Law, Informal Institutions and Trust

An experimental perspective

Recht, informele instituties en vertrouwen

Een experimenteel perspectief

Proefschrift ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam op gezag van
de rector magnificus
Prof.dr. H.A.P. Pols
en volgens besluit van het College voor Promoties

De openbare verdediging zal plaatsvinden op
maandag 14 december 2015 om 14.30 uur
door

Huojun Sun
geboren te Zhejiang, P.R. China

Erasmus University Rotterdam

Promotiecommissie

Promotoren:	Prof.dr. M.G. Faure LL.M.
	Prof.dr. M. Bigoni
Overige leden:	Prof.dr. C.W. Engel
	Prof.dr. W.H. van Boom
	Prof.dr. M. Rizzoli
Co-promotor:	Dr. A.M.I.B. Vandenberghe

This thesis was written as part of the European Doctorate in Law and Economics programme

A collaboration between


ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA


Universität Hamburg


ERASMUS UNIVERSITEIT ROTTERDAM

Contents

List of Figures

List of Tables

Introduction

Sophia is a graduate student from Shanghai University (China). She came to the University of Siena (Italy) to study the history of Renaissance art through the Marco Polo Programme. She communicates frequently with her culture-loving friends in Shanghai about things in Italy through WeChat (the Chinese version of Whatsapp). She enjoys being the local guide for Chinese travelers who visit Sienna and even gives them some hospitality.

One day, a friend in Shanghai asked Sophia if it was possible to import some Sassicaia wine to China. Sophia found some wine agents through searching engines, one of them is a small company set up by some Chinese students in Italy. The service they provide includes the purchase, customs declarations and logistics of importing Italian wine to China. The company also provides competitive prices in order to attract more business. Sophia contacted the company's sales representative, who was very keen to make the deal. Sophia was required to pay half the value of the contract as a deposit. She was not sure how much she should trust a firm set up by some students.

Sophia used to purchase some products through Wenzhou Merchants in Italy. With a long history of doing business in Italy, businesspeople from Wenzhou have established an informal network of gossip and social sanctions, so those who cheat on their clients may loss their reputation and even suffer from group boycotts enforced by their peers. However, the firm set up by Chinese students doesn't work the same way. Eventually, Sophia and this firm reached an agreement that she should pay the deposit to Paypal. Therefore, the deposit will only be transferred to the firm when the transaction is completed. A third-party payment system solves the trust issue.

Inspired by this little story, I produce three independent chapters to investigate three relevant questions in the dissertation. In the first Chapter, I study whether formal legal enforcement, such as using an impartial third-party payment system to enforce a payment agreement, can enhance the levels of trust and trustworthiness of contracting parties. The second Chapter examines how people who lack the protection of an effective legal system can establish informal institutions that depend on social sanctions to facilitate mutually advantageous exchanges. The third Chapter investigates whether highly trustful people, such as the heroine of the above story, are more or less sensitive than skeptical ones to cues on potential violations of trust (i.e. deliberate breach of contract).

The three chapters contribute to our understanding of how law, informal institutions and trust affect economic transactions. Simultaneous commercial exchange, which avoids disputes between parties over the date of payment and the conditions of delivery, leaves no space for legal intervention (Volckart and Mangels, 1999). Not many transactions, however, are carried out simultaneously, especially in the context of globalization where, on the one hand, contracting parties might never meet in person, and, on the other hand, they might not even be subject to the same legislative body. In sequential transactions, such as the story described above, after receiving goods or money, the second party may not give the first party something in return. Anticipating the potential exploitation, the first party may never enter the contractual relationship, and potential gains from trade are lost. To facilitate mutually advantageous transactions, contracting parties must find effective mechanisms *ex ante* to restrain temptation to renege *ex post* (Greif, 2006).

The medieval Law Merchant provides a good historical example of how merchants in transnational exchange use private-ordering institutions to mitigate commitment problems. In the tenth, eleventh, and twelfth centuries, while long-distance trade was emerging in Europe, merchants created, without the intervention of the state, an effective system of private enforcement – *lex mercatoria* (the *Law Merchant*) – to secure their transactions (Trakman, 1983).[1] In the absence of a centralized public authority with effective coercive power, merchant courts were established in several merchant communities or merchant fairs (the Champagne Fairs, for instance) to resolve legal disputes arising between merchants. The law merchants in courts kept account of those merchants who defaulted, and disseminated this information about cheating publicly. Their decisions could initiate group boycotts that collectively punished merchants with a public record of cheating, and "the threat of boycott of all future trade 'proved, if anything more effective than physical coercion'"(Benson, 1989, p.649). As a consequence, the *Law Merchant*, "far from being substitutes for the reputation mechanism, is to make the reputation system more effective as a means of promoting honest trade" (Milgrom et al., 1990, p.3). Besides medieval merchants, law-and-norm scholars find that other social groups also establish similar social sanction norms that depend on gossip networks to regulate their members' behavior: American fish wholesalers (Ellickson, 1989), Shasta County's ranchers (Ellickson, 1986; 1991), Mexican California's gold miners (Clay and Wright, 2005; Zerbe and Anderson, 2001), Wisconsin businesspeople (Macaulay, 1963) and many other merchant communities develop effective social norms (Bernstein, 1992, 2001; Greif, 1993; Landa, 1981; Richman, 2006).[2]

[1] Based on their historical investigation, Kadens (2012; 2015) and Sachs (2006) reveal that local authorities also contribute to the enforcement of commercial contracts.

[2] In law and economics literature, scholars usually use customary law, social norms and informal institutions interchangeably.

Social norms are widely recognized as effective mechanisms of social control alternative to a centralized coercive authority (Blocher, 2012). They are "not merely regularities of behavior, but obligatory regularities, the deviation from which incurs disapproval and other (non-legal) sanctions" (McAdams, 1996, p.2241).[3] This definition of social norm can reach back to the Medieval Ages, in which the medieval jurists believed that customary norm "consisted of a repeat behavior to which the relevant majority of the community had tacitly consented to be bound to perform" (Kadens, 2012, p.1163). The repetition of behavior, and "*tacitus consensus*" (i.e. "tacit consent"), which is later reinterpreted as "*opinion iuris*" (i.e. "a normative obligation", see Parisi, 1998), now become two essential elements necessary to construct a social norm in a variety of social science disciplines (Bicchieri, 2006; Cooter, 1998).

In recent years, social norms have received widespread attention among law-and-norm scholars,[4] and economic analysis of norms has provided several important insights. One might be termed the "shadow of the law" hypothesis dating back to Macaulay's (1963) work: law and norms are alternative means of social control, and contracting parties maximize their welfare within the law's shadow (Richman, 2012). In his seminal study of economic transactions between Wisconsin companies, Professor Macaulay (1963) finds that businesspeople tend to build trust-based relationships and resolve their disputes through non-legal means, particularly if potential litigation is much more expensive. Following Macaulay's work, many law-and-norm scholars investigate whether informal norms or institutions are more efficient than legal mechanisms. While some scholars advocate that, since informal norms or institutions spontaneously evolve as a result of merchants' dealings or industry consensus, they can respond more quickly to changing commercial environments and are thereby preferable to formal legal enforcement (Cooter, 1996, 1997; Kraus, 1997), other authors take a more critical view of norms and provide examples to justify the implementation of formal legal rules (Feldman, 2006; Mahoney and Sanchirico, 2001; Posner, 1996).

In addition to the "shadow of the law" hypothesis, law-and-norm scholars present another extralegal mechanism of informal norms or institutions – "order without law" – originating from Ellickson's (1986; 1991) work. In his influential anthropological field

[3] Posner and Rasmusen (1999) divide non-legal sanctions into six categories: "automatic sanctions", "guilt", "shame", "informational sanctions", "bilateral costly sanctions", and "multilateral costly sanctions".

[4] A series of prominent symposia is a good sign of the growing scholarly interest in this subject. *See, e.g.*, Themed Issue, *Social Norms: Theory and Evidence from Laboratory and Field*, Journal of the European Economic Association, Volume 11, Issue 3 (2013); Symposium Issue, *Custom*, Texas International Law Journal, Volume 46, Issue 3 (2013); Special Symposium Issue, *Custom and Law*, Duke Law Journal, Volume 62, Number 3 (2012); Symposium Issue, *Social Norms, Social Meaning, and the Economic Analysis of Law*, Journal of Legal Studies, Issue 27, Number S2 (1998); Symposium, *Law, Economics, and Norms*, University of Pennsylvania Law Review, Volume 144, Number 5 (1996).

study on the cattle-control norms in rural Shasta County, California, Ellickson (1986; 1991) shows that informal norms may work as an effective mechanism of social control: ranchers reject the county's formal legal rules and resolve their disputes through an informal network of gossip and social sanctions in which those who violate the community's norms suffer from social disdain and ostracism. Inspired by Professor Robert Ellickson's work, many researchers study important historical examples, which include *Medieval Iceland* (Hadfield and Weingast, 2013), *Gold rush California* (Clay and Wright, 2005) and *Medieval merchants* (Greif, 1993; Milgrom et al., 1990), in order to examine how alternative extralegal mechanisms achieve economic governance and social order at a time when centralized coercive nation-states do not exist. They find that in close-knit communities there exists a publicly accessible information center - such as the *Law Merchant* in medieval merchants' communities, and it helps to coordinate collective sanctions in order to secure deterrence of norm violation and promote cooperation. Similar social sanction norms are also observed in modern-day communities in developing countries where the governments fail to provide a sufficiently strong system of contract enforcement, and even abuse their authority to engage in profit-seeking punishment (Fafchamps, 1996; McMillan and Woodruff, 1999).

In much of the world, formal legal institutions that impartially enforce contract performance and impose sanctions on breaching party mitigate opportunism problems in non-simultaneous exchanges. Such institutions are costly to build and cumbersome to enforce, however, and parties frequently seek non-legal mechanisms to enforce their agreements even though they know that there are formal legal rules on which they can rely (Dixit, 2004). In addition, when transacting parties lack the protection of an effective legal system, they tend to establish informal norms or institutions to facilitate exchange. Two research questions are naturally arising: how do legal and non-legal mechanisms interact to sustain economic transactions? And how do social norms emerge in commercial communities where no formal legal institutions exist?

The "shadow of the law" hypothesis mainly focuses on the relative efficiencies of norms and legal rules for regulating contracting behavior. The issue of how formal legal mechanisms affect informal norms or institutions, however, is much less explored. The first Chapter, "*Law and Trust*," contributing to this literature, explores how formal contract enforcement affects norms of good conduct, such as trust and trustworthiness. This chapter first shows the macro-economic evidence that both good legal rules and high trust are crucial for economic exchange and development, but how legal mechanisms and trust interact is much less clear. It then considers the relationship between trust and formal contracts at the firm levels, finding that firms are inclined to adopt formal contracts when their performance is verifiable *ex post* but not necessarily observable *ex ante*, while contracting parties tend to build trust-based relationships and resolve their disputes through non-legal means when their performance is observable but costly to

verify. Since firms can freely choose trust-based relationships or formal contracts to do business with their partners, the causal effect of formal contracts on trust is still difficult to identify based on happenstance data from firms.

To identify a causal link between formal contract enforcement and trust, this chapter also surveys the evidence from relevant experimental studies. The experimental results generally show that formal contracts initiated by the principal crowd out the agent's intrinsic trustworthiness and reduce her beneficial behavior towards the principal. However, when the content of formal contracts is mutually agreed or recognized as legitimate by parties, or when the parties are involved in a highly heterogeneous market where no predominant norm of fairness exists, formal legal mechanisms work as complements for non-legal means. In addition, when commercial environment becomes extremely uncertain, parties are inclined to rely on both legal and non-legal mechanisms to facilitate their economic transactions, thereby increasing their economic welfare. Many subjects seem to anticipate the possible perverse effects of formal contracts, and as a consequence, they deliberately include indefinite clauses in their contracts to mitigate possible detrimental outcomes.

While many theoretical and empirical studies on "order without law" try to uncover the mechanism of how existing social norms coordinate collective punishment on norm-violators in order to secure social cooperation, few studies explore how social norms emerge in the absence of formal legal institutions. The second Chapter, "*A Fine Rule from a Brutish World? An Experiment on Endogenous Punishment Institution and Trust*," filling the gap in this literature, examines whether people who lack the protection of formal legal institutions are willing to endogenously adopt a collective punishment institution, and whether the endogenous adoption of collective punishment mechanism can help a society to coordinate an efficient outcome, characterized by high levels of trust and trustworthiness. This chapter first introduces a theoretical analysis of the consequences of the introduction of a collective punishment institution, which is based on Anderlini and Terlizzese (2012). In the model, trustees can choose to breach trust and grab the entire surplus, or to repay trust with a trustworthy action. Their decision depends on the "cost of cheating", which has two components: one is idiosyncratic; the other one increases with the number of trustworthy trustees present in the society, and can be interpreted as reflecting a social norm, which is exogenously given. The introduction of the mechanism transforms the trust game into a coordination game with a high-trust and a low-trust equilibrium, which are Pareto-ranked.

We build on this model, to study whether and how the endogenous adoption of the mechanism through majority voting can affect the equilibrium outcome. Our model predicts that all subjects, regardless of their preferences and expectations, should vote in favor of collective punishment, hence the mechanism should be endogenously introduced.

As a consequence, the outcome of the vote cannot be interpreted as a signal of others' intentions, and it should not matter whether the mechanism is exogenously imposed or endogenously adopted. An alternative, behavioral hypothesis is that voting can instead work as a coordination device. The endogenous adoption of collective punishment mechanism that punishes untrustworthy behavior could be taken as a signal for the general willingness to coordinate on a high-trust, high-trustworthiness equilibrium.

The theoretical model informs our empirical analysis, which is based on a laboratory experiment. The experiment comprises three games. The first is a binary trust game, in which the only equilibrium strategy is not to trust, and not to reciprocate. The second game is identical to the first one, but we exogenously introduce a collective punishment mechanism under which cheating is sanctioned and the severity depends on the number of other trustees in society who choose not to cheat. This creates a coordination game with a second, Pareto superior equilibrium with full trust and full trustworthiness. The third game is designed to study whether the possibility of endogenously adopting collective punishment by means of a majority-voting system facilitates coordination on the efficient equilibrium.

In line with the model, we find that the introduction of the collective punishment mechanism induces a significant increase in the levels of trustworthiness, and to a lesser extent also of trust. The endogenous introduction of the mechanism by means of a majority-voting rule does not significantly improve coordination on the efficient equilibrium. In contrast with our theoretical predictions, not all subjects seem to be able to anticipate the change in behavior induced by the introduction of collective punishment, and a majority of them vote against it. Subjects seem to be unable to endogenously adopt an institution which, when exogenously imposed, proves to be efficiency enhancing.

Besides well-functioning formal institutions and effective social norms, individual characteristics, particularly subtle psychological states, may also be crucial for establishing social trust, and even for recovering trust from a deliberate breach of contract. The standard efficient breach hypothesis, originating from the Holmesian "option view" of contract, suggests that a contractual obligation is merely an option: the promisor can freely choose the option to breach or pay damages equal to the difference between the value of performance and the contract price, leaving no space for psychological factors (Markovits and Schwartz, 2011; 2012). A series of experimental studies conducted by law and economics scholars, however, have shown that individual heuristics and moral intuitions can significantly influence contract compliance (Eigen, 2012), contract breach (Wilkinson-Ryan, 2010; 2011), the performance of assigned contracts (Wilkinson-Ryan, 2012), and the selection of alternative remedies for contract breach (Bigoni et al., 2014; Depoorter and Tontrup, 2012; Rachlinski and Jourden, 1998; Wilkinson-Ryan and Baron, 2009). In line with these experimental results, social psychologists have also found that

psychological mechanisms contribute to the enforcement of contracts (Robinson and Rousseau, 1994) and trust recovery following a breach of contract (Kuwabara et al., 2014; Lount et al., 2008; Schilke et al., 2013).

While previous experimental studies have considered how individual heuristics and moral intuitions affect economic transactions in short term, few studies focus on the effect of psychological factors on individuals' dynamic contracting behavior. To fill the gap in this literature, the third Chapter, "*Who Are More Naïve? High or Low Trustors*," explores whether high-trustors are less susceptible to the anticipated aversive emotions aroused by the potential betrayal and engage in acquiring useful information about their partners, thereby predicting others' trustworthiness more correctly than low-trustors. Butler et al. (2012) find that people tend to form beliefs on others' trustworthiness based on their own, which implies that high-trustors who hold relatively optimistic default expectation of others' trustworthiness easily trust more than they should and thereby are often cheated. In contrast with Butler et al.'s (2012) findings, Yamagishi (2011) experimentally show that, instead of being gullible, high-trustors are more sensitive to information that potentially reveals others' trustworthiness or untrustworthiness.

Yamagishi (2011) presents two potential explanations for a positive relationship between generalized trust and social intelligence. In the first hypothesis, he assumes that social intelligence is inherently heterogeneous between individuals in a society. Those who are socially intelligent can afford to expect that most people are trustworthy since they are highly sensitive to untrustworthiness cues, while socially unintelligent people who are less sensitive are better off assuming that unknown others are generally untrustworthy. His second hypothesis is that high-trustors tend to take more social risks and are, therefore, more vulnerable to exploitation, which pushes them to invest cognitive resources in cultivating social intelligence for detecting others' trustworthiness. After acquiring social intelligence to discern others' trustworthiness, they can afford to have a high level of generalized trust. In contrast, those who have not made such cognitive investments are slow in detecting the cues of untrustworthiness in their partners and thus are frequently betrayed in trust relations. The frequent experience of misplaced trust can lead to a progressive withdrawal from potentially fruitful, but risky interactions. As a result, they will be trapped in an "equilibrium of mistrust", thereby maintaining low default expectations of the trustworthiness of others.

We investigate this issue by means of a trust game experiment in which subjects repeatedly face opponents belonging to a high- or a low-trustworthiness group (i.e. either group A or B). We manipulate the feedback subjects receive on their partner's behavior, varying across treatments. In the *Baseline* treatment, we allow subjects to receive feedback from their partner unconditionally after they decide whether or not to trust. In the *Free Endogenous Feedback* treatment, subjects are allowed to decide whether to

acquire feedback about their partner's action after they decide whether to trust. In the *Contingent Feedback* treatment, subjects could be informed their partner' choice only if they decide to trust their partner. In all these three treatments, subjects are told whether their partner belongs to group A or B, but they don't know which of the groups contains the higher proportion of trustworthy people. In the *Ex-ante Feedback* treatment, however, trustors are told whether their partner belongs to the high or low trustworthiness group, before they make their choice.

This setup allows us to examine whether high-trustors are better than low-trustors at predicting others' trustworthiness, and identify the underlying mechanism that generates the behavioral difference between high- and low-trustors. The experimental results show that high- and low-trustors are equally able to distinguish which group is more trustworthy, and to condition their trust accordingly. Moreover, compared to their counterparts, high-trustors learn whom to trust or distrust faster not because they are better at processing the trustworthiness-related information, or that they deliberately collect differentiating social data through trusting more, but only because they are less susceptible to the anticipated aversive emotions aroused by the potential betrayal and thereby are more keen to acquire the useful information about their partner's actions.

Chapter 1 is single-authored. Chapters 2 and 3 have been written together with co-authors. My personal contributions to the co-authored Chapters are summarized in Table A.

Table A: Personal contribution to co-authored Chapters

	Chapter 2	Chapter 3
Idea	Leading	Leading
Experimental Design	Proportional	Leading
Data Collection	Proportional	Minor
Data Analysis	Proportional	Leading
Writing	Leading	Leading

Reference:

Anderlini, L. and Terlizzese, D., 2012, "Equilibrium Trust", EIEF working paper.

Benson, B., 1989, "The Spontaneous Evolution of Commercial Law," *Southern Economic Journal*, 55, 644-661.

Bernstein, L., 1992, "Opting Out of the Legal System: Extralegal Contractual Relations in the Diamond Industry," *Journal of Legal Studies*, 21, 115-157.

Bernstein, L., 2001, "Private Commercial Law in the Cotton Industry: Creating Cooperation through Rules, Norms, and Institutions," *Michigan Law Review*, 99, 1724-1790.

Bicchieri, C., 2006, *The Grammar of Society: The Nature and Dynamics of Social Norms*, Cambridge University Press.

Bigoni, M., Bortolotti, S., Parisi, F., and Porat, A., 2014, "Unbundling Efficient Breach," Coase-Sandor Institute for Law and Economics Working Paper.

Blocher, J., 2012, "Order without Judges: Customary Adjudication," *Duke Law Review*, 62, 579-605.

Butler, J., Giuliano, P., & Guiso L., 2012, "The Right Amount of Trust." EIEF Working Paper.

Clay, K., and Wright, G., 2005, "Order without Law? Property Rights during the California Gold Rush," *Explorations in Economic History*, 42, 155-183.

Cooter, R., 1996, "Decentralized Law for a Complex Economy: the Structural Approach to Adjudicating the New Law Merchant," *University of Pennsylvania Law Review*, 144, 1643-1696.

Cooter, R., 1997, "Normative Failure Theory of Law," *Cornell Law Review*, 82, 947-979.

Cooter, R., 1998, "Expressive Law and Economics", *Journal of Legal Studies*, 27, 585–608.

Depoorter, B., and Tontrup, S., 2012, "How Law Frames Moral Intuitions: The Expressive Effect of Specific Performance," *Arizona Law Review*, 54, 673-717.

Dixit, A., 2004, *Lawlessness and Economics: Alternative Modes of Governance*, Princeton, NJ: Princeton University Press.

Eigen, Z., 2012, "When and Why Individuals Obey Contracts: Experimental Evidence of Consent, Compliance, Promise, and Performance," *Journal of Legal Studies*, 41, 67-93.

Ellickson, R., 1986, "Of Coase and Cattle: Dispute Resolution Among Neighbors in Shasta County," *Stanford Law Review*, 38, 623-687.

Ellickson, R., 1989, "A Hypothesis of Wealth-Maximizing Norms: Evidence from the Whaling Industry," *Journal of Law, Economics, and Organization*, 5, 83-97.

Ellickson, R., 1991, *Order without Law: How Neighbors Settle Disputes*, Harvard University Press.

Fafchamps, M., 1996, "The Enforcement of Commercial Contract in Ghana," *World Development*, 24, 427-448.

Feldman, E., 2006, "The Tuna Court: Law and Norms in the World's Premier Fish Market," *California Law Review*, 94, 313-370.
Greif, A., 1993, "Contract Enforceability and Economic Institutions in Early Trade: the Maghribi Traders' Coalition", *American Economic Review*, 83, 525–48.
Greif, A., 2006, *Institutions and the Path to the Modern Economy: Lessons from Medieval Trade*, Cambridge, MA: Cambridge University Press.
Hadfield, G., and Weigast, B., 2013, "Law without the State," *Journal of Law and Courts*, 1, 3-34.
Kadens, E., 2012, "The Myth of the Customary Law Merchant," *Texas Law Review*, 90, 1153-1206.
Kadens, E., 2015, "The Medieval Law Merchant: The Tyranny of a Construct," *Journal of Legal Analysis*, forthcoming.
Kraus, J., 1997, "Legal Design and the Evolution of Commercial Norms," *Journal of Legal Studies*, 26, 377-411.
Kuwabara, K., Vogt, S., Watabe, M., and Komiya, A., 2014, "Trust, Cohesion, and Cooperation After Early Versus Late Trust Violations in Two-Person Exchange," *Social Psychology Quarterly*, 77, 344-360.
Landa, J., 1981, "A Theory of the Ethnically Homogeneous Middleman Group: An Institutional Alternative to Contract Law," *Journal of Legal Studies*, 5, 349-362.
Lount, R., Zhong, C., Sivanathan, N., and Murnighan, K., 2008, "Getting Off on the Wrong Foot: The Timing of a Breach and the Restoration of Trust," *Personality and Social Psychology Bulletin*, 34, 1601-1612.
Macaulay, S., 1963, "Non-contractual Relations in Business: A Preliminary Study." *American Sociological Review*, 28, 55-67.
Mahoney, P., and Sanchirico, C., 2001, "Competing Norms and Social Evolution: Is the Fittest Norm Efficient?" *University of Pennsylvania Law Review*, 149, 2027-2062.
Markovits, D., and Schwartz, A., 2011, "The Myth of Efficient Breach: New Defenses of the Expectation Interest," *Virginia Law Review*, 97, 1939-2008.
Markovits, D., and Schwartz, A., 2012, "The Expectation Remedy Revisited," *Virginia Law Review*, 98, 1093-1107.
McAdams, R., 1996, "Group Norms, Gossip, and Blackmail," *University of Pennsylvania Law Review*, 144, 2237-2292.
McMillan, J., and Woodruff, C., 1999, "Dispute Prevention without Courts in Vietnam," *Journal of Law, Economics, & Organization*, 15, 637-658.
Milgrom, P., North, D., and Weingast, B., 1990, "The Role of Institutions in the Revival of Trade: The Law Merchant, Private Judges, and the Champagne Fairs," *Economics and Politics*, 2, 1-23.
Parisi, F., 1998, "Customary Law", in *The New Palgrave Dictionary of Economics and the Law*, ed. Peter Newman, Palgrave Macmillan.
Posner, E., 1996, "Law, Economics, and Inefficient Norms," *University of Pennsylvania Law Review*, 144, 1697-1744.

Rachlinski, J., and Jourden, F., 1998, "Remedies and the Psychology of Ownership," *Vanderbilt Law Review*, 51, 1541-1582.
Richman, B., 2006, "How Community Institutions Create Economic Advantage: Jewish Diamond Merchants in New York," *Law & Social Inquiry*, 31, 383-420.
Richman, B., 2012, "Norms and Law: Putting the Horse before the Cart," *Duke Law Journal*, 62, 739-766.
Robinson, S., and Rousseau, D., 1994, "Violating the Psychological Contract: Not the Exception but the Norm," *Journal of Organizational Behavior*, 15, 245-259.
Sachs, S., 2006, "From St. Ives to Cyberspace: The Modern Distortions of the Medieval Law Merchant," *American University International Law Review*, 21, 685-812.
Schilke, O., Reimann, M., and Cook, K., 2013, "Effect of Relationship Experience on Trust Recovery Following a Breach," *Proceedings of the National Academy of Science*, 110, 15236-15241.
Trakman, L., 1983, *The Law Merchant: The Evolution of Commercial Law*, Fred B. Rothman & Co.
Volckart, O., and Mangels, A., 1999, "Are the Roots of the Modern *Lex Mercatoria* Really Medieval?" *Southern Economic Journal*, 65, 427–450.
Wilkinson-Ryan, T., 2010, "Do Liquidated Damages Encourage Breach? A Psychological Experiment," *Michigan Law Review*, 108, 633-672.
Wilkinson-Ryan, T., 2011, "Breaching the Mortgage Contract: The Behavioral Economics of Strategic Default," *Vanderbilt Law Review*, 64, 1547-1583.
Wilkinson-Ryan, T., 2012, "Transferring Trust: Reciprocity Norms and Assignment of Contract," *Journal of Empirical Legal Studies*, 9, 511-535.
Wilkinson-Ryan, T., and Baron, J., 2009, "Moral Judgment and Moral Heuristics in Breach of Contract," *Journal of Empirical Legal Studies*, 6, 405-423.
Yamagishi, T., 2011, *Trust: The Evolutionary Game of Mind and Society*, Springer.
Zerbe, R., and Anderson, L., 2001, "Culture and Fairness in the Development of Institutions in the California Gold Fields," *Journal of Economic History*, 61, 114-143.

Chapter 1

Law and Trust

Abstract. This survey addresses the question of whether formal legal enforcement crowds out or crowds in the amount of trust in a society. Based on a review of relevant empirical studies in the literature on macroeconomics, inter-firm cooperation and laboratory experiments, it can be concluded that find that formal legal mechanisms, especially formal contracts backed by a powerful authority, normally work as substitutes for trust, rather than complements, except when they are perceived as legitimate, or when there are no strong social norms of fairness (i.e. the population in a society is considerably heterogeneous), or when the environment in which repeated commercial relationships take place becomes highly uncertain.

Key words: Crowding-out, Legal Rule, Contract Enforcement

This chapter is based on Sun, H., forthcoming, "Law and Trust," *International Journal of Applied Behavioral Economics*. I am grateful for valuable comments and suggestions by Maria Bigoni, Ann-Sophie Vandenberghe, and participants of the EMLE Midterm Meeting 2014 at the University of Bologna.

1.1 Introduction

Interpersonal trust is an essential feature of social life, which pervades friendship relations, family relations, and commercial relations. During the past decades, trust has received widespread attention across disciplines,[5] and researchers have shared consensus on the importance of trust in the conduct of human affairs.

Indeed, trust is the keystone for successful economic development. Nobel laureate economist Kenneth Arrow (1972, p.357) has emphasized that "virtually every commercial transaction has within itself an element of trust," and that "much of the economic backwardness in the world can be explained by the lack of mutual confidence." A growing body of literature has revealed that aggregate measures of trust at the country level are positively correlated with important economic variables such as the GDP growth, the provision of public goods, or the size of firms. Knack and Keefer (1997) find positive correlations between a country's average annual GDP and a measure of trust from the World Values Survey for a sample of 29 market economies between 1980 and 1992. Within a specific country - the U.S., Dincer and Uslaner (2010) find a robust relationship between trust and economic growth across American states. Recently, a series of studies conducted by Guiso and his coauthors provide fruitful microeconomic evidence on the role of trust in economic activities. Guiso et al. (2004, 2008a) show that a larger share of trusting people is positively correlated with the development of financial market across countries. Less trustful individuals are less likely to participate in stock market and, conditional on buying stock, they purchase less, which limits the size of a country's stock market. Guiso et al. (2009) use data on bilateral trust between European countries, and find that higher bilateral trust tends to breed more trade between two countries. In addition, they also find that the effect is stronger for more trust-intensive goods.

While there is a clear consensus in the literature that trust is crucial to economic success, the question of how a society achieves a high level of trust is less clear. In closed communities or specific industries, high trust and trustworthiness levels are easily self-sustained between parties without legal interventions because long-term payoffs conditional on cooperation within the existing relationship exceed gains from short-term defection. In his seminal study of agreements between Wisconsin companies, Macaulay (1963, 1985) finds that many agreements are non-contractual, with no legal enforcement. Businesspeople trust and honor each other because they want to sustain their long-term business relationships. By contrast, the problem of

[5] In the past decades, trust, especially among strangers, has become a vital topic in sociology (Coleman, 1990; Fukuyama, 1995; Gambetta, 1988), psychology (Deutsch, 1958, 1962; Sullivan and Transue, 1999), economics (Arrow, 1974; Zak and Knack, 2001), political science (Putnam, 1993; Uslaner, 2002), neuroscience (King-Casas et al., 2005; Kosfeld, et al., 2005; Zak et al., 2004), medical and bioethics studies (Hall, 2002; Hall et al., 2004), management science (Zaheer and Venkatraman, 1995; Zaheer, et al., 1998, 2003) and legal studies (Blair and Stout, 2001; Mitchell, 2001; Rose, 1995).

trust in others is relatively pronounced in large, anonymous societies, where effective contract enforcement is highly demanded in order to promote mutually advantageous transactions. As Nobel laureate economic historian Douglass North argues, the ability of societies to develop effective, low-cost enforcement of contracts is the most important source of national prosperity (North, 1990). But, how or in what way do formal legal mechanisms, in particular formal contract enforcement, affect the levels of trust and trustworthiness in a society? How do law and trust interact to sustain economic exchanges?

Many economic transactions are non-simultaneous. After receiving goods or money, the second party is supposed to give the first party something in return. In the absence of formal contract enforcement, the second party has incentive to renege. Anticipating the potential exploitation, the first party may never enter the exchange relationship, and potential gains from trade are lost. Trust and formal contracts are two fundamental approaches to solve the commitment problem and to facilitate mutually advantageous transactions. In particular, parties may reach a mutual agreement on an exchange and enforce the agreement without centralized coercive enforcement, because the threat of losing future transaction opportunities or informal sanctions makes parties comply with the mutually agreed arrangement. Alternatively, parties may rely on formal legal mechanisms to enforce their partners' performance and to impose remedies in case of a breach. The expectation of enforcement of remedies for breaches induces parties more likely to be trustful and trustworthy (Baird, 1990). Besides the rules on remedies for contract breach, there are other types of contract rules that may secure profitable transactions, like the rules on fraud: a contract is void when a seller lied about the quality of the goods. A buyer will be more willing to enter a transaction when he knows that the law will void the contract when it turn out that the seller was dishonest.

While some scholars advocate that relying on legal mechanisms to regulate business relationships may signal lack of trust and reduce the intrinsic trustworthiness (Macaulay, 1963; Gulati, 1995), other researchers argue that formal legal rules can establish a mutually shared belief about contract obligations between contracting parties and facilitate to coordinate their behavior, which enhances the levels of trust of parties (Das and Teng, 1998). This chapter contributes to the literature on law and trust, and aims to address the question of whether and on what conditions formal contract enforcement crowds out or crowds in the amount of trust in an economy for the purposes of facilitating efficient economic transactions and economic growth.

To achieve the research goal, I first present the basic concept of trust on which our main discussion is based in the following section. In Section 1.3, I then investigate how trust and formal legal enforcement influence economic growth and development, and how they interact to sustain economic exchanges. Besides the macroeconomic evidence, I also discuss the firm-level evidence on the interaction effect of trust and formal contracts in Section 1.4. Since the existing empirical evidence can not provide

precise causal link between legal enforcement and trust, in Section 1.5 I show how relevant experimental evidence can help us to solve the current ambiguity in the literature. Section 1.6 provides concluding remarks.

1.2 The Concept of Trust

1.2.1 Definition

Before entering the debate about law and trust, one must first answer the question of how "trust" is to be defined and measured, which is not an easy task.[6] The literature in social sciences offers many definitions of trust, depending on the specific context and content of the study. In his influential book, Uslaner (2002) specifies two types of trust: strategic trust and moralistic trust. Strategic trust occurs when people have past experience with their current partner and also the expectation of payoffs conditional on long term cooperation, or when some effective external enforcement mechanism (such as an arbitrator, or the courts) regulates the behavior of transaction parties. Strategic trust reflects people's expectations about how others will behave under legal and non-legal mechanisms. According to the perspective of strategic trust, people trust others if and only if it is in their material self-interest to do so, which can also been defined as "calculative trust" (see Williamson, 1993). Besides the strategic or calculative aspect of trust, Uslaner (2002) suggests that trust also has a moral dimension. When involved in an economic transaction, people treat their partners as part of the moral community to which they belong. As a result, they are morally required to hold optimistic beliefs about others' trustworthiness and thereby trust undoubtedly. In the psychological literature, moralistic trust is also called "affect-based trust" or "altruistic trust" (see McAllister, 1995).

To capture an effective definition of trust, which accommodates two essential aspects of trust discussed above, I follow Rousseau et al. (1998, p.395) and define trust as "a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or the behavior of another,"[7] a definition based on a review of definitions in different social sciences. According to this definition, the non-contractible expectation of positive reciprocity on the part of the trustee is essential to the concept of trust. In particular, forming an entirely new exchange relationship can create the possibility of mutual benefit if the trustee is trustworthy; a decision to trust also implies the risk of substantial loss to the trustor if the trustee acts in an untrustworthy manner. In addition, the willingness to take such

6 Although there are different kinds of trust (for example, trust in government, or trust in strangers) in the literature, I only concern about the interpersonal trust in this paper. In the interaction between firms, I presume that the decision to trust is made by real persons (CEO, for instance), based on their expectations about others' trustworthiness.

7 Similarly, Mayer et al. (1995, p.712) define trust as "the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party."

social risks is also required for trust to emerge, which may be related to one's attitude toward general risk (Eckel and Wilson, 2004; Karlan, 2005; Schechter, 2007).[8]

1.2.2 Measures

Although Rousseau et al. (1998)'s definition of trust well captures the behavior people engaged in social interactions on fiduciary issues, it does not answer the question of how to precisely measure trust. In the last two decades, experimental method has become the basic tool to measure trust and trustworthiness. A trust game designed by Berg and his coauthors (Berg et al., 1995) has come to dominate the field. In the trust game experiment, two parties are involved in a sequential exchange in which there is no contract to enforce agreements. Subjects are endowed with \$10, anonymously matched and assigned to either the role of trustor or trustee. At stage one of the game, the trustor may either pass nothing, or any portion x of the endowment ($0 \leq x \leq 10$) to the trustee. The trustor then keeps $10 - x$, and the experimenter triples the remaining money so that 3x is passed onto the trustee. In stage two, the trustee may either pass nothing, or pass any portion y of the money received ($0 \leq y \leq 3x$) back to the trustor.[9] Given the fact that the trustee in the game is under no obligation to return anything, the amount sent by the trustor has been seen as a proxy for "trusting" behavior; the greater is the amount sent, the more trustful is the trustor. Similarly, the amount returned to the trustor by the trustee is used as a measure of "trustworthiness".[10]

This behavioral measure captures two main dimensions of trust - the non-contractible expectation about an unknown partner's trustworthiness and the willingness to be vulnerable to possible exploitation. At its core, the trust game mirrors a basic moral hazard problem, and reflects many different real world settings such as making investments in a company, or lending money to someone. The non-simultaneous exchange relationship can be represented by the sequence of choices in this game: the first player, like the promisee, has to decide whether to enter the relationship, not knowing whether the second player will perform. The second player, like the promisor, has already received the benefit of the deal and must decide whether to reciprocate the first player's trust. In the literature, this two-stage trust game has become a popular and frequently replicated measure of trust and trustworthiness.[11]

---

[8] Recent experimental evidence (Bohnet and Zeckhauser, 2004; Bohnet et al., 2008) shows that the decision to trust is not only determined by risk aversion, but also by betrayal aversion, that is, the fear of being betrayed or exploited by another in a social exchange.

[9] Variations of this game are common. For example, Fetchenhauer and Dunning (2012) adopt a binary format, as above, in which the trustor must keep or give all the endowment and the trustee must keep all or give half back.

[10] Recently, legal scholars (e.g. Wilkinson-Ryan, 2012) have adopted the trust game experiment to model contractual exchange, and then to test the effect of reciprocity norms on contract performance.

[11] Johnson and Mislin (2011) collect data from 162 replications of the Berg et al. (1995)'s trust game involving more than 23,000 subjects, and then conduct a meta-analysis of these games to identify the effect of experimental protocols and geographic variation on this behavioral measure of trust and trustworthiness, finding robust evidence that subjects always give some amount to their partner and are

Albeit reliable, experimental measures of trust suffer from major limitations: they are time consuming and expensive to conduct, and consequently they can involve only a small and often non-representative sample of the population. This makes them unsuitable for empirical studies based on large datasets, such as those in the literature on economic growth. An alternative approach is that of relying on attitudinal survey questions, such as those asked in the American General Social Survey (GSS) which has measured trust annually since 1972, the World Values Survey (WVS) which has been widely used to measure cross-cultural differences in trust, and the German Socioeconomic Panel (SOEP) (see Glaeser, et al., 2000; Sapienza, et al., 2013; Naef and Schupp, 2008). The trust question in the GSS or in the WVS, which has been most widely used, asks: *Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?* The survey respondents can answer in a binary way to this question by agreeing either with "*Most people can be trusted*" or with "*Can't be too careful*." This trust measure has been seriously criticized by many social science scholars (Miller and Mitamura, 2003; Fehr, 2009) pointing out that a risk-averse or cautious person may share the view that "Most people can be trusted" but that at the same time prudence or risk aversion may induce the person to say "Can't be too careful" because the person engages in avoiding small probability risks that have large payoff consequences.

To rule out having reasonable people agree with both answer categories, many social science scholars turn to "one-dimensional" questions that directly distinguish between trust and distrust (Miller and Mitamura, 2003; Yamagishi et al., 1998; Yamagishi and Yamagishi, 1994). For example, Yamagishi and his colleagues ask questions such as "do you agree that most people are trustful of others" with five answer categories (agree fully, agree somewhat, neutral, disagree somewhat, disagree fully). In view of the problems inherent in the GSS question, these new measures of trust are likely to be better.

The positive answers to trust questions in the survey are always interpreted as generalized trust, that is, the default expectation of other people's trustworthiness in a society (Rotter, 1980). And participants in the trust game decide whether to trust their partner often based on their general expectation of the trustworthiness of unknown others. Therefore, subjects' trustful behavior in the game should be positively correlated with measures of generalized trust towards strangers. Indeed, Sapienza et al. (2013) observe a positive relationship between stated beliefs in the trust game and responses to trust questions in the survey. In addition, several experimental studies have also found that trust as measured by the survey is a significant predictor of trusting behavior in the trust game (Fehr et al., 2003; Bellemare and Kroeger, 2007).[12]

---

positively reciprocated.

[12] However, Glaeser et al. (2000) find that the association between survey and behavioral measures of trust is weak. One possibility is that student subjects used in the experiment already knew each other before playing the game while trust questions measure their generalized trust towards strangers.

1.3 The Role of Trust, Law in Economic Growth

In the macroeconomic literature on development, while there is a clear consensus that both good formal institutions and high societal trust are crucial to promote trade and development, the question of how legal institutions and trust interact and co-evolve is much less clear. In this section I briefly introduce evidence about the roles of trust and legal institutions in economic performance, then I focus on how their interaction affects economic growth.

1.3.1 Trust and Growth

Recent empirical studies have yielded evidence of a sizable variability in the extent to which people trust others across countries as well as within countries (Algan and Cahuc, 2013; Tabellini, 2010). In general, northern European countries lead the ranking with high average levels of interpersonal trust, whereas people in African and south American countries are less likely to trust others. Substantial variability in the levels of trust also emerges within countries, for example, the trust level of Italians is almost twice as high in Trentino Alto Adige as it is in Sicilia (Tabellini, 2010). A main stylized fact from the studies of trust and economic development across countries is that income per capita is positively correlated with the level of trust in a country, where trust is measured by survey techniques (Guiso et al., 2008b; Knack and Keefer, 1997; La Porta et al., 1997; Tabellini, 2008).

In this literature, trust is argued to improve economic development through several channels. Markets in the real world always face many trading frictions since contracts are not perfectly enforceable, which generates demand for regulation even when people realize that the government is corrupt (Aghion et al., 2010). High levels of trust and trustworthiness work as a lubricant to reduce social frictions without investing resources in monitoring and contract enforcement, and thus permit more investment in production (Knack and Keefer, 1997; La Porta et al., 1997). In addition, high social trust also expands the scope of exchange and increases efficiency by diverting trade to less connected but more efficient traders (Guiso et al., 2009; La Porta et al., 1997).

Interestingly, in the seminal paper by Knack and Keefer (1997) the authors unexpectedly find that the interaction effect of trust and initial income per capita on economic growth is statistically significant, indicating that trust has a stronger effect on growth in poor countries than in higher income countries. One explanation is that low-income countries lack credit markets and enforceable legal rules, and thus trust is especially important for growth. The studies on the relatively new market economies of Russia and Eastern Europe have confirmed this hypothesis (Hendley et al., 2000; Johnson et al., 2002): as the law becomes prevalent in dealing with commercial transactions, the importance of long-term relationship between parties disappears, i.e.

relation-based trust is less necessary, implying that trust and law are substitutes.

While exogenous environmental factors such as climate variability (Durante, 2009) and the quality of institutions such as legal enforcement (Buggle, 2013; Becker et al., forthcoming) are likely to shape trust, both of them have also been found to affect economic outcomes (Acemoglu et al., 2001; Hall and Jones, 1999). Therefore, identifying the causal relationship between trust and economic growth by field data becomes difficult. The most common strategy to establishing this causality in the literature is using instrumental variables for trust.[13] Some instruments have included the degree of "ethno-linguistic homogeneity" in a country (Knack and Keefer, 1997), the prevalence of "hierarchical religions", such as Catholicism, in a country (La Porta et al., 1997), or historical events such as quality of education and past political institutions in the history of Europe (Tabellini, 2010) and slave trade in the history of Africa (Nunn, 2008). While all of these variables are strongly correlated with trust, they may plausibly have a direct impact on growth, which induces us to interpret the previous results with some caution. A more recent approach is using the trust levels of different waves of immigrants to the US, as time varying instruments for trust in the home countries of the immigrants (Algan and Cahuc, 2010). This approach makes the instruments reasonable, because the trust levels between immigrants and the peers in their home countries are highly correlated but the trust levels of immigrants are less likely to directly affect the economic performance in their home countries.

1.3.2 Legal Systems and Growth

Another branch of growth literature emphasizes the impact of institutions such as legal systems on economic performance. There is much evidence that different past legal systems do have long-term effects on economic development (Acemoglu et et., 2001).[14] Specifically, La Porta et al. (1998, 1999) document empirically that common law countries have better property rights and more developed financial markets than countries with civil law, and thus achieve more successful economic performance. They argue that common law is more likely to protect minority shareholders strongly and to allow entrepreneurs to establish a business easily, and thereby intrinsically superior to civil law.[15]

Next step to be addressed is to examine the role of trust in the relationship between legal systems and economic growth. To exploit this point, researchers often focus on the effect of law on trust. Cross (2005) uses the International Country Risk Guide (ICRG) to estimate the "rule of law" for nations. After controlling for other

[13] In economics, the method of instrumental variables (IV) is used to estimate causal relationships when controlled experiments are not feasible. Generally, there are two main requirements for using an IV: first, the IV must be correlated with the explanatory variable(s) concerned by researchers; second, the IV must have no direct impact on the explained variable.

[14] For recent survey papers, see Ogilvie and Carus (2014) and Spolaore and Wacziarg (2013).

[15] According to Posner (1973), case law in the common law framework allows for dynamic adaptation to innovation in economic activities, thereby achieving economic performance more successfully than civil law.

demographic factors, it shows that the rule of law is associated with significantly higher levels of generalized trust. This result is consistent with a series of studies on trust across countries (Algan and Cahuc, 2014) showing that trust has a strong positive correlation with the quality of the legal system, and that the correlation is robust to using different measures of institutional quality often used in economic literature, such as the rule of law, the strength of property protections, the control of corruption, and the enforcement of contracts.

Based on macro data, however, it is still difficult to identify whether high trust is the cause or the consequence of a good legal system. One possibility is that high trust may simply reflect better institutional design, because the traditional survey questions on trust are usually context-free, i.e. without mentioning legal protection in a country, and people are highly likely to estimate the level of trustworthiness in a society based on their heuristics about legal enforcement (Guiso et al., 2011; Ho and Huffman, 2013). Another opposite possibility is that trust allows parties to exploit extra surplus that would otherwise need to be spent on monitoring (Posner, 2002), and these resources can in turn be used not only for investment in physical capital, but also potentially for investment in better legal environments. With more efficient and effective legal institutions, higher trust levels will be supported. Therefore, more evidence is needed to provide a definitive answer.

1.4 The Role of Trust, Law in Inter-firm Cooperation

Since firms work as engine for economic growth, it is interesting to investigate the relationship between trust and legal enforcement at the firm level. In this section I briefly summarize the evidence of the active roles of trust and formal contracts in inter-firm cooperation [16] and examine the conditions under which these two enforcement strategies are substitutes or complements.

1.4.1 Trust and Formal Contracts

Relation-based trust and formal contracts are two fundamental approaches to promote the inter-firm cooperation. In business relationships, contracting parties usually choose between these two approaches to encourage performance of contract obligations. The relative efficiency of these two approaches depends on the specific circumstances. The formal contracts rely on enforcement of the written agreements by courts and threaten contractually specified damages against the breaching party for non-performance. To achieve this goal, courts must collect reliable information required to rule on contractual disputes and adjudicate claims lawfully, thus properly reflecting the parties' rights and duties. Since courts cannot directly observe the

[16] During the past decades, there has been many studies discussing inter-firm cooperation, especially in management research, see Arrighetti et al. (1997), Badawi (2010), Das and Teng (1998, 2001), Dyer (1997), and Mellewigt et al. (2007).

detailed interactions between firms, they have to rely on information offered by the firms. In practice, therefore, the costs of formal verification limit the applicable scope of contractual-based approach.[17] The trust-based approach, in contrast, depends entirely on private behavior - each party's ability to directly observe the other's actions and willingness to punish misbehavior and/or reward good deeds directly as soon as they are noted.

In the literature, two different and competing views can be distinguished: trust and formal contracts are substitutes (Macaulay, 1963; Gulati, 1995) or trust and formal contracts are complements (Schmidt and Schnitzer, 1995; Poppo and Zenger, 2002). According to the "substitution" view, greater trust would manifest itself through less complete contracts, whereas according to the "complementation" view, greater trust would manifest itself through more detailed contracts. There are two possible reasons why having access to state-backed contracts could crowd out trust. One is the preference-based explanation: since parties involved in a commercial relationship are interdependent, they may be highly altruistic towards each other and value their partner's payoff seriously. Introducing a detailed contract could signal lack of trust and generate negative reciprocity, thereby crowding-out the pre-existing trust in their relationship. The other is the norm-based explanation: there may exist a norm of social sanctions in a specific industry, and people are willing to punish those who abuse their commercial partner's trust through ostracism or boycott (Greif et al., 1994; Greif, 2006). Requesting to sign formal contracts enforced by coercive authorities may destroy the prevalent norm, thereby spreading distrust among parties. Alternatively, in the perspective of complementarity, the detailed contracts reflect mutual expectations, and establish a common understanding about actions that are deemed wrongful, thereby serving as a coordinating device to complement trust (Das and Teng, 1998). In addition, formal contracts can also reduce the gains from short-term defection and secure the value of relation-based transactions (Baker et al., 1994; Baker and Choi, 2015).

1.4.2 Empirical Evidence

To examine how the relation-based trust and contract damages are likely to interact, Badawi (2010) collects a sample of 89 franchise documents from the Uniform Franchise Offering Circulars (UFOCs) in 2007. The empirical evidence supports the view of substitution, indicating that when trust-based approach is likely to be effective, franchisors tend to forgo the enforcement of formal contracts, and vice versa. Specifically, when relational punishments successfully induce the franchisees to comply with contract obligations and save transaction costs compared to formal enforcement, the franchisors do not heavily rely on credible damage threats – in his sample, only 20 of the 89 contracts include liquidated damages. Liquidated damages

[17] For example, as Ben-Shahar and Bernstein (2000) argued, secrecy consideration may deter aggrieved parties from offering detailed information to courts, leading to under-deterrence of parties in breach.

become attractive only when trust-based mechanisms are difficult to be implemented. The author finds that the contract damages are most prevalent in motel and real estate brokerage franchises, because in these industries the franchisees can easily switch their assets to another franchise, which undermines the franchisor's ability to threaten his/her franchisees by potentially informal punishments (for example, terminating the relationships).

Contrary to Badawi's findings, Mellewigt et al. (2007) use survey data on human resource management from 600 randomly selected companies of a variety of industries, and demonstrate that contractual control does not crowd out trust. In their study, they find that the parties often use formal contracts to respond to outside contingencies. In addition, formal contracts also serve to clarify parties' duties and coordinate their behavior. As a consequence, formal contracts and relation-based trust are complementary and mutually reinforcing. This result reflects the recent emerging commercial practice called "braiding" (Gilson et al., 2010): in order to improve the collaboration in the contemporary commercial environment characterized by high holdup risk and outcome uncertainty, firms tend to write complex contracts that integrate formal and informal arrangements in a way that allows each to assess the capability and characteristics of the other, thereby facilitating both parties to respond better to unforeseen circumstances.[18]

The issue of how trust impacts the business world becomes, without doubt, very important, especially in the context of globalization where, on the one hand, contracting parties might never meet in person, and, on the other hand, they might not even be subject to the same legislative body. There is a need for identifying the causal effect of formal contracts on trust between firms. The previous empirical evidence based on happenstance data that is collected through the survey or firms' records, however, is ambiguous. In the next section, it is investigated whether the ambiguity in the literature can be disentangled by looking at experimental evidence.

1.5 Lessons from Experimental Evidence

Both formal contract enforcement and trust are very important to economic development and inter-firm cooperation, however, the interaction effects of these two mechanisms on economic performance are still highly debated: while many studies show that formal contract enforcement may enhance trust in a economy, other scholars find that formal contracts tend to reduce intrinsic trustworthiness and lead to less trust and a worse outcome. Identifying the causal effect of formal contract enforcement on trust with happenstance data is difficult because not only contract enforcement can influence the levels of trust in societies, but also people's trust

[18] Recently, Barnett (2015) have found that, in the Hollywood motion-picture industry, a hybrid instrument intertwining formal and informal contracts is widely adopted to respond to the transactional hazards of an environment where neither formal contract nor reputation effects can effectively enforce parties to perform their contract obligations.

beliefs can induce a demand for legal intervention.

Economic experiments allow us to disentangle the competing explanations and identify the detailed conditions under which formal contracts may crowd in or out societal trust. In this section, I review most of relevant experimental evidence, and find that using formal contracts to increase trust may have the opposite effect except when there are no salient social norms of fairness (i.e. the population in an environment is highly diverse), or when formal contracts are perceived as legitimate, or when the environment in which repeated commercial relationships take place becomes highly undetermined, which is consistent with the empirical evidence described in the previous sections.

1.5.1 Basic Evidence of Crowding-out Effect

The role of formal contract enforcement can be introduced into the trust game in different ways, such as punishments, rewards, or monitoring, but has typically been done so through restrictions on the choice set of the trustee. In their influential study, Falk and Kosfeld (2006) make use of a two-stage principal-agent game. The agent invests her effort in a productive activity, which is costly to her but beneficial for the principal. Before the agent makes her decision, the principal determines the agent's choice set: he can either restrict the agent's choice set, in which case he requires the latter to invest at least a minimal level of effort, or he can leave the choice set unrestricted. In this experimental setting, leaving the decision completely up to the agent mirrors the principal's trust as defined by Rousseau et al. (1998): a decision about whether to become vulnerable to another person's possible exploitation. The authors find that control entails hidden costs caused by the existence of agents who choose a lower level of effort if controlled than otherwise, and the hidden costs outweigh the benefit of control.[19] This substitution effect, that contracts enforcing trustworthy behavior may undermine the intrinsic trustworthiness of the trustees, has received the name "motivation crowding out" in economics (Frey and Jegen, 2001).

Other studies have also replicated the results of crowding-out found by Falk and Kosfeld (2006) in a variety of other contractual settings. Fehr and Gaechter (2002) study whether explicit contract incentives may create a hostile environment of threat and distrust undermining the agent's reciprocity. They conduct a standard principal-agent experiment under two treatments, a trust treatment (TT) and an incentive treatment (IT). In TT, subjects play a three-stage game: in the first stage, the principal makes a contract offer, which consists of a fixed wage and a desired effort demand; in the second stage agents decide whether to accept the offer; if agents accept the offer, they enter the third stage and then choose their effort levels. The IT is identical to the TT, with the exception that the principal can punish shirking agents. They find that principals in TT offer higher wages and demand higher effort levels

[19] Ziegelmeyer et al. (2012) replicate the Falk and Kosfeld's experiment and confirm the existence of hidden costs of control.

than in IT, resulting in higher actual average effort. This result implies that the explicit punishment scheme could undermine agents' intrinsic motivation.

Fehr and Rockenbach (2003) examine the trustee's negative response to the fines imposed by the investor in the trust game. In their experiment, German students in the role of "investors" are given the opportunity to transfer an amount to the other player, called the "trustee". When the investor transfers money to the trustee, he or she also specifies a desired level of back-transfer, which is non-binding. The transferred amount is then tripled by the experimenter. Knowing the investor's choice and desired level of back-transfer, the trustee could in turn "back-transfer" some (or all, or none) of this tripled amount. Two treatments are conducted. In the fine treatment, the investor could impose a fine if the trustee's back-transfer is less than the desired amount. The investor could also decline the use of the fine, the choice of using or declining the fine option being known to the trustee and taken prior to the trustee's decision. In the trust treatment, no such incentives are available to the investor. The authors find that trustees reciprocate generous initial transfers by investors with greater back-transfers, but that the use of the fine reduces back-transfers conditional on the investor's transfer, while renouncing the use of the fine when it is available to the investor increases return transfers. Overall, their study implies that refraining from the threat of fine, although the threat is available, could itself be perceived as a kind intention, which induces trustees to increase their reciprocity, while using the sanction to enforce an unfair distribution of income may be perceived as a bad action, inducing trustees to respond negatively.[20]

The issue of whether monitoring may induce the crowding-out effect on agent effort and thus backfire on the principal is addressed in Dickinson and Villeval (2008). To explore this principal-agent issue, they consider an employer-worker type relationship where the worker engages in a real-effort task. An employer can monitor a worker by choosing the probability with which the worker's output is audited. After being informed about the employer's monitoring choice, the worker then performs the task. They run this principal-agent game under two different conditions. In the stranger treatment, the authors match subjects as strangers in each round and preserve the anonymity of the pairs over the experiment, while in the partner treatment, they use a partner matching protocol and the same matching pairs play the game for ten rounds. They find evidence in the partner treatment that increased monitoring crowds-out agents' effort, implying that the crowding-out effect is more likely to occur in close relationships.[21]

[20] To separate the roles of incentives and negative intentions in undermining cooperation in Fehr and Rockenbach's data, Houser et al. (2008) add another treatment where trustees face threats of fines imposed (or not) by nature, and find that sanctions have statistically indistinguishable effects on trustees regardless of whether trustees are threatened intentionally by investors or randomly by nature, implying that the detrimental effect is mainly driven by the incentive itself.

[21] Similar results are also found in Masella et al. (2014), showing that contract incentives may crowd out pro-social behavior more strongly among subjects who share the same group identity.

1.5.2 Robustness and Extension of the Crowding-out Effect

1.5.2.1 Subject Pool

The basic experimental evidence of the crowding-out effect of the explicit contract discussed above is typically based on observing the behavior of undergraduate students. This is often criticized because students' behavior may not be representative of behavior in naturally occurring environments where many important commercial decisions are normally made by corporate managers. Fehr and List (2004) replicate the Fehr and Rockenbach's protocol (2003) using two types of subject pool, Chief Executive Officers (*CEOs*) and students. They replicate evidence of the hidden cost of punishment in both types of subject pool: the majority of *CEO* and student principals use the punishment option in the incentive condition, as a result, they obtain lower back-transfers by agents than do their peers who do not adopt the punishment option even when it is available. Interestingly, it is also found that *CEO* principals transfer more money and use the punishment option less often than students. Moreover, for any given transfer level, *CEO* agents pay back more money than students. Consequently, *CEOs* consistently achieve higher efficiency levels.

1.5.2.2 Framing

Fehr and Gaechter (2002) have shown that the punishment option adopted by the principal crowds out agents' pro-social behavior. One possible reason is that people may naturally dislike the punishment since it evokes negative feelings. To test whether the negative effect of explicit incentives is mainly driven by a "natural aversion" to punishment, Fehr and Gaechter introduce an additional bonus treatment (BT) where a shirking agent, instead of paying a fine, does not receive a bonus if their actual efforts do not reach the desired levels demanded by the principal. The incentive structure is exactly the same in IT and BT. Overall, it is found that with the material incentive framed as bonus, effort levels are significantly higher than when it is framed as punishment, although the efficiency in BT is still lower than in TT (i.e. the hidden cost of incentive still exists in BT). Similar contract framing manipulation is also implemented in other experimental studies (e.g. Brooks et al., 2012; Hannan et al., 2005; Hossain and List, 2012), suggesting that effort provision is very sensitive to incentive framing.

Schnedler and Vadovic (2011) build on Falk and Kosfeld's findings (2006) by using the basic experimental design as their baseline treatment (BT) and introduce an additional condition, endowment treatment (ET). In this new treatment, the game is formally equivalent to that in BT. However, they relabel the principal's strategies: control by the principal is now labeled as preventing stealing. Their results show that in ET the share of agents that punish the principal for controlling is lower than in BT, and their reduction in effort is smaller too. Interestingly, in ET principals control more often than in BT. Since hidden costs of control are significantly lower in ET,

principals are not worse off when they control than when they do not control in ET.

The hidden cost of control may be affected by the mere wording in the experimental instructions. To test this, Hagemann (2007) replicates Falk and Kosfeld's experiment (2006) as her baseline treatment (BT) and complements it with two new treatments. In Falk and Kosfeld's protocol, they use the phrase "participant B can decide to force participant A to give at least 10 points or to leave him completely free to decide". In fact, these instructions strongly accentuate the negative meaning if the principal decides to control the agent and therefore might influence the agents' transfer decision. Avoiding the wording "to force" and "to leave him completely free to decide", in the constrain treatment (CT) the author rephrases the wording: the principal now has the possibility "to constrain or not to constrain the agent", while in the neutral treatment (NT), the author lets the principal just "offer one out of two kinds of contract that allow the agent to choose his transfer from different ranges". The negative impact of control is replicated in BT. However, this negative impact of control disappears in CT. Furthermore, a hidden benefit of control is surprisingly observed in NT, where the principal obtains more profits when deciding to control than when deciding not to control. Hence, this study implies that instructions can trigger a demand effect that pushes the participants' attention in a certain direction, naturally generating the hidden cost of control.

1.5.2.3 Social Norms

Most of the previous studies revealing the hidden cost of control usually find high effort in the absence of explicit incentives, suggesting that a strong social norm may govern behavior in these settings (Sliwka, 2007). To make the social norm salient, Kessler and Leider (2014) allow subjects to make a non-binding agreement on playing the mutually beneficial actions before being assigned the roles in the principal-agent game. If the agreement is formed, a fairness norm is established. Then the authors introduce four treatments. In the baseline treatment (BT), the roles of principal and agent are assigned immediately after the players are told whether they have made an agreement. After that, the principal is given the option of whether to impose control. After the principal chooses, the choice is revealed to the agent. The agent then chooses his effort. In the mutual minimum treatment (MT), before assigning the roles of principal and agent, the authors randomly give one of the players the option to impose control on whichever player becomes the agent (i.e. control is imposed symmetrically). Once the player decides whether to impose control, the researchers assign the roles of principal and agent and ask subjects to play the principal-agent game. In the unknown agent treatment (UT), before assigning the roles, the authors randomly give one of players the option to impose control on the other player if that other player becomes the agent (i.e. control is imposed asymmetrically). Then the subjects are assigned the roles and play the principal-agent game. In the last treatment, the consent treatment (CT), before assigning the roles, the authors allow both players to suggest whether or not control should be imposed on

whichever player becomes the agent. Thus, each player can suggest control or no control and control works only if both players suggest control. After making a decision, the players are told who suggests the control and whether control works. Then the players are assigned their roles and play the principal-agent game.

Their study shows that most subjects (i.e. about 85%) are strongly in favor of having an agreement across all four treatments, with very little difference between treatments; offering to make an agreement significantly increases agents' effort levels in all treatments. When there is no agreement, i.e. the fairness norm is relatively weak, there is no hidden cost of control and imposing control is even profitable for principals. However, when there is an agreement, i.e. the fairness norm is salient, the cost to the principal of imposing control depends on the treatment. Cost is very high in BT and UT, but it is eliminated in the MT and is reversed in CT. Therefore, this new study implies that the hidden cost of incentive schemes is very sensitive to the social norm variation. When the social norms of fairness become more salient, it is more costly to impose control on the agent.[22] However, the mutual consent between principal and agent could legitimize the control, and make it seem less distrustful, unexpectedly generating the hidden benefit of control.[23]

In the real world, new employees usually are not aware of the prevalent norms in their corporation, but they could infer the existing work norms in their organization based on owners or managers' incentive schemes, because explicit contracts can signal the private information held by the principal. To explore this idea, Danilov and Sliwka (2013) implement a simple one-shot principal-agent game. In the baseline treatment (BT), the principal can choose between a fixed wage contract and a performance-based contract. After the principal chooses one of contracts, the agent then determines his effort. They elicit the agents' efforts for both contract types using the strategy method. In the norms treatment (NT), they replicate the BT with one addition: they show the principals a table containing the effort levels chosen by participants in a preceding baseline session and inform the agents that their principal has seen such a contributions table (the agents do not know its content). Hence, the agents do not know the behavior of others, but they are aware that the principals had this information prior to the contract choice.

Their results show that when a fixed wage is chosen by an informed principal in NT, effort levels are nearly 50% higher than in BT even though the incentive structures for agents are completely identical in both treatments. However, when the informed principal selects the performance-based contract in NT, the agent responds with lower effort levels than in BT. Furthermore, the authors find that the agents' beliefs about the prior information of their principal are substantially affected by the

[22] Similar results are observed in Kessler and Leider (2012), finding that mandatory minimum rules produce worse results than do the unenforceable handshake contracts when the norms are salient.
[23] Related results have also been found by Saaksvuori (2013). The author presents results from trust games run among college students in Germany, and reports that endogenously formed centralized sanctioning institutions significantly increase trust and trustworthiness.

principal's contract choice, i.e. agents correctly gain an understanding of the prevalent norms through their principal's choice, which in turn dramatically affect their own decisions.[24]

1.5.2.4 Repeated Games

Previous experimental studies have confirmed the crowding-out effect of explicit incentives in the non-repeated principal-agent settings; they find that incentive contracts usually undermine intrinsic motivation, and thus entail the hidden cost. Few experimental studies examine the effect of formal contracts in repeated settings. To fill the gap, Lazzarini et al. (2004) study how a formal contract interacts with relation-based trust to affect individual behavior in repeated exchanges. In their experiment, subjects are randomly matched to play repeated principal-agent games where after each period the ongoing relationships between paired subjects continue with specific probabilities and the probabilities vary among the different pairs. In each period, after knowing the probability of game continuation, the principal decides whether or not to choose the formal contract in which the payment is contingent on the agent's choice. The agent then makes the decision. Their study shows that the principals are more likely to choose the formal contracts when the probability of continuation across periods becomes very low. By enforcing the formal agreements, contracts facilitate the self-enforcement of informal agreements. This complementarity effect is particularly important when repeated interactions with partners are unlikely and thus self-enforcement is very difficult.

People usually believe that strong contract enforcement induces the trustee to reciprocate her partner, and that cooperative behavior in the present then reinforces an expectation of cooperation in the future even when contract enforcement is removed (Poppo and Zenger, 2002). To test this idea, Malhotra and Murnighan (2002) conduct a binary trust game under four treatments. In the baseline treatment (BT), all subjects act in the role of trustors, and play the game with the same partner for four periods. In each period, the trustor chooses to "trust" or "not trust". If the trustor chooses not to trust, the game ends; if the trustor chooses to trust, the trustee has the option of honoring or exploiting the trust. Unbeknown to the subjects, their trustee is a computer program that always honors the trust. In the partial-contract treatment (PT), in the first two periods the trustee (i.e. the computer program) proposes a binding contract in which the computer exogenously enforces the option of "trust" without the subject's intention and the trustee is mandated to honor the trust, and then the trustor decides whether or not to accept this proposed contract; in the last two periods, the contract is not available to the trustee and the subjects play the game presented in the BT. The allowed-but-not-chosen treatment (AT) is identical to the BT, with the

[24] Similar experimental evidence is found in other studies (Cardinaels and Yin, 2014; Galbiati et al., 2013). For example, Galbiati et al. (2013) compare the effect of sanction mechanism that is enforced exogenously by the experimenter to the same sanction mechanism that is selected by a subject who has superior information about the previous behavior of the other players, finding that the endogenous sanction mechanism is perceived as a negative signal by subjects and is thereby counterproductive.

exception that the trustees are allowed to propose contracts in all four periods but are programmed not to activate them in the last two periods. The mentioned treatment (MT) is also identical to the BT, with the exception that the possibility for contracts is described but the trustees are never allowed to propose them.

The results indicate that when contracts are no longer allowed, after having been allowed previously, trust drops dramatically in PT compared to in BT. Also, trust drops even more dramatically in treatment AT, when the trustee chooses not to propose a contract, after having proposed contracts twice previously, implying that the intention matters. However, merely mentioning but not allowing binding contracts does not have a significant effect on trust. Hence, this study implies that trust could not develop during the cooperative but contractually mandated interactions, and that strong contracts not only impede the development of trust but also diminish the existing trust. A similar "Removing The Incentive" paradigm is used by Mulder et al. (2006), finding that participants who have experienced the presence of a sanctioning system trust fellow group members less than participants who have not.

In repeated exchanges, the type and extent of contract incentives not only affect individuals' intrinsic motivation in the short term, but may "also influence the process of preference-updating by which individuals acquire new tastes or social norms that will persist over long periods" (Bowles and Polania-Reyes, 2012, p.374). As a result, in the perspective of long periods, explicit contracts may reinforce trust or trustworthiness by updating people's preferences. To explore this idea, Bohnet et al. (2001) randomly match participants and ask them to play a two-person contract game in which the trustor has to decide whether she wants to enter a contract without knowing whether the trustee will perform; if the trustee breaches, a chance move decides whether he is held liable for the cost of the breach. There are three types of contract enforcement probabilities: low, medium and high. The experiment consists of two blocks: the first block has three periods, while the second block has six. In order to create different legal regimes, the authors vary the contract enforcement probability of the first block across sessions. However, subjects in all sessions undergo weak contract enforcement during the second block. In both blocks, after each period, aggregate information on outcomes is provided, that is, both trustors and trustees know how many contracts were offered and performed in the previous rounds. The experimental results show that, in the first block, subjects achieve the highest degree of efficiency when contract enforcement probability is high; in the second block, the differential effects of prior enforcement gradually vanish and all sessions converge to a high level of cooperation when all subjects enter into the low-probability environment. Hence, this study implies that, in the short term, strong contract enforcement help people to establish a relatively high level of cooperation; in the long term, when people enter into a weak enforcement environment, they are very sensitive to their partner's previous performance rate and engage in screening of potential partners. Naturally, the intrinsic trustworthiness of trustees becomes a key variable in the transaction. Consequently, people's preferences are updated and trustworthiness is

crowded in with weak enforcement.[25]

1.5.2.5 Rational Response

Previous studies have shown that exerting control, monitoring, and other explicit incentives can be counterproductive for principals. Alternatively, delegating decision rights to agents (Charness et al., 2012) is often perceived as friendly and helps principals to achieve higher profitability. Principals seem aware of the importance of agents' reciprocity motivation, but the reduction in the intensity of explicit incentives when their payoff depends primarily on agents' effort levels is not statistically significant. For instance, only 20% to 30% of principals choose the delegation option in Charness et al. (2012), although delegating the wage decision significantly enhances agent performance and increases the earnings of principals. In the real world, however, it is found that contracting parties often deliberately include incomplete and usually unenforceable terms in their contracts.[26]

To examine whether principals correctly anticipate the detrimental effect of explicit contracts and then rationally respond to it, Sloof and Sonnemans (2011) ask subjects to play three repeated trust games. For each repeated trust game, subjects will be randomly matched with a distinct partner. Then the subject in the role of trustor moves first and decides whether or not to trust trustee: if she chooses to trust, the trustee then decides whether or not to honor trust; if she does not trust, the existing explicit contract will be enforced. The probability of repetition for each repeated trust game is different across treatments. And there are two types of explicit contracts, one gives trustor high payoff (i.e. "good" contract) and the other gives low payoff (i.e. "bad" contract). For each session, in the first repeated trust game a good (or bad) explicit contract is exogenously determined and always applicable during this game. In the second game another explicit contract (i.e. good or bad contract) exogenously replaces the initial one and applies for this game. In the last game the good or bad contract is endogenously chosen by the trustor. The main findings are that cooperation is more likely when there are more repetitions of the game and only bad explicit contracts are available in the repeated trust game. Anticipating this, the majority of subjects choose bad explicit incentives to facilitate cooperation.

Recently, another type of informal incentive has attracted researchers' attention –

[25] Similar argument can be found in Scott (2000, p.1632), claiming that "our putative moral defective observes that she loses opportunities because she cannot make credible commitments. The motivation to increase her opportunity set stimulates the necessary characterological changes in values. Out of this process emerges a 'new person.' New and better preferences and values - honesty, loyalty, trustworthiness - now form part of the individual's stock of traits."

[26] In legal studies, for example, Scott (2003) analyzes a large sample of courts cases litigated between 1998 and 2002 in the U.S., and finds that contracts that specify an up-front payment plus an "indefinite" promise of a bonus payment in case of satisfactory performance are quite common in the business world, although these contracts are incomplete and unenforceable in the view of the courts. Scott concludes that these deliberately incomplete contracts allow trustors to signal their trusting intentions.

discretionary incentives, where the game structure allows principals to sanction or reward agents discretionarily after observing agents' effort levels; this implicit incentive is not credible and enforceable, and is also costly to principals. However, as many experimental studies suggest, the discretionary incentives perform better than explicit incentives as the latter are often perceived as a hostile act and crowd out intrinsic trustworthiness (Fehr et al., 2007; Fehr and Schmidt, 2007).

In Fehr et al. (2007), the authors introduce three types of contracts to subjects. In the explicit contract the principal offers a wage, a required effort level, and a fine to be paid if the agent is caught shirking. In the trust contract, the principal offers a fixed wage to the agent and asks for high effort in return. Lastly, the discretionary contract is similar to the trust contract, except that the principal announces that she might pay a bonus if the agent exerts more effort than required. They then ask principals to choose among these three contracts and play the game. In line with the previous evidence, discretionary incentives perform better from the firm's perspective than the other two contracts. Interestingly, principals seem to know the crowding-in effect of discretionary incentives because this contract is chosen much more often than other two contracts. In a subsequent paper, Fehr and Schmidt (2007) examine whether combining a discretionary and explicit contract helps to improve efficiency. In their new experiment, principals can choose between a purely discretionary contract and a combined contract. The data show that the vast majority of principals select the purely discretionary contract, which also turns out to be more efficient.

1.5.3 Evidence from Field Experiments

Compared to econometric or statistical techniques that rely heavily on naturally occurring data to answer causal questions, laboratory experimental methodology can precisely identify causation via randomization. Even though controlled laboratory experimentation provides important insights on causation, the generalizability of laboratory experimental outcomes is still highly criticized. One main critique is that college students are disproportionately employed as subjects in lab experiments, which could not help us make inferences about the behavior of other groups of people in the real world. Also, the artificial lab context may potentially bias behavior. For example, the nature and extent of the scrutiny associated with the lab may induce subjects to distort their real preferences in order to please the experimenter (Levitt and List, 2007). To mitigate these methodological problems, field experiments randomly implementing an intervention in the real world rather than in the laboratory are emerging in economics and other social sciences (Harrison and List, 2004).

Belot and Schroder (forthcoming) study the effects of monitoring and contract incentives on work quality using a field experiment where the subjects do not know that they are participants in an experiment. In their experiment, subjects are employed to identify the value and country of origin of euro coins that are collected in different countries in the euro zone. Subjects have one day to finish the task and are asked to

return the coins by an exact deadline. Three treatments are conducted in the experiment. In the trust treatment (TT), a fixed wage is offered and no requirement of work quality is mentioned. In the monitoring and weak incentives treatment (MW), a tolerated number of mistakes is specified, and subjects are informed that a moderate penalty deducting from the fixed wage will incur if the number of mistakes exceeds the tolerated number. The monitoring and strong incentives treatment (MS) is identical to the MW, with the exception that the penalty here is heavier than the one in the MW.

The authors find that weak incentives do not reduce the number of mistakes significantly while strong incentives improve work quality. In addition, they reveal a negative effect of monitoring. Specifically, they find that when monitoring is implemented the fraction of participants who return the coins later increase dramatically in both incentive treatments. These findings imply that deliberately implementing monitoring in contractual relationships may signal the distrust of principals, and thus induce agents to retaliate it with negative reciprocity.

While the evidence on control aversion presented in Belot and Schroder (forthcoming) is pronounced, it is not full clear how well they extend to more realistic markets where the contracting parties are formal legal entities. Bengtsson and Engstrom (2014) conduct a field experiment to investigate whether implementing monitoring crowds out the intrinsic motivation of Swedish non-profit organizations. In Sweden, at the beginning of each year various proposals of non-profit organizations are submitted to the Swedish foreign aid agency (Sida). Once the proposals are approved, Sida will sign contracts with the organizations and distribute funds to them. Traditionally, the contracts are based on trust and self-regulation. In their experiment, the authors randomly select a sample of non-profit organizations and assign threats of audits to them. Specifically, the selected organizations are informed that Sida will audit their financial documentation at the end of the fiscal year and that they will risk losing future funds if Sida detects any irregularities, while non-selected organizations receive no information about Sida's upcoming audit at all.

They find that non-profit organizations who are monitored significantly reduce their expenditures and are more likely to return unused funds to Sida than non-monitored organizations. In addition, the reduction in expenditures does not reduce the performance of treated organizations. Specifically, organizations in the treatment group extend their outreach more widely and are reported by local media more often compared to non-treated ones, implying that monitoring does not crowd out the pro-social behavior but actually improves economic efficiency.

How could we reconcile these controversial findings on monitoring? While many studies reveal that strong contract enforcement crowds out intrinsic motivation to cooperation and undermines trust (e.g. see Gneezy and Rustichini, 2000), few papers examine how the nature of power exercised by authorities (or principals) influences

the crowding-out effect. According to Turner (2005), there are two kinds of power: coercive and legitimate power. Legitimate power can enhance trust, while coercive power always reduces trust. Power adopted to improve efficiency may be perceived as legitimate rather than coercive (Gangl et al., 2015). Therefore, it is convincing that contract control enforced by Sida, a social oriented principal, is more likely to be perceived as legitimate than the one implemented by a self-interested principal, thereby generating hidden benefits of control.

Cassar et al. (2014) conduct field experiments in Italy and Kosovo to identify the causal effect of formal enforcement on trust and trustworthiness and also to study how legal enforcement and preexisting trust interact to influence cooperation. Their experiments consist of four stages. In the first stage, subjects are randomly matched to play a one-shot trust game (Berg et al., 1995). In the second stage, subjects enter a market game of 10 rounds where subjects decide whether to behave honestly, cheat, or stay out, in the absence of any legal enforcement. In the third stage, two treatments are introduced: in the partial enforcement system (PES) treatment, subjects participate 10 rounds of the market game in which weak contract enforcement is implemented, while in the impartial enforcement system (IES) treatment, subjects play 10 rounds of the market game with strong enforcement. In the last stage, subjects play the one-shot trust game again with a randomly selected partner.

The authors show that both PES and IES treatments enhance trust, but the increase is more pronounced in the IES treatment. For trustworthiness, the IES treatment significantly increases trustworthiness while the PES treatment decreases it. This suggests that strong contract enforcement aiming to improve cooperation may be perceived as legitimate and therefore trigger internalized norms of cooperation.[27] In addition, the authors find that the subjects' preexisting trust is negatively correlated with their dishonest behavior in the market game, but only for those who do not experience an impartial institution. It implies that trust may act as an alternative to formal institutions in promoting mutually advantageous transactions, but only in the absence of strong legal enforcement.

1.6 Conclusions

The prevailing view in economics is that a well functioning and impartial legal system is the key to development. If governments provide sufficiently strong systems of contract enforcement that promote investment and encourage trade, prosperity follows. Trust between investors and entrepreneurs (or between firms) is also an important facilitator of investment and production. There is strong evidence that

[27] Similar findings are reported in Mironova and Whitt (2013). The authors conduct a field experiment in Kosovo and allow subjects to play a repeated trust game with intervention, which sometimes punishes dishonest subjects (Charness et al., 2008). They find that the possibility of third-party punishment increases the levels of trust and trustworthiness and that the positive effect persists even after the enforcement mechanism is removed.

countries' average levels of trust are positively correlated with per capital income and economic growth, suggesting that trust may generate enduring increases in gains from trade. While there is a consensus that both good legal rules and high trust are crucial for trade and development, the relationship between them is much less clear. On the one hand, formal contract enforcement reduces the benefits of opportunistic behavior within the contractual relationship, directly promoting societal trust. On the other hand, formal legal mechanisms may undermine the intrinsic trustworthiness of people, thereby leading to a low level of trust.

In order to understand how formal contract enforcement affects trust, I offer a survey of the literature on trust and legal mechanisms. I first focus on the macro evidence, and find that good quality legal institutions have a positive effect on trust. However, a possible limitation to causal identification in these studies is that legal institutions are themselves the outcome of societal trust. A recent experimental paper (Campos-Ortiz et al., 2012) has confirmed this idea, showing that subjects coming from countries with higher levels of trust devote more resources to public good production and are more likely to pass a binding majority vote on establishing a formal legal institution that facilitates cooperation in a laboratory environment.

Since most investment and production activities occur between firms, I also consider the relationship between trust and formal contracts in inter-firm cooperation. In order to regulate each other's behavior, firms are willing to engage in a formal contracting relationship, or enter into an incompletely specified collaboration to establish repeated exchanges where the effect of reputation or some combination of legal or non-legal mechanisms works. The actual adoption of commercial strategies by firms depends on specific conditions. Normally, a formal contract is adopted when performance is verifiable *ex post* but not necessarily observable *ex ante*, while an informal arrangement has an advantage when performance is observable but costly to verify. As the contemporary commercial environment is becoming increasingly uncertain, many firms establish long term collaboration but with formal contracts. Since firms can freely choose between trust-based relationships and formal contracts to do business with their partners, it is even more difficult to identify the causal effect of formal contracts on trust using happenstance data from firms.

Identifying a causal link between contract enforcement and trust with happenstance data is exceedingly difficult because both may be co-determined in the real world. I therefore turn to the experimental studies, which manipulate the exogenous adoption of formal contracts and measure their effect on trust. In the experimental literature, subjects are usually involved in a standard (or modified) principal-agent game, where the principal can introduce incentives to constrain the agent's behavior or trust and delegate decision rights to the agent. The experimental evidence generally shows that incentives initiated by the principal crowd out the agent's intrinsic trustworthiness and reduce her beneficial behavior towards the principal. However, when the content of formal incentives is mutually agreed or recognized as legitimate by the involved

parties, or the parties are involved in a highly heterogeneous market where no predominant norm of fairness exists, then formal contracts are preferred and found to enhance efficiency. In addition, when environmental factors become extremely uncertain during repeated commercial relationships, parties always use formal contracts to facilitate their informal arrangements, thereby increasing their economic welfare. Many subjects seem to anticipate the possible perverse effects of formal incentives, and as a consequence, they deliberately include indefinite clauses in their contracts to mitigate possible detrimental outcomes.

Understanding the relationship between trust and legal rules has important implications for understanding how contract enforcement interacts with societal trust, and further affects economic development. It also helps us understand why certain contracting institutions work in some market environments but not others. Since trust beliefs may be formed based on the individuals' heuristics about the general legal environment in a country, it is necessary to manipulate subjects' initial trust levels in future experiments. In this way, not only can we understand whether contracting institutions influence societal trust, but we are also able to unravel the underlying mechanism of how legal institutions shape the causal effect of trust on economic performance.

Reference

Acemoglu, D., Robinson, J., and Johnson, S., 2001, "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review*, 91, 1369-1401.

Aghion, P., Algan, Y., Cahuc, P. and Shleifer, A., 2010, "Regulation and Distrust." *Quarterly Journal of Economics*, 125(3), 1015-1049.

Algan, Y., and Cahuc, P., 2010, "Inherited Trust and Growth." *American Economic Review*, 100, 2060-2092.

Algan, Y., and Cahuc, P., 2013, "Trust and Growth", *Annual Review of Economics*, vol (5), 521-549.

Algan, Y., and Cahuc, P., 2014, "Trust and Human Development: Overview and Policy", *Handbook of Economic Growth,* eds Aghion, P., and Durlauf, S., 49-120.

Arrighetti, A., Bachmann, R., and Deakin, S., 1997, "Contract Law, Social Norms and Inter-firm Cooperation", *Cambridge Journal of Economics*, 21, 171-195.

Arrow, K., 1972, "Gifts and Exchanges," *Philosophy and Public Affairs*, 1, 343-362.

Arrow, K., 1974, *The Limits of Organization*. New York: Norton.

Badawi, A., 2010, "Relational Governance and Contract Damages: Evidence from Franchising." *Journal of Empirical Legal Studies*, 7 (4), 743-785.

Baird, D., 1990, "Self-Interest and Cooperation in Long-term Contract." *Journal of Legal Studies*, 19, 583-596.

Baker, S., and Choi, A., 2015, "Contract's Role in Relational Contract," *Virginia Law Review*, 101, 559-607.

Baker, G., Gibbons, R., and Murphy, K., 1994, "Subjective Performance Measures in Optimal Incentive Contracts." *Quarterly Journal of Economics*, 109, 1125-1156.

Barnett, J., 2015, "Hollywood Deals: Soft Contracts for Hard Markets," *Duke Law Journal*, 64, 605-669.

Becker, S., Boeckh, K., Hainz, C., and Woessmann, L., forthcoming, "The Empire Is Dead, Long Live the Empire!" *Economic Journal*.

Bellemare, C. and Kroeger, S., 2007, "On Representative Social Capital," *European Economic Review*, 51, 183-202.

Belot, M., and Schroder, M., forthcoming, "The Spillover Effects of Monitoring: A Field Experiment," *Management Science*.

Bengtsson, N., and Engstrom, P., 2014, "Replacing Trust with Control: A Field Test of Motivation Crowd Out Theory," *Economic Journal*, 124, 833-858.

Ben-Shahar, O., and Bernstein, L., 2000, "The Secrecy Interest in Contract Law," *Yale Law Journal*, 109, 1885-1925.

Berg, J., Dickhaut J., and McCabe K., 1995, "Trust, Reciprocity and Social History," *Games and Economic Behavior*, 10, 122-142.

Blair, M., and Stout, L., 2001, "Trust, Trustworthiness and the Behavioral Foundations of Corporate Law," *University of Pennsylvania Law Review*, 149 (6), 1735-1810.

Bohnet, I., Fery, B., and Huck, S., 2001, "More Order with Less Law: On Contract Enforcement, Trust, and Crowding", *American Political Science Review*, 95 (1), 131-144.
Bohnet, I., Grieg, F., Herrmann, B., and Zeckhauser, R., 2008, "Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States", *American Economic Review*, *98*(1), 294– 310.
Bohnet, I., and Zeckhauser, R., 2004, "Trust, Risk and Betrayal", *Journal of Economic Behavior and Organization*, *55*, 467–484.
Bowles, S., and Polania-Reyes, S., 2012, "Economic Incentives and Social Preferences: Substitutes or Complements?" *Journal of Economic Literature*, 50 (2), 368-425.
Brooks, R., Stremitzer, A., and Tontrup, S., 2012, "Framing Contracts: Why Loss Framing Increases Effort," *Journal of Institutional and Theoretical Economics*, 168, 62-82.
Buggle, J., 2013, "Law and Social Capital: Evidence from the Code Napoleon in Germany," working paper.
Campos-Ortiz, F., Balafoutas, L., Putterman, L., Batsaikhan, M., Ahn, T., and Sutter, M., 2012, "Security of Property as a Public Good: Institutions, Socio-Political Environment and Experimental Behavior in Five Countries," working paper.
Cardinaels, E., and Yin, H., 2014, "Think Twice before Going for Incentives: Social Norms and the Principal's Decision on Compensation Contracts." Working paper.
Cassar, A., d'Adda, G., and Grosjean, P., 2014, "Institutional Quality, Culture, and Norms of Cooperation: Evidence from Behavioral Field Experiments", *Journal of Law and Economics*, 57, 821-863.
Charness, G., Cobo-Reyes, R., and Jimenez, N., 2008, "An Investment Game with Third-party Intervention," *Journal of Economic Behavior & Organization*, 68, 18-28.
Charness, G., Cobo-Reyes, R., Jimenez, N., Lacomba, J., and Lagos, F., 2012, "The Hidden Advantage of Delegation: Pareto Improvement in a Gift Exchange Game", *The American Economic Review*, 102(5), 2358-2379.
Coleman, J., 1990, *Foundations of Social Theory*. The Belknap Press of Harvard University Press.
Cross, F., 2005, "Law and Trust", *Georgetown Law Journals*, 93 (5), 1457-1545.
Danilov, A., and Sliwka, D., 2013, "Can Contracts Signal Social Norms? Experimental Evidence," Unpublished Paper, University of Cologne.
Das, T., and Teng, S., 1998, "Between Trust and Control: Developing Confidence in Partner Cooperation in Alliances." *Academy of Management Review*, 23, 491-512.
Das, T., and Teng, S., 2001, "Trust, Control, and Risk in Strategic Alliances: An Integrated Framework." *Organization Studies*, 22, 251-283.
Deutsch, M., 1958, "Trust and Suspicion," *Journal of Conflict Resolution*, 2 (4), 265-279.
Deutsch, M., 1962, "Cooperation and Trust: Some Theoretical Notes," *Nebraska Symposium on Motivation*, 275-320.
Dickinson, D., and Villeval, M., 2008, "Does Monitoring Decrease Work Effort? The Complementarity between Agency and Crowding-out Theories." *Games and*

Economic Behavior, 63, 56-76.
Dincer, O., and Uslaner, E., 2010, "Trust and Growth", *Public Choice*, 142, 59-67.
Durante, R., 2009, "Risk, Cooperation and the Economic Origins of Social Trust: An Empirical Investigation", Working Paper.
Dyer, J., 1997, "Effective Inter-firm Collaboration: How Firms Minimize Transaction Costs and Maximize Transaction Value", *Strategic Management Journal*, 18 (7), 535-556.
Eckel, C., and Wilson, R., 2004, "Is Trust a Risky Decision?" *Journal of Economic Behavior and Organization*, 55, 447-465.
Falk, A., and Kosfeld, M., 2006, "The Hidden Costs of Control." *American Economic Review*, 96 (5), 1611-1630.
Fehr, E., 2009, "On the Economics and Biology of Trust," *Journal of the European Economic Association*, 7, 235-266.
Fehr, E., Fischbacher, U., von Rosenbladt, B., Schupp, J., and Wagner, G., 2003, "A Nationwide Laboratory: Examining Trust and Trustworthiness by Integrating Behavioral Experiments into Representative Surveys," CESifo Working Paper 866.
Fehr, E., and Gaechter, S., 2002, "Do Incentive Contracts Undermine Voluntary Cooperation?" unpublished paper, University of Zurich.
Fehr, E., Klein, A., and Schmidt, K., 2007, "Fairness and Contract Design", *Econometrica*, 75, 121-154.
Fehr, E., and Schmidt, K., 2007, "Adding a Stick to the Carrot? The Interaction of Bonuses and Fines", *The American Economic Review*, 97(2), 177-181.
Fehr, E., and Rockenbach, B., 2003, "Detrimental Effects of Sanctions on Human Altruism", *Nature*, 422, 137-140.
Fetchenhauer, D., and Dunning, D., 2012, "Betrayal Aversion versus Principled Trustfulness-How to Explain Risk Avoidance and Risky Choices in Trust Games", *Journal of Economic Behavior and Organization*, 81, 534-541.
Frey, B., and Jegen, R., 2001, "Motivation Crowding Theory." *Journal of Economic Surveys*, 15, 589-612.
Fukuyama, F., 1995, *Trust: Social Virtues and the Creation of Prosperity*. New York: Free Press.
Galbiati, R., Schlag, K., and van der Weele, J., 2013, "Sanctions that Signal: An Experiment," *Journal of Economic Behavior and Organization*, 94, 34-51.
Gambetta, D., 1988, *Trust: Making and Breaking Cooperative Relations*. New York: Basil Blackwell.
Gangl, K., Hofmann, E., and Kirchler, E., 2015, "Tax Authorities' Interaction with Taxpayers: A Conception of Compliance in Social Dilemmas by Power and Trust," *New Ideas in Psychology*, 37, 13-23.
Gilson, R., Sabel, C., and Scott, R., 2010, "Braiding: The Interaction of Formal and Informal Contracting in Theory, Practice, and Doctrine", *Columbia Law Review*, 110 (6), 1377-1447.
Glaeser, E., Laibson, D., Scheinkman, J., and Soutter, C., 2000, "Measuring Trust", *Quarterly Journal of Economics*, 115 (3), 811-846.
Gneezy, U., and Rustichini, A., 2000, "A Fine is a Price," *Journal of Legal Studies*,

29, 1-17.
Greif, A., 2006, *Institutions and the Path to the Modern Economy*, London: Cambridge University Press.
Greif, A., Milgrom, P., and Weingast, B., 1994, "Coordination, Commitment, and Enforcement: The Case of the Merchant Guild," *Journal of Political Economy*, 102, 745-776.
Guiso, L., Sapienza P., and Zingales, L., 2004, "The Role of Social Capital in Financial Development," *American Economic Review*, 94, 526-556.
Guiso, L., Sapienza P., and Zingales, L., 2008a, "Trusting the Stock Market," *Journal of Finance*, 63(6): 2557-2600.
Guiso, L., Sapienza P., and Zingales, L., 2008b, "Alfred Marshall Lecture: Social Capital as Good Culture." *Journal of European Economic Association*, 6, 295-320.
Guiso, L., Sapienza P., and Zingales, L., 2009, "Culture Biases in Economic Exchange?" *Quarterly Journal of Economics*, 124, 1095-1131.
Guiso, L., Sapienza P., and Zingales, L., 2011, "Civil Capital as the Missing Link," *Handbook of Social Economics*, 417-480.
Gulati, R., 1995, "Does Familiarity Breed Trust? The Implications of Repeated Ties for Contractual Choice in Alliances." *Academy of Management Journal*, 38, 85-112.
Hagemann, P., 2007, "What's in a Frame? On Demand Effects and Trust in Experimental Studies," Unpunished paper, University of Cologne.
Hall, M., 2002, "Law, Medicine, and Trust," *Stanford Law Review*, 55 (2), 463-527.
Hall, M., Thom, D. and Pawlson, G., 2004, "Measuring Patients' Trust in Physicians When Assessing Quality of Care," *Health Affairs*, 23(4), 124-132.
Hall, R., and Jones, C., 1999, "Why Do Some Countries Produce So Much Output per Worker than Others?" *Quarterly Journal of Economics*, 114, 83-116.
Hannan, R., Hoffman, V., and Moser, D., 2005, "Bonus versus Penalty: Does Contract Frame Affect Employee Effort?" *Experimental Business Research: Economic and Managerial Perspectives*, Vol. II, edited by A. Rapoport and R. Zwick, Netherlands: Springer.
Harrison, G., and List, J., 2004, "Field Experiments," *Journal of Economic Literature*, 42, 1009-1055.
Hendley, K., Murrell, P., and Ryterman, R., 2000, "Law Works in Russia: The Role of Law in Interenterprise Transaction", in P. Murrell (ed.), *Assessing the Value of Law in Transition Economies*, University of Michigan Press.
Ho, B., and Huffman, D., 2013, "Trust and the Law," unpublished paper.
Hossain, T., and List, J., 2012, "The Behavioralist Visits the Factory: Increasing Productivity Using Simple Framing Manipulations," *Management Science*, 58, 2151-2167.
Houser, D., Xiao, E., McCabe, K., and Smith, V., 2008, "When Punishment Fails: Research on Sanctions, Intentions and Non-cooperation", *Games and Economic Behavior*, 62, 509-532.
Johnson, N., and Mislin, A., 2011, "Trust Games: A Meta-analysis." *Journal of Economic Psychology*, 32, 865-889.
Johnson, S., McMillan, J., and Woodruff, C., 2002, "Courts and Relational

Contracts." *Journal of Law, Economics, and Organization*, 18, 221-277.
Karlan, D., 2005, "Using Experimental Economics to Measure Social Capital and Predict Financial Decisions." *American Economic Review*, 95, 1688-1699.
Kessler, J., and Leider, S., 2012, "Norms and Contracting." *Management Science*, 58 (1), 62-77.
Kessler, J., and Leider, S., 2014, "Procedural Fairness and the Cost of Control." Unpublished Paper, University of Pennsylvania.
King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., and Montague, P. R., 2005, "Getting to Know You: Reputation and Trust in a Two-Person Economic Exchange." *Science*, 308, 78–83.
Knack, S., and Keefer, P., 1997, "Does Social Capital Have an Economic Payoff: a Cross-country Investigation," *The Quarterly Journal of Economics*, 112(4).
Kosfeld, M., Heinrichs, M., Zak, P., Fischbacher, U., and Fehr, E., 2005, "Oxytocin Increases Trust in Humans," *Nature*, 435, 673-676.
La Porta, R., Lopez de Silanes F., Shleifer A., and Vishny R., 1997, "Trust in Large Organizations," *American Economic Review*, 87(2), 333-338.
La Porta, R., Lopez-de-Silanes F., Shleifer A. and Vishny R., 1998, "Law and Finance," *Journal of Political Economy*, 106, 1113-55.
La Porta, R., Lopez-de-Silanes F., Shleifer A. and Vishny R., 1999, "The Quality of Government," *Journal of Law, Economics and Organization*, 15, 222-279.
Lazzarini, S., Miller, G., and Zenger, T., 2004, "Order with Some Law: Complementarity versus Substitution of Formal and Informal Arrangements", *Journal of Law, Economics, and Organization*, 20(2), 261-298.
Levitt, S., and List, J., 2007, "What Do Laboratory Experiments Measuring Social Preferences Reveal about the Real World?" *Journal of Economic Perspectives*, 21, 153-174.
Macaulay, S., 1963, "Non-contractual Relations in Business: A Preliminary Study." *American Sociological Review*, 28, 55-67.
Macaulay, S., 1985, "An Empirical View of Contract." *Wisconsin Law Review*, 1985, 465-482.
Malhotra, D., and Murnighan, K., 2002, "The Effects of Contracts on Interpersonal Trust", *Administrative Science Quarterly*, 47 (3), 534-559.
Masella, P., Meier, S., and Zahn, P., 2014, "Incentives and Group Identity", *Games and Economic Behavior*, 86, 12-25.
Mayer, R., Davis J., and Schoorman, F., 1995, "An Integrative Model of Organizational Trust." *Academy of Management Review*, 20, 709-734.
McAllister, D., 1995, "Affect- and Cognition-Based Trust as Foundations for Personal Cooperation in Organizations", *The Academy of Management Journal*, 38, 24-59.
Mellewigt, T., Madhok, A., and Weibel, A., 2007, "Trust and Formal Contracts in Interorganizational Relationships: Substitutes and Complements", *Managerial and Decision Economics*, 28, 833-847.
Miller, A. S., and Mitamura, T., 2003, "Are Surveys on Trust Trustworthy?" *Social Psychology Quarterly*, 66, 62-70.

Mironova, V., and Whitt, S., 2013, "International Peacekeeping and Micro-foundations for Positive Peace: Lab-in-the-Field Evidence from Kosovo", working paper.
Mitchell, L., 2001, "The Importance of Being Trusted," *Boston University Law Review*, 81, 591-617.
Mulder, L., van Dijk, E., De Cremer, D., and Wilke, H., 2006, "Undermining Trust and Cooperation: The Paradox of Sanctioning Systems in Social Dilemmas," *Journal of Experimental Social Psychology*, 42, 147-162.
Naef, M., and Schupp, J., 2008, "Measuring Trust: Experiments and Surveys in Contrast and Combination." Working paper, Royal Holloway College, London.
North, D., 1990, *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.
Nunn, N., 2008, "The Long Term Effects of Africa's Slave Trades." *Quarterly Journal of Economics*, 123,139-176.
Ogilvie, S., and Carus, A. W., 2014, "Institutions and Economic Growth in Historical Perspective," *Handbook of Economic Growth*, eds Aghion, P., and Durlauf, S., 403-513.
Poppo, L., and Zenger, T., 2002, "Do Formal Contracts and Relational Governance Function as Substitutes or Complements?" *Strategic Management Journal*, 23, 707-725.
Posner, R., 1973, *Economic Analysis of Law*, Little Brown and Company.
Posner, E., 2002, *Law and Social Norms*, Harvard University Press.
Putnam, R., 1993, *Making Democracy Work: Civic Traditions in Modern Italy*, Priceton, NJ: University Press.
Rose, C., 1995, "Trust in the Mirror of Betrayal," *Boston University Law Review*, 75, 531-558.
Rotter, J., 1980, "Interpersonal Trust, Trustworthiness, and Gullibility." *American Psychologist*, 35, 1-7.
Rousseau, D., Sitkin, S., Burt, R., and Camerer, C., 1998, "Not So Different after All: A Cross-discipline View of Trust," *Academy of Management Review*, 23, 393-404.
Saaksvuori, L., 2013, "Voluntary Formation of Centralized Sanctioning Institutions." *The Journal of Socio-Economics*, 44, 150-159.
Sapienza, P., Toldra, A., and Zingales, L., 2010, "Understanding Trust", *The Economic Journal*, 123(573), 1313-1332.
Schechter, L., 2007, "Traditional Trust Measurement and the Risk Confound: An Experiment in Rural Paraguay", *Journal of Economic Behavior and Organization*, 62, 272-292.
Schmidt, K., and Schnitzer, M., 1995, "The Interaction of Explicit and Implicit Contracts," *Economic Letters*, 48, 193-199.
Schnedler, W., and Vadovic, R., 2011, "Legitimacy of Control", *Journal of Economics and Management Strategy*, 20 (4), 985-1009.
Scott, R. E., 2000, "The Limits of Behavioral Theories of Law and Social Norms," *Virginia Law Review*, 86, 1603-1647.

Scott, R. E., 2003, "A Theory of Self-Enforcing, Indefinite Agreements", *Columbia Law Review*, 103, 1641-1699.
Sliwka, D., 2007, "Trust as a Signal of a Social Norm and the Hidden Costs of Incentive Schemes," *American Economic Review*, 97, 999-1012.
Sloof, R., and Sonnemans, J., 2011, "The Interaction between Explicit and Relational Incentives: An Experiment," *Games and Economic Behavior*, 73, 573-594.
Spolaore, E., and Wacziarg, R., 2013, "How Deep Are the Roots of Economic Development?" *Journal of Economic Literature*, 51 (2), 325-369.
Sullivan, J. L., and Transue J. E., 1999, "The Psychological Underpinnings of Democracy: A Selective Review of Research on Political Tolerance, Interpersonal Trust, and Social Capital," *Annual Review of Psychology*, 50, 625-650.
Tabellini, G., 2008, "The Scope of Cooperation: Values and Incentives." *Quarterly Journal of Economics*, 123, 905-950.
Tabellini, G., 2010, "Culture and Institutions: Economic Development in the Regions of Europe." *Journal of European Economic Association*, 8, 677-716.
Turner, J., 2005, "Explaining the Nature of Power: A Three-process Theory," *European Journal of Social Psychology*, 35, 1-22.
Uslauner, E., 2002, *The Moral Foundations of Trust*. New York: Cambridge University Press.
Wilkinson-Ryan, T., 2012, "Transferring Trust: Reciprocity Norms and Assignment of Contract", *Journal of Empirical Legal Studies*, 9 (3), 511-535.
Williamson, O., 1993, "Calculativeness, Trust, and Economic Organization." *Journal of Law and Economics*, 36, 453-486.
Yamagishi, T., Cook, K. S., and Watabe, M., 1998, "Uncertainty, Trust, and Commitment Formation in the United States and Japan", *The American Journal of Sociology*, 104(1), 165-194.
Yamagishi, T., and Yamagishi, M., 1994, "Trust and Commitment in the United States and Japan", *Motivation and Emotion*, 18(2), 129-166.
Zaheer, A., and Venkatraman, N., 1995, "Relational Governance as an Inter-organizational Strategy: An Empirical Test of the Role of Trust in Economic Exchange," *Strategic Management Journal*, 16, 373-392.
Zaheer, A., McEvily, B., and Perrone, V., 1998, "Does Trust Matter? Exploring the Role of Inter-organizational and Interpersonal Trust on Performance," *Organization Science*, 9, 141-159.
Zaheer, A., McEvily, B., and Perrone, V., 2003, "Trust as an Organizing Principle," *Organization Science*, 14, 91-103.
Zak, P., and Knack, S., 2001, "Trust and Growth," *The Economic Journal*, 111, 295-321.
Zak, P., Kurzban, R., and Matzner, W., 2004, "The Neurobiology of Trust," *Annals New York Academy of Sciences*, 1032, 224-227.
Ziegelmeyer, A., Schmelz, K., and Ploner, M., 2012, "Hidden Costs of Control: Four Repetitions and an Extension", *Experimental Economics*, 15, 323-340.

Chapter 2

A Fine Rule From a Brutish World?

An Experiment on Endogenous Punishment Institution and Trust

Huojun Sun, Maria Bigoni

Abstract. By means of a laboratory experiment, we study whether the endogenous adoption of a collective punishment mechanism can help a society coordinating on an efficient outcome, characterized by high levels of trust and trustworthiness. The experiment comprises three games. The first is a binary trust game, in which the only equilibrium strategy is not to trust, and not to reciprocate. The second game is identical to the first one, but we exogenously introduce a collective punishment mechanism under which cheating is sanctioned and the severity depends on the number of other trustees in society who choose not to cheat. This creates a coordination game with a second, Pareto superior equilibrium with full trust and full trustworthiness. The third game is designed to study whether the possibility of endogenously adopting collective punishment by means of a majority-voting system facilitates coordination on the efficient equilibrium. In theory, most subjects, regardless of their preferences and expectations, should vote in favor of collective punishment. As a consequence, the outcome of the vote cannot be interpreted as a signal of others' intentions, and it should not matter whether collective punishment is exogenously imposed or endogenously adopted. An alternative, behavioral hypothesis is that voting can work as a coordination device. We find that the introduction of the punishment mechanism induces a significant increase in the levels of trustworthiness, and to a lesser extent also of trust. The endogenous introduction of the mechanism by means of a majority-voting rule does not significantly improve coordination on the efficient equilibrium. In contrast with our theoretical predictions, not all subjects seem to be able to anticipate the change in behavior induced by the introduction of collective punishment, and a majority of them vote against it. Subjects seem to be unable to endogenously adopt an institution which, when exogenously imposed, proves to be efficiency enhancing.

Keywords: Coordination, Majority Voting, Social Sanctions, Trust Game

2.1 Introduction

At least since Aristotle's time, there has been a general consensus among legal scholars that the law defined as an obligation backed by powerful state coercion can create and maintain social order, such as enforcing property rights, adjudicating disputes, and providing an efficient level of public goods through adequately collecting a variety of taxes. Both theoretical and empirical studies have shown that a well-functioning and impartial legal system largely enhances societal trust, thereby promoting trade and economic development (Algan and Cahuc, 2013; Guiso, et al., 2008; Tabellini, 2008). Particularly in a standard contractual relationship, better enforcement, it is typically assumed, can increase the likelihood of contract performance by increasing the probability of the sanction and the cost of breach, naturally stimulating all manner of reliance investments that have specific value in the contractual relationship (Polinsky and Shavell, 2008).[28] Nearly half of the world's governments, however, fail to provide a sufficiently strong system of contract enforcement (Leeson and Williamson, 2009), and even abuse their authority to engage in profit-seeking punishment, which is detrimental to the country's economic performance (Xiao, 2013). Therefore, it becomes of paramount importance to understand how people who lack the protection of an effective legal environment can establish private-order institutions (or norms) to facilitate mutually advantageous exchanges.

In his influential anthropological field study on the cattle-control norms in rural Shasta County, California, Ellickson (1986, 1991) shows that social norms may work as effective mechanisms of social control. He argues that, when the social fabric is sufficiently dense and connected, social norms might supersede the legal rules, even if transaction costs are high – or precisely for that reason, as it is argued. Social norms have long been recognized as having great influence on individual behavior in social sciences, such as economics (Elster, 1989), sociology (Hechter and Opp, 2001), social psychology (Cialdini, et al., 1990; Schultz et al., 2007) and legal studies (Posner, 1997; Posner and Rasmusen, 1999). Nonetheless, the definition of a social norm is still controversial. One can consider two different meanings of the concept of social norm: *descriptive* norm, and *injunctive* norm (Cialdini, et al., 2006). The former is often adopted by social scientists, and refers to what most people do, to the commonly observed behavior, in contrast to what deviants do. The latter, commonly adopted by philosophers, refers to what one ought to do in order to gain social approval and to be rewarded, or to avoid censure and informal punishment (Cooter, 1998).[29] While Cooter (1998) places more emphasis on the second type of concept, Bicchieri's (2006)

[28] Introducing a third-party intervention into an investment game, Charness et al. (2008) reveal that the incentives (i.e. sanctions or rewards) implemented by an independent third-party significantly increase trust and trustworthiness in the investment game.

[29] Krupka and Weber (2013) empirically show that differences in injunctive norms – which they elicit by means of a novel approach based on incentivized coordination games – may explain the observed behavioral differences that emerge across several previous experimental dictator games.

formal definition of "social norm" encompasses both aspects, by stating that a behavioral rule is a social norm if (i) people are aware of rule existing and know that it applies to the situation under analysis (contingency condition), (ii) they expect that the others will conform to the rule (empirical expectations condition), and (iii) they believe others to think that people ought to obey the rule (normative expectations condition). It will soon become clear that the second condition is the one playing the most crucial role in our study.

Anderlini and Terlizzese (2013) theoretically study the introduction of a social norm into a standard contractual relationship, by letting the promisor's behavior be constrained by the average behavior of other promisors in a society. More specifically, in their model they represent a bilateral contractual relationship in the absence of contract enforcement as a one-shot binary trust game. Think for instance of an investor and an agent, strangers to each other. The investor lends some money to the agent, who makes an investment, and this investment generates a surplus proportional to the invested sum. The agent then decides whether to cheat and keep the entire surplus, or to share it with the investor. Cheating entails a cost, characterized by two components: one component is idiosyncratic and depends on the exogenously given "type" of the agent, while the second component is socially determined and common to all agents, and depends on the total number of transactions in a society that go through without cheating.[30] Hence, the stronger the norm of trustworthiness in a society, the higher the cost of cheating for the agents. Anderlini and Terlizzese (2013) note that the norm-driven component of the cheating cost can be interpreted as reflecting psychological remorse when the agent's action deviates from average behavior (Huang and Wu, 1994), or as resulting from a collective punishment mechanism, whose effectiveness depends on average behavior. Our experimental design adopts the second perspective, potentially inflicting a sanction on the dishonest agents. The introduction of this norm-driven component of the cost of cheating transforms the trust game into a coordination game with high-trust and low-trust equilibria, which are Pareto-ranked.

Existing experimental evidence indicates that norms of trustworthiness may differ across societies (Buchan et al., 2002), and such a difference might affect individual behavior, inducing the emergence of one or other of the equilibria. The issue of how social norms emerge in societies, however, remains largely unexplored. Anderlini and Terlizzese (2013) assume that the "social sensitivity" to the norm-driven component of the cheating cost is exogenously given. In this study we take a further step, and investigate the effects of the endogenous adoption of a collective punishment mechanism whose intensity is proportional to the strength of the norm of trustworthiness in society. More specifically, we investigate whether the adoption of such mechanism through majority voting can help a society in coordinating on an

[30] Previous experimental studies have revealed that individuals involved in social dilemmas are heterogeneous in terms of social preferences (Blanco et al., 2011). Anderlini and Terlizzese (2013) assume that there are two types of agents, high-type and low-type agents, who differ in their preference for honesty and the magnitude of the psychological cost they suffer when abusing their partner's trust.

efficient equilibrium, characterized by high levels of trust and trustworthiness.[31] Starting from a simplified version of Anderlini and Terlizzese's model, we theoretically show that most subjects, regardless of their preferences and expectations, vote in favor of the punishment mechanism, hence this mechanism will be endogenously introduced. As a consequence, a majority vote in favor of collective punishment cannot be interpreted as a signal of subjects' intentions, and it should not matter whether collective punishment is exogenously imposed or endogenously adopted. This theoretical prediction contrasts with the findings of recent experimental studies, which revealed that the endogenous adoption of institutions induces higher cooperation levels in social dilemma situations, relative to the case in which the same institutions are exogenously implemented; scholars refer to this phenomenon as "the dividend of democracy" (Dal Bo et al., 2010; Markussen et al., 2014; Sutter et al., 2010; Tyran and Feld, 2006).[32]

The theoretical model informs our empirical analysis, which is based on a laboratory experiment. In our experiment, each subject plays three one-shot games with three different partners. The first game is a standard binary trust game. In the second game, a collective punishment mechanism is exogenously introduced, under which cheating is sanctioned with a severity that depends on the trustworthiness of the others. In the third part of the experiment they have to choose whether to play according to the rules of the first, or of the second game, by means of a majority voting mechanism. To reduce the risk of spillover effects, the outcomes of these three games are not revealed to the subjects until the end of the session. In half of the sessions the sequence of the first and the second game is reversed, to control for possible order effects. This design allows us to test whether subjects are willing to opt for having a collective punishment mechanism in place, and to study how the endogenous adoption of such mechanism affects individual beliefs and behavior.

We report four main findings. First, in line with the model, we find that the introduction of collective punishment induces a significant increase in the levels of trustworthiness, and to a lesser extent also of trust. Second, the endogenous introduction of the punishment mechanism by means of a majority-voting rule does not significantly change behavior, with respect to what is observed when the mechanism is exogenously imposed. Third, in contrast with our theoretical predictions, not all subjects seem to be able to anticipate the change in behavior induced by the introduction of collective punishment, and a majority of them vote against it. We also find that subjects with higher cognitive abilities and with a

[31] In real world, we rarely observe that the norm is established through a voting mechanism. However, people in a community could publicly express their attitudes towards a specific norm (Kadens and Young, 2013). Therefore, we use the voting mechanism as a simple way to capture the essential dimension of the public expression of the norm.

[32] Vollan et al. (2013) replicate Tyran and Feld's (2006) study using a sample of Chinese people. They observe that the cooperation rate is higher under an exogenously imposed institution than under a democratically selected rule. Their analyses show that this result is mainly driven by the fact that the Chinese culture attributes a high importance to obeying authorities.

background in statistics are more likely to vote in favor of the punishment mechanism. Finally, in an additional treatment, we provide information about the aggregate behavior with and without collective punishment; we find that on average this additional information does not increase the likelihood of the mechanism being adopted.

The paper has the following structure: Section 2 discusses how our work relates to the existing literature. Section 3 presents our theoretical model and testable predictions, and describes the experimental design and procedures; Section 4 illustrates the main results of the experiments; Section 5 concludes.

2.2 Related Literature

Our paper builds upon a considerable number of studies on the effects of informal institutional arrangements on individual behavior in social dilemma situations, in the absence of a powerful state (Ostrom, 1990). A variety of decentralized governance institutions have emerged in remarkably diverse environments (Bernstein, 1992, 2001; Greif, 2006).

In early trade, Greif (1989, 1993) portrays a well-defined and cohesive group based on Jewish religion and family origins in the Maghreb, the "Maghribi traders" who engage in long-distance, large-scale trading across the whole Muslim Mediterranean. Lacking effective legal institutions, these merchants rely on informal sanctions based on collective relationships within an exclusive coalition. Members of the Maghribi traders' coalition always recruit agents from their own coalition, convey information about their agent's misbehavior swiftly to other members, and collectively ostracize agents who abused their principal's trust, thereby successfully resolving the problem of commitment in one-shot bilateral contractual relationships, even in the absence of binding contracts. Similar social sanction institutions also proved to work well in Mexican California before the time of the gold rush in 1848-1949 (Clay, 1997; Clay and Wright, 2005) and in the practice of group lending in the developing countries (Besley and Coate, 1995).

These anthropological studies on informal sanctioning institutions emphasize the role of information-sharing among the investors in regulating the agents' behavior.[33] By contrast, our research adopts an alternative approach: in our set-up, in order to gain the investors' trust, agents are allowed to adopt a collective punishment mechanism whose severity depends on the average behavior of all agents' in the society. Therefore, the effectiveness of our mechanism relies on the agents' and the investors' beliefs, rather than on information-sharing.

[33] In Kimbrough and Rubin (2015), subjects play the trust game under a highly anonymous set-up, where the investors only know the group identity of their agents. When the investors are allowed to share their transaction experience with other investors, the groups with high percentages of dishonest agents are collectively boycotted, which secures the high efficiency of the market.

Secondly, our paper is also related to the literature on expressive law (Cooter, 1998; McAdams, 2000a, 2000b; Posner, 1998, 2000). The classic "law and economics" approach focuses on deterrence: a law enforced by a sanction increases the expected costs of the illegal activity and thereby induces compliance (Becker, 1968; Polinsky and Shavell, 2000). Despite its success in many cases, this view can hardly explain why most people obey legal rules even in a situation where they could improve their material payoffs if they violate an obligation (Tyler, 1990).

The expressive law theories provide several possible explanations. One potential reason is that the legitimate rules may influence individual preferences by letting people realize which behavior is legally prohibited. Another possible reason is that even though legal rules are mild, they may act as coordination devices that help people predict what others will do. Announcing an expressive legal rule that does not change the equilibrium is a form of "cheap talk". Despite being "cheap", some forms of talk, especially announced by a powerful authority or determined by a majority voting mechanism, have been found to actually coordinate individuals' behavior in social dilemma situations.[34]

These theories have increasingly gained momentum among theoretical scholars. However, only a handful of experimental studies have examined how mild rules actually influence individual behavior (Bohnet and Cooter, 2003; Galbiati and Vertova, 2008; McAdams and Nadler, 2005; Tyran and Feld, 2006). Our experimental study contributes to this literature in two aspects. First, the social sanction in our experiment is not always a deterrent but works only if the majority behaves honestly. Therefore, the socially shared beliefs are crucial to affect individual behavior. Second, instead of a powerful authority announcing the rule, the rule in our paper is determined by a voting mechanism, which enhances the legitimacy of the rule and may influence individual behavior through changing people's preferences or coordinating their beliefs. Our study is also related to the experimental literature on the trust game with punishment (Fehr and Rockenbach, 2003; de Quervain et al., 2004; Vollan, 2011), however, it departs substantially from that strand of literature, in that the activation and size of the punishment in our case depends on the behavior of the society as a whole, and not on the individual decision of a trustor, who may sanction an untrustworthy trustee.

A closer relation emerges between our work and the literature concerning the endogenous adoption of institutions. Recent experimental studies have revealed that an institution established endogenously (e.g. through a voting mechanism) can induce higher cooperation levels in social dilemma situations, compared to the same

[34] In Kamei (2014), subjects are more likely to contribute to cooperation in the public good game when a mild sanction rule is collectively selected even without altering the equilibrium of full free riding. Unexpectedly, the author also finds that the positive effect of endogenous selection of the institution does not disappear even when subjects enter into an exogenous setting with an identical institution.

institution implemented exogenously on an otherwise identical group (Dal Bo et al., 2010; Markussen et al., 2014; Sutter et al., 2010; Tyran and Feld, 2006).

Broadly speaking, there are two approaches to endogenous institution formation. Under the first approach, groups are fixed, and members of each group are asked to vote for a specific scheme or to choose one from a broad menu of schemes (Kosfeld et al., 2009; Sutter et al., 2010). Previous experimental results indicate that the endogenous adoption of informal sanctioning (Tyran and Feld, 2006; Ertan et al., 2009) or rewarding (Sutter et al., 2010) institutions largely enhances the levels of cooperation, relative to the case in which the same institutions are imposed exogenously. In addition, subjects tend to converge on the most efficient institutions as they gain experience over a course of multiple votes (Putterman et al., 2011).

The second approach is the "voting by feet" mechanism in open communities (Gurerk et al., 2006, 2014; Fehr and Williams, 2013) where subjects can choose between different institutions and endogenously form groups with other members who also select the same institution. They find that prosocial individuals adopting efficient punishment institutions under endogenous selection quickly establish a cooperative culture. These institutions increasingly attract other types of subjects to migrate to these more cooperative groups and to comply with the prevailing norms. Therefore, endogenously chosen institutions induce the whole group to coordinate on high cooperation levels, so that in practice there is little or no need to recur to punishment.

Most experimental papers on endogenous formation of institutions are based on the framework of public good games, except Dal Bo et al. (2010) who use a prisoner's dilemma game. To the best of our knowledge, no existing empirical research addresses the effect of endogenous adoption of social sanction mechanisms on individual behavior in the trust game. Compared to the previous studies, our peculiar design, i.e. within-subject design without feedback across games, allows us to identify the important role of ex-ante beliefs of subjects in equilibrium selection. Furthermore, since subjects are exposed to the trust game with and without the collective punishment mechanism before voting for the preferred rule governing their interactions, we can investigate how different experiences of the effects of collective punishment affect individual's voting behavior. Finally, in line with what argued by Markussen et al. (2014), that "the dividend of democracy" is driven by the signaling function of voting which promotes coordination on high-contribution outcomes, our design also allows us to test whether the endogenous adoption of the punishment mechanism could be taken as signal of the general willingness to coordinate on a high trust and high trustworthiness equilibrium.

2.3 Materials and Methods

In this section, we first present the theoretical model that informs our experimental design, and derive the predictions, which will be empirically tested in Section 4. Then

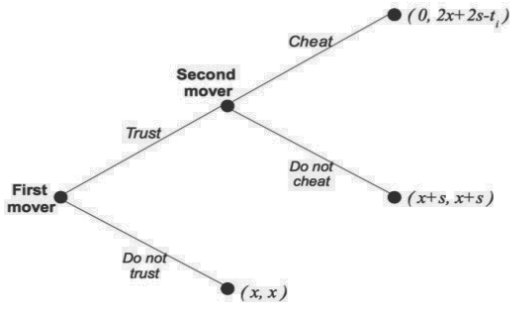
we describe the experimental design and procedures.

2.3.1 Theoretical model

As a baseline situation, we consider the binary investment (or trust) game depicted in **Figure 2.1**. Each player is initially given an endowment $x > 0$. The first mover decides whether to trust the second mover or not. If she chooses not to trust her partner, both of them keep their endowments and leave the transaction. If instead she chooses to trust and transfers her endowment, the second mover efficiently invests the money he received, together with his own endowment, to generate a total of $2x + 2s$, with $s > 0$. The second mover now has to choose whether to cheat on the first mover, and keep the entire amount leaving the first mover with nothing, or to split it equally with her, so that each party gets $x + s$. We further assume that, in the society, all players face equal chances of playing the game in the role of the first or second mover.

Following the Anderlini and Terlizzese's (2013) approach, we assume that there are two types of players in the society, "high" (*H*) and "low" (*L*). *H*-type players have a preference for honesty and suffer a psychological cost $t_H > 0$ when abusing their partner's trust, and the idiosyncratic cost of cheating for the *H*-type players is so high that they will never cheat: $t_H > x + s$. *L*-type players instead are only interested in (expected) monetary payoffs (i.e. $t_L = 0$), so they will always cheat when in the role of second movers. For simplicity, we also assume that players are risk neutral.

Figure 2.1: the basic trust game.

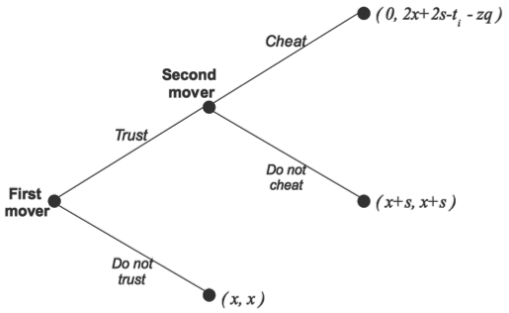


At the beginning of the stage game, all players are randomly assigned to the role of first or second mover, and matched in pairs. Players choose their strategy before knowing their role, and the strategy determines their action both as the first and as the second mover. Let *p* represent the proportion of *H*-type players in the society, which is assumed to be common knowledge. It is straightforward to verify that, regardless of his type, a player will trust as a first mover if $p > \frac{x}{x+s}$. Let us denote this threshold Θ.

a). A collective punishment mechanism

Now consider the introduction of a collective punishment mechanism into the trust game, as depicted in **Figure 2.2**. In this new game, besides possibly suffering the psychological cost t_i, the player who cheats faces the risk of being punished by his peers. This potential punishment zq depends on two elements: the strength z of the sanction implemented collectively by the players who do not cheat - which is exogenously given - and the fraction q of transactions in society where cheating does not take place. The behavior of the *H*-type players as second movers is not affected by the sanction, as they would never cheat, in any case. The behavior of the *L*-type players instead might change, as they may choose not to cheat either, if $q_i \geq \frac{x+s}{z}$, where q_i represents player *i*'s beliefs about q. Let us denote this second threshold θ.

Figure 2.2: the trust game with an exogenous collective punishment mechanism.



In the following, we assume that $0 < \theta < \Theta < 1$, which is consistent with the parameters we adopt in the experiment. If the proportion p of *H*-types in society is larger than the threshold θ then in the game with collective punishment, any player *i* will never cheat as a second mover and will always trust as a first mover, regardless of his own type. If instead $p < \theta$, this game becomes a coordination game with two Pareto-ranked equilibria. In the low-efficiency equilibrium, *L*-type players cheat in the role of second mover, and nobody trusts as a first mover. In the high-efficiency equilibrium, instead, neither *L*-types nor *H*-types cheat as second movers, and everybody trusts as a first mover. There exists, however, the risk of miscoordination, as subjects cannot be certain of the strategy the others will adopt.

Let β_i be player *i*'s belief about the fraction of the other players who adopt the cooperative strategy *(trust, do not cheat)* in the trust game with a collective punishment mechanism. Then, we could obtain the belief q_i about the total number of players who will not cheat, which depends on two elements: the proportion of

intrinsically trustworthy players p, and the belief β_i.

$$q_i = p + (1 - p)\beta_i$$

To summarize, for any value of p, the introduction of a collective punishment mechanism does not decrease trustworthiness with respect to the baseline scenario, and might increase both trust and trustworthiness, if the proportion of *H*-types p is high enough, or if a sufficiently high number of players have high beliefs β_i about the fraction of the other players who adopt the cooperative strategy.[35]

Hypothesis 1: *In presence of a collective punishment mechanism, the levels of trust and trustworthiness are equal or higher than in the baseline scenario.*

b). Endogenous adoption of the collective punishment mechanism

We now consider the case in which, prior to playing the game (and before roles are assigned), players express their preference on whether to have or not a collective punishment mechanism in place. More specifically, we consider the case in which the implementation of the punishment mechanism is determined by a majority voting rule. The main question we would like to pursue is whether this mechanism can affect the beliefs q_i, thus serving as a coordination device to drive the society towards the efficient equilibrium.

Let us consider again the behavior of player i in the game with a collective punishment mechanism in place. Depending on the player *i*'s belief q_i, we can envisage five possible cases based on the types of players. For the *L*-type, i.e. selfish players, there are three possible scenarios:

(i.) $q_i \leq \theta < \Theta \leq 1$: the player chooses the strategy *(do not trust, cheat)*;
(ii.) $\theta < q_i < \Theta \leq 1$: the player chooses the strategy *(do not trust, do not cheat)*;
(iii.) $q_i \geq \Theta$: the player chooses the strategy *(trust, do not cheat)*.

For the *H*-type, i.e. intrinsically trustworthy players, there are two possible scenarios:

(iv.) $q_i < \Theta \leq 1$: the player chooses the strategy *(do not trust, do not cheat)*;
(v.) $q_i \geq \Theta$: the player chooses the strategy *(trust, do not cheat)*.

However, these boil down to the first three scenarios, as (ii) and (iv) coincide, as well as (iii) and (v). Let us now calculate the player's expected profit in the trust game with collective punishment, under these three alternative scenarios. Remember that in

[35] An alternative, behavioral hypothesis is that the exogenous introduction of a punishment mechanism could crowd out intrinsic motivations for trustworthiness (Bowles and Polania-Reyes, 2012; Fehr and Rockenbach, 2003).

the basic trust game, when $p < \Theta < 1$, player *i*'s expected profit is equal to x, no matter what, while if $p > \Theta$, then in the basic trust game player i would trust as a first mover, and everyone else does the same. In this case his expected payoff depends on his type.

Scenario (i). As a first mover, player i will not trust, hence he will be sure to earn x. As a second mover he will earn x if his partner does not trust, and $2x + 2s - zq_i$ if his partner chooses to trust. Because β_i is player *i*'s belief about the fraction of other players who adopt the cooperative strategy *(trust, do not cheat)*, he will expect the former event to take place with probability $1 - \beta_i$, and the latter with probability β_i. Hence, the expected profit a player can obtain in the game with collective punishment is:

$$E(\pi^s) = \frac{1}{2}x + \frac{1}{2}[x(1-\beta_i) + (2x + 2s - zq_i)\beta_i] = x + \frac{\beta_i}{2}(x + 2s - zq_i)$$

The expected profit above is greater than x if $q_i < \frac{x+2s}{z}$, which is true for every $q_i \leq \theta = \frac{x+s}{z}$. Hence, a selfish player with belief $q \leq \theta$ will prefer to have the punishment mechanism in place.

Scenario (ii). As a first mover, the player i will not trust, hence he will be sure to earn x. As a second mover he will earn x if his partner does not trust, which happens with probability $1 - \beta_i$, and $x + s$ if his partner chooses to trust, which happens with probability β_i. Hence, the expected profit in the game with collective punishment is:

$$E(\pi^s) = \frac{1}{2}x + \frac{1}{2}[x(1-\beta_i) + (x + s)\beta_i] = x + \frac{\beta_i}{2}s \geq x$$

Hence, both a selfish player and an intrinsically trustworthy player with beliefs $\theta < q_i < \Theta$ will prefer to have the collective punishment mechanism in place.

Scenario (iii). As a first mover, player i will trust, hence he will earn $x + s$ with probability q_i and 0 with probability $1 - q_i$. As a second mover he will earn x if his partner does not trust, which happens with probability $1 - \beta_i$, and $x + s$ if his partner trusts, which happens with probability β_i. Hence, the expected profit a player can obtain in the game with collective punishment is:

$$E(\pi^s) = \frac{1}{2}q_i(x + s) + \frac{1}{2}[x(1-\beta_i) + (x + s)\beta_i] = \frac{1}{2}[q_i(x + s) + x + \beta_i s]$$

In this case, however, the expected payoff $E(\pi^b)$ in the basic trust game depends on player i's type, and on whether $p > \Theta$. If $p < \Theta \leq q_i$ then $E(\pi^b) = x < E(\pi^s)$ and player i will vote in favor of the punishment mechanism. Indeed, the expected

profit in presence of collective punishment is grater than x if $q_i(x+s)+\beta_i s > x$, which holds for every $q_i \geq \Theta = \frac{x}{x+s}$. Hence, both a selfish player and an intrinsically trustworthy player with $p < \Theta \leq q_i$ will prefer to have the punishment mechanism in place.

If instead $p > \Theta$, the preferences of *H*-type and *L*-type players will differ. If player i is an *H*-type, in the basic trust game as a first mover he will trust, hence expecting to earn $x+s$ with probability p and 0 with probability $1-p$. As a second mover he earns $x+s$ because all first movers should trust. Hence, the expected profit a player can obtain is:

$$E(\pi^b) = \frac{1}{2}p(x+s) + \frac{1}{2}(x+s) = \frac{1+p}{2}(x+s)$$

Consider also that if $p > \Theta$ then $q_i = \beta_i = 1$ for all players. Hence $E(\pi^s) = x + s \geq E(\pi^b)$: when $p > \Theta$, *H*-type players will always vote in favor of collective punishment.

By contrast, if player i is an *L*-type, in the basic trust game as a first mover he will trust, hence he will earn $x+s$ with probability p and 0 with probability $1-p$. As a second mover he earns $2(x+s)$ because all first movers should trust, and he will cheat. Hence, the expected profit a player can obtain is:

$$E(\pi^b) = \frac{1}{2}p(x+s) + (x+s) = \frac{2+p}{2}(x+s) > x + s = E(\pi^s)$$

Hence, when $p > \Theta$, *L*-type players will vote against collective punishment.

Hypothesis 2: *H-type players will always vote in favor of the introduction of a collective punishment mechanism; L-type players will also vote in favor of it, unless the proportion of H-types is sufficiently high to induce them to trust in the Baseline ($p > \Theta$).*

As a consequence, the collective punishment mechanism will always be adopted if $\Theta \geq 0.5$, which is the case in our experiment. Hence, we can state the following hypothesis on the effects of the vote on trust and trustworthiness.

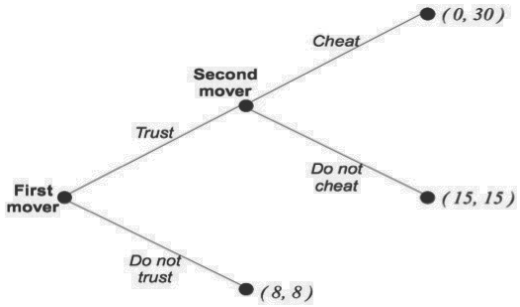
Hypothesis 3: *a majority vote in favor of the collective punishment mechanism does not reveal anything on the distribution of types and beliefs, hence it should not affect trust and trustworthiness levels, as compared to those observed when the mechanism is exogenously introduced.*

2.3.2 Experimental design

Our experimental treatments were based on variants of the binary-choice trust game (Bohnet et al., 2008) introduced in the previous section. We adopted a

within-subject design, in which each participant was exposed to three treatments: *Baseline*, *Exogenous* and *Voting*. At the beginning of the session each subject was assigned into a group of six. In each treatment, subjects were paired with one of their group's members, to play a one-shot game. Matching across treatments was done so to ensure that no two subjects would meet more than once.[36] The group composition was kept constant during the whole session.

Figure 2.3: the basic trust game, with the parameterization adopted in the *Baseline* treatment.



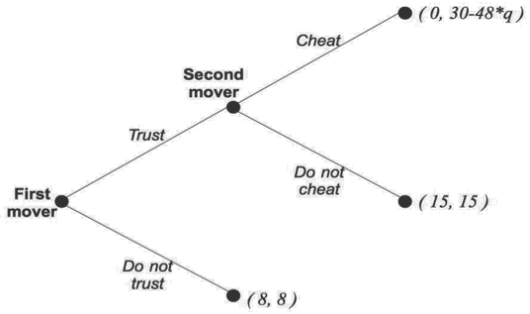
In the *Baseline* treatment, subjects were asked to play the binary trust game (i.e. Baseline game), as parameterized and represented in **Figure 2.3**. We adopted the strategy method (Brandts and Charness, 2011): all subjects had to choose their action both as a first mover and as a second mover, before knowing which role they would actually be assigned. Once all subjects had made their two choices, roles were randomly assigned and subjects were matched in pairs. In each pair, payoffs were determined by the choice each of the two players had made for the role he was actually assigned.

In the *Exogenous* treatment, the strategic environment, the information structure and the options subjects had to choose were the same as in the *Baseline* game but, here, a collective punishment mechanism was exogenously introduced, under which cheating was sanctioned and the severity depended upon the number of subjects in the group, who chose not to cheat as second movers (i.e. Exogenous game, see **Figure 2.4**).[37]

[36] With the exception of the Voting-IF treatment, as illustrated below.

[37] In order to be consistent with the theoretical model, in **Figure 2.4** the size of the sanction ($48*q$) is expressed in terms of the fraction q of subjects who choose not to cheat, in a group of six. In fact, in the experimental instructions, we expressed that variable as a function ($8*N$) of the number N of trustworthy players (see ***Appendix 2***). With the parameters adopted in our set up, we have that $\theta=0.35$ and $\Theta=0.53$. This implies that trusting is profitable even in the *Baseline* treatment, if the proportion of *H-types* in the society is higher than 0.53, while if this proportion is as high as 0.35, in the *Exogenous*

Figure 2.4: the trust game with an exogenously imposed collective punishment mechanism, with the parameterization adopted in the *Exogenous* treatment.



After experiencing these two variations of the trust game, subjects entered the third treatment (*Voting*). At the beginning of this last treatment, before roles were assigned, subjects were asked to vote for implementing either the *Baseline* or the *Exogenous* game, then a majority voting mechanism determined which of the two variations of trust games would have been ultimately played within the group, in this final phase. Abstention was not allowed. Before playing this third trust game, subjects were informed of the number of their group members who voted in favor of either option.

To reduce the risk of spillover effects, the outcomes of these three games were not revealed to the subjects, until the end of the session.[38] In addition, to control for possible order effects, in four sessions subjects were exposed to the *Baseline* treatment first, then they played the *Exogenous* treatment and finally the *Voting* treatment, while in other four sessions the order of the first two treatments was reversed.[39]

In order to examine whether having information about the aggregate behavior with and without collective punishment affected the individual voting behavior, in four of the sessions we introduced one additional treatment, after the *Voting* treatment. This treatment, denoted *Voting-IF*, was identical to the *Voting* treatment, with two exceptions. First, before voting subjects received information on the aggregate behavior of their group members in the Baseline and Exogenous treatments. More specifically, they were shown the number of subjects who chose either option, as a first and as a second mover, in each of the two treatments. Second, subjects were told that their partner might have been be the same person as in one of the previous three games.

---

treatment not cheating becomes more profitable than cheating.

[38] Each part of the instructions was distributed and read just before subjects started to play the corresponding game, which implies that subjects had no prior knowledge about the next part of the experiment.

[39] For more information on the treatments and sessions, please refer to Table A in the **Appendix 1**.

Since our experiment was relatively complex, to ensure full understanding of the instructions, subjects were asked to complete a comprehension quiz with calculations and questions before making decisions in each stage game (see **Appendix 2**). Subjects were rewarded with €0.40 for each question they answered correctly at the first try. There were six questions per treatment (no questions before the *Voting-IF* treatment), hence subjects could earn in total €7.20 for the comprehension quiz.

At the end of the session, all subjects had to fill in a questionnaire including questions on their individual characteristics (gender, age, education, social status), general trust, risk attitudes, social preferences and cognitive abilities (see the **Appendix 3** for the complete text of the questionnaire). These questions allowed us to study how personal characteristics may affect the voting behavior, as well as the impact of the endogenous/exogenous introduction of collective punishment on individual behavior.

The experiment involved 96 subjects, divided in 8 sessions (see Table A in **Appendix 1**) and was conducted at the Bologna Laboratory for Experiments in Social Sciences (BLESS). Subjects were mostly undergraduate students at the University of Bologna, and were recruited through ORSEE (Greiner, 2015). About 53 percent of the subjects were male; nobody took part in more than one session. The experiment was programmed and implemented using the software z-Tree (Fischbacher, 2007). For each session, after showing up to the lab at the pre-scheduled session time, the 12 participants were randomly assigned to cubicles to avoid eye contact, and no communication was allowed during the experiment. The average session lasted about 1 hour and 15 minutes. Subjects were paid privately in cash at the end of the session and earned on average 18.25 Euros, including the earnings from the comprehension quiz. No show-up fee was given.[40]

2.4 Results

In this section we carry out four steps of analysis. First, we juxtapose data from the *Baseline* and the *Exogenous* treatments, in order to analyze whether exogenously introducing collective punishment enhances the levels of trust and of trustworthiness in society. Second, we study subjects' voting behavior, and test whether a majority of subjects vote in favor of collective punishment as predicted in our theoretical model. We also investigate who are the subjects who vote in favor of the punishment mechanism, and whether they differ from those who vote against it, along any significant dimension. Third, we examine whether the endogenous introduction of a collective punishment mechanism promotes efficiency by boosting trust and trustworthiness with respect to the case in which such a mechanism is exogenously

[40] For each session we recruited 15 subjects, to take into account possible no-show-ups, but only 12 students were randomly selected to participate in the experiments. Supernumerary subjects were paid 5 Euros and had to leave before the session started.

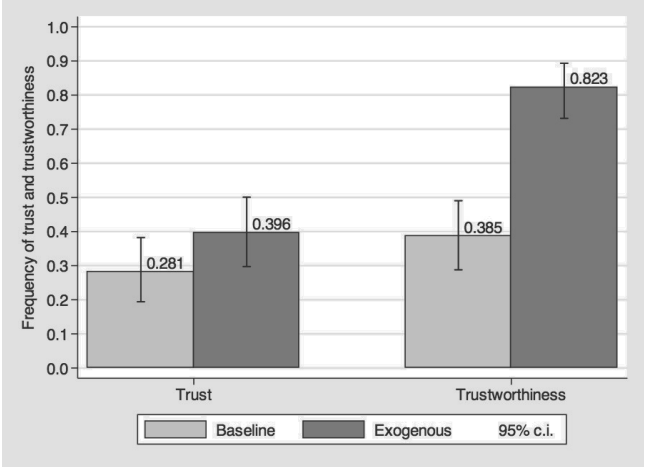
imposed. We also study whether the endogenous choice not to adopt collective punishment depresses trust and trustworthiness, as predicted by our model. Finally, we examine whether the exposure to information about the aggregate behavior of their members in the *Baseline* and *Exogenous* treatments affects a subject's decision to vote in favor of the punishment mechanism.

2.4.1 Effects of collective punishment, when exogenously imposed

The main difference between *Baseline* and *Exogenous* games lies in the way the payoff of the player in the role of a second mover (i.e. trustee) depends on the other trustees' behavior, in case he chooses to abuse his partner's trust. This manipulation has a direct effect on trustworthiness and only an indirect effect on trust, because the player in the role of a first mover (i.e. trustor) will change her behavior only if she expects collective punishment to have a (direct) effect on the others' levels of trustworthiness. For this reason, we first present the results about trustees' behavior and then illustrate trustors' behavior.

As shown in **Figure 2.5**, the fraction of the trustworthy actions is larger when the collective punishment mechanism is exogenously imposed. More specifically, only 38.5% of subjects in the role of trustee reciprocate trust in the *Baseline* treatment while 82.3% of trustees in the *Exogenous* treatment behave trustworthily. The difference is strongly significant ($p<0.001$). If not specified otherwise, comparisons across treatments are performed by means of logit regressions, where the only explanatory variable is a treatment dummy, and standard errors are robust for clustering at the subject's level. Two-tailed z-tests using each subject as an independent observation always confirm the results.

Figure 2.5: frequency of trustful and trustworthy choices in the *Baseline* and *Exogenous* treatments.



Notes: One observation per subject, per treatment. The whiskers represent 95% confidence intervals.

The impact of collective punishment on trustees' behavior emerges regardless of the order in which subjects are exposed to the *Baseline* and the *Exogenous* treatment, the level of trustworthiness being almost twice as high in the latter than in the former ($p<0.001$ in both cases, Table B in the **Appendix 1**). In addition, when we compare behavior across subjects, and focus exclusively on the first game played in each session, we observe that the difference in trustworthiness remains highly significant ($p<0.001$, Table B in the **Appendix 1**).

Figure 2.5 also shows that the overall level of trust is higher in the *Exogenous* than in the *Baseline* treatment. Specifically, while the average level of trust in the *Baseline* game is 28.1%, it reaches 39.6% in the *Exogenous* game, and the difference is statistically significant ($p=0.040$). However, if we control for the order effect, we find that when *Baseline* is implemented first the exogenously imposed punishment mechanism does not significantly enhance the trust ($p=0.784$). Conversely, when the punishment mechanism is implemented first but removed afterwards, the level of trust drops dramatically ($p=0.012$, see Table C in **Appendix 1**). We can summarize our results as follows.

Result 1: *the presence of a collective punishment mechanism significantly increases trustworthiness, and to a lesser extent also trust.*

2.4.2 Endogenous adoption of collective punishment

Our theoretical model predicts that, in the *Voting* treatment, *H*-types would always vote in favor of the collective punishment mechanism, while L-types would vote against it only if the proportion of H-types in society is very high (Hypothesis 2). Our data reveal instead that only a minority of subjects (30.2%) vote in favor of the mechanism, and that subjects' voting behavior does not seem to depend on their preferences or beliefs. This result does not depend on the order of the first two treatments: 29.2% of subjects vote in favor of the mechanism when the *Baseline* treatment is first played, while 31.2% opt for the punishment mechanism when subjects are first exposed to the *Exogenous* treatment, and the difference is not statistically significant ($p=0.825$).

Since we adopt the strategy method in the experiment, for every subject we observe both choices (as a trustor and a trustee) in each treatment. By looking at subjects' behavior as trustees in the *Baseline* treatment, we can classify subjects as *L*-types and *H*-types: by definition, those who do not cheat in the *Baseline* are *H*-types. Information on the choice as first movers is also relevant in order to predict voting behavior. Indeed, according to our model, *L*-type players would vote against the introduction of the punishment mechanism only if they trust in the *Baseline*. **Table 2.1** reports the distribution of subjects, along these two dimensions.

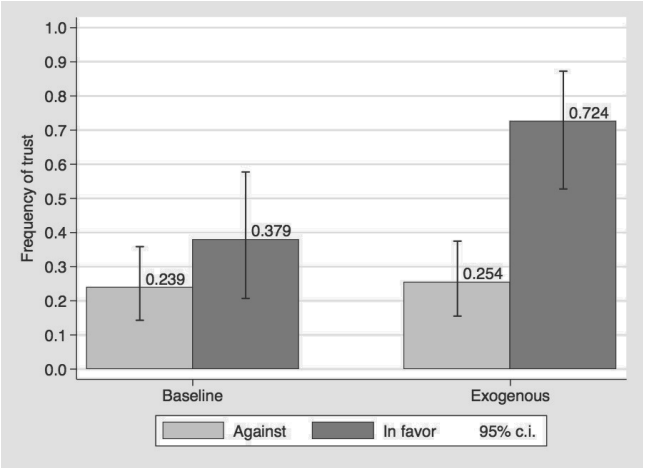
Table 2.1: subjects' behavior in the *Baseline* treatment.

	Trust in *Baseline*		
Reciprocate in *Baseline*	Yes	No	Total
Yes (*H-type)*	20.8%	17.7%	38.5%
No (*L-type)*	7.3%	54.2%	61.5%
Total	28.1%	71.9%	100.0%

Table 2.1 reveals that, according to our predictions, only 7.3% of the subjects would have voted against the adoption of collective punishment, in the *Voting* treatment, while in our experiment this proportion was much higher.

To better understand the source of this discrepancy between our results and the theoretical predictions, we now investigate the determinants of subjects' voting decision. First, we divide subjects into two categories, depending on their voting decisions: against collective punishment and pro-punishment. We find that these two categories of subjects have similar levels of trust and trustworthiness in the *Baseline* treatment, implying that there is no difference in the preferences or ex-ante beliefs between them (p=0.165 for the difference in trust level, and p=0.409 for the difference in trustworthiness level). P-values in this paragraph are obtained by means of logit regressions where the only explanatory variable is a dummy taking value one for subjects who voted in favor of the punishment mechanism in the *Voting* treatment, and with standard errors robust for clustering at the subject's level. Results are always confirmed by two-tailed z-tests using each subject as an independent observation.

Figure 2.6: differences in trust between subjects who voted in favor and against collective punishment.



Notes: One observation per subject, per treatment. The whiskers represent 95% confidence intervals.

In the *Exogenous* treatment, as revealed in **Figure 2.6**, subjects who vote in favor of collective punishment are more likely to trust their partners than others (72.4% vs. 25.4%, $p<0.001$). We also find that these pro-punishment subjects react more to the introduction of the punishment mechanism, i.e. they are more likely to increase their level of trust from the *Baseline* to the *Exogenous* game, as compared to the subjects who voted against the mechanism ($p=0.002$).

Table 2.2: Logit regressions on the determinants of subjects' voting behavior.

Dependent variable: Vote	**Model 1**	**Model 2**	**Model 3**
Trust-BL	0.126		0.021
	(0.109)		(0.086)
Trustworthiness-BL	0.022		0.069
	(0.106)		(0.111)
Trust-EX		0.342***	0.334***
		(0.054)	(0.033)
Trustworthiness-EX		0.146	0.089
		(0.120)	(0.104)
Controls	No	No	Yes
Number of Observations	96	96	96

Notes: Marginal effects from logit regressions (Standard errors robust for clustering at the session level are reported in parentheses). Trust-BL (Trustworthiness-BL) equals 1 for subjects choosing to trust (reciprocate) in the *Baseline* treatment; Trust-EX (Trustworthiness-EX) equals 1 for subjects choosing to trust (reciprocate) in the *Exogenous* treatment; *Controls* indicates the presence of fourteen regressors, aimed at controlling for subjects' individual characteristics. These include all the variables listed in **Table 2.3**. The symbols *, **, and *** indicate significance at the 10%, 5% and 1% level, respectively.

To dig deeper into these differences, we run a series of logit regressions (**Table 2.2**). The dependent variable indicates whether the subject voted in favor of collective punishment. In Model 1 we introduce subjects' choices in the *Baseline* as explanatory variables, finding that subjects' preferences and their ex-ante beliefs about others do not affect their voting behavior. In Model 2, we instead use their choices in the *Exogenous* treatment as explanatory variables. Our result shows that the probability that a subject votes in favor of collective punishment is 34.2% higher when she chose to trust in the *Exogenous*. This strongly significant difference reappears in the Model 3 where we include all four choices of subjects in both the *Baseline* and *Exogenous* treatments, which suggests that only those who can anticipate the impact of collective punishment on others' trustworthiness, and react to it with a higher level of trust, are inclined to vote in favor of it.

Result 2. *Only about 30% of subjects vote in favor of the collective punishment mechanism, and the voting behavior does not depend on subjects' preferences and beliefs.*

Our next step is to explore the question of whether subjects' individual characteristics affect their voting behavior. **Table 2.3** reveals that subjects who vote in favor of the punishment mechanism have higher cognitive abilities than the others,

as supported by an ordered logit regression on the number of correct answers given to the three questions of the Cognitive Reflection Test. The result is confirmed if we look at the IQ test to measure subjects' cognitive abilities, which also reveals that subjects who vote in favor of collective punishment are significantly more likely to answer correctly. Results in **Table 2.3** also indicate that, although our experimental design is relatively complicated, subjects could answer most of the control questions correctly before playing the game and, on average, the subjects who voted against or in favor of the punishment mechanism could provide a similar number of right answers. This implies that all subjects could well understand the instructions, and that differences in the voting behavior are not driven by comprehension problems.

Table 2.3: Individual characteristics and voting.

Individual characteristics	**Against (N=67)**	**In favor (N=29)**	**Significance of the difference**
Male	49.3%	62.1%	$p>0.1$[b]
Age	25.6	24.1	$p=0.079$[a]
Higher education	67.2%	44.8%	$p=0.049$[b]
CRT	1.1	1.6	$p=0.093$[a]
IQ	1.2	1.7	$p=0.002$[a]
Economics	50.7%	48.3%	$p>0.1$[b]
Statistics	44.8%	58.6%	$p>0.1$[b]
Game theory	28.4%	20.7%	$p>0.1$[b]
Trust	17.9%	17.2%	$p>0.1$[b]
Altruism	7.8	8	$p>0.1$[a]
Risk aversion	5.8	5.2	$p>0.1$[a]
RightAnswerBL	5.4	5.2	$p>0.1$[a]
RightAnswerEXO	5.1	5.2	$p>0.1$[a]
RightAnswerVOTE	5.6	5.6	$p>0.1$[a]

Notes: *Male* is a dummy taking value 1 for males and 0 for females; *Age* indicates subjects' age; *Higher education* equals 1 for those who have obtained at least a bachelor degree, and 0 otherwise; *CRT* ranges between 0 and 3 and is calculated by a three-item cognitive reflection test introduced by Frederick (2005); *IQ* ranges between 0 and 3 and is calculated by a three-item IQ test; *Economics*, *Statistics*, and *Game theory* are dummies taking value 1 for those who have taken at least one course in economics, statistics, or game theory, respectively; *Trust* equals 1 for those whose answer to the WVS on generalized trust is positive, and 0 otherwise; *Altruism* corresponds to our questionnaire-based measure of altruism; *Risk aversion* indicates subjects' answer to the risk attitude question; *RightAnswerBL*, *RightAnswerEXO*, and *RightAnswerVOTE* indicate the number of the correct answers to the control questions in the *Baseline, Exogenous*, and *Voting* treatment, respectively.

The symbols *, **, and *** indicate significance at the 10%, 5% and 1% level, respectively.
[a] Ordered logit regression, with standard errors robust for clustering at the subject's level.
[b] Logit regression, with standard errors robust for clustering at the subject's level.

Table 2.4 reports results from three logit regressions providing further support for this result. The dependent variable is a dummy taking value one for the subjects who voted in favor of collective punishment. Model 1, where the only explanatory variable is *CRT*, indicates that cognitive abilities measured by the Cognitive Reflection Test are not significantly correlated with subjects' voting behavior. When we measure

cognitive abilities based on the IQ questions in Model 2, instead, we find that the probability of voting for the punishment mechanism is 24.2% larger among subjects with higher cognitive abilities relative to other subjects. In Model 3 we include as regressors three dummy variables meant to capture the academic background of the subjects. Results indicate that subjects who have some prior knowledge of statistics are more likely to vote in favor of the punishment mechanism. These significant results still hold in Model 4 where we introduce additional controls for individual characteristics (listed in **Table 2.3**) and for subjects' choices in the *Baseline* and *Exogenous* treatments (listed in **Table 2.2**). These regressions suggest that only subjects who have higher cognitive abilities, or have a background in statistics, are able to fully anticipate the consequences of the introduction of collective punishment, hence its profitability.

Table 2.4: Voting behavior and individual characteristics

Dependent variable: Vote	Model 1	Model 2	Model 3	Model 4
CRT	0.076			-0.003
	(0.048)			(0.041)
IQ		0.242**		0.176*
		(0.102)		(0.091)
Economics			0.010	-0.076
			(0.060)	(0.092)
Statistics			0.188***	0.169*
			(0.078)	(0.086)
Game theory			-0.124	-0.123
			(0.067)	(0.089)
Trust-BL				0.021
				(0.086)
Trustworthiness-BL				0.069
				(0.111)
Trust-EX				0.334***
				(0.033)
Trustworthiness-EX				0.089
				(0.104)
Controls	No	No	No	Yes
N. Obs.	96	96	96	96

Notes: Marginal effects from logit regressions (standard errors robust for clustering at the session level are reported in parentheses). *CRT* ranges between 0 and 3 and is calculated by a three-item cognitive reflection test introduced by Frederick (2005); *IQ* ranges between 0 and 3 and is calculated by a three-item IQ test. *Economics*, *Statistics*, and *Game theory* are dummies taking value 1 for those who have taken at least one course in economics, statistics, or game theory, respectively; *Controls* indicates the presence of the remaining nine controls for individual characteristics (see **Table 2.3**).

The symbols *, **, and *** indicate significance at the 10%, 5% and 1% level, respectively.

2.4.3 Effects of the endogenous adoption or rejection of the punishment mechanism

The existing experimental literature on public good games has shown that there is a "dividend of democracy" in the sense that institutions endogenously chosen through voting can be more efficient than the same institutions being exogenously imposed on decision makers (Dal Bo et al., 2010; Sutter et al., 2010). One possible reason is that voting for the deterrent (or non-deterrent) institutions that punish uncooperative subjects credibly signals an intention to establish a high level of cooperation and thereby induces other group members to do the same. Consequently, the voting mechanism promotes coordination on the efficient, cooperative outcome (Markussen et al., 2014). In this part, we investigate whether "the dividend of democracy" can also be observed in our setting. Specifically, we investigate whether the punishment mechanism, when endogenously chosen, could significantly increase the levels of trust and trustworthiness relative to the case in which it is exogenously imposed.

Result 3. *When subjects vote for (not) introducing collective punishment, the levels of trust and trustworthiness are not significantly different from the case in which collective punishment is exogenously (not) introduced.*

In our study, only three groups endogenously adopt the collective punishment mechanism, while the other thirteen groups play the baseline trust game in the Voting treatment. Consider the behavior of subjects in the role of trustee first. When the majority of the group members vote against the implementation of collective punishment, the average level of trustworthiness does not change substantially, decreasing from 34.6% to 33.3%, relative to the *Baseline* treatment. Similarly, in groups where collective punishment is endogenously adopted trustworthiness levels decreased from 100% to 94.4%, relative to the *Exogenous* treatment. Neither difference is statistically significant ($p=0.318$ for the former comparison, and $p=0.225$ for the latter).

Similar results emerge if we focus on trusting behavior: when subjects vote for not introducing the punishment mechanism, compared to the *Baseline*, the level of trust drops from 25.6% to 23.1%, while the fraction of trustful behavior remains stable at 55.6% in groups where the collective punishment mechanism is determined by the majority voting mechanism. These two differences are also not statistically significant ($p=0.286$ for the former comparison, and $p=0.190$ for the latter).

While "the dividend of democracy" has been often observed in previous experimental papers, our study fails to find any positive effect of the voting mechanism on the society's ability to coordinate on an efficient outcome. Part of the reason is that the exogenously imposed punishment mechanism had induced a higher level of trust and trustworthiness among those who endogenously adopt it, hence there is little space for improvement. Another possible reason is that, when the mechanism is endogenously chosen, not all subjects positively react to it, but only those who voted in favor of it.

Indeed, we find that the three groups where the punishment mechanism was endogenously activated achieved a higher level of trustworthiness in the *Exogenous* treatment: the average level of trustworthiness is 100% in these three groups and 78.2% in the other thirteen groups, and the difference is significant ($p<0.001$). These three groups also exhibit higher levels of trust than other groups in the *Exogenous* treatment. The average level of trust is 55.6% in the three groups where the punishment mechanism is endogenously imposed and 35.9% in the other thirteen groups, and the difference is significant ($p=0.009$). In this paragraph, comparisons are based on logit regressions where the only explanatory variable is a dummy taking value one for subjects belonging to the three groups who adopted the collective punishment mechanism in the *Voting* treatment, and standard errors robust for clustering at the subject's level.

In addition, within these three groups, we could not find that all subjects positively react to the collectively determined punishment mechanism. Our results suggest that the endogenously chosen mechanism makes those who prefer its activation act more trustfully, while other subjects who vote against collective punishment seem to be immune to it. In fact, when collective punishment is endogenously chosen, those who vote in favor of it increase their trust level from 70% to 80% respect to the *Exogenous* treatment, while others reduce their trust level from 37.5% to 25%. Due to the limited sample, however, we cannot detect whether these differences are statistically significant.

2.4.4 Effects of information about others' behavior on voting

We now turn to the question of whether feedback about the aggregate behavior in the group, with and without collective punishment, could help subjects understand the effectiveness of the punishment mechanism, thereby changing their voting behavior. In the last 4 experimental sessions, we added a fourth game, where subjects received information on the aggregate behavior of their group members in the *Baseline* and *Exogenous* treatments before deciding whether to vote for or against collective punishment (see Section 3).

Result 4. *Even though exposed to feedback about others' past behavior, the large majority of subjects do not change their vote. Only information about others' trust levels in the Exogenous game positively affects a subjects' decision to vote in favor of the punishment mechanism.*

Among the 48 subjects who took part in these additional sessions, only 8 (i.e. 16.7%) changed their vote after observing the aggregate information about the first two treatments. Of them, five subjects voted in favor of collective punishment in the *Voting-IF* treatment, and three voted against it. A logit regression indicates that there is no difference in the voting behavior between in the *Voting* and *Voting-IF* treatments ($p=0.438$). Only two groups endogenously adopted the collective punishment mechanism in the last treatment. To explore subjects' voting behavior in more depth,

we run two logit regressions, whose results are reported in **Table 2.5**.

The dependent variable is a dummy taking value one when the subject voted in favor of collective punishment. *N. Trust-BL* (*N. Trustworthiness-BL*) indicates the number of the other group members who are trustful (trustworthy) in the *Baseline* treatment; *N. Trust-EX* (*N. Trustworthiness-EX*) indicates the number of the other group members who are trustful (trustworthy) in the *Exogenous* treatment; *Pro-punishment Vote* equals 1 if the subject voted in favor of the punishment mechanism in the third game. Model 1 shows that observing an additional trustful group member in the *Exogenous* game increased the probability of a subject voting for the punishment mechanism by 17.5%. It also highlights the high persistency of voting behavior: the probability that a subject votes in favor of the punishment mechanism in the *Voting-IF* treatment is 47.1% higher when s/he preferred to vote for the punishment mechanism rather than against it in the *Voting* treatment. In order to examine whether the subjects who voted in favor of the punishment mechanism in *Voting* are more sensitive to the feedback on others' trust levels in the *Exogenous*, we include the interaction term into Model 2, finding that the pro-punishment subjects are not better than the others at using the aggregate information. To sum up, these results imply that the additional information could not help subjects to understand the effectiveness of the punishment mechanism, regardless of their voting behavior in the third game.

Table 2.5: Voting behavior and feedback information.

Dependent variable: Vote	**Model 1**	**Model 2**
N. Trust-BL	-0.043	-0.043
	(0.032)	(0.032)
N. Trustworthiness-BL	0.013	0.013
	(0.081)	(0.081)
N. Trust-EX	0.175**	0.166***
	(0.072)	(0.070)
N. Trustworthiness-EX	0.079	0.079
	(0.076)	(0.076)
Pro-punishment Vote	0.471***	0.424***
	(0.105)	(0.095)
N. Trust-EX × Pro-punishment Vote		0.034
		(0.028)
N. Obs.	48	48

Notes: Marginal effects from logit regressions (standard errors robust for clustering at the session level are reported in parentheses).
The symbols *, **, and *** indicate significance at the 10%, 5% and 1% level, respectively.

2.5 Discussion and Conclusions

In this paper, we explore whether the endogenous adoption of a collective punishment mechanism can help a society coordinate on an efficient outcome, characterized by high levels of trust and trustworthiness. We first introduce a theoretical analysis of the consequences of the introduction of a collective punishment mechanism, which largely builds upon Anderlini and Terlizzese's (2013) work. We then design and run an experiment to empirically test the theoretical predictions we previously derived.

We find that subjects exhibit significantly higher levels of trust and trustworthiness when a collective punishment mechanism is imposed exogenously. In contrast with the previous studies on the "dividend of democracy", however, we fail to observe that the punishment mechanism induces higher level of cooperation when it is democratically chosen compared to the case in which it is exogenously activated. One potential explanation is that in most previous studies based on the public game the subjects could directly inflict punishment on low contributors to enforce the endogenously determined rule, or the punishment was fixed and determined ex-ante by the experimenter. By contrast, in our trust game, even when the social sanction rule is democratically introduced, the severity of punishment depends on the average behavior in society, which makes it more unpredictable from the subjects' perspective; hence a higher cognitive effort is necessary to anticipate how others will react to the rule, and to predict its overall effects on profits and welfare. Further experimental studies are needed to more precisely pin down the mechanisms driving these differences in results.

Another important finding is that a majority of subjects vote against the collective punishment mechanism, even though from an ex post perspective it would have paid off, on average, to vote in favor of it. Previous experimental studies have shown that subjects are reluctant to choose a punishment institution when facing alternative options. In Sutter et al. (2010), subjects are allowed to vote for a voluntary contribution mechanism (VCM), an institution with reward possibility and an institution with punishment possibility. The authors report that under unanimous voting, the punishment option is rarely selected. A similar behavior pattern is also observed in Botelho et al. (2007). After having experienced both the VCM and the VCM with the punishment option, subjects decide to choose the governing institution for the final period. Botelho et al. (2007) find that in their experiment 77.8% of subjects vote against the punishment institution. One possible reason is that subjects may naturally dislike the punishment since it evokes negative feelings. To test whether opting against the sanction is mainly driven by a "natural aversion" to punishment, in future research we plan to run a follow-up experiment where we reframe the game without changing the incentives, and substitute penalties with rewards. Another potential explanation is that cognitive limitations may refrain subjects from anticipating the positive effect of the introduction of collective punishment. Putterman et al. (2011) find that intelligence predicts subjects' votes on

efficient schemes when they are permitted to vote over a menu of sanction rules. Our study also confirms that subjects with high cognitive abilities are more likely to anticipate the effectiveness of collective punishment and therefore vote in favor of it.

In an additional treatment, we investigate whether the information about the others' aggregate behavior with and without collective punishment affects subjects' voting choices, finding that subjects hardly change their votes respect to the no-feedback condition. In Gurerk (2013), before a voting phase in which they choose among alternative institutions governing the public good provision, subjects are provided with the complete history of a punishment institution which was actually implemented in a previous experiment. The author finds that social information significantly induces more subjects to accept the punishment option and reach full contributions more quickly over time. Our study fails to replicate the positive effect of social information, a result which is in line with some previous studies, showing that a high percentage of subjects are reluctant to select a relatively efficient mechanism even when they are exposed to the complete information on subjects' behavior under the alternative institutional regimes (Dal Bo et al., 2010; Gurerk, et al., 2006; Hilbe, et al., 2014). One possible reason is that subjects may need repetition to fully understand the change in incentives introduced by the collective punishment mechanism, and its effects on others' behavior; we see this as an interesting route for future research. Another possible way of promoting the endogenous adoption of an efficiency-enhancing institution is group communication. Alm et al. (1999) investigate the effect of voting on a social norm of tax compliance by letting subjects vote via majority rule on different aspects of the fiscal system. They find that, without communication, subjects vote against an increase in the levels of sanction enforcement imposed on tax evaders. However, when subjects are allowed to communicate before voting, they are more likely to select a greater level of enforcement, achieving an overall increase in efficiency. Along these lines, we could also expand our set-up and examine the question of whether group communication before the voting phase facilitates the acceptance of the collective punishment institution. All this, however, is left for future research.

Acknowledgement

We thank Stefania Bortolotti, Marco Casari, Davide Dragone, Diego Gambetta and Paolo Vanin for insightful comments. This paper also benefited from comments received by seminar participants at the University of Bologna, the European University Institute, and the European School on New Institutional Economics. The usual disclaimer applies. We gratefully acknowledge financial support from the Law and Economics Research Center of Zhejiang University (RG201310004), and from the Italian Ministry of Education (grant FIRB-Futuro in Ricerca no. RBFR084L83).

References

Algan, Y., and Cahuc, P., 2013, "Trust and Human Development: Overview and Policy", *Handbook of Economic Growt*h, ed. by Philippe Aghion and Steven Durlauf.

Alm, J., McClelland, G., and Schulze, W., 1999, "Changing the Social Norm of Tax Compliance by Voting." *Kyklos*, 52, 141-171.

Anderlini, L. and Terlizzese, D., 2013, "Equilibrium Trust", Georgetown University, mimeo.

Becker, G. S., 1968, "Crime and Punishment: An Economic Approach", *Journal of Political Economy*, 76, 169-217.

Bernstein, L., 1992, "Opting out of the Legal System: Extralegal Contractual Relations in the Diamond Industry", *Journal of Legal Studies*, 21, 115–157.

Bernstein, L., 2001, "Private Commercial Law in the Cotton Industry: Creating Cooperation through Rules, Norms, and Institutions", *Michigan Law Review*, 99, 1724–90.

Besley, T., and Coate, A., 1995, "Group Lending, Repayment Incentives and Social Collateral." *Journal of Development Economics*, 46(1), 1-18.

Bicchieri, C., 2006, *The Grammar of Society: The Nature and Dynamics of Social Norms*, Cambridge University Press.

Blanco, M., Engelmann, D., and Normann, H. T., 2011, "A Within-subject Analysis of Other-regarding Preferences." *Games and Economic Behavior*, 72, 321-338.

Bohnet, I., and Cooter, R., 2003, "Expressive Law: Framing or Equilibrium Selection?" KSG Working Paper No. RWP03-046; and UC Berkeley Public Law Research Paper No. 138. http://ssrn.com/abstract=452420.

Bohnet, I., Grieg, F., Herrmann, B., and Zeckhauser, R., 2008, "Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States." *American Economic Review*, 98(1), 294– 310.

Botelho, A., Harrison, G., Costa Pinto, L., and Rutstrom, E., 2007, "Social Norms and Social Choice." Working paper 05-23, Department of Economics, College of Business Administration, University of Central Florida.

Bowles, S., and Polania-Reyes, S., 2012, "Economic Incentives and Social Preferences: Substitutes or Complements?" *Journal of Economic Literature*, 50, 368-425.

Brandts, J., and Charness, G., 2011, "The Strategy versus the Direct-Response Method: A First Survey of Experimental Comparisons", *Experimental Economics*, 21, 1-24.

Buchan, N. R., Croson, R. T., & Dawes, R. M., 2002, "Swift Neighbors and Persistent Strangers: A Cross-Cultural Investigation of Trust and Reciprocity in Social Exchange." *American Journal of Sociology*, 108(1), 168-206.

Charness, G., Cobo-Reyes, R., and Jimenez, N., 2008, "An Investment Game with Third-Party Intervention." *Journal of Economic Behavior & Organization*, 68, 18-28.

Cialdini, R., Reno, R., and Kallgren, 1990, "A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places." *Journal of*

Personality and Social Psychology, 58, 1015-1026.
Cialdini, R. B., Demaine, L. J., Sagarin, B. J., Barrett, D. W., Rhoads, K., & Winter, P. L., 2006, "Managing Social Norms for Persuasive Impact." *Social Influence*, 1, 3-15.
Clay, K., 1997, "Trade without Law: Private-order Institutions in Mexican California", *Journal of Law, Economics, and Organization*, 13, 202–231.
Clay, K., and Wright, G., 2005, "Order without Law? Property Rights during the California Gold Rush", *Explorations in Economic History*, 42, 155-183.
Cooter, R., 1998, "Expressive Law and Economics", *Journal of Legal Studies*, 27, 585–608.
Dal Bo, P., Foster, A., and Putterman, L., 2010, "Institutions and Behavior: Experimental Evidence on the Effects of Democracy", *American Economic Review*, 100, 2205-2229.
de Quervain, D., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., and Fehr, E., 2004, "The Neural Basis of Altruistic Punishment." *Science*, 305, 1254-1258.
Ellickson, R., 1986, "Of Coase and Cattle: Dispute Resolution among Neighbors in Shasta County." *Stanford Law Review*, 38, 623-687.
Ellickson, R., 1991, *Order Without Law: How Neighbors Settle Disputes*, Cambridge, Harvard University Press.
Elster, J., 1989, "Social Norms and Economic Theory." *Journal of Economic Perspectives*, 3(4), 99-117.
Ertan, A., Page, T., and Putterman, L., 2009, "Who to Punish? Individual Decisions and Majority Rule in Mitigate the Free Rider Problem." *European Economic Review*, 53, 495-511.
Fehr, E., and Williams, T., 2013, "Endogenous Emergence of Institutions to Sustain Cooperation." Working Paper, University of Zurich.
https://sites.google.com/site/tonywilliamsresearch/Tony_Williams_JOB_MARKET_PAPER.pdf
Fehr, E., and Rockenbach, B., 2003, "Detrimental Effects of Sanctions on Human Altruism." *Nature*, 422, 137-140.
Fischbacher, U., 2007, "z-Tree: Zurich Toolbox for Ready-made Economic Experiments." *Experimental Economics*, 10, 171-178.
Frederick, S., 2005, "Cognitive Reflection and Decision Making." *Journal of Economic Perspective*, 19(4), 25-42.
Galbiati, R., and Vertova, P., 2008, "Obligations and Cooperative Behavior in Public Good Games." *Games and Economic Behavior*, 146-170.
Greif, A., 1989, "Reputation and Coalitions in Medieval Trade: Evidence on the Maghribi Traders", *Journal of Economic History*, XLIX, 857–82.
Greif, A., 1993, "Contract Enforceability and Economic Institutions in Early Trade: the Maghribi Traders' Coalition", *American Economic Review*, 83, 525–48.
Greif, A., 2006, *Institutions and the Path to the Modern Economy: Lessons from Medieval Trade*, Cambridge University Press.
Greiner, B., 2015, "Subject Pool Recruitment Procedures: Organizing Experiments

with ORSEE." Journal of the Economic Science Association 1 (1), 114-125.
Guiso, L., Sapienza, P., and Zingales, L., 2008, "Social Capital as Good Culture." *Journal of the European Economic Association*, 6(2-3), 295-320.
Gurerk, O., 2013, "Social Learning Increases the Acceptance and the Efficiency of Punishment Institutions in Social Dilemmas." *Journal of Economic Psychology*, 34, 229-239.
Gurerk, O., Irlenbusch, B. and Rockenbach, B., 2006, "The Competitive Advantage of Sanctioning Institutions." *Science*, 312, 108–111.
Gurerk, O., Irlenbusch, B. and Rockenbach, B., 2014, "On Cooperation in Open Communities." *Journal of Public Economics*, 120, 220-230.
Hechter, M., and Opp, K., 2001, *Social Norms*. New York: Russell Sage Foundation.
Hilbe, C., Traulsen, A., Rohl, T., and Milinshi, M., 2014, "Democratic Decisions Establish Stable Authorities That Overcome the Paradox of Second-order Punishment." *Proceedings of the National Academy of Sciences*, 111, 752-756.
Huang, P., and Wu, H., 1994, "More Order without More Law: A Theory of Social Norms and Organizational Cultures." *Journal of Law, Economics, and Organization*, 10(2), 390-406.
Kadens, E., and Young, E., 2013, "How Customary Is Customary International Law?" *William & Mary Law Review*, 54, 885-920.
Kamei, K., 2014, "Democracy and Resilient Pro-Social Behavioral Change: An Experimental Study", Available at SSRN: http://dx.doi.org/10.2139/ssrn.1756225.
Kimbrough, E., and Rubin, J., 2015, "Sustaining Group Reputation." *Journal of Law, Economics, & Organization*, 31, 599-628.
Kosfeld, M., Okada, A., and Riedl, A., 2009, "Institution Formation in Public Good Games." *American Economic Review*, 1335-1355.
Krupka, E., and Weber, R., 2013, "Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary?" *Journal of the European Economic Association*, 11(3), 495-524.
Leeson, P., and Williamson, C., 2009, "Anarchy and Development: An Application of the Theory of Second Best." *Law and Development Review*, 2(1), 76-96.
Markussen, T., Putterman, L., and Tyran, J., 2014, "Self-Organization for Collective Action: An Experimental Study of Voting on Sanction Regimes." *Review of Economic Studies*, 81, 301-324.
McAdams, R., 2000a, "An Attitudinal Theory of Expressive Law." *Oregon Law Review*, 79, 339–390.
McAdams, R., 2000b, "A Focal Point Theory of Expressive Law." *Virginia Law Review*, 86, 1649–1731.
McAdams, R., and Nadler, J., 2005, "Testing the Focal Point Theory of Legal Compliance: The Effect of Third-Party Expression in an Experimental Hawk/Dove Game." *Journal of Empirical Legal Studies*, 2(1), 87-123.
Ostrom, E., 1990, *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge University Press, New York.
Polinsky, A. M. and Shavell, S., 2000, "The Economic Theory of Public

Enforcement of Law." *Journal of Economic Literature*, 38, 45-76.
Polinsky, A. M. and Shavell, S., 2008, "Economic Analysis of Law", *The New Palgrave Dictionary of Economics*, ed. by Steven N. Durlauf and Lawrence E. Blume.
Posner, E., 1998, "Symbols, Signals, and Social Norms in Politics and the Law." *Journal of Legal Studies*, 27, 765–789.
Posner, E., 2000, *Law and Social Norms*, Harvard University Press, Cambridge, MA.
Posner, R., 1997, "Social Norms and the Law: An Economic Approach." *American Economic Review*, 87(2), 365-369.
Posner, R., and Rasmusen, E., 1999, "Creating and Enforcing Norms, with Special Reference to Sanctions." *International Review of Law and Economics*, 19, 369-382.
Putterman, L., Tyran, J., and Kamei, K., 2011, "Public Goods and Voting on Formal Sanction Schemes." *Journal of Public Economics*, 95, 1213-1222.
Schultz, P. W., Nolan, J. M., Cialdini, R., B., Goldstein, N. J., and Griskevicius, V., 2007, "The Constructive, Destructive, and Reconstructive Power of Social Norms." *Psychological Science*, 18(5), 429-434.
Sutter, M., Haigner, S., and Kocher, M., 2010, "Choosing the Carrot or the Stick? Endogenous Institutional Choice in Social Dilemma Situations." *Review of Economic Studies*, 77, 1540-1566.
Tabellini, G., 2008, "The Scope of Cooperation: Norms and Incentives." *Quarterly Journal of Economics*, 123(3), 905-950.
Tyler, T., 1990, *Why People Obey Law*, Yale University Press.
Tyran, J., and Feld, L., 2006, "Achieving Compliance when Legal Sanctions are Non-deterrent." *Scandinavian Journal of Economics*, 108(1), 135-156.
Vollan, B., 2011, "The Difference between Kinship and Friendship: (Field-) Experimental Evidence on Trust and Punishment." *The Journal of Socio-Economics*, 40, 14-25.
Vollan, B., Zhou, Y., Landmann, A., Hu, B., Herrmann-Pillath, C., 2013, "Cooperation and Authoritarian Norms: An Experimental Study in China." Working Papers in Economics and Statistics, University of Innsbruck.
Xiao, E., 2013, "Profit –Seeking Punishment Corrupts Norm Obedience." *Games and Economic Behavior*, 77, 321-344.

Appendix 1

Table A: Treatments and sessions

Session type	**Baseline-first**	**Exogenous-first**	**Baseline-first +Information**	**Exogenous first + Information**
Order	BL-EX-VT	EX-BL-VT	BL-EX-VT-VF	EX-BL-VT-VF
Session dates	Dec. 03, 2013; Dec. 10, 2013	Dec.12, 2013	March 20, 2014 March 24, 2014	March 20, 2014 March 24, 2014
N. Subjects	24	24	24	24
N. Independent observations	24	24	24	24

Notes: In the table, BL stands for *Baseline*, EX for *Exogenous*, VT for *Voting*, and VF for *Voting-IF*.

Table B: Order effects on trustworthiness

	Trustworthiness (%)		
	1st Game		**2nd Game**
BL-EXO-VOTE	43.8%	$<^{***}$	77.1%
	$\wedge^{***}$		$\vee^{***}$
EXO-BL-VOTE	77.1%	$>^{***}$	33.3%

Notes: BL stands for *Baseline* treatment, EXO for *Exogenous* treatment, VOTE for *Voting* treatment. *** indicates the significance at 1% level based on a two-tailed z-test.

Table C: Order effects on trust

	Trust (%)		
	1st Game		**2nd Game**
BL-EXO-VOTE	41.7%	~	43.8%
	~		$\vee^{***}$
EXO-BL-VOTE	35.4%	$>^{***}$	14.6%

Notes: BL stands for *Baseline* treatment, EXO for *Exogenous* treatment, VOTE for *Voting* treatment. *** indicates the significance at 1% level based on a two-tailed z-test.

Appendix 2: Experimental instructions (Baseline first + Information)

Welcome. This is a study on how people make decisions. In this study you can earn money based on how well you follow the instructions, and on the decisions made by you and by the other participants. You will be paid in private and in cash at the end of the session.
Please turn off your mobile phone. From this moment on, no form of communication among participants is allowed. If you have any question, or need assistance of any kind, please raise your hand and one of us will come to your desk to help you.
Please, follow the instructions carefully. In this study there are four parts, and for each part, we will distribute and read the corresponding instructions. In the first three parts, after having read the instructions, we will ask you to answer six questions, to verify your full understanding. For every question you answer correctly you earn €0.40. So you can earn up to €7.2 by answering correctly to all questions for Parts 1, 2 and 3 of the study. In addition you will earn money for the decisions you and the other participants will make in Parts 1, 2, 3 and 4 of the study.

Now, I will read instruction for Part 1.

Instructions for Part 1

In this part of the study, participants are randomly divided into **groups of six**. In each group, three participants will be assigned the role BLUE, while the other three will be RED, then the computer will form pairs of subjects belonging to the same group. If you are BLUE, you will be paired with a RED player, and vice versa. Your counterpart will never know your true identity, nor will you know hers/his.

Your earnings are expressed in tokens that will be converted in Euros at the rate of 1 Euro for 3 tokens.

BLUE has to make one choice: between option A and option B. RED has to make one choice: between option X and option Y. Table 1 summarizes the earnings corresponding to BLUE's and RED's choices.

Table 1: earnings in Part 1

BLUE chooses	RED chooses	Earnings
A	X	BLUE: 0 RED: 30
	Y	BLUE: 15 RED: 15
B	Irrelevant	BLUE: 8 RED: 8

If BLUE chooses option A, earnings depend on the choice made by RED:

- if RED chooses X, BLUE earns **0** tokens and RED earns **30** tokens;
- if RED chooses Y, BLUE earns **15** tokens and RED earns **15** tokens.

If BLUE chooses option B, the choice made by RED has no consequences on either BLUE's or RED's earnings:

- BLUE earns **8** tokens and RED earns **8** tokens.

We ask you to make a decision first as RED, then as BLUE. **We will inform you of the role you are actually assigned in this Part only at the end of the session.**

If you are assigned the BLUE role, your earnings from this part will depend on the choice you made as BLUE, and on the choice made by your counterpart as RED.

If you are assigned the RED role, your earnings from this part will depend on the choice you made as RED, and on the choice made by your counterpart as BLUE.

You will be informed of the results of this Part only at the end of the session.

We will now make an **example**. At the end of the example we will ask you to answer two questions, to verify you understanding of the instructions. Remember that you earn €0.40 for each question you answer correctly.

Look at your screen. You now have to make a choice as RED. Please, choose X, and confirm your choice. Good. You now have to make a choice as BLUE. Please, choose B and confirm your choice. Good. On your screen, you will now see two questions. Please, give your answers by pressing the corresponding buttons.

If you are not sure about the answer, you can re-read the instructions. Take your time and think carefully before answering the question.

[As RED, you chose X and as BLUE you chose B. You are assigned the BLUE role, and your counterpart, who is assigned the RED role, chose Y.

- *How much do you earn?*
- *How much does your counterpart earn?]*

We will now make another **example**. At the end of the example we will ask you to answer two questions, to verify you understanding of the instructions. Remember that you earn €0.40 for each question you answer correctly.

Look at your screen. You now have to make a choice as RED. Please, choose X, and confirm your choice. Good. You now have to make a choice as BLUE. Please, choose A and confirm your choice. Good. On your screen, you will now see two questions. Please, give your answers by pressing the corresponding buttons.

[As RED, you chose X and as BLUE you chose A. You are assigned the RED role, and your counterpart, who is assigned the BLUE role, chose A.

- *How much do you earn?*
- *How much does your counterpart earn?]*

You will now read on your screen the last two questions. Please, give your answers by pressing the corresponding buttons.

- *How much are 6 tokens worth, in Euros?*
- *Will you know if you are RED or BLUE before making your choice?*

If you have any doubts on the instructions, please raise your hand now. Good, then we can start with Part 1.

Instructions for Part 2

In this part of the study, participants are in **the same groups of six as in Part 1**. In each group, three participants will be assigned the role BLUE, while the other three will be RED, then the computer will form pairs of subjects belonging to the same group. If you are BLUE, you will be paired with a RED player, and vice versa. Your counterpart will never know your true identity, nor will you know hers/his. Your counterpart will **NOT** be the same person as in Part 1.

Your earnings are expressed in tokens that will be converted in Euros at the rate of 1 Euro for 3 tokens. You may also lose tokens. In the unlikely event your total earnings at the end of the study are negative, you may lose part of the money you earned by correctly answering the questions on the instructions. In any case, we guarantee you a minimum earning of €5 for your participation.

BLUE has to make one choice: between option A and option B. RED has to make one choice: between option X and option Y. Table 2 summarizes the earnings corresponding to BLUE's and RED's choices. *Earnings for RED may depend on the choices made by the other five members of the group.*

Table 2: earnings in Part 2

BLUE chooses	RED chooses	Earnings
A	X	BLUE: 0 RED: 30 – 8 x *number of others who choose Y*
	Y	BLUE: 15 RED: 15
B	Irrelevant	BLUE: 8 RED: 8

If BLUE chooses option A, earnings depend on the choice made by RED:

- if RED chooses X, BLUE earns **0** tokens. Earnings for RED depend on the choices made as RED by the other five members of the group. Notice that all members of your group make decisions both as RED and as BLUE, before knowing the role they are actually assigned.
    - If 0 of the others chooses Y, RED will get 30 tokens.
    - If 1 of others chooses Y, RED will get 22 tokens.
    - If 2 of others choose Y, RED will get 14 tokens.
    - If 3 of others choose Y, RED will get 6 tokens.
    - If 4 of others choose Y, RED will lose 2 tokens.
    - If 5 of others choose Y, RED will lose 10 tokens.
- if RED chooses Y, BLUE earns **15** tokens and RED earns **15** tokens.

If BLUE chooses option B, the choice made by RED has no consequences on either BLUE's or RED's earnings:

- BLUE earns **8** tokens and RED earns **8** tokens.

We ask you to make a decision first as RED, then as BLUE. **We will inform you of the role you are actually assigned in this Part only at the end of the session.**

If you are assigned the BLUE role, your earnings from this part will depend on the choice you made as BLUE, and on the choice made by your counterpart as RED.

If you are assigned the RED role, your earnings from this part will depend on the choice you made as RED, on the choice made by your counterpart as BLUE, *and on the choices made as RED by each of the other five members of your group*.

You will be informed of the results of this Part only at the end of the session.

We will now make an **example**. At the end of the example we will ask you to answer two questions, to verify you understanding of the instructions. Remember that you earn €0.40 for each question you answer correctly.

Look at your screen. You now have to make a choice as RED. Please, choose Y, and confirm your choice. Good. You now have to make a choice as BLUE. Please, choose B and confirm your choice. Good. On your screen, you will now see two questions. Please, give your answers by pressing the corresponding buttons.

If you are not sure about the answer, you can re-read the instructions. Take your time and think carefully before answering the question.

[As RED, you chose Y and as BLUE you chose B. You are assigned the BLUE role, and your counterpart, who is assigned the RED role, chose X. Two of the other members of your group chose Y as RED.

- *How much do you earn?*
- *How much does your counterpart earn?]*

We will now make another **example**. At the end of the example we will ask you to answer two questions, to verify you understanding of the instructions. Remember that you earn €0.40 for each question you answer correctly.

Look at your screen. You now have to make a choice as RED. Please, choose X, and confirm your choice. Good. You now have to make a choice as BLUE. Please, choose A and confirm your choice. Good. On your screen, you will now see two questions. Please, give your answers by pressing the corresponding buttons.

[As RED, you chose X and as BLUE you chose A. You are assigned the RED role, and your counterpart, who is assigned the BLUE role, chose A. Four of the other members of your group chose Y as RED.

- *How much do you earn?*
- *How much does your counterpart earn?]*

You will now read on your screen the last two questions. Please, give your answers by pressing the corresponding buttons.

- *Can your counterpart in Part 2 be the same person as in Part 1?*
- *How many people are there in each group?*

If you have any doubts on the instructions, please raise your hand now. Good, then we can start with Part 2.

Instructions for Part 3

In this part of the study, participants are in **the same groups of six as in Parts 1 and 2**. In each group, three participants will be assigned the role BLUE, while the other three will be RED, then the computer will form pairs of subjects belonging to the same group. If you are BLUE, you will be paired with a RED player, and vice versa. Your counterpart will never know your true identity, nor will you know hers/his. Your counterpart will **NOT** be the same person as in Part 1 or in Part 2.

In Part 3, you will be asked to take 3 decisions. First you will have vote in favor of either Situation 1, or Situation 2. Then you will have to make a choice as RED and as BLUE, as in Parts 1 and 2.

Situation 1 is the situation you faced in Part 1 of this study, represented in Table 3.

Table 3: Situation1

BLUE chooses	RED chooses	Earnings
A	X	BLUE: 0 RED: 30
	Y	BLUE: 15 RED: 15
B	Irrelevant	BLUE: 8 RED: 8

Situation 2 is the situation you faced in Part 2 of this study, represented in Table 4.

Table 4: Situation 2

BLUE chooses	RED chooses	Earnings
A	X	BLUE: 0 RED: 30 – 8 x *number of others who choose Y*
	Y	BLUE: 15 RED: 15
B	Irrelevant	BLUE: 8 RED: 8

When all participants have casted their vote, you will be informed of how many of your group's members voted for Situation 1, of how many of your group's members voted for Situation 2, and of the outcome of the vote.

If the majority of the members of your group vote for Situation 1, then the rules for the rest of this Part will be the same as in Part 1. If instead the majority of the members in your group vote for Situation 2, then the rules for the rest of this Part will be the same as in Part 2. If in your group three members vote in favor of Situation 1, and three members vote in favor of Situation 2, then the outcome will be randomly determined by the computer.

We ask you to make a decision first as RED, then as BLUE. **We will inform you of the role you are actually assigned only at the end of the session.**

If you are assigned the BLUE role, your earnings from this part will depend on the choice you made as BLUE, and on the choice made by your counterpart as RED.

If you are assigned the RED role, your earnings from this part will depend on the choice you made as RED, and on the choice made by your counterpart as BLUE. *In case in your group the outcome of the vote is Situation 2, earnings for RED may also depend on the choices made as RED by each of the other five members of your group*.

You will be informed of the results of this Part only at the end of the session.

We will now make an **example**. At the end of the example we will ask you to answer two questions, to verify you understanding of the instructions. Remember that you earn €0.40 for each question you answer correctly.

Look at your screen. You now have to vote either for Situation 1 or for Situation 2. Please, vote for Situation 2, and confirm your choice.

You can now see on your screen that the majority of your group members voted for Situation 1. Hence, the rules for the rest of this Part will be the same as in Part 1.

You now have to make a choice as RED. Please, choose Y, and confirm your choice. Good. You now have to make a choice as BLUE. Please, choose B and confirm your choice. Good. On your screen, you will now see two questions. Please, give your answers by pressing the corresponding buttons.

If you are not sure about the answer, you can re-read the instructions. Take your time and think carefully before answering the question.

[Situation 1 has been selected. As RED, you chose Y and as BLUE you chose B. You are assigned the BLUE role, and your counterpart, who is assigned the RED role, chose X. Four of the other members of your group chose Y as RED.

- *How much do you earn?*
- *How much does your counterpart earn?]*

We will now make another **example**. At the end of the example we will ask you to answer two questions, to verify you understanding of the instructions. Remember that you earn €0.40 for each question you answer correctly.

Look at your screen. You now have to vote either for Situation 1 or for Situation 2. Please, vote for Situation 1, and confirm your choice.

You can now see on your screen that the majority of your group members voted for Situation 2. Hence, the rules for the rest of this Part will be the same as in Part 2.

You now have to make a choice as RED. Please, choose X, and confirm your choice. Good. You now have to make a choice as BLUE. Please, choose A and confirm your choice. Good. On your screen, you will now see two questions. Please, give your answers by pressing the corresponding buttons.

[Situation 2 has been selected. As RED, you chose X and as BLUE you chose A. You are assigned the RED role, and your counterpart, who is assigned the BLUE role, chose A. Two of the other members of your group chose Y as RED.

- *How much do you earn?*
- *How much does your counterpart earn?]*

You will now read on your screen the last two questions. Please, give your answers by pressing the corresponding buttons.

- *Can your counterpart in Part 3 be the same person as in Part 1 or Part 2?*
- *If four members of your group vote for Situation 1 and two members of your group vote for Situation 2, in Part 3 your group will play according to the rules adopted in Part 1 of the study. True or False?*

If you have any doubts on the instructions, please raise your hand now. Good, then we can start with Part 3.

Instructions for Part 4

In this part of the study, participants are in **the same groups of six as in Parts 1, 2 and 3**. In each group, three participants will be assigned the role BLUE, while the other three will be RED, then the computer will form pairs of subjects belonging to the same group. If you are BLUE, you will be paired with a RED player, and vice versa. Your counterpart will never know your true identity, nor will you know hers/his. Your counterpart **may** be the same person as in **Part 1, Part 2 or in Part 3**.
Rules for Part 4 are the same as for Part 3: you will be asked to take 3 decisions. First you will have vote in favor of either Situation 1, or Situation 2. Then you will have to make a choice as RED and as BLUE, as in Parts 1, 2 and 3. **Differently from Part 3**, in Part 4, before making your decisions, you will receive **information on the choices** that you and your group members made in **Parts 1, and 2**.
At the end of this Part, you will receive information on the outcome of Parts 1, 2 3 and 4 of the study. You will know the role you have been assigned in each Part, and the earnings you obtained.

Appendix 3

Questionnaire

We kindly ask you to complete this questionnaire. The answers you give will not affect in any way your earnings. Some of these questions refer to personal information, which will help us in this study. Your identity will not be revealed under any circumstances in the presentation of the results.

Please answer carefully. Once an answer is given, you can no longer change it.

Press OK to begin. Thank you.

1. Were the instructions you have received for today's activities clear?

(1) No, not at all (2) No, not so much (3) Yes, enough (4) Yes, very much

2. Gender (press the corresponding button)

(1) Male (2) Female

3. Age (please, give your answer using the slider below and press ok to confirm)

4. Were you born in Italy?

(1) Yes (2) No

5. Education background

(1) Middle high school (2) High school (3) Bachelor degree
(4) Master degree (5) Ph.D. or postgraduate degree (6) Other

6. Occupation

(1) Student (2) Self-employed worker (3) Employee (4) Retired
(5) Jobless (6) Others

6.1 Field of studies (this question is accessed only if the subject gives answer (1) to question 6)

(1) Social sciences (2) Mathematical, Physical and Natural sciences
(3) Engineering and Architecture (4) Medicine
(5) Literature and Philosophy (6) Others

7. Have you attended courses in Economics?

(1) Yes (2) No

8. Have you attended courses in Statistics?

(1) Yes (2) No

9. Have you attended courses in Game Theory?

(1) Yes (2) No

10. Have you previously participated as a volunteer in other researches?
(choose one or more answers)

(1) Yes, in the field of economics
(2) Yes, in the field of psychology
(3) Yes, in the field of medicine or biology
(4) No

11. Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?

(1) Most people can be trusted(2) Can't be too careful (3) No idea

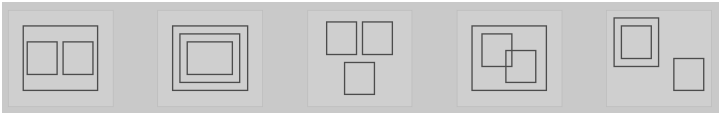
12. Are you generally a person who is fully prepared to take risks or do you try to avoid taking risk?
Please tick a box on the scale, where the value 1 means: "unwilling to take risks" and the value 10 means: "fully prepared to take risk"

1, 2, 3, 4, 5, 6, 7, 8, 9, 10

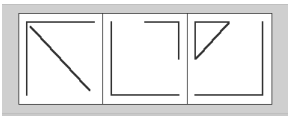
13. In general, do you think it is important to help others, and take care of their well being?
Please tick a box on the scale, where the value 1 means: "not important at all" and the value 10 means: "Maximally important"

1, 2, 3, 4, 5, 6, 7, 8, 9, 10

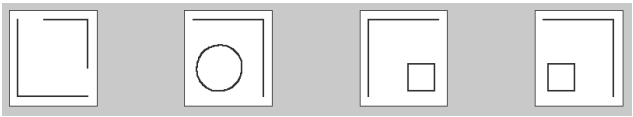
14. Which of these diagrams represents the relationship between Orange-Citrus Fruit-Fruit? Please select an answer and click OK to confirm.



15. Select the element that completes the following series.



Please select an answer and click OK to confirm.



16. A bat and a ball cost $ 1.10 in total. The bat costs $ 1.00 more than the ball. How much does the ball cost?

17. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?

18. In a pond, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire pond, how long would it take for the patch to cover half of the pond?

Chapter 3

Who Are More Naïve? High or Low Trustors

Huojun Sun, Yefeng Chen

Abstract. Whether trustful people are more or less sensitive than skeptical ones to cues on others' trustworthiness is an open question, which only recently attracted scholar's attention. We investigate this issue by means of a trust game experiment in which subjects repeatedly face opponents belonging to a high- or a low-trustworthiness group. We find that high and low trustors are equally able to distinguish which group is more trustworthy, and to condition their trust accordingly. However, when subjects can choose whether or not to receive information on the outcome of their own past interactions, high trustors learn whom to trust or distrust faster. Our interpretation is that they are less susceptible to the anticipated aversive emotions aroused by the potential betrayal and thereby are more keen to acquire useful information about their partner's behavior.

Key words: False Consensus Effect, Predict Trustworthiness, Betrayal Aversion

We are grateful for valuable comments and suggestions by Maria Bigoni, Marco Casari, and participants of the Economic Science Association World Meeting 2013 at the University of Zurich, the 26th Annual Conference of the International Association for Conflict Management in Tacoma 2013, the 9th International Meeting on Experimental and Behavioral Economics in Madrid 2013, the 2013 Florence Workshop on Behavioral and Experimental Economics, and 2013 ECORE Summer School on Governance and Economic Behavior.

3.1 Introduction

While trust in general is crucial to a country's economic success, high-trustors – or individuals with a high level of generalized trust – are often seen as gullible and naïve Pollyannas. People seem to believe that high-trustors tend to overestimate their partner's trustworthiness based on incomplete information and have a bias in processing the trustworthiness-related information. Using data from the European Social Survey, Butler et al. (2012) find that people tend to be subject to "the false consensus effect",[41] and to form beliefs on others' trustworthiness based on their own. In particular, highly trustworthy individuals think others are like them and easily tend to form beliefs that are too optimistic, which induces them to trust more than they should and thereby be often cheated. A further experimental study (Butler et al., forthcoming) confirms this evidence, suggesting that subjects playing repeated trust games (Berg et al., 1995) tend to extrapolate their opponent's preferences from their own and that subjects keep holding their initial trust beliefs even after several rounds of play, implying that high-trustors who hold overly-optimistic default expectation of others' trustworthiness fail to calibrate it through learning.

Contrary to Butler et al.'s findings, a series of experiments implemented by Yamagishi and his co-authors show that, instead of being gullible, high-trustors are more sensitive to information that potentially reveals others' trustworthiness or untrustworthiness (Kosugi and Yamagishi, 1998; Kakiuchi and Yamagishi, 1997; Yamagishi and Kakiuchi, 2000) and that high-trustors predict others' trustworthiness more accurately than low-trustors (Kikuchi, Watanabe, and Yamagishi, 1997). In agreement with Yamagishi's experimental evidence, Sturgis et al. (2010) find that standard measures of intelligence at age 10-11 can explain variability in generalized trust in early middle age, even after controlling for a large number of socio-economic variables, based on data from two British birth cohort studies. This research is replicated by at least three empirical studies using different sources of data. In particular, Carl and Billari (2014) find a strong association between generalized trust and intelligence in the U.S., Hooghe et al. (2012) find one in the Netherlands, and Oskarsson et al. (2012) find one in Sweden.

Yamagishi (2001) theorizes that generalized trust is a form of social intelligence – "the ability to understand [one's] own and other people's internal states and use that understanding in social situations" (Yamagishi, 2011; p.125), which is distinct from standard measures of intelligence (IQ, for instance). He then presents two potential explanations for a positive relationship between generalized trust and social intelligence. In the first hypothesis, he assumes that social intelligence is inherently heterogeneous between individuals in a society. Those who are socially intelligent can

[41] In Ross et al. (1977), the false consensus effect is defined as a cognitive bias whereby a person tends to overestimate how the extent to which his or her beliefs or opinions are shared by other people. For a detailed discussion, see Engelmann and Strobel (2000).

afford to expect that most people are trustworthy since they are highly sensitive to untrustworthiness cues, while socially unintelligent people who are less sensitive are better off assuming that unknown others are generally untrustworthy. His second hypothesis is that high-trustors tend to take more social risks and are, therefore, more vulnerable to exploitation, which pushes them to invest cognitive resources in cultivating social intelligence for detecting others' trustworthiness. After acquiring social intelligence to discern others' trustworthiness, they can afford to have a high level of generalized trust. In contrast, those who have not made such cognitive investments are slow in detecting the cues of untrustworthiness in their partners and thus are frequently betrayed in trust relations. The frequent experience of misplaced trust can lead to a progressive withdrawal from potentially fruitful, but risky interactions. As a result, they will be trapped in an "equilibrium of mistrust", thereby maintaining low default expectations of the trustworthiness of others.

This study examines how individuals with different degrees of generalized trust determine whom to interact with and whom to avoid, when they don't have information on others' individual reputation, they cannot rely on the incentives arising from repeated interactions, and there are no contractual mechanisms to deter opportunism. Are high-trustors naïve and credulous as suggested by Butler and his coauthors? Conversely, as argued by Yamagishi, are high-trustors more sensitive than low-trustors to information that predicts whether those with whom they interact are trustworthy, which in turn supports them to maintain high default expectations of others' trustworthiness? We conduct experiments to test these two competing theories and further to identify the underlying mechanism that generates the behavioral difference between high- and low-trustors.

We first elicit subjects' generalized trust by means of a set of non-incentivized attitudinal questions, and on the basis of their answers we classify them as high-trustors and low-trustors. Then we let subjects play a binary-choice trust game repeatedly (Camerer and Weigelt, 1988; Bohnet, et al. 2010). For each session, initially, all of 20 subjects are asked to play the role of trustee and their decisions are elicited using the strategy method[42]. Inspired by Fetchenhauer and Dunning (2012), unbeknown to subjects, we let a computer randomly divide these trustees into two groups. Specifically, in one of the groups (group A) at least 50% of subjects choose to reciprocate if trusted by their partners while in the other group (group B) the percentage of trustworthy trustees is less than 50%. Then, all of 20 subjects are asked to play the game in the role of trustor for 20 periods and for each period, each trustor is matched with one trustee randomly picked, either from group A or B. The innovation in our experiment is that we vary the feedback subjects receive on their partner's behavior across treatments. In the *Baseline* treatment, after having decided whether to trust or not, subjects always receive information on the action taken by

[42] In experimental studies, there are two different methods of eliciting decisions: one is the direct-response method, in which subjects make decisions whenever it is their time to do so; the other is the strategy method, in which subjects make contingent decisions for all nodes at which they may have to play. For a survey of these two experimental methods, see Brandts and Charness (2011).

their partner, i.e. regardless of their own choice, they always know ex post whether trusting would have been profitable. In the *Free Endogenous Feedback* treatment, subjects are allowed to decide whether to acquire feedback about their partner's action after they decide whether to trust. In the *Contingent Feedback* treatment, subjects get to know their partner' choice only if they decide to trust him. In all these three treatments, before choosing whether to trust or not, subjects are told whether their current opponent belongs to group A or B, but they don't know which of the groups contains a higher proportion of trustworthy people. In the *Ex-ante Feedback* treatment, instead, trustors are told whether their partner belongs to the high or low trustworthiness group, before making their choice.

This setup allows us to test Butler et al.'s (forthcoming) hypothesis, that subjects' trusting behavior is persistently influenced by their own trustworthiness and is insensitive to feedback revealing the distribution of the trustworthiness of trustees. In addition, our experimental design also allows us to consider Yamagishi's theory and examine three possible reasons why high-trustors are better than low-trustors at predicting others' trustworthiness. First, according to Yamagishi's first hypothesis, since high-trustors are more socially intelligent than low-trustors, even if receiving the same type of information as in the *Baseline* or in the *Ex-ante Feedback* treatments, they may be better at processing information about the reliability of their partners. As a result, high-trustors are more likely than low-trustors to recognize reliable partners in the transactions. Second, a growing body of research indicates that the decision to trust is deeply influenced by betrayal aversion, that is, the anticipated aversive emotions connected to possible betrayal or exploitation of one's own trust by another person (Aimone and Houser, 2012, 2013; Aimone et al., forthcoming; Bohnet and Zeckhauser, 2004; Bohnet et al., 2008). People with higher levels of emotional intelligence – one type of social intelligence (Grewal and Salovey, 2005) – are shown to manage their emotions more successfully and adapt their behavior in response to others' cooperativeness more quickly in the Prisoner's Dilemma game (Fernandez-Berrocal et al., 2014). We expect that high-trustors may be better than low-trustors at regulating the anticipated aversive emotions and thereby be more willing to acquire feedback about their partners when information acquisition must be intentional. As a result, they should be better at judging the trustworthiness of others in the *Free Endogenous Feedback* treatment. Third, according to Yamagishi's cognitive investments hypothesis, in order to develop skills to distinguish trustworthy from untrustworthy partners, high-trustors engage in collecting more differentiating social data and learning more relative to low-trustors. Consequently, more trusting behavior will be observed among high-trustors when subjects are allowed to obtain feedback only if they chose to trust (i.e. *Contingent Feedback* treatment).

We report three main findings. First, in line with Butler et al.'s (forthcoming) theory, we show that when no ex-ante feedback is provided, subjects' initial trusting behavior in the game is significantly correlated with their trustworthiness level. When decomposing data based on subject-types, we find that the consensus effect is only

observed among those who are classified as low-trustors on the basis of their questionnaire answers to the trust questions. Second, the consensus effect disappears very quickly, as evident by that fact that both high- and low-trustors are able to identify the low- and high-trustworthiness groups and to condition their behavior on the group of their current opponent. Finally, compared to their counterparts, high-trustors are better at predicting others' trustworthiness only because they are less susceptible to the anticipated aversive emotions aroused by the potential betrayal and thereby are keener to acquire information about others' actions.

The remainder of this paper is organized as follows: Section 3.2 lays out the related literature; Section 3.3 describes our experimental design and procedures; Section 3.4 reports the main results of the experiment; Section 3.5 concludes with legal policy implications.

3.2 Related Literature

Economic life is dominated by encounters with strangers, and market participants need to distinguish trustworthy from opportunistic partners in their transactions. To secure the mutually advantageous transactions, people may be capable of reading their partner's nonverbal cues of reciprocating intentions, which is highly related to the literature on "green beard effect" (Dawkins, 1976; Frank, 1988, 2005; West and Gardner, 2010). According to the perspective of the "green beard effect", altruists typically have observable characteristics (e.g., a green beard) that distinguish them from non-altruists, and other altruists, who also have this unique feature, can recognize them and treat them preferentially. In recent behavioral studies, target subjects are usually allowed to play a computer-mediated one-shot Prisoner's Dilemma game for real money, and their pictures are taken at the very moment of decision-making. Among these pictures, the "standard" pictures that are equated for background, brilliance and luminance, and more or less equal in size are randomly selected and presented to observers (Yamagishi, et al. 2003; Verplaetse, et al. 2007). These studies reveal that observers can accurately discriminate non-cooperative pictures from cooperative ones based on their quick nonverbal impressions.

Frank et al. (1993) and Brosig (2002) give subjects opportunities to communicate with their partners in a separate room for a certain time period ranging from 10 to 30 minutes before playing a one-shot two-person prisoner's dilemma game and ask them to predict their opponent's decisions. As a result, they find that subjects are able to predict their partner's play with an accuracy rate above chance. Stated differentially, Frank et al. (1993) find an accuracy of 11 percentage points above chance and Brosig (2002) 8 percentage points above chance. Moreover, a similar study has also revealed that, after watching a TV show on prisoner's dilemma games, female subjects who are substantially more cooperative are better at identifying cooperators in the real games (Belot, et al. 2012).

The most common paradigm to test the readability of nonverbal cues on trustworthiness is the standard trust game (Berg, et al. 1995). Using the trust game experiments, several studies have demonstrated a causal effect of facial cues (male facial width, for instance) on trusting behavior, with subjects investing significantly more real money in partners with trustworthy-looking faces who, normally, are less likely to abuse trust (van't Wout and Sanfey, 2008; Stirrat and Perrett, 2010).

In the real world, in order to predict trustworthiness correctly, people rely not only on physical cues from their partners, but also on prior social experience of social interactions, including positive or negative experience. Previous studies have shown that initial social knowledge of moral character (Delgado, et al. 2005; Mikolajczak, et al. 2010), prior direct experience with partners (Fareri, et al. 2012), or a behavioral history of partners (Rezlescu, et al. 2012) can influence subjects' capability of predicting trustworthiness. In our study we exclusively focus on the role of the feedback that subjects receive about their partner's behavior, which varies across treatments, and examine how subjects with different degrees of generalized trust learn, trial by trail, through positive or negative feedback about their partner's trustworthiness. Therefore, our paper complements the studies of the "green beard effect" that mainly concentrate on the effect of physical cues on predicting trustworthiness.

The second literature branch to which our paper contributes is the study of dynamic trust learning. To examine whether subjects can correctly recognize trustworthy partners over time, Phan and his colleagues (Sripada et al. 2009, 2013; Phan et al., 2010) randomly assign three types of opponent (20 trials with each type) to subjects and ask them to play repeated trust games with their opponents. Specifically, all active subjects in the lab are assigned to play the role of trustor, and told that they will be playing with other players who have previously participated in the same game as counterparts (i.e. trustees) and whose responses were previously recorded and now serve as counterparts' "reactions" to their decisions - this particular setting leaves no space for trustors to manipulate the beliefs of their partner by strategic actions. In addition, subjects are also told that they will play with three types of counterparts who were classified based on their previously recorded actions as: (1) type 1: "tend to split the money more than 50% of the time"; (2) type 2: "tend to split the money about 50%"; and (3) type 3: "tend to split the money less than 50% of the time". Unbeknown to the subjects, however, they actually play with the computer-simulated agent with different preprogrammed strategies. Once subjects choose their decision, regardless of their choosing to invest or keep the money, feedback about their partner's pre-recorded choice is provided immediately to the subjects. The authors find that subjects are able to distinguish trustworthy from untrustworthy partners, and that learning occurs rapidly, with differential investing based on partner type observed on average by the fifth trial, and stabilizing thereafter.

Van den Bos et al. (2011) take a further step to examine who are the subjects who are able to learn whom to trust or distrust more rapidly. They employ a similar experimental design used by Phan et al. (2010),[43] where three age groups, namely late childhood, mid-adolescence and young adulthood, are recruited to play repeated trust games with predetermined three-type trustees. It is revealed that subjects of all age groups perceive the three types of partners differing significantly in their trustworthiness, and increasingly trust the most trustworthy partner the most and the least trustworthy the least, suggesting that subjects of all ages are able to learn to trust and distrust their partner based on the feedback they received. In addition, it also finds that adults and adolescents adapt their levels of trust in response to others' trustworthiness more quickly than children do.[44]

Our paper complements these studies of dynamic trust learning and mainly examines the capability differences in predicting trustworthiness between high- and low-trustors. However, in contrast to the previous studies, particularly to the studies based on the framework of Phan et al. (2010), our experimental design allows us to achieve progress in two methodological issues. Firstly, in previous studies, in order to obtain three different types of trustees - a particular situation that may not arise naturally with high probability, researchers deliberately deceive the participant with regards to the partner with whom he or she is matched. While some kinds of deception are acceptable in psychological studies especially when the experiment itself really requires that subjects' behavior in the lab situation to resemble the behavior they might display in the real-world situation, experimental economists believe that researchers should not employ deception in the design of experiments, because deception could evoke suspicion and mistrust among participants, which may dramatically change their behavior in experiments, and even in future experiments (Hertwig and Ortmann, 2001; Jamison, et al. 2008; for a thorough review, see Ortmann and Hertwig, 2002). Instead, in our repeated trust game setting described above, we successfully create two types of trustees (i.e. low- and high-trustworthiness groups) without deceiving the subjects.

Secondly, in the framework of Phan et al. (2010), trustors are always informed about the decision of their partner immediately after they have to decide whether or not to trust that person. Based on this artificial non-contingent feedback, all types of trustors (different age groups, for example) are able to learn quickly about their partner's trustworthiness, and adapt their behavior accordingly. We extend the previous setting to cover another situation where trustors receive their partner's

[43] There are two main differences in experimental design between these two studies: contrary to Phan et al. (2010), the subjects in Van den Bos et al. (2011) are told that the other player makes his or her decision through an Internet connection in real time but in reality the choice is made by the computer program and is displayed after a variable delay 2-4 seconds; secondly, photographs of partners of the same age and gender are presented to the subjects.

[44] A series of experiments conducted by Fett and her coauthors show that subjects with low capability of metalizing such as children, or people with psychosis, are less sensitive to signals that reveal others' trustworthiness or untrustworthiness (Fett et al., 2014a, 2014b; Gromann et al., 2013).

feedback only when they decide to trust their partners, a more realistic phenomenon that better matches real world situations (Fetchenhauer and Dunning, 2010), and examine whether this new condition induces more significant behavioral differences in trust learning between high- and low-trustors. Since the informative feedback facilitates trustors to adapt their behavior, it is important to note that trustors have an incentive to make more trusting decisions in the initial periods of the game to maximize the informativeness of the feedback. Recently, a neuroimaging study (Krueger, et al. 2007) has revealed that a high activation in the paracingulate cortex (PcC), a brain region frequently implicated in conflict monitoring and cognitive control in social interactions (Baumgartner, et al. 2008), is observed among sophisticated trustors when exploiting trusting strategies more often in the initial stages of the trust game, then it gradually diminishes with experience, reflecting the behavior of the sophisticated trustors who stabilize their trusting strategies in later stages while the opposite pattern is observed among unsophisticated trustors. Therefore, we expect that, in our contingent feedback condition, more trusting behavior will be observed among high-trustors in the initial periods, which makes them learn to trust or distrust their partners more quickly.

3.3 Experimental Design

Our experiment consists of four main parts, which are depicted in **Figure 3.1**. We will first describe the trust game (part III), which represents the core of our experiment. Then we will illustrate the other three parts, in detail.

3.3.1 Trust Game

The trust game involved a trustor who had to decide whether to trust the trustee. If the trustor chose not to trust the trustee, both got 10 Yuan. If the trustor chose to trust, the trustee had to decide whether to reciprocate. When the trustee reciprocated, both got 15 Yuan; otherwise, the trustor got 8 Yuan while the trustee got 22 Yuan (see **Figure 3.2**).[45]

Fig. 3.1 The main experimental parts

[45] In our paper, the payoff structure of the trust game replicates Bohnet and Zeckhauser (2004) and Bohnet et al. (2008, 2010). The instructions use neutral framing.

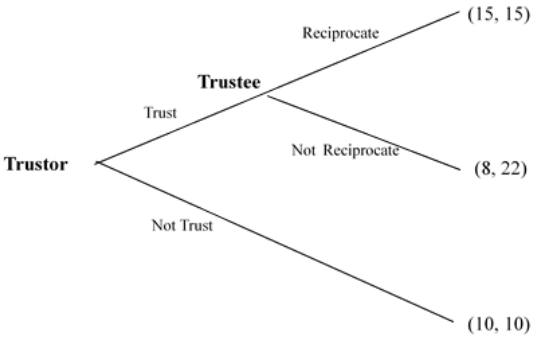


Fig. 3.2 The binary choice trust game

We implemented the trust game in two stages. In the first stage, all subjects played in the role of trustee, and were asked to decide whether to reciprocate or not, in case their opponent chose to trust (i.e. the strategy method).[46,47] After all the trustees had made their decisions, unbeknownst to them, the computer randomly formed two groups, according to the pre-programmed assignment rule (see **Appendix 3**). In group A, more than 50% of subjects chose to reciprocate, while in group B much less than 50% chose to reciprocate.[48]

In the second stage, subjects were asked to play the trust game in the role of trustor for 20 periods. Before making their decisions, they were told that the computer had randomly formed two groups of trustees based on the choices made in the previous stage: in one of the groups, at least 50% of the people chose to honor trust, while in the other group, much less than 50% of the people behaved trustworthily. For each period, each trustor was randomly matched with one trustee picked from one of the groups above. Subjects were also told that a period would be randomly selected. This would be used to determine the payments for themselves at this stage and for their partner in the previous stage. At the beginning of the period, subjects were told whether their opponent belonged to group A or to group B, and in all but the *Ex-ante*

[46] Although the strategy method has been shown to decrease the trustees' trustworthiness in the trust game, it produces similar trusting behavior compared to the direct-response method (Casari and Cason, 2009). In our paper, we mainly focus on the trustors' behavior, therefore, using the strategy method will not weaken our results.

[47] Similarly to Guillen and Ji (2011), subjects in our non-standard ordering setting were neither given information about their partners nor told that all subjects would act as trustor in the next part of the experiment. Our results on the levels of trustworthiness do not qualitatively differ from previous studies using the strategy method in a standard ordering.

[48] Note that it would have been impossible to form two groups of equal size. For each session, the final assignment of trustees is presented in **Appendix 4**.

Feedback treatment they did not know which of the two groups was characterized by higher frequency of trustworthy people.

In order to identify which factor exactly makes high-trustors adapt their behavior in response to others' trustworthiness (or untrustworthiness) more quickly than do low-trustors, four different treatments were implemented at this stage (see **Table 3.1**).

Table 3.1
Treatments and sessions

Treatment	*Baseline*	*Ex-ante Feedback*	*Contingent Feedback*	*Free Endogenous Feedback*
Feedback acquisition	Unconditional	Unconditional	Conditional	Intentional
Ex-ante information	No	Yes	No	No
Session	01/05/2013; 01/06/2013	01/05/2013; 01/06/2013	01/05/2013; 01/06/2013	01/05/2013; 01/06/2013
Subjects	40	40	40	40
Independent observations	40	40	40	40

Notes: Sessions conducted in January 2013. In the table, "Feedback acquisition" indicates that, for each period, subjects are unconditionally (or conditionally) informed of the feedback from their partners; "Ex-ante information" indicates whether subjects are exposed to ex-ante information about the trustee groups before making their decision.

In the *Baseline* treatment, for each period, subjects were allowed to observe the actual decision of their opponent immediately after they decided whether or not to trust, i.e. they received the feedback from their opponent regardless of their own choice. This treatment allows us to test one possible reason high-trustors are better than low-trustors at predicting others' trustworthiness, i.e. they process the trustworthiness-related information more quickly and effectively when receiving the same type of information.

Hypothesis 1: *even if they receive the same type of information, high-trustors learn to trust or distrust their partners more quickly than do low-trustors.*

In the *Baseline* condition, as subjects did not have any information about the trustworthiness of their potential partners at the beginning, they had to learn by trial and error. In many instances, however, people may have had prior social experience with potential partners who may have been colleagues, relatives, or business partners.

These experiences coded by initial social knowledge (Delgado et al., 2005), or prior direct experience (Fareri et al., 2012) can influence decisions about whether to trust. In the *Ex-ante Feedback* treatment, for each period, besides receiving feedback about their opponent's choice unconditionally, trustors were also told whether their opponent belonged to the high or to low trustworthiness group, before making their first decision. Previous experimental studies have shown that subjects are very sensitive to the feedback that is consistent with their prior social impressions (Fareri et al., 2012; Li et al., 2011). It implies that unbiased aggregate information could speed up the subjects' abilities to learn how to identify people who are trustworthy and thereby weaken high-trustors possibly relative advantage in processing the decentralized information on others' trustworthiness.

Hypothesis 2: *when they receive aggregate information on the degree of reliability of the trustee, high-trustors and low-trustors behave similarly.*

In the *Contingent Feedback* treatment, the setting was the same in the *Baseline* except that, for each period, subjects were informed of the choice made by their partner only when they had decided to trust that person. This condition allows us to test Yamagishi's cognitive investments hypothesis, i.e. high-trustors are better at detecting reliable partners because they deliberately trust more in the initial periods and thus collect more information about their partners. If no significant difference in trusting behavior between high- and low-trustors emerges in the *Baseline* condition but a difference emerges here, a possible conclusion is that high-trustors are better at predicting trustworthiness because they acquire more information through more trusting, but not because they are better at processing this information.

Hypothesis 3: *when feedback is contingent on trusting, high-trustors trust more than low-trustors do.*

In the *Free Endogenous Feedback* treatment, the setting was the same in the *Baseline* except that, for each period, subjects could always receive feedback on their partner's choice regardless of their trusting behavior, but the information acquisition had to be intentional. The design of this condition is motivated by a series of recent behavioral studies suggesting that subtle psychological factors, such as betrayal or regret aversion, may modulate people's trusting behavior (Aimone and Houser, 2012; Behnet et al. 2008; Fetchenhauer and Dunning, 2010).[49] Moreover, in order to elicit the maximal possibility of betrayal or regret aversion, we adopted the "opt-in" rather than "opt-out" as a default option (Dana et al. 2007; Larson and Capra, 2009; Grossman, 2010). This condition allows us to test another possible reason high-trustors are better than low-trustors at detecting reliable partners, i.e. they are

[49] Normally, when people trust another person and that person betrays their trust, they become painfully aware of that betrayal, and this "betrayal aversion" leads many trustors to avoid risk more when a person, rather than nature, determines the outcome of uncertainty; and at the same time, if people distrust another person and that person is actually a trustworthy person, trustors expect to feel regret and avoid knowing that person's choice.

keener to acquire information about others' behavior.

Hypothesis 4: *when feedback is endogenous, high-trustors ask for feedback more often than low-trustors.*

At the very end of the trust game, in all but the *Ex-ante Feedback* treatment, we invited all subjects in the session to guess which of the two groups contained the higher proportion of trustworthy people. Subjects who answered correctly would be paid 10 Yuan as a monetary reward. This task helps us examine whether subjects could correctly recognize the high-trustworthiness group after several learning periods.

3.3.2 Questionnaire

At the beginning of each session, subjects were asked to fill in a questionnaire detailing individual characteristics (gender, age, background, etc.) and measuring individual risk attitudes, time preferences, social preferences, and trust beliefs (see **Appendix 1**). In this study, we followed the literature to classify subjects into two categories (i.e. high- and low-trustors) based on their questionnaire answers to the trust questions (Carter and Weber, 2010; Yamagishi and Yamagishi, 1994). If subjects played the trust game before the questionnaire, their answers to the trust questions may be highly susceptible to the experience in the game, which weakened the reliability of the subject classification. Therefore, we chose to have the questionnaire before the trust game. In addition, to avoid demand effect and to obtain reliable trust beliefs, we introduced many questions into the questionnaire to deemphasize the questions on trust.

The most frequently used measure of trust beliefs is taken from the General Social Survey (GSS) /World Values Survey (WVS). Both surveys assess trust using the following question: "*Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?"* The survey respondents can answer in a binary way to this question by agreeing either with "*Most people can be trusted*" or with "*Can't be too careful*." This trust measure has been widely criticized by many social science scholars pointing out that a risk-averse person may share the view that "Most people can be trusted", while at the same time risk aversion may induce this person to say "Can't be too careful" because this person engages in avoiding small probability risks that have large payoff consequences (Miller and Mitamura, 2003; Fehr, 2009). To avoid the possible ambiguity, in our study we adopted "one-dimensional" questions on trust (i.e. from question 25 to question 30 in the questionnaire) developed by Yamagishi and his colleagues (Yamagishi et al. 1998; Yamagishi and Kosugi, 1999; Yamagishi and Yamagishi,

1994),[50] because this new questionnaire has been demonstrated as a more reliable instrument for the measurement of trust beliefs and has achieved highly predictive validity in several contexts (Carter and Weber, 2010).

3.3.3 Lottery Game

After the questionnaire, subjects were asked to play a lottery game, which helps us measure subjects' risk attitudes. Clearly, the decision to trust a stranger entails a risk. Uncertainty regarding a potential trustee's prosocial preference is the source of risk. This raises the important concern over whether the decision to trust a social partner is influenced by one's general attitude toward risk (Eckel and Grossman, 1996; Houser et al., 2010; Schechter, 2007). In order to control for differences in risk attitudes between high- and low-trustors, we therefore implement this lottery game.

The game is similar to Holt and Laury's (2002), which offers subjects a series of pair-wise lotteries of both safe and risky options presented in **Figure 3.3**. Take Decision 1 for example: if subjects choose Option A, they receive 10 Yuan; if they choose Option B, the risky option, they have a 10% chance of winning 25 Yuan and a 90% chance of winning nothing. In this lottery game, the safe option does not change, but the expected payoff for the risky option increases as we move down the table.

Decision	Option A	Option B	Your Choice
Decision 1	10	25 with a probability of 10% 0 with a probability of 90%	Option A Option B
Decision 2	10	25 with a probability of 20% 0 with a probability of 80%	Option A Option B
Decision 3	10	25 with a probability of 30% 0 with a probability of 70%	Option A Option B
Decision 4	10	25 with a probability of 40% 0 with a probability of 60%	Option A Option B
Decision 5	10	25 with a probability of 50% 0 with a probability of 50%	Option A Option B
Decision 6	10	25 with a probability of 60% 0 with a probability of 40%	Option A Option B
Decision 7	10	25 with a probability of 70% 0 with a probability of 30%	Option A Option B
Decision 8	10	25 with a probability of 80% 0 with a probability of 20%	Option A Option B
Decision 9	10	25 with a probability of 90% 0 with a probability of 10%	Option A Option B
Decision 10	10	25 with a probability of 100% 0 with a probability of 0%	Option A Option B

Confirm

Fig. 3.3: Lottery game

[50] The 6-item questionnaire with general statements to measure subjects' beliefs about honesty and trustworthiness of others (see **Appendix 1**) is the English translation of Yamagishi and Kosugi's (1999) Trust Belief Scale, which we copy from Carter and Weber (2010).

Subjects were asked to choose either Option A or Option B for each decision. After they made all of their decisions, the computer would randomly choose one of the decisions. For this selected decision, if subjects chose Option A, they got 10 Yuan; if subjects chose Option B, another random draw by the computer determined the payoff of the subjects in this part.

3.3.4 Bayesian Updating Game

At the end of the trust game, all subjects were asked to play a Bayesian updating game (El-Gamal and Grether, 1995). Since subjects in the trust game encountered different opponents, as they observed new feedback, they would update their beliefs about others' trustworthiness. The heterogeneity in Bayesian updating abilities may induce subjects to perform differently in judging others' trustworthiness. The Bayesian updating game allows us to take into account the possibility that high-trustors are better than low-trustors at processing the statistical information in general, which in turn makes them predict others' trustworthiness more accurately.

Subjects were asked to play the game for 20 periods. Before playing the game, they saw two "bingo cages", labeled as cage A and cage B. They were informed that both cages contained six balls, but that the composition was different. In cage A, there were four Red balls (R) and two Green balls (G) while in cage B, there were three Red balls (R) and three Green balls (G).

At the beginning of each period, the computer randomly drew two integer numbers to determine which of the cages was selected for the following task. The computer first randomly drew an integer number from the set {1, 2, 3, 4, 5, 6}, then drew another number that could be 2, 3, or 4, with equal probability. The first drawn number was not revealed to subjects while the second drawn number was announced publicly. If the first drawn number was smaller or equal to the second drawn number, cage A was selected, otherwise cage B was selected.

Once cage A (or cage B) was selected, six draws from this cage were performed (the ball being replaced each time), and the result (GGRRGG, for instance) was displayed on the subjects' computer screen. Subjects were then asked to guess which cage was used to generate the observed result. They did not receive any feedback about the correctness of each guess until the end of the last period. The subject who achieved the highest score was paid 200 Yuan at the end of the experiment. If more than one subjects got the same highest score, they shared 200 Yuan.

Hypothesis 5: *high-trustors are better at Bayesian learning than low-trustors.*

3.3.5 Experimental Procedure

The experiment involved 160 subjects, divided in 8 sessions and was conducted at the Social Science Experimental Center (SSEC) of Zhejiang University, China. Subjects were mostly undergraduate students at Zhejiang University, and were recruited through posters on university campus noticeboards. About 48 percent of the subjects were male; nobody took part in more than one session. The experiment was programmed and implemented using the software z-Tree (Fischbacher, 2007). For each session, after showing up at the lab at the pre-scheduled session time, the 20 subjects were randomly assigned to a cubicle to avoid eye contact, and no communication was allowed during the experiment. The instructions were distributed separately before each part of the experiment and were read aloud by the experimenter. Subjects' questions, if any, were answered by the experimenter in private. To make sure that all subjects understood the instructions correctly, they had to complete a comprehension quiz with calculations and questions before making decisions in each part. The average session lasted about 1.5 hours. Subjects were paid privately in cash at the end of the session and earned on average 60 Yuan (i.e. nearly 7 Euros), including the show-up fee of 10 Yuan.[51]

3.4 Experimental Results

This section reports the main results. In Section 4.1 we provide aggregate data to gain a general description of the experiment, and then examine the reliability of the data. In Section 4.2 we mainly focus on individuals' trusting behavior in repeated trust games to test the competing theories in our setting: first, we examine whether subjects' trusting behavior is persistently influenced by their own trustworthiness and is immune to feedback about their partners, which is the main implication of Butler et al.'s (2012, forth.) hypothesis; then we study whether and under which conditions high-trustors are better than low-trustors at predicting others' trustworthiness (**Hypothesis 1-4**). In Section 4.3 we juxtapose data from the questionnaire, the lottery game, the trust game and the Bayesian updating game to study whether there is a difference between high-trustors and low-trustors in terms of individual characteristics, and in particular, whether high-trustors are better than low-trustors at Bayesian updating (**Hypothesis 5**).

3.4.1 General Information

We begin by giving an overview of findings about the distributions of subject types across treatments. Following Yamagishi's approach, we first construct a trust score by averaging each subject's questionnaire answers to the trust questions, and then divide

[51]To take into account possible no-show-ups, we recruited more than 20 students for each session. Only 20 students were randomly selected to participate in the experiment, and supernumerary students were paid 20 Yuan and had to leave before the session started.

our subjects into two categories at the median trust score: those whose score is equal or above the median are labeled "H-types" (i.e. high-trustors), while the remaining subjects are labeled "L-types" (i.e. low-trustors). As shown in **Table 3.2**, the percentage of H-types ranges between 45% (in the *Contingent Feedback* treatment) and 65% (in the *Ex-ante or Free-endogenous Feedback* treatment), and no significant differences emerge across treatments (p=0.22, chi-square test with three degrees of freedom). On average 58.8% of subjects are categorized as H-types in our whole dataset.

We then analyze the data from the trust game and report the findings about the average trustworthiness across the treatments. The percentage of trustworthiness ranges between 22.5% (in *Baseline* treatment) and 37.5% (in *Free-endogenous Feedback* treatment) and no significant differences emerge across treatments (p=0.53, chi-square test with three degrees of freedom). On average, nearly 30% of the subjects in our sessions are willing to honor trust if trusted by their partners. Our results regarding our subjects' trustworthiness levels are similar to those reported by Bohnet et al. (2010) who use the same payoff structure for the trust game. There is also no significant difference in trustworthiness between H-types and L-types (p=0.14, chi-square test with one degree of freedom), with 24.24% of H-types and 35.11% of L-types reciprocating trust.

In the trust game, when assigned the role of trustor, subjects trust their partner with about 41% frequency in the *Baseline* treatment, while more trusting behavior is found in the other three treatments. Especially in the *Contingent Feedback* treatment, 74% trust is observed, which seems to imply that trustors are engaging in learning their partners' types through trusting more. Decomposing the trusting behavior based on the partner groups, we find that more than 50% trust is given to the "good" group (i.e. group A) for all treatments. Particularly in the *Contingent Feedback* treatment, when facing the opponent from the "good" group, subjects choose to trust about 96% chance. Similar trust levels are also observed in the *Ex-ante Feedback* treatment where subjects already knew which group was the "good" group. When analyzing the trust towards the "bad" group (i.e. group B), different behavior patterns are found between unconditional and conditional feedback treatments. In the *Baseline* and *Ex-ante Feedback* treatments, less than 30% trust is shown towards the "bad" group. However, in the *Contingent Feedback and Free Endogenous Feedback* treatments where obtaining feedback depends on subjects' further steps, nearly twice as much trusting behavior is observed.

Table 3.2
Average level of main variables

	Baseline	*Ex-ante Feedback*	*Contingent Feedback*	*Free Endogenous Feedback*
H-type (%)	60%	45%	65%	65%
Trustworthiness (%)	22.5%	32.5%	30%	37.5%
Trust (%)	40.9%	52.4%	74%	63.8%
Trust towards "good" group (%)	75.3%	90.8%	95.8%	86.5%
Trust towards "bad" group (%)	26.5%	22.6%	58%	43.3%

Notes: The percentage of H-type subjects is calculated using the data from the questionnaire while the percentage of the other four variables is calculated using the data from the trust game.

To ensure the reliability of the data, we check whether subjects acquired the information about the composition of the two groups from others who had played the game in the previous sessions. More specifically, we test whether the level of trust towards group A is significantly higher than towards group B in the first period, when they have no information on the levels of trustworthiness of the two groups' members.

Table 3.3
Testing the possibility of information exchange

Session	Treatment	# Observations in group A	#Observations in group B	Statistical Significance
1	Baseline	N_g=4	N_b=16	p=0.025**
2	Contingent Feedback	N_g=8	N_b=12	p=0.209
3	Ex-ante Feedback	N_g =10	N_b=10	p=0.068*
4	Endogenous Feedback	N_g =10	N_b=10	p=0.531
5	Ex-ante Feedback	N_g =8	N_b=12	p=0.001***
6	Endogenous Feedback	N_g =9	N_b=11	p=0.413

7	Baseline	N_g =9	N_b=11	p=0.095*
8	Contingent Feedback	N_g =9	N_b=11	p=0.881

Notes: Our null hypothesis is that subjects cannot distinguish group A from group B in the first period of the trust game. N_g indicates the observations in group A while N_b indicates the observations in group B. p-values reported in the last column are from a two-tailed chi-square test. The symbols *, **, and *** indicate significance at the 10%, 5% and 1% level, respectively.

As shown in **Table 3.3**, only in session 1 and session 5, the p-values indicate a significant relationship between trust and the type of partner group, and in session 3 and session 7, this relationship is weakly significant. Since subjects in session 3 and 5 (i.e. the *Ex-ante Feedback* treatment) have been told that in group A at least 50% of the trustees chose to reciprocate trust while in group B less than 50% did, it is not surprising to find that subjects can distinguish these two groups even in the first period. We can also rule out that subjects in session 1 had any information about the composition of the two groups because they are the first participants who attended this experiment.[52] Although a weakly significant difference in the levels of trust towards the two groups in the first period is observed in session 7, it vanishes dramatically in the next period (p=0.888, chi-square test with one degree of freedom), suggesting that the significant difference must have emerged by chance.

3.4.2 Testing the Competing Theories

3.4.2.1 Consensus Effect

According to Butler et al. (forth.), the consensus effect is so strong that it could remain even after several rounds of game play. As shown in the Section 4.1, when no aggregate information about trustees' groups had been offered to subjects, subjects could not distinguish high- from low-trustworthiness groups in the first period. We first examine whether subjects' trusting behavior in the first period is highly correlated to their own trustworthiness.

Finding 1. *When no ex-ante information is provided before making their decision, subjects' initial trusting behavior is highly correlated with their own trustworthiness; and the consensus effect is mainly contributed by L-types.*

[52] In the first session of the experiment, subjects do not always trust more towards group A, as evident by the fact that the significant difference disappears in period 3 (p=0.648, chi-square test with one degree of freedom).

Table 3.4a
Consensus effect among L-types

	Trustworthiness	Untrustworthiness
Trust	100%	71.8%
Distrust	0%	28.2%

Notes: "Trust" (or "Distrust") indicates that subjects in the role of trustor decide to trust (or to distrust) their partner in the first period; "Trustworthiness" (or "Untrustworthiness") denotes that subjects honor (or abuse) the trust as a trustee.

Table 3.4b
Consensus effect among H-types

	Trustworthiness	Untrustworthiness
Trust	87%	75.6%
Distrust	13%	24.4%

Notes: "Trust" (or "Distrust") indicates that subjects in the role of trustor decide to trust (or to distrust) their partner in the first period; "Trustworthiness" (or "Untrustworthiness") denotes that subjects honor (or abuse) the trust as a trustee.

In order to examine the existence of the consensus effect, researchers normally elicit subjects' trust beliefs about their partners before playing the first period of the trust game. However, many economists criticize the reliability of the belief elicitation method with or without monetary incentives (Blanco et al., 2010). To avoid the possible bias produced by the elicitation method, we analyze the correlation between subjects' trusting behavior in the first period and their own trustworthiness using the data from all but the *Ex-ante Feedback* treatment. We find a significantly positive correlation between these two behavioral variables (p=0.027, chi-square test with one degree of freedom). When decomposing the data based on subject-types, as shown in **Table 3.4a-3.4b**, we find that L-types are more likely to trust if they previously honored their partner's trust as a trustee (p=0.03, chi-square test with one degree of freedom) while H-types' trust in the first period is not significantly correlated with their previous trustworthiness (p=0.27, chi-square test with one degree of freedom). It implies that, in our experiment, only L-type subjects are susceptible to the consensus effect, generating their initial trusting behavior based on their own trustworthiness.[53]

Although subjects (more precisely, L-types) make their first decision from their own trustworthiness, this phenomenon disappears very quickly. We then examine whether trustors could correctly decode each partner type's inclination to reciprocate over time (i.e. trust learning), thereby hindering the effect of false consensus.

Finding 2. *Both H- and L-types show increasing trust for the trustworthy type partners and decreasing trust for the untrustworthy type partners over time.*

[53] When using Fisher's exact test, we also find similar results (p=0.029 for pooled data; p=0.047 for L-type subjects; and p=0.35 for H-type subjects).

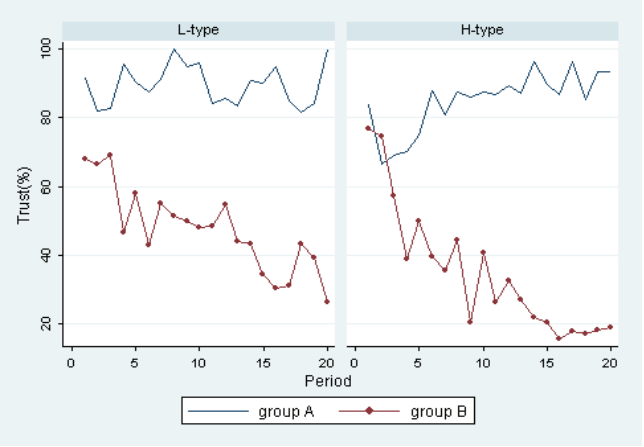


Fig. 3.4 Trust learning[54]

Table 3.5
Trust Learning

	Baseline Treatment		**Contingent Feedback Treatment**		**Free Endogenous Feedback Treatment**	
Variable	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Period	-0.100	-0.100	-0.152	-0.159	-0.114	-0.119
	(0.020)***	(0.020)***	(0.019)***	(0.020)***	(0.020)***	(0.020)***
H-type	0.305	0.297	-0.502	-0.370	-1.303	-1.374
	(0.226)	(0.228)	(0.209) **	(0.219) *	(0.232)***	(0.236)***
High-group	0.659	0.675	0.381	0.304	0.265	0.205
	(0.436)	(0.439)	(0.611)	(0.617)	(0.503)	(0.507)

[54] Based on the data from all but the *Ex-ante Feedback* treatment, the line in the figure represents the frequency of trustful choices towards group A for each period, while the connected line represents the frequency of trustful choices towards group B for each period.

H-type ×	-0.107	-0.112	1.224	1.280	0.127	0.161
High-group	(0.388)	(0.391)	(0.638) *	(0.642) **	(0.454)	(0.458)
Period ×	0.166	0.166	0.182	0.192	0.185	0.194
High-group	(0.034)***	(0.034)***	(0.051)***	(0.052)***	(0.034)***	(0.035)***
Controls	No	Yes	No	Yes	No	Yes
Observations	800	800	800	800	800	800

Notes: Standard errors reported in parenthesis. The dependent variable is a dummy indicating trust. × denotes interaction terms. *H-type* equals 1 for subjects who are H-types; *High-group* equals 1 for groups that have higher proportion of trustworthy subjects; *Controls* includes *a dummy variable* session, and three controls for individual characteristics, i.e. *Cognitive ability*[55], *Altruism* corresponding to our questionnaire-based measure of altruism, and *Risk aversion* indicating subject's risk attitude in the lottery game. The symbols *, **, and *** indicate significance at the 10%, 5% and 1% level, respectively.

After removing the data from the *Ex-ante Feedback* treatment, in the remaining treatments, as shown in **Figure 3.4**, we find that both H- and L-types are more successful at distinguishing the high trustworthiness group from the low one after obtaining more experience, and this finding is also supported by the result of the non-parametric test ($p<0.001$, Wilcoxon signed-rank test). [56] To provide additional evidence, we further implement a mixed effects logistic (MEL) regression per treatment. As revealed in **Table 3.5**, the coefficient of the interaction term "Period × High-group" is strongly significant, implying that subjects successfully achieve trust learning and recognize the good type partners quickly. Our results are not consistent with Butler et al.'s main findings. One possible reason is that, in Butler et al.'s (forthcoming) paradigm, subjects' roles for each period are randomly reassigned, which makes it more difficult for subjects to estimate the stable distribution of their partner's types, as a result, the false consensus effect is naturally and repeatedly observed in their study.

[55] To measure cognitive ability, we use a three-item cognitive reflection test (CRT) introduced by Frederick (2005). Following Oechssler et al. (2009) and Hoppe and Kusterer (2011), we classify the subjects into two types based on their CRT score: those who correctly answered zero or one of the questions are classified as low cognitive able subjects while the rest are classified as high cognitive able subjects.

[56] First, we focus on trustors in general, finding that trustors can successfully distinguish the high trustworthiness group from the low one ($p<0.001$, Wilcoxon signed-rank test). Then, we separate H-type from L-type subjects, finding similar results (for L-types, $p<0.001$, Wilcoxon signed-rank test; for H-types, $p<0.001$, Wilcoxon signed-rank test).

3.4.2.2 Difference in predicting trustworthiness

To interpret their counterintuitive finding that high-trustors are better than low-trustors at predicting others' trustworthiness, Yamagishi (2001; 2011) proposes two hypotheses. First, he argues that people in society have different levels of social intelligence, and that high-trustors who are more socially intelligent are more sensitive than low-trustors to information that reveals others' trustworthiness. Second, he theorizes that high-trustors are more willing than low-trustors to invest cognitive resources in cultivating social intelligence for detecting others' trustworthiness, consequently, they engage in collecting more differentiating social data through trusting their potential partners more. In order to test Yamagishi's theory, in this part we first examine whether H-types are better than L-types at processing information about the reliability of their partners when exogenously exposed to the same type of information (i.e. decentralized or aggregate information), and then study whether H-types are more willing to endogenously acquire more feedback about their partners, which facilitates them to predict others' trustworthiness more accurately relative to L-types.

Finding 3. *H- and L-types perform similarly on predicting others' trustworthiness both in the Baseline treatment and in the Ex-ante Feedback treatment.*

Table 3.6
Trusting behavior and performance

	Baseline	*Ex-ante Feedback*	*Contingent Feedback*	*Free Endogenous Feedback*
Trust (%) of H-type	42.5%	56.5%	71.9%	57.1%
Trust (%) of L-type	38.4%	44.6%	75.7%	76.1%
Trust group A (%) of H-type	78.2%	93.6%	97.4%	82.5%
Trust group A (%) of L-type	75%	81.5%	94.7%	93.3%
Trust group B (%) of H-type	25.3%	25.7%	53.1%	34.3%
Trust group B (%) of L-type	20.8%	15.4%	61.7%	40.6%
Guess Correctness (%) of H-type	91.7%	-	100%	96.2%

Guess Correctness (%) of L-type	100%	-	100%	85.7%

Notes: "Trust (%)" stands for the frequency of subjects' trustful actions in trust games; "Trust group A (or B)" stands for the frequency of subjects' trustful actions towards group A (or B) in trust games; "Guess Correctness (%)" stands for the percentage of correctness in the guessing task at the end of the trust game.

First, we analyze the data from the *Baseline* treatment to examine whether H-types perform better in the guessing task after finishing 20 decisions in the trust game. Unexpectedly, as shown in **Table 3.6**, H-types achieve about 92% correctness while L-types achieve 100% correctness, but the performance difference is not significant (p=0.236, chi-square test with one degree of freedom). This implies that at the very end of the trust game both types of subjects are able to detect the high trustworthiness group. We then test whether H-types learn to trust or distrust faster than L-types do, finding that H-types do not trust the high-trustworthiness-group members more (p=0.9, two tailed Wilcoxon rank-sum test, N_1=16 and N_2=24) and the low-trustworthiness-group members less (p=0.42, two tailed Wilcoxon rank-sum test, N_1=16 and N_2=24). To provide additional evidence, we further implement mixed effects logistic (MEL) regressions. As reported in **Table 3.7a -3.7b**, compared to L-types, H-types do not trust the "good" group more and the "bad" group less, over time.

Next, we focus on the data from the *Ex-ante Feedback* treatment and examine whether H-types are more adaptive than L-types to this specific game task when informed of the general distribution information about the trustworthy players in each trustee-group at the very beginning of the trust game. We could not find that H-types trust the "good" group more (p=0.32, two tailed Wilcoxon rank-sum test, N_1=14 and N_2=26) and the "bad" one less (p=0.47, two tailed Wilcoxon rank-sum test, N_1=14 and N_H=26). These results are also confirmed by MEL regressions shown in **Table 3.7a-3.7b**. In summary, the findings revealed in the *Baseline* and *Ex-ante Feedback* treatments imply that both H- and L-types can efficiently take advantage of information about their partners, as a consequence, exposure to the same type of exogenous information (i.e. decentralized or aggregate feedback) makes them perform similarly.

Table 3.7a
Trust towards the high trustworthiness group (Exogenous feedback)

	Baseline		Ex-ante Feedback	
Variable	Model 1	Model 2	Model 1	Model 2
Period	0.048	0.047	-0.047	-0.084
	(0.041)	(0.042)	(0.044)	(0.052)
H-type	-0.019	-0.052	1.219	-0.891
	(0.592)	(0.609)	(0.806)	(1.051)
H-type × Period	0.022	0.027	0.035	0.068
	(0.054)	(0.055)	(0.065)	(0.072)
Controls	No	Yes	No	Yes
Observations	260	260	360	360

*Notes: Standard errors reported in parenthesis. The dependent variable is a dummy indicating trust. Controls includes a dummy variable session, and three controls for individual characteristics, i.e. Cognitive ability, Altruism and Risk aversion. The symbols *, **, and *** indicate significance at the 10%, 5% and 1% level, respectively.*

Table 3.7b
Trust towards the low trustworthiness group (Exogenous feedback)

	Baseline		Ex-ante Feedback	
Variable	Model 1	Model 2	Model 1	Model 2
Period	-0.088	-0.088	-0.069	-0.062
	(0.033)***	(0.033)***	(0.042)	(0.044)
H-type	0.619	0.634	0.446	-0.253
	(0.451)	(0.454)	(0.516)	(0.569)
H-type × Period	-0.031	-0.033	0.030	0.024
	(0.042)	(0.042)	(0.048)	(0.050)
Controls	No	Yes	No	Yes
Observations	540	540	440	440

*Notes: Standard errors reported in parenthesis. The dependent variable is a dummy indicating trust. Controls includes a dummy variable session, and three controls for individual characteristics, i.e. Cognitive ability, Altruism and Risk aversion. The symbols *, **, and *** indicate significance at the 10%, 5% and 1% level, respectively.*

Finding 4. *In the Contingent Feedback treatment, H- and L-types acquire a similar amount of information about their partner's actions, therefore, they perform similarly on predicting others' trustworthiness.*

According to Yamagishi's cognitive investments hypothesis, we expect to see that H-types are more likely to deliberately collect information about their partners by trusting more than L-types. However, as shown in **Table 3.6**, both H- and L-types choose to trust more than 70% of the time in the *Contingent Feedback* treatment,

implying that they acquire a similar amount of feedback from their partners (p=0.16, two tailed Wilcoxon rank-sum test, N_1=22 and N_2=18).[57] Consequently, as shown in the **Table 3.6**, at the end of the trust game both H- and L-types guess the high trustworthiness group with 100% correctness. To investigate the development of trust over interactions with two trustee-groups, we analyze the change in trust across 20 periods. As shown in **Figure 3.5**, the behavioral trend of H- and L-types is similar, which is supported by a two-tailed Wilcoxon rank-sum test revealing that H-types do not trust more the "good" group (p=0.83, N_1=22 and N_2=18) and less the "bad" group (p=0.28, N_1=22 and N_2=18).

We then conduct MEL regressions to examine whether H-types' trusting behavior is more adaptive than L-types. As shown in **Table 3.8a**, while the coefficient for H-types is negative and weakly, the interaction term "H-type × Period" is significantly positive. It implies that H-types initially trust less than L-types but, after accumulating experience, they recognize the high trustworthiness group and increase their trust in it dramatically, which is consistent with the trend shown in **Figure 3.5**.

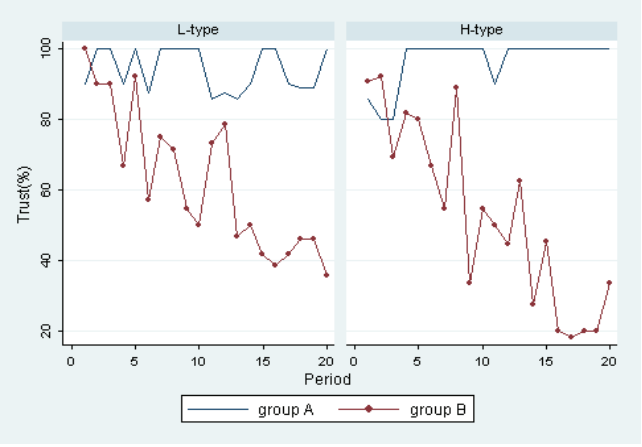


Fig. 3.5 Learning in the *Contingent Feedback*[58]

[57] Even in the first period of the trust game, the frequency of trustful choices is similar between H- and L-types (L-types: 74.2%; H-types: 73.4%).

[58] Based on the data from the *Contingent Feedback* treatment, the line in the figure represents the frequency of trustful choices towards group A by period, while the connected line represents the frequency of trustful choices towards group B by period.

Table 3.8a
Trust towards the high trustworthiness group (Endogenous feedback)

	Contingent Feedback		Free-endogenous Feedback	
Variable	Model 1	Model 2	Model 1	Model 2
Period	-0.040	-0.034	-0.017	-0.012
	(0.056)	(0.059)	(0.059)	(0.058)
H-type	-1.841	-1.905	-2.176	-2.201
	(1.115)*	(1.223)	(0.761)***	(0.766) ***
H-type × Period	0.344	0.397	0.109	0.116
	(0.162)**	(0.172)**	(0.066)*	(0.067)*
Controls	No	Yes	No	Yes
Observations	340	340	380	380

*Notes: Standard errors reported in parenthesis. The dependent variable is a dummy indicating trust. Controls includes a dummy variable session, and three controls for individual characteristics, i.e. Cognitive ability, Altruism and Risk aversion. The symbols *, **, and *** indicate significance at the 10%, 5% and 1% level, respectively.*

Table 3.8b
Trust towards the low trustworthiness group (Endogenous feedback)

	Contingent Feedback		Free-endogenous Feedback	
Variable	Model 1	Model 2	Model 1	Model 2
Period	-0.133	-0.137	-0.044	-0.046
	(0.026)***	(0.026)***	(0.032)	(0.032)
H-type	-0.038	0.105	-0.024	-0.075
	(0.477)	(0.484)	(0.491)	(0.500)
H-type × Period	-0.042	-0.042	-0.119	-0.123
	(0.039)	(0.039)	(0.042)***	(0.043)***
Controls	No	Yes	No	Yes
Observations	460	460	420	420

*Notes: Standard errors reported in parenthesis. The dependent variable is a dummy indicating trust. Controls includes a dummy variable session, and three controls for individual characteristics, i.e. Cognitive ability, Altruism and Risk aversion. The symbols *, **, and *** indicate significance at the 10%, 5% and 1% level, respectively.*

Finding 5. *In the Free Endogenous Feedback treatment, H-types show less trust towards the low trustworthiness group, and this main effect is qualified by a significant time trend.*

In the *Contingent Feedback* treatment, subjects face two groups with different distributions of trustworthy trustees, and their main task is essentially to identify

which one of the two groups has a higher frequency of trustworthy members. This specific setting easily encourages subjects to trust much more often. In the real world, besides the motivation of learning (i.e. curiosity), other psychological factors, such as betrayal or regret aversion, may also influence subjects' trusting behavior. We want to examine whether H-types are better than L-types at regulating the aversive emotions aroused by betrayal or regret, and thereby are more likely to acquire feedback about their partners, which makes them learn to trust or distrust faster.

As shown in **Table 3.6**, when information acquisition is intentional in the *Free Endogenous Feedback* treatment, H-types perform slightly better in the guessing task relative to L-types, but the difference is not significant (p=0.232, chi-square test with one degree of freedom). When we turn to observe subjects' trusting behavior in detail, as shown in **Figure 3.6**, we find that H-types learn faster to distrust the low-trustworthiness-group members (p=0.01, two tailed Wilcoxon rank-sum test, N_1=14 and N_2=26), while both types of subjects perform similarly in predicting the trustworthiness of those from the high-trustworthiness-group (p=0.21, two tailed Wilcoxon rank-sum test, N_1=19 and N_2=29).

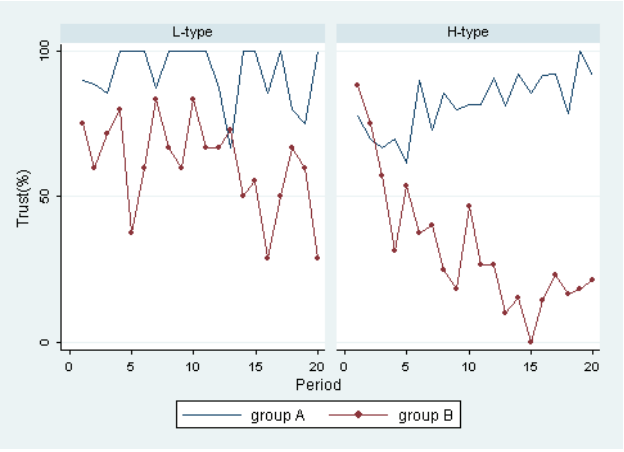


Fig. 3.6 Learning in the *Free Endogenous Feedback*[59]

In order to study whether the time trend emerging in the *Free Endogenous Feedback* treatment is statistically significant, here we introduce MEL regressions to examine whether H-types adapt their strategies towards different trustee-groups more quickly than do L-types. We first focus on subjects' trusting behavior towards the

[59] Based on the data from the *Free Endogenous Feedback* treatment, the line in the figure represents the frequency of trustful choices towards group A for each period, while the connected line represents the frequency of trustful choices towards group B for each period.

"good" group. As shown in **Table 3.8a**, while the interaction term "H-type × Period" is positive and weakly significant in MEL regressions, the coefficient for H-types is negative and highly significant. It implies that H-types catch up with L-types who start at higher levels of trust, which is also confirmed by the evidence shown in **Figure 3.6**. We then examine how subjects learn to distrust the trustees from the "bad" group. As revealed in **Table 3.8b**, the interaction term "H-type × Period" is significantly negative, implying that H-types are quicker at adapting their behavior in response to the cues of untrustworthiness.

Finding 6. *When information acquisition has to be intentional, H-types have a higher willingness to acquire information about their partners' actions as compared to L-types. The underlying reason for this phenomenon is not because H-types are less regret averse but because they are less betrayal averse than L-types.*

As shown in **Table 3.9**, H-types ask for feedback more than 90% of the time after they decide whether or not to trust while only 80% of L-types ask for feedback, and this difference is weakly significant (p=0.06, two tailed Wilcoxon rank-sum test, N_1=14 and N_2=26), suggesting that H-types are more likely to acquire information about their partners, which may help them learn faster to trust or distrust others than do L-types.

Table 3.9
Asking for information

	H-type	*L-type*
Asking for feedback (%)	94.4%	80%
Asking for feedback (%) if Not Trust	91.9%	91%
Asking for feedback (%) if Trust	96.3%	76.5%

Note: "Asking for feedback (%)" stands for the frequency of asking for information about partners' actions by subjects.

We further examine which psychological factors actually drive this phenomenon. One possibility is that when subjects choose to distrust their partner but their partner may actually behave trustworthily, they expect to feel regretful and engage in avoiding discovering their partner's choice (i.e. regret aversion); another possibility is that when subjects choose to trust their partner but their partner may abuse that trust, they choose not to know about their partner's action in order to reduce this potentially painful psychological cost (i.e. betrayal aversion). As revealed in **Table 3.9**, when

choosing to distrust their partner, L-types ask for feedback slightly more frequently than do H-types but this difference is not significant (p=0.53, two tailed Wilcoxon rank-sum test, N_1=10 and N_2=26), which is consistent with recent experimental evidence showing that choosing to distrust does not activate aversive emotions (Aimone et al., forthcoming). However, when subjects choose to trust their partner, H-types have a higher willingness to acquire feedback and the difference is significant (p=0.03, two tailed Wilcoxon rank-sum test, N_1=14 and N_2=25). This suggests that H-types may be better than L-types at regulating the anticipated negative emotions aroused by betrayal aversion and thereby are more keen to acquire the useful information about their partner's action, consequently, they learn to trust or distrust their partner more quickly.

3.4.3 Individual characteristics

As discussed in the Section of Experimental design, we follow Yamagishi's approach to classify subjects into two categories on the basis of their questionnaire answers (i.e. H- and L-types). Our main results could be seriously weakened if these two types of subjects are highly different in the terms of individual characteristics. To secure the robustness of the results, in this part, we analyze the data about individual characteristics.

Firstly, we focus on the questionnaire data and find that there is no significant difference between H- and L-types in terms of the general individual characteristics except that the individual's social status and the experience on obtaining financial aid. H-types are more likely to be student leaders and to receive financial aid from the university (for the details, see **Appendix 5**).

Secondly, we use the data from the lottery game and the Bayesian updating game, and try to examine whether H- and L-types behave differently in these two games.

Table 3.10
Risk attitudes and Bayesian learning abilities

	Baseline	*Ex-ante Feedback*	*Contingent Feedback*	*Free Endogenous Feedback*
Risk aversion index of H-type	5.1	5.1	4.8	5.2
Risk aversion index of L-type	5.1	5.4	5.8	5.2
Correctness (%) of H-type	81.7%	85.4%	86.1%	86.4%
Correctness (%) of L-type	84.7%	83.2%	84.6%	86.4%

Notes: "Risk aversion index" is calculated using the data from the lottery game, which measures the degree of a subject's risk aversion; "Correctness (%)" indicates the percentage of guess correctness in the Bayesian updating game by subjects.

Finding 7. *There is no significant difference between H- and L-types in terms of risk attitudes.*

In previous studies, there is a heated debate about whether risk attitudes can be used to predict trusting behavior in the trust game: running two experiments with a diverse set of subjects in fifteen villages of rural Paraguay, Schechter (2007) reveals that risk attitudes are highly predictive of play in the trust game while Eckel and Grossman (1996) and Houser et al. (2010) could not replicate this finding using other subject samples. In our study, we use the lottery game to measure subjects' risk attitudes[60], and try to test whether H- and L-types hold significantly different risk attitudes. As shown in **Table 3.10**, on average subjects choose the first five safe options and then switch to select the risky options. H- and L-types are similar in terms of risk attitudes (p=0.16, two tailed Wilcoxon rank-sum test, N_1=64 and N_2=88).

Finding 8. *H- and L-types are equally able to process statistical information in the Bayesian updating task.*

As shown in **Table 3.10**, we observe that on average both H- and L-types achieve more than 80% correctness in the Bayesian updating game, and that there is almost no difference in the performance between them. This result is also supported by a two-tailed Wilcoxon rank-sum test (p=0.85, N_1=66 and N_2=94), suggesting that H- and L-types have similar Bayesian learning abilities.

[60] In our study, before analyzing the data of risk attitudes, we removed 8 subjects because of their inconsistent choices in the lottery task. In our measurements higher values mean more risk aversion (i.e. the value for the extreme risk aversion is 10).

3.5 Conclusions

People seem to believe that high-trustors are gullible and tend to trust others indiscriminately, thereby performing worse than low-trustors. However, many experimental studies reveal that the reverse is true: compared to their counterparts, high-trustors are better at lie detection (Carter and Weber, 2010), and are significantly more accurate in their predictions of others' trustworthiness (Yamagishi, 2011). In this study we take a further step to identify which factors make high-trustors adapt their behavior in response to others' trustworthiness or untrustworthiness more quickly. Our main findings are that both high- and low-trustors can learn whom to trust over time, and that high-trustors are better than low-trustors at predicting others' trustworthiness not because they are better at processing the trustworthiness-related information, or that they deliberately collect differentiating social data through trusting more, but only because they are less susceptible to the anticipated aversive emotions aroused by the potential betrayal and thereby have a higher willingness to acquire the valuable information about their partner's actions.

These findings have important implications for empirical research and legal policy. Firstly, in Butler et al. (forthcoming), subjects are shown to form their trust beliefs about other people based on their own trustworthiness and this false consensus effect has a strongly persistent impact on their trusting behavior even after several periods of learning. Consequently, high-trustors who hold overly optimistic default expectations of others' trustworthiness have to suffer the substantial cost of forming biased trust beliefs. Contrary to their studies, our experiment allows subjects to learn their partner's types under a stable distribution of trustees, which helps them utilize past useful experience. Consequently, even though no ex-ante aggregate information is provided, both high- and low-trustors are able to learn whom to trust or distrust over time. It implies that when investors are involved in a stable commercial environment, they could correctly recognize their partners' types by learning. However, when the environment becomes highly undetermined, they may have great difficulty evaluating their partners' trustworthiness and tend to make their decisions mainly based on their own trustworthiness, which causes them consequently to suffer because of this. Therefore, in future work it is appropriate to manipulate the degree of uncertainty in the experimental environment in order to investigate whether the effectiveness of trust learning is dependent on this.

Secondly, in all treatments trustors could correctly trust trustworthy type partners more and trust untrustworthy type partners less over time; however, different trusting behavior between unconditional and conditional feedback treatments is still observed. When facing the low trustworthiness group, subjects trust much more in the *Contingent Feedback* and *Free Endogenous Feedback* treatments than in the other two treatments. It implies that if necessary information about their potential partners is unconditionally provided, much unprofitable trusting behavior can be avoided.

Thirdly, in terms of other legal policy implications, the results raise important questions for interventions designed to improve market efficiency, which typically focus on incentive mechanisms, such as pricing or monetary rewarding systems. Our study provides experimental evidence, suggesting that with higher degrees of generalized trust, high-trustors seem to manage the aversive emotions aroused by betrayal more successfully, which causes them to have a higher willingness to be involved in market participation. Consequently, they learn more about distinguishing trustworthy from untrustworthy partners. In order to overcome the problems generated by betrayal aversion, the decision-maker of public policy should mandatorily disclose the necessary information on the outcome of transactions and make the commercial environment transparent. Consequently, the scope of market transactions would expand, and more investors be attracted to participate in market transactions and recognize trustworthy partners quickly, thereby sustaining a higher level of market efficiency.

Reference

Aimone, J., and Houser, D., 2012, "What You Don't Know Won't Hurt You: A Laboratory Analysis of Betrayal Aversion." *Experimental Economics*, *15*(4), 571-588.

Aimone, J., and Houser, D., 2013, "Harnessing the Benefits of Betrayal Aversion." *Journal of Economic Behavior and Organization*, *89*, 1–8.

Aimone, J., Houser, D., and Weber, B., forthcoming, "Neural Signatures of Betrayal Aversion: An fMRI Study of Trust." *Proceedings of the Royal Society B – Biological Science*.

Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., and Fehr, E., 2008, "Oxytocin Shapes the Neural Circuitry of Trust and Trust Adaption in Humans." *Neuron*, 58, 639-650.

Belot, M., Bhaskar V., and van de Ven, J., 2012, "Can Observers Predict Trustworthiness." *The Review of Economics and Statistics*, 94(1), 264-259.

Berg, J., Dickhaut J., and McCabe K., 1995, "Trust, Reciprocity and Social History", *Games and Economic Behavior*, 10, 122-142.

Blanco, M., Engelmann, D., Koch, A., and Normann, H., 2010, "Belief Elicitation in Experiments: Is There a Hedging Problem?" *Experimental Economics*, 13, 412-438.

Bohnet, I., Grieg, F., Herrmann, B., and Zeckhauser, R., 2008, "Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States", *American Economic Review*, *98*(1), 294– 310.

Bohnet, I., Hermann, B., and Zeckhauser, R., 2010, "Trust and the Reference Points for Trustworthiness in Gulf and Western Countries", *Quarterly Journal of Economics*, *125*(2), 811–828.

Bohnet, I., and Zeckhauser, R., 2004, "Trust, Risk and Betrayal", *Journal of Economic Behavior and Organization*, *55*, 467–484.

Brandts, J., and Charness, G., 2011, "The Strategy versus the Direct-Response Method: A First Survey of Experimental Comparisons", *Experimental Economics*, 21, 1-24.

Brosig, J., 2002, "Identifying Cooperative Behavior: Some Experimental Results in a Prisoner's Dilemma Game", *Journal of Economic Behavior and Organization, 47,* 275-290.

Butler, J., Giuliano, P., & Guiso L., 2012, "The Right Amount of Trust." EIEF Working Paper.

Butler, J., Giuliano, P., & Guiso L., forthcoming, "Trust, Values and False Consensus." *International Economic Review*.

Camerer, C. F., and Weigelt, K., 1988, "Experimental Tests of a Sequential Equilibrium Reputation Model", *Econometrica*, LVI, 1-36.

Carl, N., and Billari, F., 2014, "Generalized Trust and Intelligence in the United States." *PLoS ONE*, 9, e91786.

Carter N., and Weber, M., 2010, "Not Pollyannas: Higher Generalized Trust Predicts Lie Detection Ability", *Social Psychological and Personality Science*, *I(3)*, 274-279.

Casari, M., and Cason, T., 2009, "The Strategy Method Lowers Measured Trustworthy Behavior", *Economics Letters*, 103, 157-159.
Dana, J., R. A. Weber, and J. X. Kuang, 2007, "Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness," *Economic Theory*, 33, 67–80.
Dawkins, R., 1976, *The Selfish Gene*. Oxford: Oxford University Press.
Delgado, M. R., Frank, R. H., and Phelps, E. A., 2005, "Perceptions of Moral Character Modulate the Neural Systems of Reward during the Trust Game", *Nature Neuroscience*, 8, 1611–1618.
Eckel, C., and Grossman, P., 1996, "Altruism in Anonymous Dictator Games", *Games and Economic Behavior*, 16, 181-191.
El-Gamal M., and Grether D., 1995, "Are People Bayesian? Uncovering Behavioral Strategies", *Journal of the American Statistical Association*, Vol. 90(432).
Engelmann, D., and Strobel, M., 2000, "The False Consensus Effect Disappears if Representative Information and Monetary Incentives Are Given", *Experimental Economics*, 3, 241-260.
Fareri, D. S., Chang, L. J., and Delgado, M. R., 2012, "Effects of Direct Social Experience on Trust Decisions and Neural Reward Circuitry", *Frontiers in Neuroscience*, 6(148), 1-17.
Fehr, E., 2009, "On the Economics and Biology of Trust", *Journal of the European Economic Association*, 7, 235-266.
Fernandez-Berrocal, P., Extremera, N., Lopes, P., and Ruiz-Aranda, D., 2014, "When to Cooperate and When to Compete: Emotional Intelligence in Interpersonal Decision-making." *Journal of Research in Personality*, 49, 21-24.
Fetchenhauer, D., and Dunning, D., 2010, "Why So Cynical? Asymmetric Feedback Underlies Misguided Skepticism Regarding the Trustworthiness of Others", *Psychological Science*, 21(2), 189-193.
Fetchenhauer, D., and Dunning, D., 2012, "Betrayal Aversion versus Principled Trustfulness-How to Explain Risk Avoidance and Risky Choices in Trust Games", *Journal of Economic Behavior and Organization*, 81, 534-541.
Fett, A., Gromann, P., Giampietro, V., Shergill, S., and Krabbendam, L., 2014a, "Default Distrust? An fMRI Investigation of the Neural Development of Trust and Cooperation." *Social Cognitive and Affective Neuroscience*, 9, 395-402.
Fett, A., Shergill, S., Gromann, P., Dumontheil, I., Blakemore, S., Yakub, F., and Krabbendam, L., 2014b, "Trust and Social Reciprocity in Adolescence – A Matter of Perspective-taking." *Journal of Adolescence*, 37, 175-184.
Fischbacher, U., 2007, "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments", *Experimental Economics*, 10, 171-178.
Frank, R. H., 1988, *Passions Within Reason: The Strategic Role of the Emotion*, W.W. Norton & Co., New York.
Frank, R. H., 2005, "Altruists with Green Beards: Still Kicking?" *Analyse & Kritik*, 27, 85-96.
Frank, R. H., Gilovich, T., and Regan, D. T., 1993, "The Evolution of One-shot Cooperation: An Experiment", *Ethology and Sociobiology*, 14, 247-256.

Frederick, S., 2005, "Cognitive Reflection and Decision Making", *Journal of Economic Perspective*, 19(4), 25-42.
Grewal D., and Salovey, P., 2005, "Feeling Smart: The Science of Emotional Intelligence." *American Scientist*, 93, 330-339.
Gromann, P., Heslenfeld, D., Fett, A., Joyce, D., Shergill, S., and Krabbendam, L., 2013, "Trust versus Paranoia: Abnormal Response to Social Reward in Psychotic Illness." *Brain*, 136, 1968-1975.
Grossman, Z., 2010, "Strategic Ignorance and the Robustness of Social Preferences", UC Santa Barbara working paper.
Guillen, P., and Ji, D., 2011, "Trust, Discrimination and Acculturation: Experimental Evidence on Asian International and Australian Domestic University Students." *Journal of Socio-Economics*, *40*, 594-608.
Hertwig, R., and Ortmann, A., 2001, "Experimental Practices in Economics: A Methodological Challenge for Psychologists?" *Behavioral and Brain Science*, 24, 383-451.
Holt, C., and Laury, S., 2002, "Risk Aversion and Incentive Effects in Lottery Choices", *American Economic Review*, 92, 1644-1655.
Hooghe, M., Marien, S., and de Vroome, T., 2012, "The Cognitive Basis of Trust: The Relation between Education, Cognitive Ability, and Generalized and Political Trust." *Intelligence*, 40, 604-613.
Hoppe, E., and Kusterer, D., 2011, "Behavioral Biases and Cognitive Reflection." *Economics Letters*, *10*, 97-100.
Houser, D., Schunk, D., and Winter, J., 2010, "Distinguishing Trust from Risk: An Anatomy of the Investment Game", *Journal of Economic Behavior and Organization*, 74:1, 72-81.
Jamison, J., Karlan, D., and Schechter, L., 2008, "To Deceive or Not to Deceive: The Effect of Deception on Behavior in Future Laboratory Experiments", *Journal of Economic Behavior and Organization*, 68, 477-488.
Kakiuchi, R., and Yamagishi, T., 1997, "General Trust and the Dilemma of Variable Interdependency." *The Japanese Journal of Experimental Social Psychology*, 12, 212-221.
Kikuchi, M., Watanabe, Y., and Yamagishi, T., 1997, "Judgment Accuracy of Other's Trustworthiness and General Trust: An Experimental Study." *The Japanese Journal of Experimental Social Psychology*, 37, 23-36.
Kosugi, M., and Yamagishi, T., 1998, "General Trust and Judgments of Trustworthiness." *The Japanese Journal of Psychology*, 69, 349-357.
Krueger, F., McCabe, K., Moll, J., Kriegeskorte, N., Zahn, R., Strenziok, M., Heinecke, A., and Grafman, J., 2007, "Neural Correlates of Trust", *Proceedings of the National Academy of Sciences of the United States of America*, 104(50), 20084-20089.
Larson, T., and Capra, C. M., 2009, "Exploiting Moral Wiggle Room: Illusory Preference for Fairness? A Comment", *Judgment Decision Making*, 4, 467–474.

Li, J., Delgado, M., and Phelps, E., 2011, "How Instructed Knowledge Modulates the Neural Systems of Reward Learning." *Proceedings of the National Academy of Sciences of the United States of America*, 108, 55-60.

Mikolajczak, M., Gross, J. J., Lane, A., Corneille, O., de Timary, P., and Luminet, O., 2010, "Oxytocin Makes People Trusting, Not Gullible", *Psychological Science*, 21(8), 1072-1074.

Miller, A. S., and Mitamura, T., 2003, "Are Surveys on Trust Trustworthy?" *Social Psychology Quarterly*, 66, 62-70.

Oechssler, J., Roider, A., and Schmitz, P., 2009, "Cognitive Abilities and Behavioral Biases." *Journal of Economic Behavior & Organization*, *72*, 147-152.

Ortmann, A., and Hertwig, R., 2002, "The Costs of Deception: Evidence from Psychology", *Experimental Economics*, 5(2), 111-131.

Oskarsson, S., Dawes, C., Johannesson, M., and Magnusson, P., 2012, "The Genetic Origins of the Relationship between Psychological Traits and Trust." *Twin Research and Human Genetics*, 15, 21-33.

Phan, K. L., Sripada, C. S., Angstadt, M., and McCabe, K., 2010, "Reputation for Reciprocity Engages the Brain Reward Center", *Proceedings of the National Academy of Sciences of the United States of America*, 107, 13099-13104.

Rezlescu, C., Duchaine, B., Olivola, C. Y., and Chater, N., 2012, "Unfakeable Facial Configurations Affect Strategic Choices in Trust Games with or without Information about Past Behavior", *PLoS, ONE*, 7(3): E34293.

Ross, L., Greene, D., and House, P., 1997, "The False Consensus Effect: An Egocentric Bias in Social Perception and Attribution Processes", *Journal of Experimental Social Psychology*, 13, 279-301.

Schechter, L., 2007, "Traditional Trust Measurement and the Risk Confound: An Experiment in Rural Paraguay", *Journal of Economic Behavior and Organization*, 62, 272-292.

Sripada, C., Angstadt, M., Banks, S., Nathan, P. J., Liberzon, I., and Phan, K. L., 2009, "Functional Neuroimaging of Mentalizing during the Trust Game in Social Anxiety Disorder", *Neuroreport*, 20(11), 984-989.

Sripada, C., Angstadt, M., Liberzon, I., McCabe K., and Phan, K. L., 2013, "Aberrant Reward Center Response to Partner Reputation during a Social Exchange Game in Generalized Social Phobia." *Depress Anxiety*, 30, 353-361.

Stirrat, M., and Perrett, D. I., 2010, "Valid Facial Cues to Cooperation and Trust: Male Facial Width and Trustworthiness", *Psychological Science*, 21, 349-354.

Sturgis, P., Read, S., and Allum, N., 2010, "Does Intelligence Foster Generalized Trust? An Empirical Test using the UK Birth Cohort Studies." *Intelligence*, 38, 45-54.

van den Bos, W., van Dijk, E., and Crone, E. A., 2011, "Learning Whom to Trust in Repeated Social Interactions: A Developmental Perspective", *Group Processes and Intergroup Relations*, 15(2), 243-256.

van't Wout, M., and Sanfey, A. G., 2008, "Friend or Foe: The Effect of Implicit Trustworthiness Judgments in Social Decision-Making", *Cognition*, 108, 796-803.

Verplaetse, J., Vanneste, S., and Braeckman, J., 2007, "You Can Judge a Book by Its Cover: The Sequel A Kernel of Truth in Predictive Cheating Detection", *Evolution*

and Human Behavior, 28, 260-271.
West, S.A., and Gardner, A., 2010, "Altruism, Spite, and Greenbeards", *Science*, 327(5971), 1341-1344.
Yamagishi, T., 2001, "Trust as a Form of Social Intelligence", In K.S. Cook (Ed.), *Trust in Society*, New York: Russell Sage.
Yamagishi, T., 2011, *Trust: The Evolutionary Game of Mind and Society*, Springer.
Yamagishi, T., Cook, K. S., and Watabe, M., 1998, "Uncertainty, Trust, and Commitment Formation in the United States and Japan", *The American Journal of Sociology*, 104(1), 165-194.
Yamagishi, T., and Kakiuchi, R., 2000, "It Takes Venturing into a Tiger's Cave to Steal a Baby Tiger: Experiments on the Development of Trust Relationships", pp.121-123 in Werner Raub and Jeroen Weesie (eds.), *The Management of Durable Relations*, Thela Thesis Publishers.
Yamagishi, T., and Kosugi, M., 1999, "Cheater Detection in Social Exchange", *Cognitive Studies*, 6(2), 179-190.
Yamagishi, T., Kikuchi M., and Kosugi M., 1999, "Trust, Gullibility, and Social Intelligence", *Asian Journal of Social Psychology*, 2, 145-161.
Yamagishi, T., Tanida, S., Mashima, R., Shimoma, E., and Kanazawa, S., 2003, "You Can Judge a Book by Its Cover: Evidence That Cheaters May Look Different from Cooperators", *Evolution and Human Behavior*, 24, 290-301.
Yamagishi, T., and Yamagishi, M., 1994, "Trust and Commitment in the United States and Japan", *Motivation and Emotion*, 18(2), 129-166.

Appendix 1

1. Gender
(1) Female (2) Male

2. How old are you?
___ years

3. Race
(1) Minority ethnic group (2) Han

4. Citizen
(1) Rural (2) Urban

5. At which year do you live in the campus?
(1) 1st year (2) 2nd year (3) 3rd year (4) 4th year (5) 5th year (6) others

6. Field of studies
(1) Economics-related (2) Noneconomics-related

7. Your highest degree goal
(1) Bachelor (2) Master (3) Ph.D.

8. Are you a member of Communist Party of China (CPC)?
(1) No (2) Yes

9. Are you a student leader (i.e. Ganbu)?
(1) No (2) Yes

10. Your father's education
(1) Primary school (2) Middle school (3) High school (4) Junior college (5) College (6) Graduate school

11. Your Mother's education
(1) Primary school (2) Middle school (3) High school (4) Junior college (5) College (6) Graduate school

12. Mother's occupation

(1) Government employee (2) Company employee (3) Self-employed worker
(4) Retired (5) Jobless (6) Others

13. Father's occupation
(1) Government employee (2) Company employee (3) Self-employed worker
(4) Retired (5) Jobless (6) Others

14. The size of your family members (including your grandfather/mother, father/mother)
_____ persons

15. Do you have siblings?
(1) No (2) Yes

16. Do you grow from a single-parent family?
(1) No (2) Yes

17. Family economics situation
(1) Very low (2) Average (4) Above average (5) Well (6) Very well

18. In order to finish your education, did you apply for any financial aid?
(1) No (2) Yes

19. Do you have some work experience?
(1) No (2) Yes

20. Did you participant the similar economic decision making experiments in the past?
(1) No (2) Yes

21. Did you participant the psychological experiments in the past?
(1) No (2) Yes

22. Suppose your earning is the only source of your family income, which the following option will you choose:
(1) Option A: your earning is constant and is equal to the sum of the current incomes of your parents; Option B: your earning is uncertain, with 50% you earn two times of the sum, with 50% you earn 66.7% of the sum.
Which option will you choose?

(2) Option A: your earning is constant and is equal to the sum of the current incomes of your parents; Option B: your earning is uncertain, with 50% you earn two times of the sum, with 50% you earn 50% of the sum.
Which option will you choose?

(3) Option A: your earning is constant and is equal to the sum of the current incomes of your parents; Option B: your earning is uncertain, with 50% you earn two times of the sum, with 50% you earn 80% of the sum.
Which option will you choose?

(4) Option A: your earning is constant and is equal to the sum of the current incomes of your parents; Option B: your earning is uncertain, with 50% you earn two times of the sum, with 50% you earn 90% of the sum.
Which option will you choose?

23. Are you generally a person who is fully prepared to take risks or do you try to avoid taking risk?
Please tick a box on the scale, where the value 1 means: "unwilling to take risks" and the value 10 means: "fully prepared to take risk"

1, 2, 3, 4, 5, 6, 7, 8, 9, 10

24. Suppose you face 7 following options in the real world, which option will you choose?
(1) with 0.1% you could obtain 100,000 Yuan otherwise obtain zero Yuan
(2) with 1% you could obtain 10,000 Yuan otherwise obtain zero Yuan
(3) with 10% you could obtain 1,000 Yuan otherwise obtain zero Yuan
(4) with 25% you could obtain 400 Yuan otherwise obtain zero Yuan
(5) with 50% you could obtain 300 Yuan otherwise obtain zero Yuan
(6) with 75% you could obtain 133 Yuan otherwise obtain zero Yuan
(7) you could certainly obtain 100 Yuan

25. Do you agree that "most people are basically honest"?
(1) Strongly disagree (2) Disagree (3) Neutral (4) Agree (5) Strongly Agree

26. Do you agree that "most people are trustworthy"?
(1) Strongly disagree (2) Disagree (3) Neutral (4) Agree (5) Strongly Agree

27. Do you agree that "most people trust a person if the person trusts them"?
(1) Strongly disagree (2) Disagree (3) Neutral (4) Agree (5) Strongly Agree

28. Do you agree that "most people are basically good-natured and kind"?
(1) Strongly disagree (2) Disagree (3) Neutral (4) Agree (5) Strongly Agree

29. Do you agree that "most people are trustful of others"?
(1) Strongly disagree (2) Disagree (3) Neutral (4) Agree (5) Strongly Agree

30. Do you agree that "generally, I am trustful"?
(1) Strongly disagree (2) Disagree (3) Neutral (4) Agree (5) Strongly Agree

31. Do you agree that “If I suffer a serious wrong, I will take revenge as soon as possible, no matter what the costs”?
Please tick a box on the scale, where the value 1 means: “strongly disagree” and the value 7 means: “strongly agree”

1, 2, 3, 4, 5, 6, 7

32. Do you agree that “If someone offends me, I will also offend him/her”?
Please tick a box on the scale, where the value 1 means: “strongly disagree” and the value 7 means: “strongly agree”

1, 2, 3, 4, 5, 6, 7

32. Are you generally a person who fully controls herself for temptation issues, e.g. quitting smoking, finishing homework in time, keeping fit, saving money for long-term plan, etc.?
Please tick a box on the scale, where the value 1 means: “very poor in self-control” and the value 10 means: “very good at self-control”

1, 2, 3, 4, 5, 6, 7, 8, 9, 10

33. Are you generally a person who cares about others’ feelings and benefits?
Please tick a box on the scale, where the value 1 means: “strong do not care about” and the value 10 means: “strongly care about”

1, 2, 3, 4, 5, 6, 7, 8, 9, 10

34. A bat and a ball cost $ 1.10 in total. The bat costs $ 1.00 more than the ball. How much does the ball cost? ____ cents

35. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? _____ minutes

36. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? _____ days

Appendix 2

Instructions (baseline)

Welcome to this study on economic decision-making!

Please, turn off your mobile phone. From this moment on, no form of communication among participants is allowed. In case you have a question, please rise your hand and the instructor will come to your desk to answer it.

This study includes two parts. In Part 1, a questionnaire will appear on your computer screen. Then followed by Part 2, which includes four experiments, and you can earn money from this part (Note that you are given the instructions for a new experiment just after the previous experiment is finished). The money you earn in the different four experiments will be paid to you in cash, and in private, at the end of experiment.

Now, we will read instructions for Part 1. At the beginning of experiments 1, 2, 3 and 4 in Part 2, we will distribute and read the corresponding instructions.

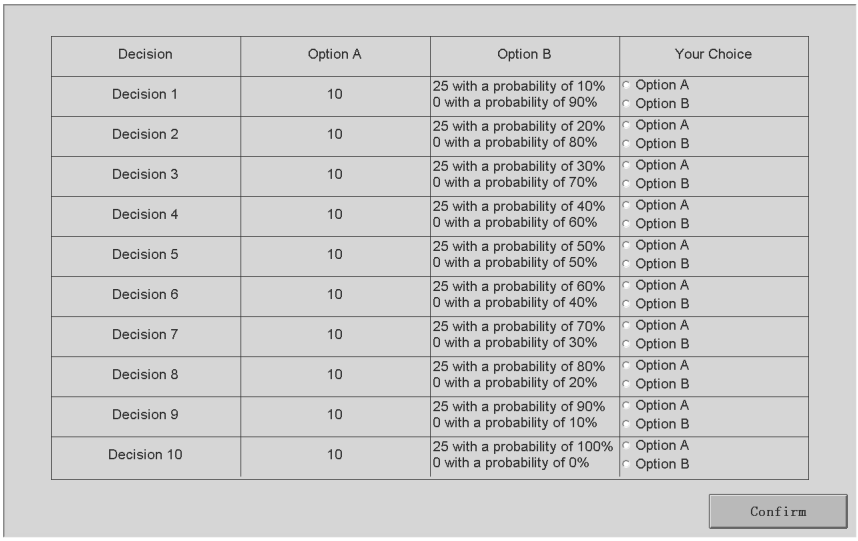
Instructions for Part 1

We now kindly ask you to fill in a questionnaire that will appear soon on your computer screen. Some of the questions concern personal information that will help us for our study. Your identity will never be revealed when results are presented. Please, answer carefully: you will not be able to change your answer, once you confirm it.

Instructions for Experiment 1

In Experiment 1, you can earn up to 25 Yuan, depending on the decisions you make and on the outcome of two random draws.

Your computer screen (Figure 1) will show ten decisions listed on the left. Each decision is a paired choice between "Option A" and "Option B." You will make ten choices and record these in the final column, but only one of them will be used in the end to determine your earnings.



Decision	Option A	Option B	Your Choice
Decision 1	10	25 with a probability of 10% 0 with a probability of 90%	Option A Option B
Decision 2	10	25 with a probability of 20% 0 with a probability of 80%	Option A Option B
Decision 3	10	25 with a probability of 30% 0 with a probability of 70%	Option A Option B
Decision 4	10	25 with a probability of 40% 0 with a probability of 60%	Option A Option B
Decision 5	10	25 with a probability of 50% 0 with a probability of 50%	Option A Option B
Decision 6	10	25 with a probability of 60% 0 with a probability of 40%	Option A Option B
Decision 7	10	25 with a probability of 70% 0 with a probability of 30%	Option A Option B
Decision 8	10	25 with a probability of 80% 0 with a probability of 20%	Option A Option B
Decision 9	10	25 with a probability of 90% 0 with a probability of 10%	Option A Option B
Decision 10	10	25 with a probability of 100% 0 with a probability of 0%	Option A Option B

Confirm

（Fig. 1）

Before you start making your ten choices, please let us explain how these choices will affect your earnings for this experiment. After you have made all of your choices, the computer will randomly draw a number between 1 and 10, to determine which of your decisions will be actually used. Another random draw will determine what your payoff is for the option you chose, A or B, for the particular decision selected. Even though you will make ten decisions, only one of these will end up affecting your earnings, but you will not know in advance which decision will be used. Obviously, each decision has an equal chance of being used in the end.

Now, please look at Decision 1 at the top. Option A pays 10 Yuan, no matter what the outcome of the second draw is. Option B yields 25 Yuan if the outcome of the second draw is 1, and it pays 0 Yuan if the throw is 2-10. The other Decisions are similar, except that as you move down the table, the chances of the higher payoff for option B increase. In fact, for Decision 10 in the bottom row, the die will not be needed since option B pays the highest payoff for sure, so your choice here is between 10 Yuan or 25 Yuan.

To summarize, you will make ten choices: for each decision row you will have to choose between Option A and Option B. You may choose A for some decision rows and B for other rows, and you may change your decisions and make them in any order. When you are finished, please click the “Confirm” button at the bottom right of the screen.

Earnings for this experiment will be added to your earnings for the next three experiments of the study, and you will be paid all earnings in cash when we finish.

So now please look at the empty boxes on the right side of the record sheet. You will have to select an option, A or B in each of these boxes, and then the random draw will determine which one is going to count. We will look at the decision that you made for the choice that counts, before throwing the die again to determine your earnings for this experiment. **You will be informed about the outcome of these draws only at the end of Part 2 of this study.**

Comprehension Questions:
Please answer following questions. Raise your hand if you need help. The instructor will come to help you and will check your answers when you are done.
1), Please read carefully the above table (Figure 1) which will appear on your screen. What is the probability of each decision (from Decision 1 to 10) that will be selected by computer?
(A) 1/2 (B) 1/10 (C) We cannot know it
2), Do you think that, for each decision (from Decision 1 to 10), you will always earn more money from option A than from option B?
(A) Yes (B) No, we cannot know it before the selection made by computer

Now you may begin making your choices!

Instructions for Experiment 2

In Experiment 2, you can earn money depending on the decisions that you and other participants will take. At the end of this study, your earnings for this experiment will be summed with the earnings from other experiments, and you will be privately paid in cash.

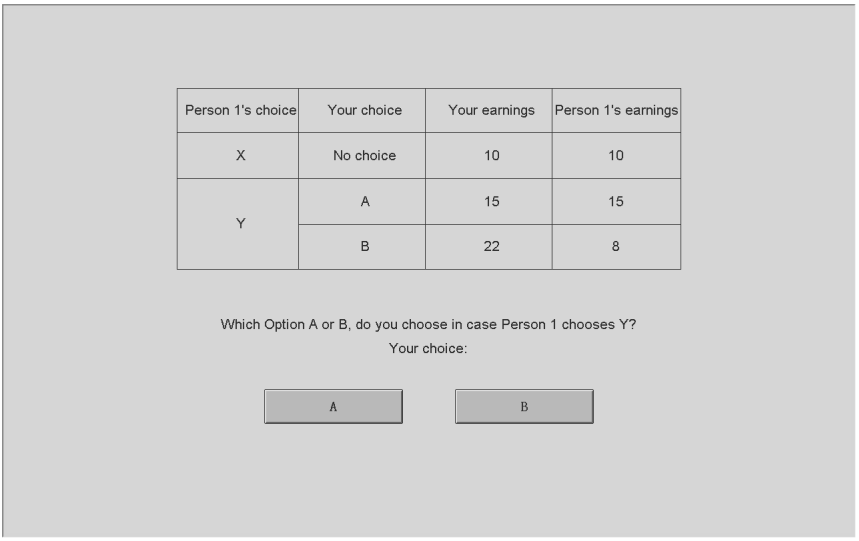
Please, follow the instructions carefully.

How the study is conducted. The study is conducted anonymously. Participants will be unidentified by other participants and experimenter, and no one will be able to associate any decisions with specific people. Participants will be randomly matched into pairs consisting of two roles: person "1" and person "2". There is no communication between you and your counterparts, and your counterpart will never know your true identity, nor will you know theirs.

What the study is about. The study seeks to understand how people decide. You and your counterpart form a pair, and your decisions will determine how much money you earn. Person 1 is confronted with two alternatives, X and Y. X gives person 1 and person 2 a payoff for sure, and person 2 does not need take action. If person 1 chooses option Y, person 2 has to choose one of two options, A or B.

You are informed the payoff structure as follows:

(1). If person 1's decision results in X, person 1 and person 2 will each get 10 Yuan.
(2). If person 1's decision results in Y and person 2 chooses A, person 1 and person 2 will each get 15 Yuan.
(3). If person 1's decision results in Y, and person 2 chooses B, person 2 will get 22 Yuan and person 1 will get 8 Yuan.



(Fig. 2)

In this experiment, all of you who participate in today's experiments will act as ***Person 2***, and you will have to make ***one decision***. Each of your counterparts will be randomly selected from the next experiment among today's participants, and you have no idea about the true identity of your counterparts. On your computer screen (Figure 2) you will see the payoff table, and will be asked to answer the following question:
" Which Option A or B, do you choose in case Person 1 chooses Y? "

To enter your decision, press the button corresponding to your choice (Option A or Option B). Later, we will randomly match your decision with the decision made by **Person 1 in the next experiment**, ultimately determining both payoffs. You will be informed of the results only **at the end of Part 2** of this study.

Comprehension Questions:
Please answer following questions. Raise your hand if you need help. The instructor will come to help you and will check your answers when you are done.
1) If your counterpart chooses X, and you choose A, how much are you paid? _____ your counterpart?_____
2) If your counterpart chooses X, and you choose B, how much are you paid?

_____ your counterpart?_____

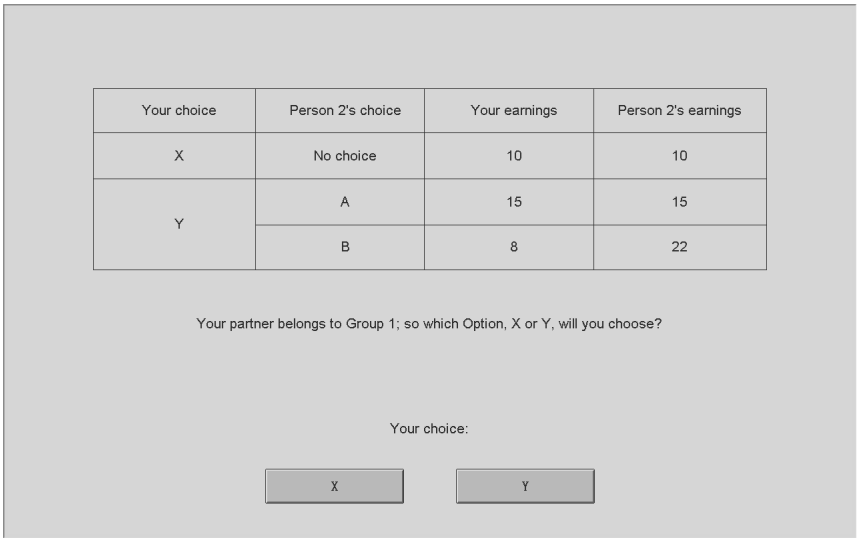
3) If your counterpart chooses Y, and you choose A, how much are you paid? _____ your counterpart?_____

4) If your counterpart chooses Y, and you choose B, how much are you paid? _____ your counterpart?_____

Now you may make your choice!

Instructions for Experiment 3 (Baseline Treatment)

How the study is conducted. In Experiment 3, you can earn money depending on the decisions that you and other participants will take. Your earning for this experiment will be summed with the earnings from the other experiments, and will be privately paid to you in cash.



(Fig.3)

What the task is about. In Experiment 3, you will face the same situation as in Experiment 2, but this time you act as **Person 1**, and play the game for **20 periods**. For each period, on your computer screen (Figure 3), you will see the payoff table, and will be asked to answer the following question:

"Your partner belongs to Group 1 (or 2), so which Option, X or Y, will you choose? "

To enter your decision, press the button corresponding to your choice (Option X or Option Y). 1), if you choose Option X, then the actual outcome of this period is complete (e.g. you will receive 10 Yuan, and person 2 will receive 10 Yuan);

2), if you choose Y, then the actual outcome of the period depends on the decision made by person 2: if person 2 chooses Option A, you will receive 15 Yuan, and person 2 will receive 15 Yuan; if person 2 chooses option B, you will receive 8 Yuan, and person 2 will receive 22 Yuan. Please check the details from the above payoff table.

For each period, you will be randomly matched with a counterpart (Person 2). Your counterpart will be one of the 20 participants in the today's experiments. You can also be possibly matched with yourself. The decision of Person 2 corresponds to the choice made by your counterpart in **Experiment 2** of this study. After this experiment, one period will be randomly selected by computer to determine the real payoffs for you and your counterpart in the previous experiment.

Please remember, for each period you are informed about the actual decision of your counterpart (i.e. person 2) immediately after you decide whether or not to choose option X (or Y), i.e. you will receive the "feedback" from your counterpart regardless of your own choice.

How your counterpart (i.e. person 2) is selected. We have randomly formed two groups of Person 2s based on the decisions made in **Experiment 2**. In one of the groups, at least 50% of the people chose Option A, while in the other group, less than 50% of the people chose Option A. We labeled these two groups "Group 1" and "Group 2". However, you do not know which is the group where at least 50% of the people chose Option A, and which is the group where less than 50% of people chose Option A.

For each period, you will be randomly matched with a person 2 picked from the Group 1 or Group 2. Before making your decision, you only know whether the current counterpart belongs to the Group 1 or the Group 2.

At the very end of this experiment, you will be asked the following question:
"Remember that there are two groups of person 2: in one of the groups, at least 50% of the people choose Option A, other choose Option B; in the other, less than 50% choose Option A. So, which group is the one where at least 50% choose Option A: Group 1 or Group 2?"
Once your guess is correct, you will win 10 Yuan and be informed this result at the very end of this study.

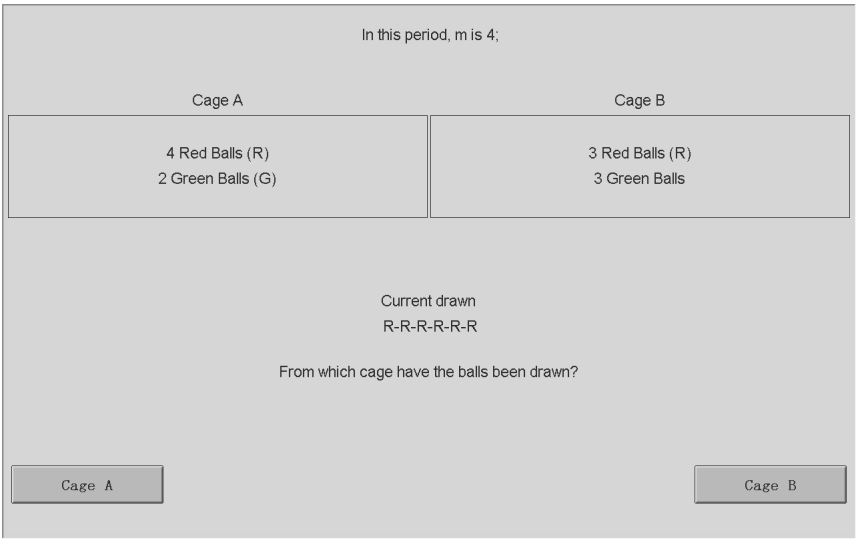
Comprehension Questions:
Please answer following questions. Raise your hand if you need help. The instructor will come to help you and will check your answers when you are done.
1), As Person 1 in this experiment, you would read the similar above table (Figure 3) in your screen. In this table, do you know which group your partner comes from?
(A) from Group 1 (B) from Group 2

2), How many people in Group 1 chose Option A?
(A) at least 50% (B) less than 50% (C) We cannot know it till now
3), If your partner chose A, and you chose X, how much are you paid? _____ your partner?_____

Now you may make your choice for the first period!

Instructions for Experiment 4

In Experiment 4, you will play for 20 periods, and will earn money depending on the decisions you make in each period. On your screen, you will see two "bingo cages", labeled **cage A, and cage B (Figure 4)**.



（Fig. 4）

At the beginning of each period, the computer will randomly draw an integer number which can be equal to 1, 2, 3, 4, 5, or 6, with equal probability, to determine whether cage A or cage B will be used for the task: if 1 through m (here m would be 2, 3, or 4 and announced later) is drawn, we use cage A for the task; otherwise we use cage B. **Please remember,** cage A contains six balls, four Red balls (R) and two Green balls (G) (i.e. **4R & 2G**); cage B also contains six balls, three Red balls (R) and three Green balls (G) (i.e. **3R & 3G**).

What the task is about. For each period, the computer first randomly selects an integer number (from 1 to 6), and then a specific value of m (m=2, 3, or 4) is randomly selected and announced, thus determining which cage (A or B) will be used. The result of this draw is not revealed to you. If cage A (or B) is selected, six draws (with replacement) from this cage are performed, and the results (e.g. GGRRGG) are displayed on your screen. You can record the outcomes of the draws on your record sheet if you wish.

Your decision for each period: which is the cage (A or B) that you believe used to generate the observations on your screen? You will not receive any feedback about the correctness of your responses until the end of the last period.

What the payment is about. After Experiment 4, we will count the number of periods in which your guess is right, and then rank the participants based on their scores. The participant with the highest score will be awarded 200 Yuan at the end of this study (if there are more than one participants who get the same highest score, the 200 Yuan prize will be shared equally among the winners.)

Comprehension Questions:

Please answer following questions. Raise your hand if you need help. The instructor will come to help you and will check your answers when you are done.

1), Please read carefully the above table (Figure 4) which may appear on your screen. Suppose m=4, what is the probability that cage A is selected?

(A) 2/6 (B) 3/6 (C) 4/6

2), If cage A is selected, what is the probability that a Red is drawn?

(A) 3/6 (B) 4/6

3), If cage B is selected, what is the probability that a Red ball is drawn?

(A) 3/6 (B) 4/6

Please make your decision seriously for each period, thank you!

END!

Appendix 3

The Pre-programmed Assignment Rule

Possible Groups	Possible number of trustworthy subjects in the session	The number of trustworthy subjects assigned to group A	The number of untrustworthy subjects assigned to group A
Groups	0	0	0
Groups	1	1	1
Groups	2	2	2
Groups	3	2	2
Groups	4	3	3
Groups	5	4	4
Groups	6	5	4
Groups	7	5	4
Groups	8	6	4
Groups	9	7	3
Groups	10	7	3
Groups	11	7	3
Groups	12	8	2
Groups	13	9	2
Groups	14	10	1
Groups	15	12	1
Groups	16	14	1
Groups	17	16	1
Groups	18	17	0
Groups	19	10	0
Groups	20	20	0

Notes: In each session, the number of trustworthy subjects could appear unexpectedly. For each possible number, we randomly assign certain number of trustworthy subjects into group A (or B), and in the meanwhile, we also randomly select certain number of untrustworthy subjects to supplement group A (or B) to make sure that at least 50% of trustees in group A honored trust while in group B strictly less than 50% did it.

Appendix 4

Final allocation of trustees across sessions

Session 1:

Trustee type	High Group	Low Group
Untrustworthy trustee	2	15
Trustworthy trustee	2	1

Session 2:

Trustee type	High Group	Low Group
Untrustworthy trustee	4	11
Trustworthy trustee	4	1

Session 3:

Trustee type	High Group	Low Group
Untrustworthy trustee	4	8
Trustworthy trustee	6	2

Session 4:

Trustee type	High Group	Low Group
Untrustworthy trustee	4	8
Trustworthy trustee	6	2

Session 5:

Trustee type	High Group	Low Group
Untrustworthy trustee	4	11
Trustworthy trustee	4	1

Session 6:

Trustee type	High Group	Low Group
Untrustworthy trustee	4	9
Trustworthy trustee	5	2

Session 7:

Trustee type	High Group	Low Group
Untrustworthy trustee	4	10
Trustworthy trustee	5	1

Session 8:

Trustee type	High Group	Low Group
Untrustworthy trustee	4	9
Trustworthy trustee	5	2

Appendix 5

Individual characteristics

Individual characteristics	H-type subjects (N=94)	L-type subjects (N=66)	Nonparametric test
Male	50%	45.5%	p>0.571[1]
Age	20.4	20.3	p=0.602[2]
Han	92.6	98.5%	p=0.090*[1]
Student leader	47.9%	31.8%	p=0.042**[1]
CPC member	25.5%	18.2%	p=0.273[1]
Financial aid	21.3%	9.1%	p=0.040**[1]
Urban citizen	43.6%	48.5%	p=0.543[1]
Single-child	53.2%	63.6%	p=0.188[1]
Working experience	63.8%	62.1%	p=0.825[1]
Economic experimental experience	63.8%	50%	p=0.171[1]
Psychological experimental experience	64.9%	63.6%	p=0.870[1]
Home income index	2.9	2.8	p=0.582[2]
Economics background	9.6%	9.1%	p=0.918[1]
High cognitive-ability	84%	77.3%	p=0.280[1]
Self control index	6.4	5.9	p=0.146[2]
Altruism index	7.9	7.4	p=0.059*[2]

Notes: "Male" equals 1 for those who are male subject; "Age" indicates subjects' age; "Han" equals 1 for those who are from Han ethnicity; "Student leader" indicates whether subjects are student leader in the university; "CPC member" indicates whether subjects are the member of Communist Party of China; "Financial aid" indicates that subjects get financial aid from the university; "Urban citizen" indicates that subjects are urban citizens; "Single-child" indicates whether subjects are the single child in their family; "Working experience" indicates whether subjects have part-time working experience; "Economic experimental experience" indicates whether subjects have some experience on economic experiments; "Psychological experimental experience" indicates whether subjects have some experience on psychological experiments; "Home income index" indicates the level of subjects' household income (the highest level is 5); "Economics background" indicates whether subjects are from the economics-related department; "High cognitive-ability" indicates whether subjects are high cognitive able; "Self control index" indicates subjects' answer to the Question 32 in Appendix 1; "Altruism index" indicates subjects' answer to the Question 33 in Appendix 1. [1]A two-tailed chi-square test, [2]A two-tailed Wilcoxon rank-sum test. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Concluding Remarks

Not many transactions are carried out simultaneously. Absent repeated interaction or binding contracts, the standard law and economic models predict that transacting parties would not be able to trade. To facilitate mutually advantageous exchanges, formal legal institutions and social norms are two common approaches to mitigate commitment problems. Besides legal and non-legal mechanisms aiming to restrain promisors' temptation to renege on their promises, individual characteristics, particularly psychological states, are also crucial for establishing trust, because they can affect promisees' reliance investments that have specific value in contractual relationships. This dissertation investigates how law, informal institutions and psychological factors affect transacting behavior. In particular, it first examines the question of how legal and non-legal mechanisms interact to sustain economics exchanges. It then investigates how social norms emerge in commercial communities where contracting parties lack the protection of formal legal institutions, and whether the endogenous adoption of an informal institution can achieve economic governance and social order. It finally studies how subtle psychological factors influence individuals' transacting behavior in a commercial environment where parties cannot rely on either long-term relationships or legal mechanisms to deter opportunism. I have tackled these pressing subquestions of the research agenda in three separate Chapters.

The first Chapter has addressed the question of how formal contract enforcement affects norms of good conduct, such as trust and trustworthiness. Based on a review of relevant empirical studies in the literature on macroeconomics, inter-firm cooperation and laboratory experiments, it can be concluded that formal legal mechanisms, especially formal contracts backed by a powerful authority, normally work as substitutes for trust, rather than complements, except when they are perceived as legitimate, or when there are no strong social norms of fairness (i.e. the population in a society is considerably heterogeneous), or when the environment in which repeated commercial relationships take place becomes highly uncertain.

These insights are very relevant to real-life legal issues. In order to encourage more welfare enhancing transactions, most legislators engage in designing the optimal legal remedies for contract breach. Their attention usually focuses on economic considerations, particularly cost-and benefit analysis. In contract cases where performance is non-verifiable by judges or unobservable by the contracting parties, trust-based commercial relationships become highly important, and legislators have to

take into account the possible effects of non-legal mechanisms when they enforce the sanctions for contract breach. In line with this argument, Cooter (1996) finds that when Judge Posner decided to make legal judgments, he also concerned himself with the prevailing norms widely shared by the contracting parties.

Another field where legislators have currently a concern for trust, is in online e-commerce. A common idea both in Europe and in the U.S. is that, due to a lack of trust by consumers, e-transactions are inhibited and inefficient. To enhance consumers' trust in distance sales, right-to-withdraw clauses are widely adopted by firms. Borges and Irlenbusch (2007) show experimentally that the adoption of withdrawal rights by firms signals their trustworthiness to potential consumers, and thereby increases trust in distance sales. Other pro-consumer mechanisms, such as impartial third-party payment systems, are also established to enhance the levels of trust in online transactions.

The second Chapter has examined whether people are willing to endogenously adopt a collective punishment institution when they lack the protection of an effective legal system, and whether the endogenous adoption of collective punishment mechanism can help a society coordinate an efficient outcome, characterized by high levels of trust and trustworthiness. The experimental results suggest that the introduction of collective punishment induces a significant increase in the levels of trustworthiness, and to a lesser extent also of trust. The endogenous introduction of the mechanism by means of a majority-voting rule does not significantly improve coordination on the efficient equilibrium. Not all participants seem to be able to anticipate the effectiveness of the mechanism, and a majority of them vote against it. In addition, this chapter also shows that participants with higher cognitive abilities and with a background in statistics are more likely to vote in favor of the mechanism.

Most law-and-norm scholars usually focus on the "close-knit" group, "a social network whose members have credible and reciprocal prospects for the application of power against one another and a good supply of information on past and present internal events" (Ellickson, 1991, p.181), when they study how informal norms or institutions govern individual behavior. Similarly, this chapter uses a small-size group where the information about cheating is publicly disclosed by the experimenter to group members and the severity of collective punishment depends on average behavior in the group. An obvious next step would be to explore whether social sanction institutions are still effective in large groups where group members' individual characteristics and beliefs are much more diverse.

The chapter has also shown that highly intelligent subjects who are able to anticipate the effectiveness of the collective punishment mechanism are more likely to vote in favor of the mechanism, which is consistent with the Ellickson's theory on "the Market for Social Norms": norm entrepreneurs (i.e. opinion leaders or activists) with an extraordinary level of social intelligence are "particularly suited to providing

the new rule and eager to have it adopted" (Ellickson, 2001, p.2); appreciative observers follow the norm after they learn the potential profitability of it. Therefore, the recommendation for lawmakers is to provide legal support to norm entrepreneurs when they are struggling to establish an effective norm or transform an inefficient norm. The enactment of the Civil Rights Acts of the 1960s in the U.S. provides a good example, showing that a good law can help norm entrepreneurs to change inefficient norms.

The third Chapter has explored whether high-trustors adapt their behavior in response to others' trustworthiness or untrustworthiness more quickly, which in turn supports them to maintain higher default expectations of others' trustworthiness relative to low-trustors. Our experimental results reveal that both high- and low-trustors are able to learn whom to trust over time, and that high-trustors are better than low-trustors at predicting others' trustworthiness not because they are better at processing the trustworthiness-related information, or that they deliberately collect differentiating social data through trusting more, but only because they are less susceptible to the anticipated aversive emotions aroused by the potential betrayal and thereby have a higher willingness to acquire the valuable information about their partner's actions.

These findings have two main important implications. First, they clearly point to the relevance of betrayal aversion – a psychological factor that has recently received significant attention in the law and economics literature. Traditional legal scholars argue that contracting parties can use contract enforcement to encourage performance of contract obligation and to threaten contractually specified damages against the breaching party for non-performance, leaving no space for psychological factors. Besides the protection of formal contracts, this chapter shows that intrinsic psychological states – the anticipated aversive emotions aroused by the potential betrayal – can also significantly influence individuals' contracting behavior and the recovery of trust after a deliberate breach of contract.

Second, the chapter has shown that when the information about the potential partners is unconditionally provided to subjects, both high- and low-trustors are able to correctly differentiate their trusting behavior according to their partner's trustworthiness, and they perform similarly in predicting others' trustworthiness. It implies that psychological factors play a minor role in a competitive market where the information about cheating flows freely and is accessible to every market participant. Therefore, lawmakers should mandatorily disclose the necessary information on the outcome of transactions and make the commercial environment transparent in order to overcome the problems generated by betrayal aversion.

Reference:

Borges, G., and Irlenbusch, B., 2007, "Fairness Crowed Out by Law: An Experimental Study on Withdrawal Rights," *Journal of Institutional and Theoretical Economics*, 163, 84-101.

Cooter, R., 1996, "Decentralized Law for a Complex Economy: the Structural Approach to Adjudicating the New Law Merchant," *University of Pennsylvania Law Review*, 144, 1643-1696.

Ellickson, R., 1991, *Order without Law: How Neighbors Settle Disputes*, Harvard University Press.

Ellickson, R., 2001, "The Market for Social Norms," *American Law and Economics Review*, 3, 1-49.

Summary

This dissertation has studied how legal and non-legal mechanisms affect the levels of trust and trustworthiness in an economy, and whether and when subtle psychological factors are crucial for establishing trust and even for recovering trust following a breach of contract. I have tackled the most pressing subquestions of this research agenda in three separate Chapters.

The first Chapter has addressed the question of whether formal legal enforcement crowds out or crowds in the amount of trust in a society. Based on a review of relevant empirical studies in the literature on macroeconomics, inter-firm cooperation and laboratory experiments, it can be concluded that formal legal mechanisms, especially formal contracts backed by a powerful authority, normally work as substitutes for trust, rather than complements, except when they are perceived as legitimate, or when there are no strong social norms of fairness (i.e. the population in a society is considerably heterogeneous), or when the environment in which repeated commercial relationships take place becomes highly uncertain.

The second Chapter has examined whether the endogenous adoption of a collective punishment institution can help a society coordinate on an efficient outcome, characterized by high levels of trust and trustworthiness. The experimental results show that the introduction of collective punishment institution induces a significant increase in the levels of trustworthiness, and to a lesser extent also of trust. The endogenous introduction of collective punishment by means of a majority-voting rule does not significantly improve coordination on the efficient equilibrium. Not all subjects seem to be able to anticipate the change in behavior induced by the introduction of the mechanism, and a majority of those who are not able to anticipate, vote against it. Subjects seem to be unable to endogenously adopt a mechanism which, when exogenously imposed, proves to be efficiency enhancing.

The third Chapter has explored whether high-trustors adapt their behavior in response to others' trustworthiness or untrustworthiness more quickly, which in turn supports them to maintain higher default expectations of others' trustworthiness relative to low-trustors. Our experimental results reveal that both high- and low-trustors are able to learn whom to trust over time, and that high-trustors are better than low-trustors at predicting others' trustworthiness not because they are better at processing the trustworthiness-related information, or that they deliberately collect differentiating social data through trusting more, but only because they are less susceptible to the anticipated aversive emotions aroused by the potential betrayal and thereby have a higher willingness to acquire the valuable information about their partner's actions.

Samenvatting

In dit proefschrift is onderzoek gedaan naar de invloed van juridische en niet-juridische mechanismen op de mate van vertrouwen en betrouwbaarheid vanuit economisch perspectief. Daarbij speelt de vraag of en onder welke omstandigheden subtiele psychologische factoren van belang zijn voor het opbouwen van vertrouwen of zelfs voor het herstel van vertrouwen wanneer dat door contractbreuk is geschaad. De belangrijkste subonderzoeksvragen zijn in drie afzonderlijke hoofdstukken behandeld.

In het eerste hoofdstuk wordt antwoord gegeven op de vraag of formele rechtshandhavingregels het in een maatschappij aanwezige vertrouwen verdringen of juist versterken. Hiervoor is gebruik gemaakt van in de literatuur beschikbaar relevant empirisch onderzoek naar macro-economische factoren, samenwerking tussen bedrijven en resultaten van laboratoriumexperimenten. Vastgesteld kan worden dat formele juridische mechanismen, met name wanneer het formele contracten betreffen die worden ondersteund door een krachtige autoriteit, eerder de plaats van het vertrouwen innemen dan dat zij het vertrouwen versterken. Dit is niet het geval wanneer dergelijke mechanismen als legitiem worden ervaren, wanneer er geen krachtige sociale billijkheidsnormen bestaan (dat wil zeggen, wanneer een gemeenschap sterk heterogeen is samengesteld) of wanneer de omgeving waarin bij herhaling commerciële relaties worden aangegaan, in ernstige mate onzeker wordt.

In het tweede hoofdstuk is onderzocht of het endogeen accepteren van een collectieve strafbepaling een maatschappij kan helpen tot een efficiënt resultaat te komen waarin een hoge mate van vertrouwen en betrouwbaarheid kenmerkende factoren zijn. Experimentele resultaten tonen aan dat de introductie van een collectieve strafbepaling een significante stijging van de betrouwbaarheidsniveaus en in mindere mate ook van vertrouwen tot gevolg heeft. De endogene introductie van collectief straffen op basis van het meerderheidsbeginsel leidt niet tot een aanmerkelijke verbetering in het realiseren van een efficiënt evenwicht. Een deel van de proefpersonen lijkt niet in staat op de gedragsverandering die door de introductie van het mechanisme teweeg wordt gebracht, te anticiperen en een meerderheid daarvan stemt tegen een dergelijke introductie. Proefpersonen lijken niet in staat endogeen een mechanisme te accepteren dat aantoonbaar efficiënt is wanneer het exogeen wordt opgelegd.

In het derde hoofdstuk staat de vraag centraal of zogenaamde high-trustors, personen die makkelijker een ander vertrouwen, hun gedrag sneller aanpassen aan de betrouwbaarheid of onbetrouwbaarheid van anderen en daardoor per definitie hogere verwachtingen hebben ten aanzien van de betrouwbaarheid van anderen dan low-trustors. Experimentele resultaten laten zien dat zowel high-trustors als low-trustors in staat zijn te leren wie ze kunnen vertrouwen. High-trustors zijn beter dan low-trustors in staat de betrouwbaarheid van anderen te voorspellen, niet omdat

zij beter zijn in het verwerken van betrouwbaarheid gerelateerde informatie of omdat ze bewust differentiërende sociale gegevens verzamelen doordat ze meer vertrouwen hebben, maar alleen omdat ze minder gevoelig zijn voor de geanticipeerde negatieve emoties van potentieel verraad en daardoor eerder bereid zijn de waardevolle informatie over de acties van hun partners te vergaren.