product/service (the material payoff) and a psychological reward from being reciprocated and having fruitfully placed his trust. He will have paid for the product, but he will be better off in terms of his utility, that will be increased with respect to the initial situation. The seller, on his side, will get the material payoff from the payment, sustain a cost in terms of sale service (i.e., customer care service) and receive a psychological reward from reciprocating and from being trusted.

At least two points need to be stressed. Firstly, the seller too has a decision to make about trust. In fact he has to trust that the buyer is going to completely and reliably pay the product he is buying. This aspect can be relevant in some market transactions. However, at least in some cases, the payment, or at least a proof of the reliability and feasibility of it, is made before the exchange is made. Internet exchanges are paid with credit cards; hotels room reservations are made giving credit card details, and so on.

Without too much simplification, we hereafter assume that the buyer is always able to pay for the product he's getting. In real world many contractual clauses and legal protections enforce the security of the seller. Nonetheless, we may consider that if the seller ships the product to an unreliable buyer, still he is loosing chances to sell the product to other potential buyers in the meantime. He is also potentially loosing money in terms of time and shipping expenses, in example.

However, we may note, the simultaneous version of the game presents this problem, although not explicitly treated. The sequential game too, with some incomplete information about buyer's trust, can also be transformed

to account for trust issues on both sides. This can be done considering a different beliefs structure for the players.

Secondly, we highlight that the game develops on the basis of two different psychological attitudes and distinct preferences. One is trust, the other is trustworthiness. We believe no simple analysis of trust is possible without a simultaneous analysis of trustworthiness. Some research on trust try to disentangle them, but most of it examines the two concepts jointly (Hardin, 2002, James, 2002, Radaev, 2002, Ashraf et al., 2006).

We believe that an agent entering an interaction on the basis of trust, is expecting a reciprocating partner on the other side. This is in line with research showing that people who trust more are generally more trustworthy (Good, 2000), and that answer to survey questions about trust are frequently based on personal experience and own trustworthiness (Glaeser et al., 2000, Holm and Danielson, 2005, Sapienza et al., 2007).

Moreover, trust is not about benevolence, but about solving a problem of uncertainty with the expectation of a positive response from the counterpart leading to a positive outcome (Mayer et al., 1995, Bhattacharya et al., 1998, Hardin, 2002). This is exactly why the trustee takes into account the trustworthiness of his counterpart, and why the game we propose is jointly based on the two concepts.

To conclude this section, we underline that the model may even apply, with or without specific extensions, to other situations. In first place, the game allows a straightforward extension to interactions involving two firms or two organisations, as well as an agent and an organisation.

Moreover, to our view the game may depict exchanges on financial markets. This is of some interest also considering the recent financial crisis. Many stressed the lack of trust on markets after the financial collapse. News, magazines and chiefs of national and international institutions claimed that trust was and still is a big issue for the recovery of the global economy. Academic research is starting to focus on this problem (Roth, 2009, Tonkiss, 2009).

Customers typically refer to a bank or a financial institution to deposit their money and eventually make some lucrative investment. For many of these customers, the financial institution is the primary counselling service for any decision of investment, and they expect positive interests on their investments.

While adequate information policies may be provided as to make customers aware of the inherent risks of every investment product, it cannot be excluded that some investment decisions are made on the basis of too little information or understanding. In addition, information may or may not be correct and transparent. One issue involving trust is then the reliability of the bank as counselling service in investment products.

Moreover, customers expect their investments to have a positive return, even without sometimes considering market contingencies and the risks and uncertainties inherent to the products they invest in. If market fails due to some problem or failure external to the financial institution itself, one should still consider the bank reliable and therefore not change his relationship with it.

While, of course, some institutions may be taken as directly responsible for their failures, some others may not be. But despite this piece of evidence, customers may not be able to discriminate and therefore trust level is lowered all over the financial market. Beliefs and expectations may play a significant role in these dynamics, both in their emergence and in the possible solutions.

After giving a general frame for our model and some possible application examples, we now turn to the description of psychological games setting, showing how one may theoretically account for those beliefs and expectations in a game theory approach. Afterwords, as said, we motivate the choice of this instrument, before moving to the formalisation of the proposed game.

3.4 Psychological games

A psychological game is one in which the payoffs of each player depend not only on what every player does, but also on his beliefs about what he thinks other players believe, what he thinks other players believe he believes, and so on (Geanakoplos et al., 1989, Rabin, 1993). The hierarchy of the beliefs may be infinite as it is the case for the definition of common knowledge in standard economic theory. The payoffs of such psychological games are endogenous, as well as strategies. In this sense, payoffs are influenced not only by the material component, but also by agents expectations and by the interpretation of what goes on in the deployment of the game.

As Geanakoplos et al. (1989) underline, the principal distinguishing feature of a psychological game is that it allows for payoffs being dependent not only on actions, but also on the hierarchy of beliefs held by every agent. Belief are probability measures over the product of other's strategies space. In equilibrium these beliefs are required to correspond to reality. Although the authors demonstrate that backward induction and some equilibrium concepts may not apply to this kind of games, they also show the existence of solution concepts, and that much of equilibrium theory can be maintained to explain this game setting. We report here the description of a normal form setting. However the considerations apply to the extensive form too.

Let N=$\{1, \ldots, n\}$ be the set of players, and for each $i \in N$ let A_i be the nonempty, finite set of actions available to player i. For any set X, $\Delta(X)$ denotes the set of (Borel) probability measures on X. Thus, $\Sigma_i = \Delta(A_i)$ is the set of mixed strategies of player i. Let $\Sigma = \times_{i \in N} \Sigma_i$ and $\Sigma_{-i} = \times_{j \neq i \in N} \Sigma_j$, $i \in N$. Each strategy profile $\sigma \in \Sigma$ induces a probability distribution P_σ over the outcome set $A = \times_{i \in N}$.

First order beliefs are defined as probability measures over the product of other players mixed strategy sets. Thus for player i, the first order belief is $B_i^1 = \Delta(\Sigma_i)$. Let $B_{-i}^1 = \times_{j \neq i} B_j^1$ and $B^1 = \times_{i \in N} B_i^1$. Endow B_i^1 with the weak topology. Since Σ_{-i} is a subset of a Euclidean space, it is a separable

metric space, and therefore B_i^1 is a separable metric space. The sets of higher order beliefs are defined for every order $k \geq 1$ by

$$B_i^{k+1} = \Delta \left(\Sigma_{-i} \times B_{-i}^1 \times \ldots \times B_{-i}^k \right),$$

$$B_i^{k+1} = \times_{j \neq i} B_j^{k+1},$$

$$B^{k+1} = \times_{i \in N} B_i^k.$$

This structure is common to games with imperfect information. Each information enters many times in the belief hierarchy. In this sense, this allows for a correlation among beliefs about others actions and beliefs about others first order beliefs. Unless beliefs are nonsensical, and since each player knows that the other players are also rational, he should not believe them to have incoherent beliefs. Coherency of beliefs is assumed to be common knowledge.

Typically in a psychological game players utility depends on the outcomes and also on beliefs. It is assumed that players seek to maximise expected utility $\bar{u}_i$. The utility $u_i(b, \sigma)$ is the payoff to player i if he believed b_i and found out that σ was actually played. No requirement is made that payoffs be linear in the beliefs. Beliefs may reflect disagreement of players over various aspects, but in equilibrium all beliefs are assumed to conform to some common view of the reality.

It follows that a psychological game $G = (A_1, \ldots, A_n; u_1, \ldots, u_n)$ in its normal form consists of a set of actions and a utility function $u_i = \bar{B}_i \times \Sigma \to \mathbb{R}$ for each player i.

An equilibrium of the game is defined as a psychological Nash equilibrium of a normal form psychological game if a pair $(b, \sigma) \in \bar{B} \times \Sigma$ satisfies

(i) $\hat{b} = \beta(\sigma)$

(ii) for each $i \in N$ and $\sigma \in \Sigma_i$, $u_i \left(\hat{b}_i, (\sigma_i, \hat{\sigma}_{-i}) \right) \leq u_i \left(\hat{b}_i, \hat{\sigma} \right).$

Geanakoplos et al. [1989] proof the existence of psychological Nash equilibrium for every psychological game under the assumption of the continuity of the utility function.

3.4.1 Why a psychological game

This section addresses the relevance of psychological games as a tool to model trust issues. The closer and more similar approach is that of Bayesian games. It is important then to underline why a psychological game approach is chosen instead of the Bayesian one.

A Bayesian game is one in which at least one of the players is uncertain about the characteristics, or some of them, of the other parties (Osborne and Rubinstein, 1994). Typically the game is modelled introducing a set of possible states of nature, each of which is a description of all the players relevant characteristics, with the set of states of nature being finite. Each player then assigns a probability measure, that is his prior belief, to every state of nature.

As a signal arrives to the player i, the player has the relevant information about the state of nature that has realised before he takes his action. Uncertainty however remains over the realised state. Since the signal may be imperfect, the player needs to choose his action based on the consideration of the possible action that the other parties would take also in other relevant states of nature. This belief may depend also on the action that the player himself would choose in other states, since the other player may also be imperfectly informed.

However, in trust dynamics not necessarily an agents observes a signal on which to base his belief about the counterpart. Indeed, his beliefs are sometimes based on previous experiences, even in totally different contexts of interaction and with different counterparts (Glaeser et al., 2000, Sapienza et al., 2007).

As defined above, a psychological game is one in which the payoffs them-

selves depend on agents beliefs. Payoffs, and strategies, become therefore endogenous and are influenced by the hierarchy of beliefs of every agent. In this sense, the game itself changes as the structure of beliefs change. There is no need to involve nature in this setting.

In a Bayesian game players are uncertain about the characteristics of their counterparts, and they assign a probability to every type of agent they may face, based on their beliefs or on observed signals. Payoffs are defined once and for all at the beginning of the game and they don't change along with the structure of agents beliefs. In Harsanyi's approach, beliefs are exogenous, and players have complete information on the utility payoff of their opponents, and the probability of facing each of these opponents.

In the psychological setting, on the contrary, agents hold internal or endogenous beliefs about other agents. These beliefs are not means to solve the uncertainty about which type of agent one is confronted to, but they define directly the payoffs that every agent has to consider in the choice of the best strategy. This is because beliefs enter the payoff functions directly, endogenising preferences and outcomes.

In fact, not only beliefs change the payoff structure, but as the game unfolds the preferences structure is also modified. While in other approaches the implementation of expectations don't change the preference-ordering of the strategies, the approach here adopted depicts the situation where beliefs modify this ordering. Instead of being fixed before the interaction, preferences in a psychological game are themselves defined by the structure of the game. Expectations of other players, and the expectation of oneself about others expectations (and so on...), influence players preferences on what to do and what to think.
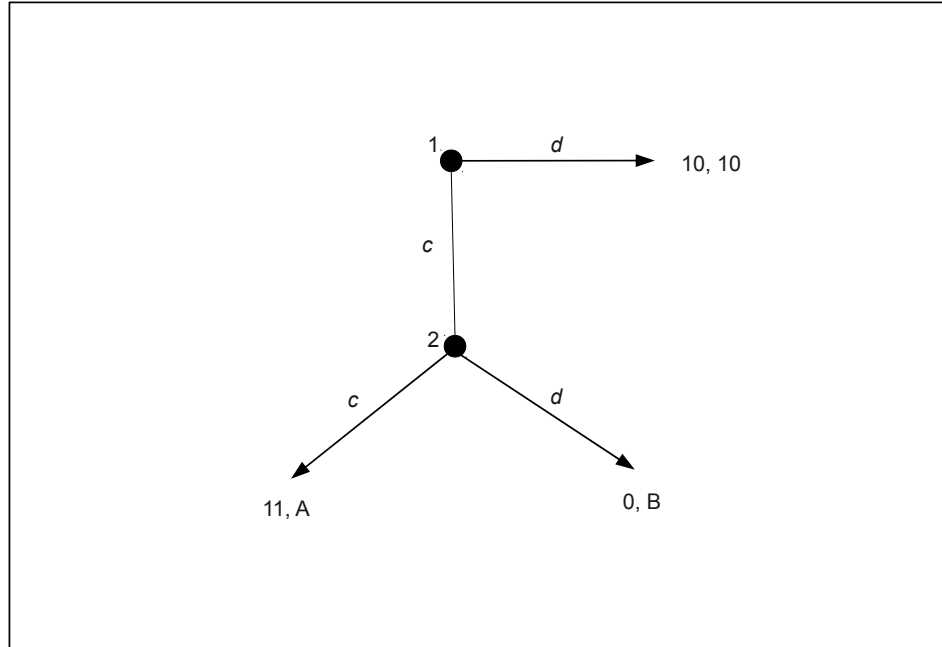
Figure 2: Psychological game: an example

In the psychological game example given in Figure 2, agent two's payoffs depend on his beliefs about agent one's decisions. Indeed, if he believes that agent one is going to play d, but then he found that c is played, his payoffs correspond to $A = 0$, $B = 2$. His preference in this occurrence is to punish player one for choosing action c, instead of the expected d. On the contrary, if agent two expects agent one to choose c, then his payoffs are $A = 5$, $B = 1$, and therefore his preference will be such that he will choose also c.

For instance, in a cooperation game an agent may hold cooperative preferences for specific expectations about others beliefs and strategies, while, for different levels of these variables, he holds uncooperative preferences. In turn, this defines a different game every time, as the game itself unfolds, by endogenising the effect of beliefs and preferences on outcomes and strategies.

We may give here another short example extracted from Heap and Varoufakis (2004) to confront the different approaches. In Figure 3a we report a normal version of a Prisoner's dilemma game. In Figure 3b we present instead a psychological version of the same game.

	c	d
c	2, 2	0, 3
d	3, 0	1, 1

(a) Normal version

	c	d
c	3-2q, 3-2r	1-2q, 2+2r
d	2+2q, 1-2r	2q, 2r

(b) Psychological version

Figure 3: Prisoner's dilemma

In the example only a second order belief is considered for each player. No first order beliefs enter the game. Differently from the previous example, both players are here influenced by their expectations about their counterparts. For instance, q and r are respectively first and second player's belief about their respective counterpart's belief about themselves choosing strategy c (cooperate). The psychological component of payoffs is here symmetric, but of course it can also be differently. In particular, in this setting players get psychological rewards from satisfying their counterparts expectations, suffering instead psychological disutility in the opposite occurrence.

We should note that for the solution of this game we need to specify q and r. However, we also note that the first game allows for only one Nash equilibrium, while in the second, we may find two of these equilibria, depending on assigned valued for q and r (in this game equilibria are found for $(q, r) = (0, 0)$ and $(1, 1)$).

What we need to impose to solve the game is consistency among the beliefs of every player, and these beliefs are required to be the equilibrium ones, in the sense of being the realised ones. Nonetheless, one can make wider considerations about the role of beliefs on agents preferences and outcomes if not interested in equilibrium itself.

The choice of this type of representation is not only due to the uncertainty linked to the trust problem. More than this, the game above shows that this approach helps define how every agent considers his payoffs to depend on the perception he has of others, on the perception he has of others perception of himself, and so on. The hierarchy of beliefs can be extended as far as needed, as in the common knowledge hypothesis.

In example, with a second order belief one considers how his own attitude, as perceived by the counterpart, may influence his own outcomes. It is not therefore an existing type of agent who is going to be the opponent, which would allow for Harsanyi's solution of assigned probabilities over the distribution of the types in the population. On the contrary, each agent strategy (or type) is decided with respect to endogenous beliefs and change based on these beliefs.

Moreover, the structure of payoffs so defined explicitly addresses the problem of psychological rewards in the trust relationship, where agents are assumed not only to get monetary satisfaction from the relationship, but also emotional rewards depending on their own and their counterparts actions (Andreoni, 1995, Fehr, 2009). This is more so where one's beliefs are confirmed after action took place. This creates a sense of satisfaction and confirms one's ability to cope with similar situations and act in the social environment (Good, 2000).

One may stress that a transformation of players utilities may allow for a similar result. In fact, one may assume that agents are somehow motivated by ethical or moral motives in the choice of their strategies, and that this motivations may change the utility received. Normally this approach has

been used to modify the payoff structure before the game takes place, in such a way to reflect players new motivations and their new utility functions.

Anyhow, this approach defines a new game, instead of confronting with the initial one, as preferences are set once and for all and the outcomes are consequently modified before action takes place.

On the contrary, psychological games allow for the modification of the payoffs as part of the game itself, endogenising the process. It is motivation itself, and the perceived beliefs about players motivation, that modify agents preferences and outcomes as the game takes place. An agent's psychology is in this setting not just dependent on an agent's moral or ethical preferences, but also on others expectations, and viceversa.

Therefore, agents don't know in advance their outcomes, but have to observe how events unfold and to confront actually played strategies with the ones expected to be played. Only at this point they identify if beliefs and reality were correctly aligned, and can evaluate their own psychological satisfaction or disappointment. In turn, this modifies actual payoffs.

We turn hereafter to the formalisation of our model.

3.5   The model

We consider two agents $i, j \in \{1, 2\}$. Player one has to choose whether to make an exchange with player two. In example, player one has to buy a product from player two without knowing all the relevant seller and product characteristics, in terms of reliability and quality. The interaction is therefore asymmetrical. This, in turn, requires player one to trust his counterpart with respect to eventual informations he may give about the product, and, most importantly, with respect to the correct price/quality ratio and service quality.

The first mover has an action space defined by two actions, that is $A_1 = \{trust, mistrust\}$. He chooses *trust* with probability p, and *mistrust* with probability $1 - p$. Player two action space is defined as $A_2 =$

$\{reciprocate, defect\}$, where *reciprocate* is chosen with probability r, while *defect* with probability $1 - r$.

Turning to the payoffs structure, outcomes are partly monetary and partly depending on psychological satisfaction or disappointment deriving from trust and trustworthiness perceptions and expectations. In particular, agents evaluate their monetary payoffs as a function of their beliefs about the other agent, and of their belief about other agent's belief on them. Therefore, we consider first and second order beliefs as part of the payoff function for each player.

We believe considering first and second order beliefs is sufficiently realistic in terms of the computational effort for involved agents. However, the literature on psychological games deepens the theoretical approach allowing for higher order beliefs to enter the payoff functions.

Beliefs are defined as probability measures. We define s as player one's belief on player two's action *reciprocate*, that is his belief on r. It follows that the second order belief, that is $\tilde{s}$, is the belief of player two on player one's belief about two playing *reciprocate*. Symmetrically, we define q as the first order belief about player one choosing *trust* among his possible actions in the set A_1, and this is player two's belief on p. The second order belief, that is $\tilde{q}$, is the belief of player one on player two's belief about one choosing *trust*. The complements to one of the defined beliefs, are themselves defining first and second order beliefs about the actions *mistrust* and *defect*.

We resume the beliefs specification in Figure 4, as they enter the payoff functions of the players.

Strategies	TRUST	MISTRUST	RECIPROCATE	DEFECT
Real probabilities	p	$1 - p$	r	$1 - r$
First order belief	q	$1 - q$	s	$1 - s$
Second order belief	$\tilde{q}$	$1 - \tilde{q}$	$\tilde{s}$	$1 - \tilde{s}$

Figure 4: Beliefs specification

As said, payoffs are function of an agent's first order belief about the other player actions, and second order belief of the other player on his actions. In example, player one, that is the trustor, consider his payoff as a function of a material payoff and of some other function of his belief about player two's trustworthiness, and of his belief about the belief of player two on his choice to trust.

We assume that player one has an initial endowment, and must choose whether to spend his money or keep them. The spent amount is sent to player two in exchange for the product he decided to buy. On the other side of the exchange, the seller receives the payment and decides to reciprocate buyer's trust, sending a good product, or to be untrustworthy and send a bad product. We enrich this basic setting defining payoff functions that depend on the previously proposed beliefs.

If player one mistrust, then he gets a payoff equal to his initial endowment, that is X. In this situation, player two, if he was holding a belief on player one choosing *trust*, may be disappointed by player one choice. His payoff then is equal to $-\varphi(q, \tilde{s})$. This function is increasing in absolute value as q and s increase, that is the disappointment is higher where the belief on player one choosing *trust* is higher, and when he strongly believes the other one thinks he will reciprocate. However, we may also think to a payoff equal to zero for the seller in this occurrence, and the analysis would remain unchanged.

If player one chooses to trust, then he sends money for the product. How much he is open to spend with a specific seller depends on his belief about player two's propensity to reciprocate, and on his belief about player two's belief on he choosing to trust. Then the amount sent is specified as a function $\alpha(s, \tilde{q})$.

Player two receives the payment sent by player one. The amount received if player one trusts is γ, that we assume is the price of the sold product. However, two things may happen at this point: either the seller ships a good

product and therefore the buyer has at least the same level of utility as with his initial endowment; or, the seller is untrustworthy and ships a bad product or gives a bad service to the buyer, leaving him with a lower utility and unsatisfied.

In case he defects, the seller simply receives the payment for the sold product and sends to the buyer a bad product, which cost we consider negligible. If he reciprocates, he sends back a good product and gives to the buyer a good sale service. This will have a cost for the seller that is expressed by $\delta\left(q, \tilde{s}\right)$, and is also increasing in both q and $\tilde{s}$.

Moreover, since player two gets satisfaction from being trusted and from being held as trustworthy, then he will also have a psychological reward from receiving one's trust and from being right in his belief about one's belief of his trustworthiness. The function defining the level of satisfaction is $\varphi\left(q, \tilde{s}\right)$. The seller is therefore satisfied by having a satisfied and trusting customer.

On the other side, if seller is untrustworthy, player one may then end up with less money and a bad product, worth nothing. If we was believing two to reciprocate, he will also have more disutility, $\eta\left(s\right)$, from recognising he was holding a wrong belief. If, instead, the seller is trustworthy, the buyer will receive a good product and a valuable service. He will evaluate this service to be a function of his belief about player two's trustworthiness, and of his belief on player two's belief about his own trust. This function is defined as $\beta\left(s, \tilde{q}\right)$. For consistency and coherency among players beliefs, also this function is increasing in both the parameters.

Similarly to the case of player two being trusted, we can define a function $\eta\left(s\right)$, that is player one gets an increased payoff when he was expecting reciprocation and he actually is reciprocated. And, as one may expect, the same function is decreasing player one's payoff when player two defects.

We resume the structure of the game and the payoffs in the next figure (see Figure 5).

We also resume the characteristics of the payoff functions hereafter. The

functions that enter the payoffs are always increasing in both the first order and second order beliefs. They get a maximum value for both the first and the second order belief being equal to one.

For player one this means that he will invest all his resources when $s = \tilde{q} = 1$. At the same time he will expect the maximum reciprocation from player two, that we assume to be a good product and a good service, that is more than what he paid for in terms of personal satisfaction. For player two, on the other side, when $q = \tilde{s} = 1$, the expectation is to get from player one all his endowment. At the same time he will ship a good product with the best service. For the satisfaction functions, the maximum is reached when $s = 1$ for player one, and when $q = \tilde{s} = 1$ for player two. On the opposite, of course, all these functions get their minimum when the beliefs of first and second order are equal to zero.
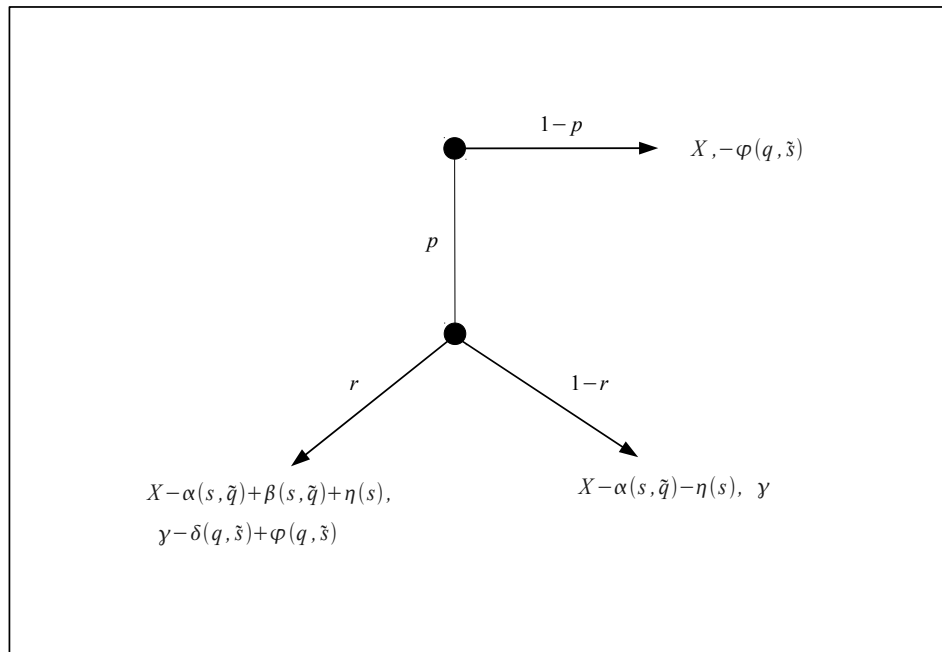

Figure 5: Game structure

We also assume that the functions are continuous and that they are increasing in their respective parameters, that is:

$\frac{\partial \alpha(s,\tilde{q})}{\partial s} \geq 0$, $\frac{\partial \alpha(s,\tilde{q})}{\partial \tilde{q}} \geq 0$, $\frac{\partial \beta(s,\tilde{q})}{\partial s} \geq 0$, $\frac{\partial \beta(s,\tilde{q})}{\partial \tilde{q}} \geq 0$ and $\frac{\partial \eta(s)}{\partial s} \geq 0$

$\frac{\partial \delta(q,\tilde{s})}{\partial q} \geq 0$, $\frac{\partial \delta(q,\tilde{s})}{\partial \tilde{s}} \geq 0$ and $\frac{\partial \varphi(q)}{\partial q} \geq 0$

We now turn to the equilibrium analysis to investigate whether the presented game has one, or more, equilibrium, and to eventually define their properties.

3.5.1 Equilibrium analysis

In the game described before players have payoffs defined by their beliefs about their counterpart behaviour and beliefs. This means that when beliefs change, the structure of the game may reveal itself different, and we cannot define a real equilibrium without investigating players expectations and perceptions.

As in the previous literature we assume that in equilibrium all beliefs are equal, that means each player must hold in equilibrium the same beliefs as the others. In example, it is not possible in equilibrium for player two to believe that player one believes he will reciprocate, and for player one to hold, at the same time, the belief that player two will not reciprocate. This of course has also a reason in the fact that, for example in the situation depicted, when the game reaches an equilibrium, then all players must hold the same knowledge about what happened, and therefore must hold same beliefs too.

This also means that in equilibrium

$\alpha(s,\tilde{q}) = \gamma$

$\beta(s,\tilde{q}) = \delta(q,\tilde{s})$

However, one can imagine some noise in the beliefs, so that first and second order beliefs are not totally aligned. In this case, even for little differences, it may happen that equilibrium is not reached, and even if the optimal solution would have been possible, it is not obtained. If this happens, i.e. player one trusts holding a sufficient belief on player two to reciprocate, but then two is not reciprocating, this may affect subsequent games among the players. Therefore, one can imagine a learning process in which every played game makes agents learn more about their counterparts, adapting their beliefs and expectations on the basis of their previous experience.

To make another example, we can imagine that player two is sure that player one will trust him. But, player one chooses to mistrust, that is not to invest. At that point, player two will change his belief about player one, but will also feel disappointed from his behaviour. In subsequent games it may well be the case that the psychological payoffs are changed, due to the previous negative experience, and then player two may possibly choose to punish player one defecting.

These learning processes, that may lead to a durable relationship or may undermine trust, will be examined in a future research, to analyse not only the effects of learning from past interactions, but also the consequences of being placed in a social environment, where learning goes on also on the basis of third party's experiences.

We may turn now to the analysis of the equilibrium.

Looking at player one we may see that we may see that he is indifferent between *trust* and *mistrust* when

$$X = s\left(X - \alpha\left(s,\,\tilde{q}\right) + \beta\left(s,\,\tilde{q}\right) + \eta\left(s\right)\right) + \left(1 - s\right)\left(X - \alpha\left(s,\,\tilde{q}\right) - \eta\left(s\right)\right)$$

that is when

$$s^*\left(\tilde{q}\right) = \frac{\alpha(s,\tilde{q}) + \eta(s)}{\beta(s,\tilde{q}) + 2\eta(s)}$$

where s is defined in its implicit form.

Turning now to player two we can see that for him to be indifferent between *reciprocate* and *defect* it may be the case where

$$\gamma - \delta\left(q,\,\tilde{s}\right) + \varphi\left(q,\,\tilde{s}\right) = \gamma$$

This is only true when

$$\delta\left(q,\,\tilde{s}\right) = \varphi\left(q,\,\tilde{s}\right)$$

that is the psychological reward from reciprocating is large enough to compensate the for a good sale service.

We analyse two distinct versions of the game, a simultaneous setting and a sequential one. In fact this may make some difference in the relevance beliefs have in the determination of outcomes.

From a qualitative point of view a typical situation involving trust is one where agents choose sequentially. One agent decides if and how much to trust his counterpart, while the trustee decides after observing agent one's action whether to reciprocate or defect. On the contrary, the simultaneous version presents some higher complication, due to the fact that player two also bases his decisions on perceived aspects of his counterpart.

We move then to the definition of the equilibria in the two settings.

Simultaneous game In the simultaneous game players both decide at the same time what action to take. We may then define three different situations and the relative sub-cases.

1. If $s < s^*$ or there is no s^*, then player one always mistrusts. This means that in equilibrium $q = \tilde{q} = p = 0$. Player one will never enter the game and no exchange is made. Payoffs are equal to $(X,\,\varphi\left(q,\,\tilde{s}\right))$. The more player two was expecting player one to trust or to believe in his trustworthiness, the lower his payoff, since he will be more disappointed.

104

2. If $s > s^*$ player one enters the game and for player two we define three sub-cases:

   (a) $q(\tilde{s}) < q^*$, that is player two will choose not to reciprocate;

   (b) $q(\tilde{s}) = q^*$, that is player two is going to play a mixed strategy with probability r;

   (c) $q(\tilde{s}) > q^*$, that is player two reciprocates. In this situation we have an equilibrium where $q = \tilde{q} = p = 1$ and $s = \tilde{s} = r = 1$.

3. If $s = s^*$ then player one plays a mixed strategy with probability p, while for player two we define again three sub-cases:

   (a) $q(\tilde{s}) < q^*$, that is player two will choose not to reciprocate;

   (b) $q(\tilde{s}) = q^*$, that is player two is also going to play a mixed strategy with probability r;

   (c) $q(\tilde{s}) > q^*$, that is player two reciprocates when trusted.

Now, for equilibrium to be stable we need to verify that beliefs are consistent and aligned to the realised values. We need therefore to exclude all the situations in which beliefs are not aligned and consistent among players, as those are not equilibria.

The problem arises when player one trusts, which implies, in equilibrium, that $q = \tilde{q} = p = 1$. Some considerations need now be taken into account.

- If a q^* does not exist, then player two would defect. Since in this situation $s = \tilde{s} = r = 0$, this is a possible equilibrium only if $s^* = 0$. This however is an anomalous situation in which player one always decides to trust, because the expected payoff from cooperation is always equal to the payoff from defection. We may note that this occurrence would mean that no higher welfare is at stake from risk taking and cooperation, therefore implying that trust is an irrelevant matter in

this situation. This leads us to the exclusion of a trust equilibrium such as that in 2.a and, for similar reasons, that in 3.a.

- If instead there exist a q^*, then $q(\tilde{s}) = 1 \geqslant q^*$ for hypothesis. This again subdivides in two cases:

  - if $q^* < 1$, player two will choose to reciprocate, therefore $q = \tilde{q} = p = 1$ and $s = \tilde{s} = r = 1$;

  - if $q^* = 1$ player two will play a mixed strategy with probability $\bar{r}$. However, the mixed strategy is a possible equilibrium only if $s = \tilde{s} = r = \bar{r} > s^*$, which, again, is the condition for player one to trust. This better specification is requested for 2.b and 3.b to be possible equilibria.

- Finally, the situations under 3. need better specification. In fact, if player one plays a mixed strategy with probability $\bar{p}$, then in equilibrium $q = \tilde{q} = p = \bar{p}$. The above considerations apply to the sub-cases, but q^* need exist for an equilibrium to be relevant, and $q^* \geqslant \bar{p}$ for an assessment of the conditions under which an equilibrium is possible.

Sequential game   In the sequential game we assume that player one chooses in the first step, while player two observes first mover's choice and decides what to do based on this information. As underlined above, in this situation the game results simplified and we find less cases.

In particular:

1. If $s < s^*$ or there is no s^*, then player one always mistrusts. This means that in equilibrium $q = \tilde{q} = p = 0$. Player one will never enter the game and no exchange is made. Payoffs are equal to $(X, \varphi(q, \tilde{s}))$. The more player two was expecting player one to trust or to believe in his trustworthiness, the lower his payoff, since he will be more disappointed.

106

2. If $s > s^*$ then player one always trusts, while for player two we define two sub-cases:

   (a) $q(\tilde{s}) = q^*$, that is player two is going to play a mixed strategy with probability r.

   (b) $q(\tilde{s}) > q^*$, that is player two reciprocates. In this situation we have an equilibrium where $q = \tilde{q} = p = 1$ and $s = \tilde{s} = r = 1$.

3. If $s = s^*$ then player one plays a mixed strategy with probability p, while for player two we define the following strategies:

   (a) $q(\tilde{s}) = q^*$, that is player two is going to play a mixed strategy with probability r.

   (b) $q(\tilde{s}) > q^*$, that is player two reciprocates. In this situation we have an equilibrium where $q = \tilde{q} = p = 1$ and $s = \tilde{s} = r = 1$.

The same considerations we made in previous section also need to be made for the sequential version. Therefore, equilibria under 2.a and 3. need better specifications to assess whether they are possible equilibria or they may lead to misaligned inconsistent beliefs, implying they are no longer equilibria.

However, we note that in the sequential game there is no situation in which player two may happen to have defection as a unique strategy. This is because if psychological rewards are relevant, then the strategic advantage of seeing that player one has trusted leads to at least a mixed strategy. Therefore, unless he is indifferent among his strategies, a second player who has been trusted will surely reciprocate in the sequential game.

3.6 Conclusions

Trust is a necessary prerequisite of interactions developing in uncertain environments. Contracts and pre-commitment practises play an important role in economic interactions and exchanges, allowing for a more or less strict

definition of rules and objectives. Moreover, those are as well a warranty of enforceability by law of deviations of one or more participants to an economic exchange. However the system of laws and institutions that is due to enforce these instruments has to be commonly held as secure and trusted by all the participants. More than this contracts are always necessarily incomplete, while pre-commitment practises may sometimes be an imperfect signal.

These are the conditions under which something else is needed to allow for interactions and economic exchanges. For instance, trust is generally relevant in situations where one agent has to place his own resources at risk before obtaining a potential better outcome, as a result of the expected cooperation of the counterpart, while not being able to perfectly observe and control others actions.

Therefore, trust seems to be important in most of economic interactions involving a delay between an agent choice and involvement and his counterpart response in terms of cooperation or defection, a monitoring problem, asymmetric information.

In the proposed model we analysed the effects of beliefs about trust and trustworthiness, to describe how they jointly work as to allow cooperation to take place. We showed how trust might allow for exchanges to take place even in one shot interactions and in settings characterised by Prisoner's dilemma like or 'market for lemons' like situations.

Trust is defined in line with a behavioural approach, and is based on beliefs. The representation of trust as a belief, allows to link the concept to some experimental and theoretical results of recent research on the topic. The psychological game setting bridges experimental results and our game theoretic approach.

Instead of requiring perfect knowledge of the outcomes and of the counterpart, we have defined a one shot game where first order and second order beliefs enter directly agents payoff functions, determining endogenously both preferences and outcomes. This, in turn, defines the best strategies agents

choose.

We observed that many different equilibria may arise. In particular, a cooperative outcome is possible even in this one shot setting. Some of these equilibria are typically excluded by standard game theory, in contrast with experimental results where cooperation is observed also in anonymous and not repeated interactions.

This framework gives a picture of how agents perceptions and expectations influence their preferences and choices. Preferences and outcomes ordering is not fixed, but changes as agents beliefs change. The model allows therefore for an interaction and context-dependant representation of trust. Moreover, we believe it sheds light on how a cooperative outcome may be reached without abandoning a rational representation of agents decision processes.

References

G. A. Akerlof. The market for" lemons": quality uncertainty and the market mechanism. *The quarterly journal of economics*, pages 488–500, 1970.

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

K. J. Arrow. *The limits of organization.* Norton New York, 1974.

N. Ashraf, I. Bohnet, and N. Piankov. Decomposing trust and trustworthiness. *Experimental Economics*, 9(3):193–208, 2006.

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

S. Ba and P. A Pavlou. Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior. *MIS quarterly*, 26(3):243–268, 2002.

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

V. Bala and S. Goyal. Learning from neighbours. *The Review of Economic Studies*, 65(3):595–621, 1998.

V. Bala and S. Goyal. Self-Organization in communication networks. 1997.

V. Bala and S. Goyal. A theory of learning with heterogeneous agents. *International Economic Review*, 36(2):303–23, 1995.

P. Bateson. The biological evolution of cooperation and trust. In D. Gambetta, editor, *Trust: Making and Breaking Cooperative Relations, electronic edition. Department of Sociology, University of Oxford*, pages 14–30. 2000.

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

N. Berggren and H. Jordahl. Free to trust: Economic freedom and social capital. *Kyklos*, 59(2):141–169, 2006.

S. Beugelsdijk, H. L. F. de Groot, and ABTM van Schaik. *Trust and economic growth: a robustness analysis*, volume 56. Oxford Univ Press, 2004.

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.

S. Birkhchandani and D. Hirshleifer. Learning from the behaviour of others: Conformity, fads, and informational cascades. *Journal of Economic Perspectives*, 12(3):151–170, 1998.

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.

S. Bowles. Group competition, reproductive leveling, and the evolution of human altruism. *Science*, 314(5805):1569–1572, 2006.

S. Braynov and T. Sandholm. Contracting with uncertain level of trust. *Computational Intelligence*, 18(4):501–514, 2002.

L. Bruni and R. Sugden. Moral canals: trust and social capital in the work of hume, smith and genovesi. *Economics and Philosophy*, 16(01):21–45, 2000.

S. Buchegger and J. Y. Le Boudec. The effect of rumor spreading in reputation systems for mobile ad-hoc networks. volume 4, page 4.1, 2003.

V. Buskens. The social structure of trust. *Social Networks*, 20(3):265–289, 1998.

V. Buskens. Trust in triads: effects of exit, control, and learning. *Games and Economic Behavior*, 42(2):235–252, 2003.

V. Buskens and W. Raub. Embedded trust: Control and learning. *Group Cohesion, Trust and Solidarity*, 19:167–202, 2002.

V. Buskens and J. Weesie. Cooperation via social networks. *Analyse & Kritik*, 22(1):44–74, 2000.

J. V. Butler. Trust, truth, status and identity an experimental inquiry. *Not published*, 2008.

C. Camerer and E. Fehr. Measuring social norms and preferences using experimental games: A guide for social scientists. *Forthcoming in: Foundations of Human Sociality - Experimental and Ethnographic Evidence from 15 Small-Scale Societies, edited by Henrich, Boyd, Bowles, Camerer, Fehr, Gintis and McElreath*, 2002.

J. Carpenter. Measuring social capital: Adding field experimental methods to the analytical toolbox. *Social capital and economic development: Well-being in developing countries*, pages 119–137, 2002.

M. R Carter and M. Castillo. The economic impacts of altruism, trust and reciprocity: An experimental approach to social capital. *Wisconsin-Madison Agricultural and Applied Economics Staff Papers*, 448, 2002.

C. Castelfranchi and R. Falcone. Trust is much more than subjective probability: Mental components and sources of trust. volume 6, pages 10–12, 2000.

A. Chaudhuri, B. Sopher, and P. Strand. Cooperation in social dilemmas, trust and reciprocity. *Journal of Economic Psychology*, 23(2):231–249, 2002.

J. S Coleman. *Foundations of Social Theory*. Harvard University Press, 1990.

J. S Coleman. Social capital in the creation of human capital. *American Journal of Sociology*, 94(S1):95, 1988.

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

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

J. H. Davis, F. D. Schoormann, R. C. Mayer, and H. H. Tan. The trusted general manager and business unit performance: empirical evidence of a competitive advantage. *Strategic Management Journal*, page 563, 2000.

A. Diekmann and S. Lindenberg. Cooperation: Sociological aspects. *International Encyclopedia of the Social and Behavioral Sciences (4). Oxford: Pergamon-Elsevier*, 2000.

K. T Dirks. The effects of interpersonal trust on work group performance. *Journal of applyed psychology*, 84:445–455, 1999.

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

R. Dur and J. Sol. Social interaction, Co-Worker altruism, and incentives. *CESIFO Working Paper*, 2476, November 2008.

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

E. Fehr, U. Fischbacher, and M. Kosfeld. Neuroeconomic foundations of trust and social preferences: Initial evidence. *American Economic Review*, 95 (2):346–351, 2005.

D. P. Ford. Trust and knowledge management: The seeds of success. *Handbook on Knowledge Management*, 2002.

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

J. Geanakoplos, D. Pearce, and E. Stacchetti. Psychological games and sequential rationality. *Games and Economic Behavior*, 1(1):60–79, 1989.

E. L Glaeser, D. I Laibson, J. A Scheinkman, and C. L Soutter. Measuring trust. *Quarterly Journal of Economics*, 115(3):811–846, 2000.

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

M. Granovetter. Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, 91(3):481, 1985.

M. Granovetter. The impact of social structure on economic outcomes. *Journal of Economic Perspectives*, 19(1):33–50, 2005.

R. Guha, R. Kumar, P. Raghavan, and A. Tomkins. Propagation of trust and distrust. pages 403–412, 2004.

L. Guiso, P. Sapienza, and L. Zingales. People's opium? religion and economic attitudes. *Journal of Monetary Economics*, 50(1):225–282, 2003.

L. Guiso, P. Sapienza, L. Zingales, and V. D. Macelli. Cultural biases in economic exchange. *NBER working paper*, 2004.

L. Guiso, P. Sapienza, and L. Zingales. Does culture affect economic outcomes? *The Journal of Economic Perspectives*, pages 23–48, 2006.

L. Guiso, P. Sapienza, and L. Zingales. Trusting the Stock-Market. Technical report, January. Mimeo, 2007.

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.

R. Hardin. *Trust and trustworthiness*. Russell Sage Foundation Publications, 2002.

S. H Heap and Y. Varoufakis. *Game theory: a critical text*. Routledge, 2004.

J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, and R. McElreath. In search of homo economicus: Behavioral experiments in 15 small-scale societies. *American Economic Review*, pages 73–78, 2001.

J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky. BAYESIAN MODEL AVERAGING: a TUTORIAL. *STATISTICAL SCIENCE*, 14(4): 382–417, 1999.

K. Hoff and P. Pandey. Belief systems and durable inequalities: An experimental investigation of indian caste. *World*, 2004.

K. Hoff and P. Pandey. Discrimination, social identity, and durable inequalities. *American Economic Review*, 96(2):206–211, 2006.

H. J. Holm and A. Danielson. Tropic trust versus nordic trust: Experimental evidence from tanzania and sweden. *The Economic Journal*, 115(503):505–532, 2005.

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

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.

H. S. Jr James. The trust paradox: a survey of economic inquiries into the nature of trust and trustworthiness. *Journal of Economic Behavior and Organization*, 47(3):291–307, 2002.

A. Josang, C. Keser, and T. Dimitrakos. Can we manage trust. 2005.

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.

A. P Kirman and N. J Vriend. Learning to be loyal. a study of the marseille fish market. *Interaction and market structure: essays on heterogeneity in economics*, page 33, 2000.

S. Knack. Trust, associational life and economic performance. volume The Contribution of Human and Social Capital to Sustained Economic Growth and Well-being: International Symposium Report, Human Resources Development Canada and OECD, pages 172–202, 2001.

S. Knack and P. Keefer. Does social capital have an economic payoff? a Cross-Country investigation*. *Quarterly Journal of Economics*, 112(4): 1251–1288, 1997.

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

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.

D. Z. Levin and R. Cross. The strength of weak ties you can trust: The mediating role of trust in effective knowledge transfer. *Management Science*, pages 1477–1490, 2004.

S. Lindenberg. It takes both trust and lack of mistrust: The workings of cooperation and relational signaling in contractual relationships. *Journal of Management and Governance*, 4(1):11–33, 2000.

M. Lorenzen. Information cost, learning, and trust: lessons from cooperation and higher-order capabilities amongst geographically proximate firms. *DRUID Working Papers*, 98-21, 1998.

G. Lu, J. Lu, S. Yao, and Y. J Yip. A review on computational trust models for multi-agent systems. *The Open Information Science Journal*, 2:18–25, 2009.

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

R. Marimon, J. P Nicolini, and P. Teles. *Competition and reputation*. European University Institute, 1999.

Y. Mass and O. Shehory. Distributed trust in open multi-agent systems. *Lecture notes in computer science*, pages 159–174, 2001.

R. C. Mayer, J. H. Davis, and F. D. Schoorman. An integrative model of organizational trust. *Academy of management review*, pages 709–734, 1995.

K. A McCabe, M. L Rigdon, and V. L Smith. Positive reciprocity and intentions in trust games. *Journal of Economic Behavior and Organization*, 52(2):267–275, 2003.

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

A. K Mishra, R. Kramer, and T. Tyler. Organizational responses to crisis: The centrality of trust. In *Trust in organizations*, pages 261–287. London: Sage Publications, 1996.

L. Mui, M. Mohtashemi, C. Ang, P. Szolovits, and A. Halberstadt. Ratings in distributed systems: A bayesian approach. 2001.

L. Mui, M. Mohtashemi, and A. Halberstadt. A computational model of trust and reputation. pages 188–196, 2002.

B. Nooteboom. Innovation and inter-firm linkages: new implications for policy. *Research policy*, 28(8):793–805, 1999.

A. Ortmann, J. Fitzgerald, and C. Boeing. Trust, reciprocity, and social history: A re-examination. *Experimental Economics*, 3(1):81–100, 2000.

M. J. Osborne and A. Rubinstein. A course in game theory. *Cambridge/MA*, 1994.

P. A Pavlou and A. Dimoka. The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation. *Information Systems Research*, 17(4):392, 2006.

R. La Porta, F. Lopez-de-Silanes, A. Shleifer, and R. W Vishny. Trust in large organizations. *The American Economic Review*, 87(2):333–338, 1997.

R. La Porta, F. Lopez-de-Silanes, A. Shleifer, and R. W Vishny. Trust in large organizations. *Social Capital: A Multifaceted Perspective*, 1999.

R. D. Putnam. Tuning in, tuning out: the strange disappearance of social capital in america. *PS: Political Science and Politics*, pages 664–683, 1995.

M. Rabin. Incorporating fairness into game theory and economics. *The American Economic Review*, pages 1281–1302, 1993.

V. Radaev. How trust is established in economic relationships when institutions and individuals are not trustworthy (The case of russia). *Collegium*, 22:23, 2002.

H. Rainer and T. Siedler. Does democracy foster trust? *Journal of Comparative Economics*, 37(2):251–269, 2009.

M. Raiser, European Bank for Reconstruction, and Development. *Trust in Transition*. European Bank for Reconstruction and Development, 1999.

F. Roth. The effect of the financial crisis on systemic trust. *Intereconomics*, 44(4):203–208, 2009.

P. Sapienza, A. Toldra, and L. Zingales. Understanding trust. *NBER working paper*, 2007.

R. Sethi and E. Somanathan. Understanding reciprocity. *Journal of Economic Behavior and Organization*, 50(1):1–27, 2003.

J. B. Slemrod and P. Katuscak. Do trust and trustworthiness pay off? *NBER Working Paper*, 2002.

J. M Smith. *Evolution and the Theory of Games*. Cambridge University Press: Cambridge, 1982.

G. Tabellini. Culture and institutions: Economic development in the regions of europe. *Working paper*, 2005.

M. Tanis and T. Postmes. A social identity approach to trust: Interpersonal perception, group membership and trusting behaviour. *European Journal of Social Psychology*, 35(3), 2005.

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

F. Tonkiss. Trust, confidence and economic crisis. *Intereconomics*, 44(4): 196–202, 2009.

R. L Trivers. The evolution of reciprocal altruism. *The Quarterly Review of Biology*, 46(1):35, 1971.

G. Weisbuch, A. Kirman, and D. Herreiner. Market organisation and trading relationships. *The economic journal*, 110(463):411–436, 2000.

A. Whitby, A. J\textbackslashtextbackslashosang, and J. Indulska. Filtering out unfair ratings in bayesian reputation systems. *The Icfain Journal of Management Research*, 4(2):48–64, 2005.

B. Williams. Formal structures and social reality. In D. Gambetta, editor, *Trust: Making and Breaking Cooperative Relations, electronic edition, Department of Sociology, University of Oxford.* 2000.

B. Yu, M. P. Singh, and K. Sycara. Developing trust in large-scale peer-to-peer systems. pages 1–10, 2004.

A. Zaheer, B. McEvily, and V. Perrone. Does trust matter? exploring the effects of interorganizational and interpersonal trust on performance. *Organization Science*, pages 141–159, 1998.

4 Learning trust from interactions in social networks

Abstract

Trust is often described as an important variable in social interactions. This is true not only for disciplines such as sociology, psychology or anthropology, but also for biology, computer science and economics. Scholars in the economic domain have mainly referred to trust as a background characteristic, allowing exchanges and potentially influencing economic performance.

Most of research calling upon trust has done this without any formalization of trust role in economic dynamics. In the last decades however, better defined measures of trust and more insights coming from multi-disciplinary studies, opened new possibilities to evaluate its effects, also on economic outcomes. Simultaneously, more studies tried to deepen the understanding of trust antecedents at the interaction level, in dyads and triads.

In this paper we investigate how trust dynamics may be modeled in a social environment. We present a model implementing some recent insights from experimental research, and we investigate how agents interactions based on trust may influence both individuals and population performance. Our focus is primarily on linking a definition of trust based on economic primitives, such as beliefs and preferences, to the ability of agents to learn about the interaction environment as to choose the most fruitful relationships.

4.1 Introduction

A system where trust lacks may even shrink below a critical threshold level necessary for its own reproduction at a certain level of development (Luhmann, 2000). Trust appears therefore as a necessary element of social interactions and also essential for the development and sustainability of economic

performance.

Many models in the literature deal with dyadic interactions, both in theoretical and experimental approaches (Axelrod and Hamilton, 1981, Berg et al., 1995, Buskens and Raub, 2002, Chaudhuri et al., 2002, Buskens, 2003, McCabe et al., 2003, Hardin, 2003). Instead, this paper aims at shedding more light on trust building and diffusion in social networks.

Computer scientists generally deal with network models related to trust in terms of reputational systems (Bhattacharya et al., 1998, Mui et al., 2001, Ba and Pavlou, 2002, Mui et al., 2002, Buchegger and Boudec, 2003, Yu et al., 2004, Guha et al., 2004, Huynh et al., 2006, Pavlou and Dimoka, 2006, Lu et al., 2009). These studies focus on computer or mobile networks, or on e-commerce websites. These scholars model mechanisms for the evaluation of network nodes, whether they are artificial (computers) or human (sellers and buyers), and for the selection, evaluation and aggregation of scores for the creation of referral systems.

On the other hand, economists are generally linking social environments and trust along three main different perspectives. The first perspective is more qualitative in nature, and refers generally to social capital literature (Granovetter, 1985, Dasgupta, 2000, Berggren and Jordahl, 2006). Secondly, managerial literature focuses on the link between trust dynamics in groups and organizational performance (Mayer et al., 1995, Mishra et al., 1996, Zaheer et al., 1998, Dirks, 1999, Davis et al., 2000, Dirks and Ferrin, 2001, McEvily et al., 2003, Dur and Sol, 2008). Finally, an increasing number of studies focuses on the link between macro-economic performance and cultural aspects of the relevant social group, typically a country (Knack and Keefer, 1997, Porta et al., 1997, 1999, Raiser et al., 1999, Glaeser et al., 2000, Knack, 2001, Radaev, 2002, Guiso et al., 2004, Tabellini, 2005, Guiso et al., 2006, 2007, Bohnet et al., 2008, Fehr, 2009, Rainer and Siedler, 2009).

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. For instance, we focus on the social construction of trust and its diffusion among a group of interacting agents.

Trust is often described as an important variable in the context of social interactions (Dasgupta, 2000, Good, 2000, Luhmann, 2000, Granovetter, 2005). The social environment is also the place where learning on others goes on, via exchange and pooling of available and discoverable information, allowing for the diffusion and the evolution of trust (Coleman, 1990, Bala and Goyal, 1995, Birkhchandani and Hirshleifer, 1998, Buskens, 1998, Buskens and Raub, 2002, Buskens, 2003, Buchegger and Boudec, 2003, King-Casas et al., 2005). And, again, a social context gives better incentives and rules (i.e. reward and punishment) that are necessary part of trust dynamics (Buskens and Raub, 2002, Buskens, 2003, Granovetter, 2005).

We investigate the relevance of a dynamic feedback in trust emergence and evolution. In fact, on one side trust allows for social interactions to take place. On the other side, the outcome of these interactions allows for an update of trust itself. 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. To this objective, we model a network of individuals interacting on the basis of trust and trustworthiness attributes. We link our model to recent developments in trust research showing a strong impact of social preferences and beliefs on trust dynamics (Cox, 2004, Fehr et al., 2005, King-Casas et al., 2005, Kosfeld et al., 2005, Fehr, 2009).

The rest of the paper is structured as follows. Section 2 gives an overview on trust related research, highlighting the fundamental results on the topic. Section 3 describes the formalization of the model. Section 4 describes the simulation process and the results. Finally section 5 highlights some possible conclusions.

4.2 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 (Dasgupta, 2000).

As said in the introduction, trust is claimed to shape economic outcomes and performance, both at the individual and the macro levels. It has been analyzed for its strategic implications in agents choices.

However, to understand trust one needs not only to assess its consequences, but also to analyze its antecedents. In particular, for a better understanding of trust dynamics research has to clarify whether trust is a well-defined attribute and eventually disentangle it from other attributes. More than this, research has to link trust to micro-economic theory concepts, such as preferences and/or beliefs. It is exactly these steps that, in turn, may motivate the impact of trust on agents decision process and on economic outcomes.

In fact, some authors 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 (Castelfranchi and Falcone, 2000). It is underlined how a richer view and description of trust is needed to make it a valuable concept for economic analysis, and not just a mere word to make a more user-friendly theory or a redundant and useless concept.

The major claim is that a cognitive theory of trust is needed as part of an enriched description of the 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.

Castelfranchi and Falcone (*ibidem*) also suggest a decomposition of trust, as partly defined by the trust in the trustee, and partly by a generalized trust in the context of the interaction. However, to our view, context is better described in terms of uncertainty or risk. One doesn't trust in any way the

context, while for instance, trust in the trustee has to be contextualized. As in Hardin's definition (Hardin, 2002), trust describes in fact a three-part relation: among the trustor (1) and the trustee (2) in a well-defined context (3). Anyhow, the context is already accounted, we believe, by agents preferences such as risk aversion.

Nonetheless, the concept of risk is strongly linked to trust. But 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 (Luhmann, 2000). In recent neuro-economic studies this socially embedded risk is defined as betrayal aversion (Bohnet et al., 2008).

In addition, risk has an intimate relationship with interest in trust dynamics, in the form of a threshold of the interest/risk rate, over which trust is placed, and under which no trust is possible. This level is both objective, i.e. context dependent, and subjective (Gambetta, 2000). In other words, disposition to cooperate is cost sensitive, so there is a cost threshold, possibly changing over time, that defines when cooperation is feasible for an agent (Williams, 2000).

These aspects are better understood when one looks to recent studies, linking trust and its mental components, in terms of beliefs and preferences. An approach referred to as neuro-economics is used, that brings together methods from neuroscience and economics, to investigate how human decisions in social and economic context are generated in the brain.

In particular, scholars analyze some evidence on other-regarding behaviors. The setting is similar to the one adopted in game theoretical approaches, and usually a trust game or an investment game are the context under which the study is conducted.

In a recent paper, Fehr (2009) documents the accumulation of strong evidence – neuro-biological (Kosfeld et al., 2005), genetic, and behavioral (Cox, 2004) – 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.

Kosfeld et al. (2005) study on oxytocin, an hormone that has been studied as an important enhancer of pro-social and affiliation behavior in many different non-human mammals, demonstrate that this substance may increase pro-social behavior of humans. This doesn't affect in any way reciprocity, hence trustworthiness, but only the component of preferences that is linked to trusting others.

Interestingly, even without oxytocin subjects, although they have a clear view of the free-riding problem in cooperation and trust, choose to enter the game out of forms of kindness, but for specific preferences for cooperation (Andreoni, 1995).

Kosfeld et al. (2005) also investigate whether trust is the same as risk taking, and the evidence suggest that more specifically trust requires a risk that is generated in a social environment and interaction, where, indeed, the specific risk is linked to agents choices. Fehr (2009) reports more research evidence that the risk present in trusting is of a different kind than the general case of risk taking. Trusting is risk taking involving social interactions, and studies in biology and evolution are demonstrating that this behavior is codified and appears in specific ways.

It is curious to note that not only humans, but even animals appear to be sensitive to others behavior, and this observation may suggest an explanation for the evolution of the mental state that we recognize as trust in ourselves (Bateson, 2000). Scholars have shown that trust may appear as a reciprocating successful strategy (Trivers, 1971, Smith, 1982).

One possible way trust may have emerged is, for instance, a tit-for-tat strategy (Axelrod and Hamilton, 1981). Biologists believe that once this

strategy spreads sufficiently in the population and demonstrates itself as a successful one, it is codified in genes and transferred to future generations, as a mean to undergo specific situations (Bowles, 2006).

We don't need however to consider the process of biological evolution to get some idea of how this mechanism might work. Some scholars demonstrate, using an experiment based on the trust game and functional magnetic resonance, that a neural process develops and changes as far as trust and interactions are developed (King-Casas et al., 2005). Decision making processes of subjects entering a trust relation are strongly influenced in magnitude and timing by the signals of trustworthiness coming from their counterpart. This seems to demonstrate that the structure of preferences may change over time, as soon as useful information is gained from the interaction, so that the preference towards trusting becomes strongly predominant.

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, it allows agents to better understand and predict the environment in which they act. As Good (2000) suggests, the reaction of another agent to one's own action is important in confirming prior experience, and this confirmation (predictability) is necessary to make the social world intelligible and seemingly knowable.

It may appear clear that some prerequisites are necessary for these dynamics to be at work. Firstly, since trust is depleted through not being used (Gambetta, 2000), agents need to interact in a trust environment for it to be valuable and develop. Secondly, agents need some capabilities in discriminating among their potential counterparts and a sufficiently developed information processing ability and flexibility (Kurzban, 2003) as to make it possible to choose the best partners to place trust in. In example, these characteristics allow agents to understand others intentions.

An interesting result on this aspect is reported in McCabe, Rigdon and Smith (2003), where the authors investigate the difference between outcomes-

based and intention-based approaches to trust. Outcomes-based approaches would predict that only the pay-offs of a game matter to guide agents choices and behaviors. On the other side, intention-based approaches, explore the possibility that, depending on the available alternatives, identical outcomes may be interpreted differently.

It is suggested that when the set of available choices differs, the same behavior by an agent may be interpreted differently by his counterparts. The authors find that a different structure of the game gives players different signals about the motives underlying the observed choices, whether those are based on trusting or not. This in turn enhances or lowers the preference towards reciprocating. We may note that, however, the structure of the game, that is the tree of available choices, is not always common knowledge in trust interactions.

Therefore, signaling and the ability to decode correctly and profitably these signals, are 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 and informations diffused along the network.

In the next paragraph we briefly add some considerations about reciprocation and reputation building as important signals in trust dynamics, taking as reference research mostly appeared in computer science. We believe this is an important point, because some of the insights emerging from this literature are useful also for the model presented in this paper. In example, problems related to generation, aggregation and pooling of informations; or, again, aspects of the learning mechanism; and, finally, relevance of a social evaluation system and of information sharing.

4.2.1 Reputation systems and social learning

Reputation is a matter of great relevance in the context of trust studies. And it encompasses the research on both trust and trustworthiness. Lot of recent

quantitative and theoretical research on this topic is linked to computer science and the study of networks as means to assess the trustworthiness of the nodes and to transfer reputation scores along the network itself.

The problem has a strong economic appeal, and it is the focus in the research area of markets, products quality and firm reliability. Economics scholars have tried to model the incentives for reputation. Failures similar to the 'market for lemons' depicted by Akerlof (1970) may obstacle the development of a market unless there is an efficient structure of incentives or well developed contract and pre-commitment practices. On the other side, whereas these instruments may not be available, trust and trustworthiness are pre-conditions to the development of the market itself, and reputation building on the seller's side is considered strongly important.

In most of these studies, market or exchange modeling is the primary focus, and matters of trustworthiness, information diffusion among agents and reputation building practices are the means to solve market failures. This approach is related to problems such as the persistence of low quality products; the reliability of firms on a competitive market to sell quality products at the right price (Tirole, 1996); the information possibly gathered from consumers from prices; the information gathered from agents in ongoing and past interactions; the diffusion of the information along the network among different agents; the persistence of corruption and the problem of partner selection and reliability (Tirole, 1996); efficiency gains from trustworthiness and reputation in competitive markets (Marimon et al., 1999, Braynov and Sandholm, 2002).

Reputation is also strongly conditioning trust behaviour, expected trustworthiness and observed trustworthiness. In an interesting experiment (Fehr, 2009) a gift exchange game conducted with and without a reputation building mechanism shows interesting results. Allowing for reputation building increases not only trust behaviour expressed by employers wage offer levels, but also reciprocity from the counterparts, expressed by workers effort levels.

Therefore, higher effort levels that at first sight might appear as caused by higher wages, are for instance a consequence of higher reciprocity inducted by an informal system of reputation building.

Some of these points are addressed also in non-economic research, but even these studies mostly rely on economic exchanges (such e-commerce) to motivate their interest on the topic or to test their hypothesis. Some scholars analyze the problem from a biological perspective and find out that reputational mechanism even work at the level of our brain, shortening the timing of our decisions in an interaction game, and demonstrating a reputation building effect (King-Casas et al., 2005).

A growing research related to trust and reputation is going on in computer science. Most of the times this literature is linked to Bayesian learning and goes under the broad categorization of reputation/referral systems studies or trust management (Josang et al., 2005). In peer to peer networks, notably in e-commerce websites, but even in the structure of informatic communication among servers and nodes in the world wide web or in mobile networks, trust and reputation are topics of strong interest for many scholars.

Researchers discuss important questions related to scores generation, discovery and aggregation (Yu et al., 2004). The problem at hand is the possibility to assess the trustworthiness of every node in the network, by means of scoring from other nodes in the same network, as to create a reputation building mechanism. To do so, one needs to structure an evaluation scheme/mechanism, and the necessary rules and incentives to foster its reliability in the correct evaluation of the individual nodes.

There is indeed a secondary problem that is linked to the reliability of the evaluating nodes, that is to their own reputation, and to the possibility for subsequent agents to assess not only the reputation of the agent with whom they may have an opportunity for valuable interaction, but also the reputation and reliability of previous partners that gave that agent an evaluation. Moreover, it has to be noted that the possible sources of information may

not be known from previous interaction, and there is no direct way to assess their identity and motivations.

Some authors underline that a trust model needs to take into account many sources of information, as to be robust against some 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. (Huynh et al., 2006). This seems to be an important prerequisite for a reputation system to work properly: on one side, the more the communication and the information sources, the more efficient the system in isolating misbehaving agents; on the other side, the sooner misbehaving agents are recognized and isolated, the better the information obtained and the more stable the system. In these contexts, misbehavior is not only intended in the sense of not respecting the rules of the interaction, but, more interestingly, not respecting the rules of a fair scoring mechanism (Whitby et al., 2005).

We consider some of these insights really important for a model of trust. Therefore, we try to account for them in the model we are going to present in the next sections. We underline, however, that the model hereafter is a starting point to analyze trust dynamics, and will surely need future development and refinement to consider all the relevant matter we presented in this brief research overview.

Anyhow, the analysis develops on these building block, with some important additions. Firstly, it develops a model of trust in a social environment, widening the scope of the model well beyond a dyadic or triadic interaction context. Secondly, it uses insights from different disciplines and approaches, both for agents characterization and for rules of interaction definition and implementation. Lastly, its simple implementation gives the chance to extend the same model further for deeper evaluations.

4.3 Model

We consider a network of N agents, indexed as $i = 1, 2, \ldots, N$. Time is discrete, indexed with $t = 0, 1, \ldots$. For each agent we define a number of characteristics, over which we model interaction and learning dynamics.

We define individual endowments in terms of trust and trustworthiness. Both trust belief p_{ij}^t and trustworthiness level w_i are set in the interval $[0, 1]$. However, while trust attribute is defined over all the possible dyads of partners in the network, trustworthiness is an individual generic attribute non interaction-dependent.

Every agent is also endowed with exogenous preferences, and in particular we give each agent an attribute in terms of the social component of preferences. This is a threshold value $\theta_i \in [0, 1]$ over which each individual is willing to interact with potential partners. Under the same value an agent prefers not to form links to other agents.

As we consider exogenous preferences, these thresholds are given and do not change over the considered period. However, it may well be the case that preferences evolve over time, in response to new information and new experience (Good, 2000, Hardin, 2002, King-Casas et al., 2005).

Moreover, our formalization resumes two insights from recent research about agents preferences. Firstly, they represent an indicator of the pro-social attitude of each agent, that plays a key role in trust dynamics (Cox, 2004, Kosfeld et al., 2005, Fehr, 2009). Secondly, they represent an indicator for agents risk aversion. In particular, we talk about betrayal aversion, that is related to socially-embedded risk taking (Bohnet et al., 2008, Fehr, 2009).

We define a generic payoff γ from cooperation and δ from defection. The decision to form a link is made confronting the expected payoff from an interaction, based on the trust belief an agent has in each of his potential partners, and the threshold payoff, dependent on the social preferences threshold value. If

$$p_{ij}^t * \gamma + \left(1 - p_{ij}^t\right) * \delta \geqslant \theta_i * \gamma + (1 - \theta_i) * \delta$$

then an agent is willing to link to his potential partner. An interaction among two agents is defined if a link is formed with the consensus of both parties.

Once an interaction is created, each agent decides whether to interact cooperatively or defect. The strategy is chosen based on each agent's own level of trustworthiness. Therefore, three different outcomes are possible: both agents cooperate, both defect, one cooperates and the other defects.

If an agent gets a positive outcome strictly higher than zero, then he records a success. On the opposite, he records a failure. However, only cooperation from both agents assures a successful interaction.

Information about the outcome of the interaction is shared among all common partners, and diffused along the network. This allows for an update of the individual levels of trust, on the basis of both direct and indirect experience. We present the update mechanism in the next subsection, describing before own information and afterwords third-party information treatment.

The learning mechanism we present is based on so-called Bayesian model averaging (Hoeting et al., 1999). Briefly, agents have different 'models' of reality, expressed by their own beliefs about others trustworthiness. Since these are shared after each period of interaction, agents need to assess which 'model' better reflects and fits available evidence and therefore all models are weighted against each round outcomes.

Moreover, we link this approach with some of the insights about score aggregation and pooling as emerged in computer science literature on trust. In this model, third-party experience is weighted giving similar weight to all the relevant sources. However, we are aware that this aspect may need further investigation. In particular, different sources may be trusted differently, therefore implying trust-dependent weights on the various information channels. We leave this aspect to future development, and we move to the description of learning dynamics.

4.3.1 Own information

Own information is embedded in a β-distributed belief $p_{ij}^t \sim \beta \left(u_{ij}^t, \, v_{ij}^t \right)$, that represents the information about the outcomes of the interactions among agent i and agent j. In our definition of the beta function, u_{ij}^t and v_{ij}^t define respectively successes and unsuccesses.

Note that this information is potentially asymmetric. In fact, if only one agent was cheating, then he registers a success (he was able to cheat without being cheated simultaneously), while the non-cheater registers an unsuccess. On the contrary, if both agents were cheating, then both register an unsuccess. This consideration is important, since it may allow for potentially contrasting information to flow in the network. In the present model we anyhow consider only symmetric information, since only the outcome of the interactions is communicated to common partners.

Based on this function, the expected likelihood of success in a partnership between i and j is $p_{ij}^t = u_{ij}^t / \left(u_{ij}^t + v_{ij}^t \right)$.

Every period partnerships are formed based on agents beliefs, just defined, and preferences. As stated above, we define the social component of individual preferences as a threshold value $\theta_i \in [0, 1]$. This determines the level of trust belief over which each agent is willing to propose or accept an interaction.

Once partners are selected and interactions are made, a network g^t of all the links ij for which $E\left[p_{ij}\right] \geq E\left[\theta_i\right]$ and $E\left[p_{ji}\right] \geq E\left[\theta_j\right]$, where the former in each inequality defines the expected payoff given agents beliefs, and the latter the expected payoff given agents preferences. If the outcomes from a successful and unsuccessful interaction are respectively $\gamma = 1$ and $\delta = 0$, then each agent i will accept and propose only links for which $p_{ij}^t \geq \theta_i$.

The interaction fails with probability depending on the level of trustworthiness of the counterparts. Trustworthiness w_i is also defined in the interval [0, 1] and is stable over time. However, also on this point, it is to underline that trustworthiness may be subject to an evolution process due to the envi-

ronmental conditions and the partnering and success history of each agent.

The subsets o^t, $s^t \subseteq g^t$ define, respectively, the indicators of positive outcomes for each agent, and of successful interactions. Note that, by defining these two different subsets, we allow for interactions to be successful if and only if the outcome is positive for both the interacting agents. Otherwise, if for i, $j \in o^t$, $o_{ij}^t \neq o_{ji}^t$, then $s_{ij}^t = s_{ji}^t = 0$ and $s^t \subset o^t$.

From one period to the next, the belief p_{ij}^t is updated according to Bayes rule, using the information embedded in s^t. After period t has taken place, positive and negative outcomes are recorded and used to update individual beliefs about partners. Therefore, $p_{ij}^{t+1} \sim \beta \left(u_{ij}^t + s_{ij}^t, \; v_{ij}^t + 1 - s_{ij}^t \right)$.

The condition for a link to be created at each period t is thus

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

This condition is only depending on i and j's interactions. Although we want the model to include other k's experiences. This is where third party information pooling comes in.

4.3.2 Third-party information: pooling and weighting

We assumed that every agent holds a belief about other agents in the network in the form of a variable $p_{ij}^t \sim \beta \left(u_{ij}^t, \; v_{ij}^t \right)$. Agents generally update their beliefs on the basis of direct experience. But, for instance, a network may allows for the diffusion of third-party information. If agents share common partners, then they may be able to exchange information about their own experiences with these partners. Moreover, they may diffuse information about their counterparts even when these are not common partners.

To generate learning from others experiences we use a methodology defined Bayesian model averaging (Hoeting et al., 1999). This is a way 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 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

$$Pr\left\{\Delta \mid D\right\} = \sum_{k=1}^{K} Pr\left\{\Delta \mid D,\, M_k\right\} Pr\left\{M_k \mid D\right\}$$

This is an average of the posterior distributions of each considered model weighted by their posterior model probability. For each model M_k, the posterior probability is given by

$$Pr\left\{M_k \mid D\right\} = \frac{Pr\left\{D \mid M_k\right\} Pr\left\{M_k\right\}}{\sum_k Pr\left\{D \mid M_k\right\} Pr\left\{M_k\right\}}$$

where

$$Pr\left\{D \mid M_k\right\} = \int Pr\left\{D \mid M_k,\, \theta_k\right\} Pr\left\{\theta_k \mid M_k\right\} d\theta_k$$

is the integrated likelihood of model M_k, θ_k is the vector of parameters for model M_k, $Pr\left\{\theta_k \mid M_k\right\}$ is the prior density of θ_k under model M_k, $Pr\left\{D \mid M_k,\, \theta_k\right\}$ is the likelihood, and $Pr\left\{M_k\right\}$ is the prior probability that M_k is the true model (given that one of the models is true). 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 about j, that is p_{kj}^t. When considering also third-party information, thus, the

posterior belief at time t is:

$$E\left[p_{ij}^t \mid g^{t-1},\, s^{t-1}\right] = E\left[p_{ij}^t \mid g_{ij}^{t-1},\, s_{ij}^{t-1}\right] Pr\left\{M_i \mid g_{ij}^{t-1},\, s_{ij}^{t-1}\right\} +$$

$$+ \sum_{k \in g_i, g_j} E\left[p_{kj}^t \mid g_{kj}^{t-1},\, s_{kj}^{t-1}\right] Pr\left\{M_k \mid g_{kj}^{t-1},\, s_{kj}^{t-1}\right\}$$

where the first terms are as defined above, while the second terms are as defined in Bayesian model averaging.

There are two important assumptions regarding the pooling of informations. Firstly, own information is weighted against third-party information, through a parameter $\alpha \in [0,\, 1]$. This allows agents, i.e., to take into account direct experience more than indirect one. Secondly, third-party information is weighted as to give a higher value to more trusted third parties. This is in line with research on reputation systems, where information is typically aggregated in a way to reflect the confidence level in the different sources (Buchegger and Boudec, 2003). Therefore

$$p\left(j\right) = \sum_{k=1}^{m} w_k p_{kj}$$

where

$$w_k^t = \frac{p_{ik}^t}{\sum_{k \in g_i, g_j}^{m} p_{ik}^t}$$

In this way, we allow for information coming from more trusted agents to be considered more relevant and safe. However, in this paper we leave this formalization and we investigate the learning dynamics giving the same

weight to all third-parties. Weight is therefore the same over all the information sources, which also implies and allows for an easier computational effort by the involved agents.

Moreover, since we allow only information about interaction outcomes to flow in the network, and we assume that there is no strategic use of information by agents, all the sources are equally trustable, by definition. In subsequent research we will investigate strategic use of information and flows of potentially contrasting results by the communication of the individual outcomes sets and we will therefore need to treat sources differently.

4.4   Performance measures and results

We run some numerical experiment to investigate the effects of learning and preferences on some performance measure. Specifically, we define performance measures for the evaluation of learning efficiency and for the analysis of how the threshold value of agents social preferences influences both learning efficiency and the number of interactions and successes in the population.

For learning efficiency we define two measures. The first is average trust distance, and it is defined as the average of all agents trust belief distance to real trustworthiness values after each period. It gives an indication of how the population trust level is moving with respect to the trustworthiness characterizing the population itself. The formalization is therefore

$$\Delta_{AV}^t = \frac{\sqrt{\sum_{i,j}\left(p_{ij}^t - w_j\right)^2}}{\sqrt{(N-1)}}$$

The second measure instead considers a similar formalization of distance, in which we only consider agents who interacted in the current period. Therefore, the average distance for interacting agents is defined as

$$\Delta_{INT}^t = \frac{\sqrt{\sum_{i,j\in g^t}\left(p_{ij}^t - w_j\right)^2}}{\sqrt{\left(\sum_{j\in g_i^t} j\right)}}$$

138

We also consider two other measures to evaluate the percentage of interactions with respect to the possible ones, and the percentage of successes with respect to ongoing interactions. We compute these measures for each period.

On one side we investigate how learning is affected by and affects the characteristics of the population of agents. Learning efficiency is important for two different reasons. In first place, the faster the learning process, the better the chances to interact only with potentially good partners, lowering therefore the chance of unsuccesses. Secondly, the faster agents learn, the better the information flowing in the network and the more relevant for the detection of untrustworthy counterparts.

On the other side, social preferences are described in the model as a threshold value, over which agents are likely to interact. This, in turn, directly influences the number of possible interactions in the network. Our question is therefore how this preferences structure interacts with learning dynamics and average trustworthiness in the population. We evaluate the effects of these characteristics on the interactions patterns and on the ability to choose more successful partners.

We discuss some of the results and possible implications, showing also some graphical evidence from our numerical experiments. We start from learning efficiency.

In first step we analyze how preferences may influence learning efficiency for individuals and population. Our results suggest that higher preferences allow for better individual learning, that is $\Delta_{INT} \simeq 0$. Although the number of interacting agents is lower due to the higher preferences, these agents show a low average distance to real trustworthiness values.

This is valid both for low (see Figure 6) and high (see Figure 7) average population trustworthiness. We observe in fact that when average trustworthiness is low the number of interactions is expected to decrease fast, therefore implying less interactions and influencing the possibility of agents

to learn from third-parties. However this effect persists also for high average trustworthiness in the population making the findings more robust. Results seem to hold also independent of the weight agents give to direct and indirect experience.
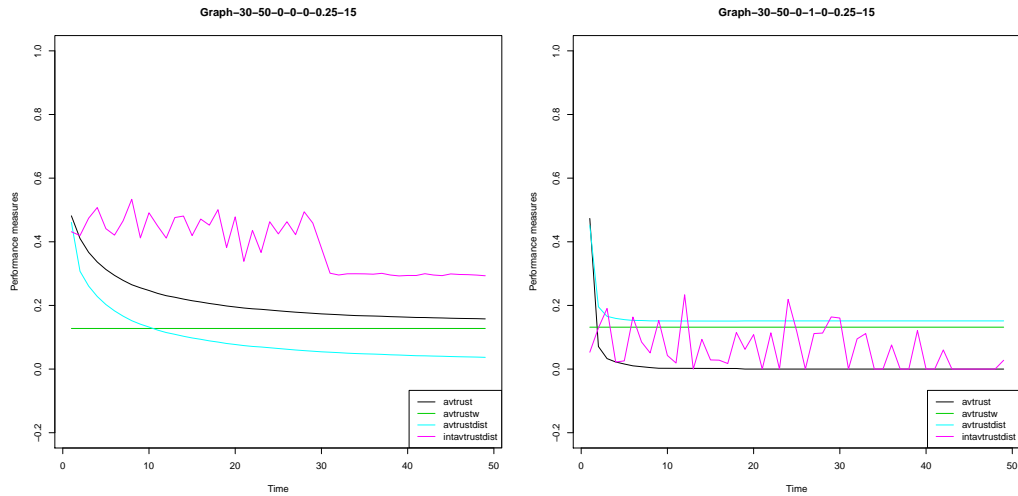

Figure 6: Higher individual learning at lower average trustworthiness, increasing preferences

As preferences increase, since the number of interaction is lower we expect higher average trust distance Δ_{AV}. Generally this is anyway decreasing. However, for higher preferences, the relation between preferences and distance may become less evident. For lower weights on direct experience and higher average trustworthiness in the population this distance is even increasing (see Figure 8).

Figure 7: Higher individual learning at higher average trustworthiness, increasing preferences
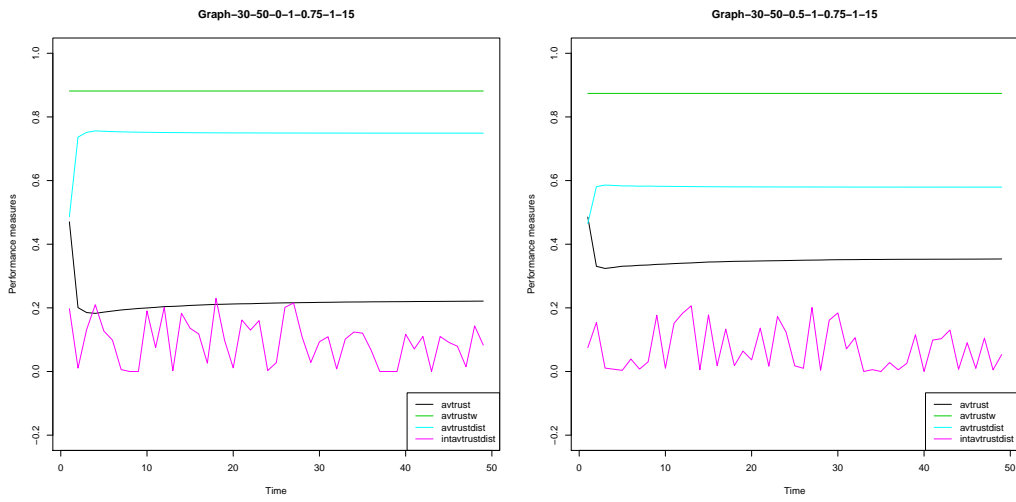

Figure 8: Increasing Δ_{AV} for lower α and higher average trustworthiness

On the other side, we observe that the lower the preferences threshold, the lower also Δ_{AV}, and therefore the closer the average trust beliefs in the population get to real average trustworthiness. An higher number of interactions due to a lower preferences threshold assures better results for the population in its complexity, although at the individual level we observe more variance in learning efficiency, as Δ_{INT} is generally higher in this setting.

We also investigate how learning efficiency varies depending on direct versus indirect experience weight, and on the average level of population trustworthiness. We observe that for low preferences threshold our results show no or little effect (see Figure 9). For higher preferences we observe on the contrary some effects.

In particular, if average population trustworthiness is low, an equal weight on direct and indirect experience seems to give the best learning results (see Figure 10). In fact for $\alpha = 0$ or $\alpha = 1$ we observe strong underestimation or overestimation of real trustworthiness in the population. For high average trustworthiness, we observe a general underestimation of the real values, with even increasing average distance (as observed before). Slightly better results are observed for higher weight on direct experience (see Figure 11).
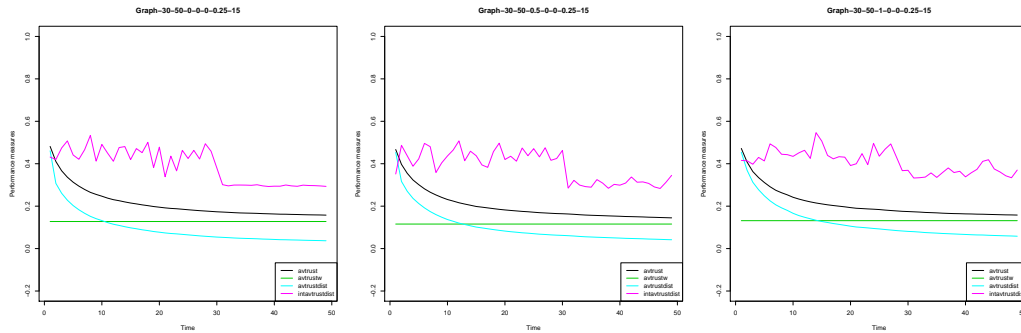

Figure 9: Learning efficiency with $\alpha = 0$ (sx), $\alpha = 0.5$ (center) and $\alpha = 1$ (dx) for low preferences
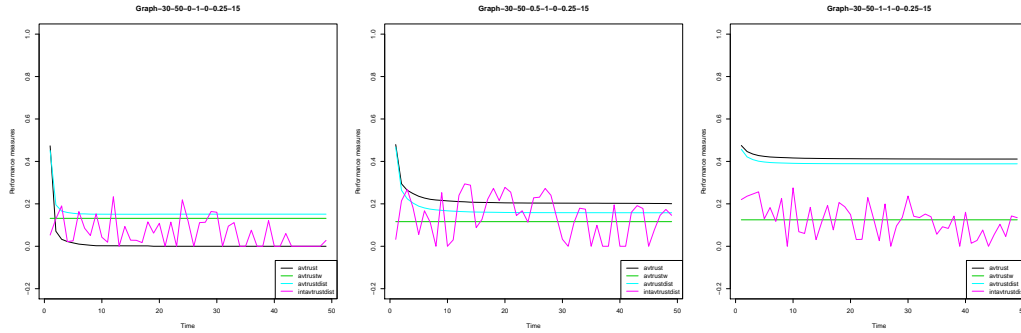
Figure 10: Learning efficiency with $\alpha = 0$ (sx), $\alpha = 0.5$ (center) and $\alpha = 1$ (dx) for high preferences and low trustworthiness
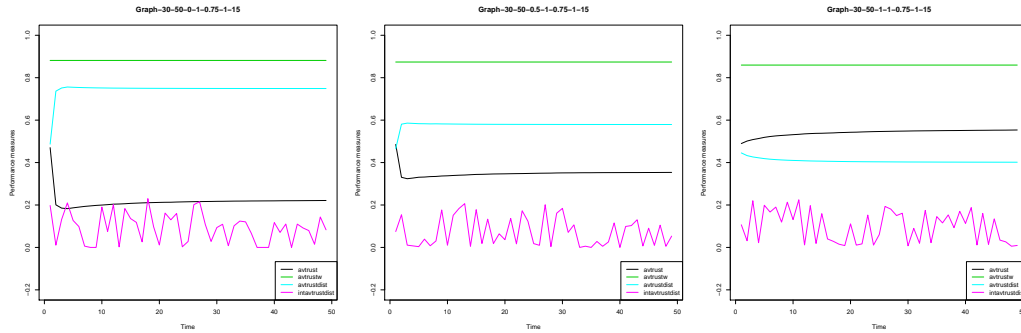

Figure 11: Learning efficiency with $\alpha = 0$ (sx), $\alpha = 0.5$ (center) and $\alpha = 1$ (dx) for high preferences and high trustworthiness

An important element to observe is that preferences threshold strongly conditions the ability to learn efficiently from both direct and indirect experience, as well as the ability to interact fruitfully even in trustworthy environments. This may well be due in big part to the assumption of stable preferences.

143

Learning efficiency is always higher if agents have low preferences thresholds, that is we observe lower average trust distance, and average trust level approaches real average trustworthiness. Anyhow, this may come at the cost of too many interactions with respect to the rate of success (see Figure 12a).

In untrustworthy environments only high thresholds assure that the percentage of interactions is in line with the rate of success, that is the few interactions that eventually remain are the successful ones (see Figure 12b). However most of the times this comes at the expense of no interaction at all, although some partnership may still be eventually profitable.

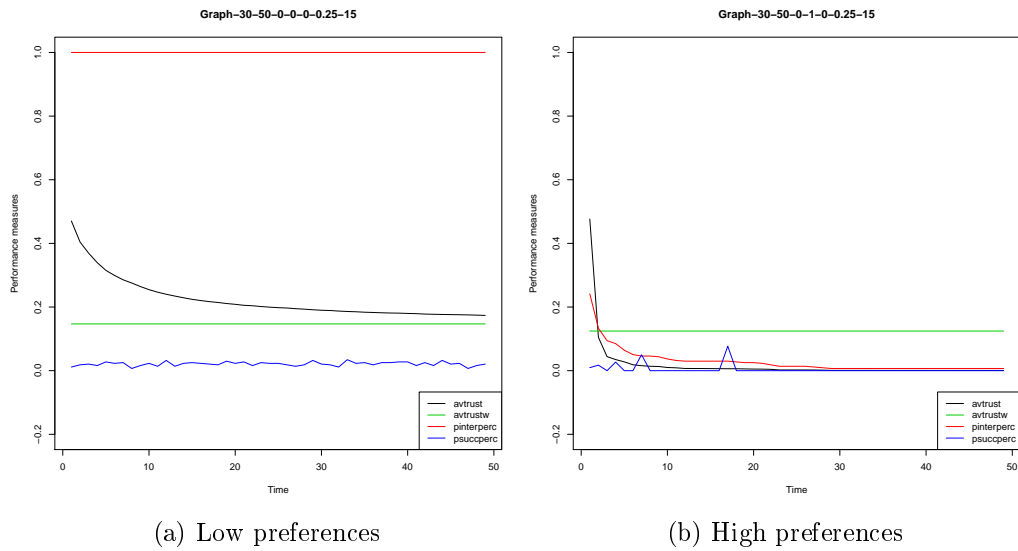(a) Low preferences   (b) High preferences

Figure 12: Low average trustworthiness - Interactions and rate of success

On the contrary, if preferences threshold is too low the process of adjustment may take long time and too many unsuccessful interactions may persist over time. We also note that even in trustworthy environments too

low preferences may sustain over time a number of interactions well beyond the successful ones (see Figure 13a).

Nonetheless, already at low levels of the preferences threshold agents are able to obtain a better performance in terms of percentage of successes, and the percentage of interactions is in line with the real rate of success (see Figure 13b).

When preferences get higher, though, the percentage of interactions gets too low and agents loose most of the potentially successful interactions they may have in their environment (see Figure 13c).

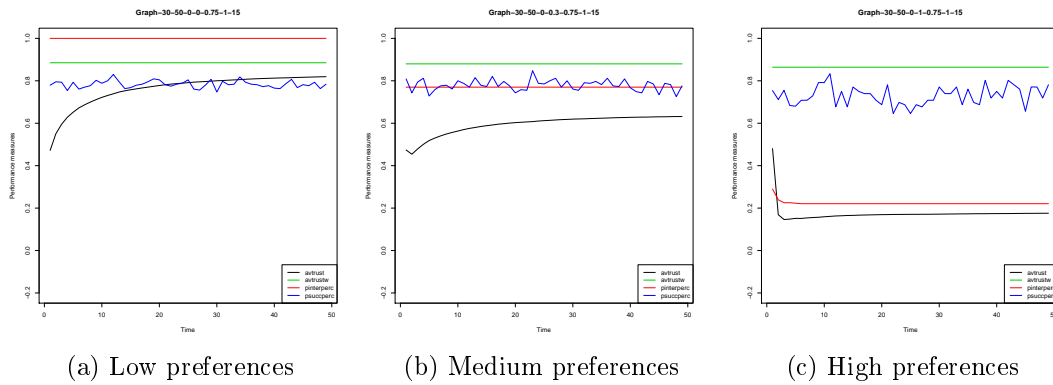(a) Low preferences   (b) Medium preferences   (c) High preferences

Figure 13: High average trustworthiness - Interactions and rate of success

We may stress at this point that trustworthiness is also a relevant characteristic to take into account for the evaluation of trust dynamics, and in this model we have imposed a fix level of trustworthiness. However, we have showed evidence about some trust dynamics and of some regularities for different trustworthiness levels in the population.

Most importantly, preferences affect both in positive and negative ways the outcomes in terms of performance, as well as the evolution of trust in

the population. A further step would now imply to evaluate how preferences themselves may evolve over time based on the available information and on past interactions. This would allow to investigate trust mechanisms more deeply.

4.5   Conclusions

To understand trust one needs not only to assess its consequences, but also to analyze its antecedents. In particular, for a better understanding of trust dynamics research needs to clarify whether trust is a well-defined attribute and eventually disentangle it from other attributes. More than this, research have to link trust to micro-economic theory concepts, such as preferences and/or beliefs. It is exactly these steps that, in turn, may motivate the impact of trust on agents decision process and on economic outcomes.

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

Even if the model relies on simplified assumption, results show some interesting insight on trust dynamics. In particular, we analyzed the role of preferences on individuals and population performance, both in terms of successful and fruitful interactions, and in terms of learning efficiency. Our interest has been on the social component of preferences, discovered to be strongly linked to trust relations.

We showed that preferences may strongly affect individuals and population performance both positively and negatively. On one side, stricter preferences, that means a higher interaction threshold, may protect individuals from failures in untrustworthy environments. On the other side, it may prevent from proper learning and fruitful interaction in trustworthy environments. Similarly, lower interactions thresholds may increase learning efficiency but simultaneously sustain a number of interactions well beyond the fruitful ones.

We believe more research is needed to investigate the effects of varying preferences. At the same time, an extension of the model would be needed to analyze the co-evolution of trust and trustworthiness.

References

G. A. Akerlof. The market for" lemons": quality uncertainty and the market mechanism. *The quarterly journal of economics*, pages 488–500, 1970.

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

K. J. Arrow. *The limits of organization*. Norton New York, 1974.

N. Ashraf, I. Bohnet, and N. Piankov. Decomposing trust and trustworthiness. *Experimental Economics*, 9(3):193–208, 2006.

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

S. Ba and P. A Pavlou. Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior. *MIS quarterly*, 26(3):243–268, 2002.

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

V. Bala and S. Goyal. Learning from neighbours. *The Review of Economic Studies*, 65(3):595–621, 1998.

V. Bala and S. Goyal. Self-Organization in communication networks. 1997.

V. Bala and S. Goyal. A theory of learning with heterogeneous agents. *International Economic Review*, 36(2):303–23, 1995.

P. Bateson. The biological evolution of cooperation and trust. In D. Gambetta, editor, *Trust: Making and Breaking Cooperative Relations, electronic edition. Department of Sociology, University of Oxford*, pages 14–30. 2000.

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

N. Berggren and H. Jordahl. Free to trust: Economic freedom and social capital. *Kyklos*, 59(2):141–169, 2006.

S. Beugelsdijk, H. L. F. de Groot, and ABTM van Schaik. *Trust and economic growth: a robustness analysis*, volume 56. Oxford Univ Press, 2004.

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.

S. Birkhchandani and D. Hirshleifer. Learning from the behaviour of others: Conformity, fads, and informational cascades. *Journal of Economic Perspectives*, 12(3):151–170, 1998.

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.

S. Bowles. Group competition, reproductive leveling, and the evolution of human altruism. *Science*, 314(5805):1569–1572, 2006.

S. Braynov and T. Sandholm. Contracting with uncertain level of trust. *Computational Intelligence*, 18(4):501–514, 2002.

L. Bruni and R. Sugden. Moral canals: trust and social capital in the work of hume, smith and genovesi. *Economics and Philosophy*, 16(01):21–45, 2000.

S. Buchegger and J. Y. Le Boudec. The effect of rumor spreading in reputation systems for mobile ad-hoc networks. volume 4, page 4.1, 2003.

V. Buskens. The social structure of trust. *Social Networks*, 20(3):265–289, 1998.

V. Buskens. Trust in triads: effects of exit, control, and learning. *Games and Economic Behavior*, 42(2):235–252, 2003.

V. Buskens and W. Raub. Embedded trust: Control and learning. *Group Cohesion, Trust and Solidarity*, 19:167–202, 2002.

V. Buskens and J. Weesie. Cooperation via social networks. *Analyse & Kritik*, 22(1):44–74, 2000.

J. V. Butler. Trust, truth, status and identity an experimental inquiry. *Not published*, 2008.

C. Camerer and E. Fehr. Measuring social norms and preferences using experimental games: A guide for social scientists. *Forthcoming in: Foundations of Human Sociality - Experimental and Ethnographic Evidence from 15 Small-Scale Societies, edited by Henrich, Boyd, Bowles, Camerer, Fehr, Gintis and McElreath*, 2002.

J. Carpenter. Measuring social capital: Adding field experimental methods to the analytical toolbox. *Social capital and economic development: Well-being in developing countries*, pages 119–137, 2002.

M. R Carter and M. Castillo. The economic impacts of altruism, trust and reciprocity: An experimental approach to social capital. *Wisconsin-Madison Agricultural and Applied Economics Staff Papers*, 448, 2002.

C. Castelfranchi and R. Falcone. Trust is much more than subjective probability: Mental components and sources of trust. volume 6, pages 10–12, 2000.

A. Chaudhuri, B. Sopher, and P. Strand. Cooperation in social dilemmas, trust and reciprocity. *Journal of Economic Psychology*, 23(2):231–249, 2002.

J. S Coleman. *Foundations of Social Theory*. Harvard University Press, 1990.

J. S Coleman. Social capital in the creation of human capital. *American Journal of Sociology*, 94(S1):95, 1988.

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

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

J. H. Davis, F. D. Schoormann, R. C. Mayer, and H. H. Tan. The trusted general manager and business unit performance: empirical evidence of a competitive advantage. *Strategic Management Journal*, page 563, 2000.

A. Diekmann and S. Lindenberg. Cooperation: Sociological aspects. *International Encyclopedia of the Social and Behavioral Sciences (4). Oxford: Pergamon-Elsevier*, 2000.

K. T Dirks. The effects of interpersonal trust on work group performance. *Journal of applyed psychology*, 84:445–455, 1999.

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

R. Dur and J. Sol. Social interaction, Co-Worker altruism, and incentives. *CESIFO Working Paper*, 2476, November 2008.

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

E. Fehr, U. Fischbacher, and M. Kosfeld. Neuroeconomic foundations of trust and social preferences: Initial evidence. *American Economic Review*, 95 (2):346–351, 2005.

D. P. Ford. Trust and knowledge management: The seeds of success. *Handbook on Knowledge Management*, 2002.

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

J. Geanakoplos, D. Pearce, and E. Stacchetti. Psychological games and sequential rationality. *Games and Economic Behavior*, 1(1):60–79, 1989.

E. L Glaeser, D. I Laibson, J. A Scheinkman, and C. L Soutter. Measuring trust. *Quarterly Journal of Economics*, 115(3):811–846, 2000.

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

M. Granovetter. Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, 91(3):481, 1985.

M. Granovetter. The impact of social structure on economic outcomes. *Journal of Economic Perspectives*, 19(1):33–50, 2005.

R. Guha, R. Kumar, P. Raghavan, and A. Tomkins. Propagation of trust and distrust. pages 403–412, 2004.

L. Guiso, P. Sapienza, and L. Zingales. People's opium? religion and economic attitudes. *Journal of Monetary Economics*, 50(1):225–282, 2003.

L. Guiso, P. Sapienza, L. Zingales, and V. D. Macelli. Cultural biases in economic exchange. *NBER working paper*, 2004.

L. Guiso, P. Sapienza, and L. Zingales. Does culture affect economic outcomes? *The Journal of Economic Perspectives*, pages 23–48, 2006.

L. Guiso, P. Sapienza, and L. Zingales. Trusting the Stock-Market. Technical report, January. Mimeo, 2007.

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.

R. Hardin. *Trust and trustworthiness*. Russell Sage Foundation Publications, 2002.

S. H Heap and Y. Varoufakis. *Game theory: a critical text*. Routledge, 2004.

J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, and R. McElreath. In search of homo economicus: Behavioral experiments in 15 small-scale societies. *American Economic Review*, pages 73–78, 2001.

J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky. BAYESIAN MODEL AVERAGING: a TUTORIAL. *STATISTICAL SCIENCE*, 14(4): 382–417, 1999.

K. Hoff and P. Pandey. Belief systems and durable inequalities: An experimental investigation of indian caste. *World*, 2004.

K. Hoff and P. Pandey. Discrimination, social identity, and durable inequalities. *American Economic Review*, 96(2):206–211, 2006.

H. J. Holm and A. Danielson. Tropic trust versus nordic trust: Experimental evidence from tanzania and sweden. *The Economic Journal*, 115(503):505–532, 2005.

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

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.

H. S. Jr James. The trust paradox: a survey of economic inquiries into the nature of trust and trustworthiness. *Journal of Economic Behavior and Organization*, 47(3):291–307, 2002.

A. Josang, C. Keser, and T. Dimitrakos. Can we manage trust. 2005.

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.

A. P Kirman and N. J Vriend. Learning to be loyal. a study of the marseille fish market. *Interaction and market structure: essays on heterogeneity in economics*, page 33, 2000.

S. Knack. Trust, associational life and economic performance. volume The Contribution of Human and Social Capital to Sustained Economic Growth and Well-being: International Symposium Report, Human Resources Development Canada and OECD, pages 172–202, 2001.

S. Knack and P. Keefer. Does social capital have an economic payoff? a Cross-Country investigation*. *Quarterly Journal of Economics*, 112(4): 1251–1288, 1997.

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

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.

D. Z. Levin and R. Cross. The strength of weak ties you can trust: The mediating role of trust in effective knowledge transfer. *Management Science*, pages 1477–1490, 2004.

S. Lindenberg. It takes both trust and lack of mistrust: The workings of cooperation and relational signaling in contractual relationships. *Journal of Management and Governance*, 4(1):11–33, 2000.

M. Lorenzen. Information cost, learning, and trust: lessons from co-operation and higher-order capabilities amongst geographically proximate firms. *DRUID Working Papers*, 98-21, 1998.

G. Lu, J. Lu, S. Yao, and Y. J Yip. A review on computational trust models for multi-agent systems. *The Open Information Science Journal*, 2:18–25, 2009.

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

R. Marimon, J. P Nicolini, and P. Teles. *Competition and reputation*. European University Institute, 1999.

Y. Mass and O. Shehory. Distributed trust in open multi-agent systems. *Lecture notes in computer science*, pages 159–174, 2001.

R. C. Mayer, J. H. Davis, and F. D. Schoorman. An integrative model of organizational trust. *Academy of management review*, pages 709–734, 1995.

K. A McCabe, M. L Rigdon, and V. L Smith. Positive reciprocity and intentions in trust games. *Journal of Economic Behavior and Organization*, 52(2):267–275, 2003.

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

A. K Mishra, R. Kramer, and T. Tyler. Organizational responses to crisis: The centrality of trust. In *Trust in organizations*, pages 261–287. London: Sage Publications, 1996.

L. Mui, M. Mohtashemi, C. Ang, P. Szolovits, and A. Halberstadt. Ratings in distributed systems: A bayesian approach. 2001.

L. Mui, M. Mohtashemi, and A. Halberstadt. A computational model of trust and reputation. pages 188–196, 2002.

B. Nooteboom. Innovation and inter-firm linkages: new implications for policy. *Research policy*, 28(8):793–805, 1999.

A. Ortmann, J. Fitzgerald, and C. Boeing. Trust, reciprocity, and social history: A re-examination. *Experimental Economics*, 3(1):81–100, 2000.

M. J. Osborne and A. Rubinstein. A course in game theory. *Cambridge/MA*, 1994.

P. A Pavlou and A. Dimoka. The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation. *Information Systems Research*, 17(4):392, 2006.

R. La Porta, F. Lopez-de-Silanes, A. Shleifer, and R. W Vishny. Trust in large organizations. *The American Economic Review*, 87(2):333–338, 1997.

R. La Porta, F. Lopez-de-Silanes, A. Shleifer, and R. W Vishny. Trust in large organizations. *Social Capital: A Multifaceted Perspective*, 1999.

R. D. Putnam. Tuning in, tuning out: the strange disappearance of social capital in america. *PS: Political Science and Politics*, pages 664–683, 1995.

M. Rabin. Incorporating fairness into game theory and economics. *The American Economic Review*, pages 1281–1302, 1993.

V. Radaev. How trust is established in economic relationships when institutions and individuals are not trustworthy (The case of russia). *Collegium*, 22:23, 2002.

H. Rainer and T. Siedler. Does democracy foster trust? *Journal of Comparative Economics*, 37(2):251–269, 2009.

M. Raiser, European Bank for Reconstruction, and Development. *Trust in Transition*. European Bank for Reconstruction and Development, 1999.

F. Roth. The effect of the financial crisis on systemic trust. *Intereconomics*, 44(4):203–208, 2009.

P. Sapienza, A. Toldra, and L. Zingales. Understanding trust. *NBER working paper*, 2007.

R. Sethi and E. Somanathan. Understanding reciprocity. *Journal of Economic Behavior and Organization*, 50(1):1–27, 2003.

J. B. Slemrod and P. Katuscak. Do trust and trustworthiness pay off? *NBER Working Paper*, 2002.

J. M Smith. *Evolution and the Theory of Games*. Cambridge University Press: Cambridge, 1982.

G. Tabellini. Culture and institutions: Economic development in the regions of europe. *Working paper*, 2005.

M. Tanis and T. Postmes. A social identity approach to trust: Interpersonal perception, group membership and trusting behaviour. *European Journal of Social Psychology*, 35(3), 2005.

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

F. Tonkiss. Trust, confidence and economic crisis. *Intereconomics*, 44(4): 196–202, 2009.

R. L Trivers. The evolution of reciprocal altruism. *The Quarterly Review of Biology*, 46(1):35, 1971.

G. Weisbuch, A. Kirman, and D. Herreiner. Market organisation and trading relationships. *The economic journal*, 110(463):411–436, 2000.

A. Whitby, A. J\textbackslashtextbackslashosang, and J. Indulska. Filtering out unfair ratings in bayesian reputation systems. *The Icfain Journal of Management Research*, 4(2):48–64, 2005.

B. Williams. Formal structures and social reality. In D. Gambetta, editor, *Trust: Making and Breaking Cooperative Relations, electronic edition, Department of Sociology, University of Oxford*. 2000.

B. Yu, M. P. Singh, and K. Sycara. Developing trust in large-scale peer-to-peer systems. pages 1–10, 2004.

A. Zaheer, B. McEvily, and V. Perrone. Does trust matter? exploring the effects of interorganizational and interpersonal trust on performance. *Organization Science*, pages 141–159, 1998.