

# KIER DISCUSSION PAPER SERIES

## KYOTO INSTITUTE OF ECONOMIC RESEARCH

Discussion Paper No.1001

“The Cultural Transmission of  
Trust and Trustworthiness”

Akira Okada

September, 2018



KYOTO UNIVERSITY

KYOTO, JAPAN

# The Cultural Transmission of Trust and Trustworthiness

Akira Okada\*

September, 2018

## Abstract

We consider the cultural transmission of trust and trustworthiness in a trust game with spatial matching á la Tabellini. Players are assumed to enjoy psychological benefits from good conducts. The equilibrium probability that an investor trusts a receiver is a monotonically decreasing function of social distance, and the one that the receiver behaves in a trustworthy manner is non-monotonic. Parents with imperfect empathy transmit their own values to their children through education, and the ratio of individuals with good values globally converges to a stationary point with heterogeneity if educational costs are sufficiently small. Trust and trustworthiness are influenced by institutions in different ways. A better “intermediate” enforcement crowds out trust and crowds in trustworthiness.

*JEL classification:* C72, D02, D64, D91

Key words: crowding effect, cultural transmission, random matching game, social distance, trust, trustworthiness

---

\*Institute of Economic Research, Kyoto University, Yoshida-Honmachi, Sakyo-Ku, Kyoto 606-8501, JAPAN. E-mail: aokada@kier.kyoto-u.ac.jp.

# 1 Introduction

Trust and trustworthiness underlie human relationships in almost all aspects. The two concepts are widely studied in various fields including social psychology, sociology, management science, political science, biology, and economics. During the last two decades, trust research has grown rapidly in economics beyond traditional approach of rational choice under selfish preference. Many empirical studies show that trust plays important roles in economic performance: economic growth (Knack and Keefer, 1997), government effectiveness (La Porta et al. 1997), financial development (Guiso et al. 2004), and international trade and investment (Guiso et al. 2009). The trust game (Berg et al., 1995) is employed as a standard framework for the study of trust and trustworthiness as the prisoner's dilemma is for that of cooperation. Given that trust and trustworthiness are fundamental for better economic outcomes, important questions are how the values of trust and trustworthiness are inherited between generations,<sup>1</sup> and how one can increase them in a society. The aim of this paper is to consider cultural transmission of trust and trustworthiness and their relationship to institution with external enforcement.

Recent empirical studies indicate the relevancy of cultural transmission of trust in economics. Dohmen et al. (2012) provide empirical supports for intergenerational transmission of risk and trust attitudes. Based on the data from the 2003 and 2004 waves of the German Socio-Economic Panel Study, they find that these attitudes are strongly positively correlated between parents and children, after controlling personal and environmental characteristics. They also find a positive impact of prevailing attitudes in local environments on child attitudes. Their empirical findings indicate that socialization plays an important role in the transmission process beyond a genetic process.<sup>2</sup> Guiso

---

<sup>1</sup>Guiso et al. (2008) consider intergenerational transmission of priors about the trustworthiness of others.

<sup>2</sup>There exists a considerable volume of the literature on genetic transmission of trust. For example, Kosfeld et al. (2005) provide neurobiological evidence of trust using oxytocin. Cesarini et al. (2008) find that genetic differences play a significant role for behavior in the

et al.(2006, p.23) define culture as “those customary beliefs and values that ethnic, religious, and social groups transmit fairly unchanged from generation to generation,” and document a strong effect of ethnic origin on trust which an American has toward others. Using data on bilateral trust between European countries, Guiso et al.(2009) document that the bilateral trust is affected by cultural aspects such as the history of conflicts of two countries and their religious similarities, and that lower bilateral trust leads to less trade and less portfolio and direct investment. They write: “trust among European countries differs in systematic ways, which are correlated to their different cultural heritages” (Guiso et al. 2009, p.1128). Tabellini (2008a) documents that the value of generalized morality has widely diffused in societies that were ruled by non-despotic political institutions in the distant past.

In a seminal paper, Bisin and Verdier (2001) present a model of intergenerational cultural transmission which incorporates two channels of transmission, “direct vertical” socialization inside family and “oblique” socialization by society. They show that if the two types of transmissions are “culturally substitute” so that parents have less incentives to educate their kids the more widely dominant are their types in the population, then the dynamics of cultural transmission has a globally stable stationary point with heterogeneous types. Focusing on the vertical socialization, Tabellini (2008b) explicitly formulates parents’ educational efforts to transmit the good values to their children. In the prisoner’s dilemma with spatial random matching, he proves that “good” parents with a higher level of generalized morality make positive efforts to educate their children and “bad” parents with a lower level of it do not. Based on the parents’ optimal educational decision, he further proves that a better local enforcement which applies to transactions between closely “connected” agents in small groups and communities hurts the diffusion of good values, while a better legal enforcement enhances it.

In this paper, we extend the analysis of cultural transmission by Tabellini

---

trust game experiment played by monozygotic and dizygotic twins.

(2008b) in the prisoner’s dilemma to the trust game of Berg et al. (1995). We do this exercise by the following reasons. Tabellini derives several implications of his model of cooperation, and relates them to many empirical studies on trust and trustworthiness in the literature. Although trust and trustworthiness may be broadly interpreted as cooperative behavior, they are not exactly identical to cooperation. Cooperation is the outcome of trust and trustworthiness. Moreover, many empirical studies show that trust and trustworthiness are different notions (Glaeser et al. 2000, Buchan et al. 2002, Cox 2004, Ashraf et al. 2006, Bohnet and Baytelman 2007, Kiyonari et al. 2006, Ben-Ner et al. 2010, for example). The prisoner’s dilemma and the trust game have different incentive structures. We cannot automatically expect that the same properties hold in both games.

We summarize the results of this paper. The equilibrium configurations are more complex in the (asymmetric) trust game with spatial matching than in the (symmetric) prisoner’s dilemma. While the equilibrium characterised by Tabellini (2008b) has three phases (all players cooperate, only good types do, and no players do), depending on social connection between players, the equilibrium of the trust game has nine phases in general, due to the possibility that players choose mixed strategies. While the equilibrium probability that an investor (trustor) trusts a receiver (trustee) is a monotonically decreasing function of a social distance between the two players, the one that a receiver behaves in a trustworthy manner is non-monotonic. Although the values of trust and trustworthiness are crowded out by better local enforcements which apply to transactions between closely connected players, these values are influenced in different manners by institutions. A better “intermediate” enforcement enhances trustworthiness, but deteriorates trust. In the following sections, we discuss empirical studies which are consistent with our results.

The remainder of the paper is organized as follows. Section 2 presents the trust game with spatial matching. Section 3 characterizes an equilibrium of the game where players enjoy psychological benefits from good conducts. Section

4 examines the cultural transmission of trust and trustworthiness. Section 5 considers the effects of institution on them. Sections 6 has concluding remarks. All proofs are given in Appendix.

## 2 The Model

We consider a trust game introduced by Berg et al. (1995).<sup>3</sup> There are two types of players called an investor (player 1 and “she”) and a receiver (player 2 and “he”). A continuum of each type of players is uniformly distributed on the circumference of a circle of size  $2S$ . The maximum distance of two players is  $S$ . Each player is randomly matched with the other type of players at distance  $y$  with a probability density function  $g(y) > 0$ .<sup>4</sup> The probability that a pair of an investor and a receiver is matched between 0 and  $S$  is equal to one, that is,  $\int_0^S g(y)dy = 1$ . Distance  $y$  could be not only geographical one but also socioeconomic differences. Examples include kinship, ethnicity, language, religion, and class. A larger circle with higher  $S$  implies a more heterogeneous society along a relevant spatial dimension.

Two matched players observe their distance and play the game as follows. An investor (trustor) decides whether she invests her endowment  $\omega > 0$  or not. If she invests, then the investment earns a return  $\gamma\omega$  ( $\gamma > 1$ ). A receiver (trustee) decides how to share  $\gamma\omega$  with the investor. Specifically, he decides whether he returns the amount  $\beta\gamma\omega$  ( $0 < \beta < 1$ ) to the investor, or not. If he returns it, the investor receives material payoff  $\beta\gamma\omega$  and himself does  $(1-\beta)\gamma\omega$ . If not, the receiver exploits the total amount  $\gamma\omega$ . If the investment is zero, then the game immediately stops and the investor’s payoff is  $\omega$  and that of the receiver is zero. The material payoffs of two players are summarized in Table 1.<sup>5</sup>

---

<sup>3</sup>While Berg et al.(1995) give the name of an investment game, we call it a trust game, following the literature.

<sup>4</sup>This model of random matching over a circle is originally considered by Dixit (2004) in the context of the repeated prisoner’s dilemma.

<sup>5</sup>In the experiment of Berg et al. (1995), parameters are such that  $\omega = 10$  and  $\gamma = 3$ .

In Table 1, each player has two actions  $C$  and  $N$  where  $C$  and  $N$  mean “invest” and “not invest” for the investor (row player), respectively, and “repay” and “not repay” for the receiver (column player), respectively. It is natural to interpret the investment as the trust that an investor places in a receiver, and the repayment as a trustworthy action of the receiver (Glaeser et al. 2000).<sup>6</sup> In what follows, parameters  $\omega$ ,  $\gamma$  and  $\beta$  are fixed. We assume that  $\beta\gamma > 1$  so that an investor is better off by repayment than in the case of no investment.

	$C$	$N$
$C$	$\beta\gamma\omega, (1 - \beta)\gamma\omega$	$0, \gamma\omega$
$N$	$\omega, 0$	$\omega, 0$

Table 1. Material payoffs in the trust game

The trust game in Table 1 has only the Nash equilibria where the investor chooses  $N$  and the receiver may choose  $N$  with any probability not less than  $1 - 1/\beta\gamma$ . Thus, the maximization of material payoffs cannot explain positive investment and repayment which have been observed in most experiments of one-shot trust games. Recent experimental studies show that social preferences play a key role in trusting behavior (Fehr 2009). Sapienza et al. (2013)

---

Choice variables are continuous. An investor chooses any investment amount  $x$  with  $0 \leq x \leq \omega$  and a receiver does any amount  $y$  with  $0 \leq y \leq \gamma\omega$  for return. A large volume of experimental studies shows that, on average, investors send about 50 percent and receivers repay about 95 percent of what has been invested (Camerer 2003). Besides the original version of the trust game, binary versions of it are frequently used in experimental studies (Bohnet and Zeckhauser 2004, and Ermisch and Gambetta 2010, for example). The result of this paper is not affected in any critical manner if an investor has an option to invest a fixed portion of the endowment  $\omega$ , or nothing.

<sup>6</sup>There, however, is no consensus on the proper definition of trust in the vast literature of trust research (Fehr 2009). Recent literature scrutinizes whether a positive investment in the trust game appropriately measures the trust of an investor, and whether a positive repayment does the trustworthiness of a receiver. See Bohnet and Zeckhauser (2004), Cox (2004), Ashraf et al. (2006), Kiyonari et al. (2006), Fehr (2009), Ben-Ner and Halldorsson (2010), Sapienza et al. (2013) among others.

argue that the quantity sent in the trust game is a combination of two components: preferences (risk aversion, inequality aversion, altruism) and beliefs in receivers' trustworthiness. By cross-cultural experiments in Russia, South Africa and the United States, Ashraf et al. (2006) show that trust is based on beliefs on trustworthiness and on unconditional kindness, and that trustworthiness is related to unconditional kindness and reciprocity.

In view of these results, we assume that individuals are motivated not only by material payoffs but also by the value of good conducts. Specifically, following Tabellini (2008b), we assume that besides the material payoffs, each player  $i (= 1, 2)$  enjoys a noneconomic (psychological) benefit  $d_i$  whenever she or he plays  $C$ , irrespective of the opponent's action. The noneconomic benefit of player  $i$  decays with distance at exponential rate  $\theta_i > 0$ . Player  $i$  receives noneconomic benefit  $d_i e^{-\theta_i y}$  by choosing  $C$  in a match with distance  $y$ .

The formulation above of individuals' motivation reflects the view that the intrinsic value of trust and trustworthiness depend on social connection. Within a circle of more familiar individuals, the norm of trust and trustworthiness can be applied with stronger force, but it becomes weaker among less familiar individuals. Coleman (1990) discusses that familiarity breeds trust. Individuals trust others more familiar to themselves than those unfamiliar. Based on the General Social Survey (GSS) for the United States in the years 1974-94, Alesina and La Ferrara (2002) find that interpersonal trust is lower in more racially heterogeneous communities and in those with higher income inequality. They argue that the effect of heterogeneity on trust is largely due to the fact that individuals trust those more similar to themselves. Hoffman et al, (1996) provide experimental evidence that decreasing social distance increases other-regarding behavior in dictator games.

For parameter values, we assume:

**Assumption 1.**  $d_1 > \omega, d_2 > \beta\gamma\omega > \omega$ .



This assumption means that when two individuals play the trust game at zero distance ( $y = 0$ ),  $C$  is the dominant action of every player and thus the trust game has a unique Nash equilibrium  $(C, C)$ .

There are two types of each player, a “good” type and a “bad” type. They differ in their rates at which noneconomic benefit of choosing  $C$  decays with distance. For  $i = 1, 2$ , let  $\theta_i^g$  be the rate of player  $i$ 's good type and  $\theta_i^b$  be that of bad type. We assume that  $0 < \theta_i^g < \theta_i^b$  for every  $i = 1, 2$ . In the population of each player  $i (= 1, 2)$ , let  $n_i (0 < n_i < 1)$  denote the fraction of good types at any point in the circle. When individuals play the trust game, they observe distance  $y$  but not the types of their partners. In the next section, we will characterize a Nash equilibrium of the trust game with random matching when the fractions  $n_i$  are fixed for every  $i = 1, 2$ .

### 3 Equilibrium

For every player  $i (= 1, 2)$ , a (mixed) strategy for type  $j (= g, b)$  is a function  $\sigma_i^j$  from  $[0, S]$  to  $[0, 1]$  where  $\sigma_i^j(y)$  is the probability that player  $i$  of type  $j$  may choose  $C$  in a match with distance  $y$ .

We first consider the behavior of an investor. Let  $\pi_2(y)$  be the probability that a receiver may choose  $C$  in a match with distance  $y$ . It holds that  $\pi_2(y) = n_2\sigma_2^g(y) + (1 - n_2)\sigma_2^b(y)$ . If the investor of type  $j (= g, b)$  chooses  $C$ , she obtains the benefit  $\beta\gamma\omega\pi_2(y) + d_1e^{-\theta_1^j y}$ , and otherwise she obtains the benefit  $\omega$ . Thus, the investor of type  $j$  is indifferent between playing  $C$  and  $N$  if

$$\beta\gamma\omega\pi_2(y) + d_1e^{-\theta_1^j y} = \omega. \quad (3.1)$$

(3.1) is rewritten as

$$y = T_1^j(\pi_2(y)) \equiv \frac{1}{\theta_1^j} \ln \frac{d_1}{\omega(1 - \beta\gamma\pi_2(y))}. \quad (3.2)$$

An investor of type  $j$  chooses  $C$  against a receiver at distance  $y$  if  $y < T_1^j(\pi_2(y))$ ,

and  $N$  if  $y > T_1^j(\pi_2(y))$ .<sup>7</sup> Given  $\pi_2$ , the value  $T_1^j(\pi_2(y))$  represents the threshold of distance such that an investor of type  $j$  chooses  $C$  in a match if the distance  $y$  is less than it. We call  $T_1^j(\pi_2(y))$  the *trust threshold* for an investor of type  $j$ . A bad type has the smaller trust threshold than a good type since the bad type's noneconomic benefit for playing  $C$  decays more rapidly than that of the good type as the distance becomes greater.

We next consider the behavior of a receiver. Let  $\pi_1(y)$  be the probability that an investor may choose  $C$  in a match with distance  $y$ . Similarly to  $\pi_2(y)$ , it holds that  $\pi_1(y) = n_1\sigma_1^g(y) + (1 - n_1)\sigma_1^b(y)$ . If the receiver of type  $j (= g, b)$  chooses  $C$ , he obtains the benefit  $(1 - \beta)\gamma\omega\pi_1(y) + d_2e^{-\theta_2^j y}$ , and otherwise he obtains the benefit  $\gamma\omega\pi_1(y)$ . Thus, the receiver of type  $j$  is indifferent between playing  $C$  and  $N$  if

$$(1 - \beta)\gamma\omega\pi_1(y) + d_2e^{-\theta_2^j y} = \gamma\omega\pi_1(y). \quad (3.3)$$

(3.3) is rewritten as

$$y = T_2^j(\pi_1(y)) \equiv \frac{1}{\theta_2^j} \ln \frac{d_2}{\beta\gamma\omega\pi_1(y)}. \quad (3.4)$$

A receiver of type  $j$  chooses  $C$  against an investor at distance  $y$  if  $y < T_2^j(\pi_1(y))$ , and  $N$  if  $y > T_2^j(\pi_1(y))$ .<sup>8</sup> Similarly to the case of an investor, we call  $T_2^j(\pi_1(y))$  the *trustworthiness threshold* for a receiver of type  $j$ . A receiver of type  $j$  chooses  $C$  against an investor if the distance between two players is less than the threshold  $T_2^j(\pi_1(y))$ . A bad type has the smaller trust threshold than a good type.

It is useful to remark here that the trust game (without or with noneconomic benefits) has a strategic property different from the prisoner's dilemma

---

<sup>7</sup>If  $1 - \beta\gamma\pi_2(y) < 0$ , we set  $T_1^j(\pi_2(y)) = +\infty$  for every  $j$ . In this case, an investor of every type chooses  $C$  against a receiver at any distance.

<sup>8</sup>If  $\pi_1(y) = 0$ , we set  $T_2^j(\pi_1(y)) = +\infty$  for every  $j$ . In this case, a receiver of every type chooses  $C$  against an investor at any distance.

considered in Tabellini (2008b). The investor's net expected material gain from playing  $N$  rather than  $C$  is given by  $\omega(1 - \beta\gamma\pi_2(y))$ , which is strictly decreasing in  $\pi_2(y)$ . The investor's incentive to cooperate is greater if the opponent is more likely to cooperate. On the other hand, the receiver's net expected material gain from playing  $N$  rather than  $C$  is given by  $\beta\gamma\omega\pi_1(y)$ , which is strictly increasing in  $\pi_1(y)$ . The receiver's incentive to cooperate is smaller if the opponent is more likely to cooperate. These facts show that the trust game has neither strategic complementarity nor strategic substitutes while the prisoner's dilemma has strategic complementarity.

We are now ready to characterize a Nash equilibrium of the trust game with noneconomic benefits. For  $i, k = 1, 2 (i \neq k)$ , the fact that  $T_i^b(\pi_k(y)) < T_i^g(\pi_k(y))$  for any value of  $\pi_k(y)$  implies that whenever a bad type of player  $i$  chooses  $C$  with positive probability, a good type of player  $i$  chooses  $C$  with probability one. Thus, in a pure-strategy equilibrium, the probability  $\pi_i(y)$  may take three values, 0,  $n_i$  and 1. The three values correspond to the cases that (1) no types cooperate, (2) only the good types cooperate, and (3) all types cooperate, respectively. In a mixed-strategy equilibrium, there are two possible cases,  $0 < \pi_i(y) < n_i$  and  $n_i < \pi_i(y) < 1$ . In the first case, good types may cooperate with positive probability and bad types do not. In the second case, good types cooperate and bad types may with positive probability.

Corresponding to three possible values, 0,  $n_i$  and 1, of the equilibrium probability  $\pi_i(y)$  of player  $i$ 's choosing  $C$ , we have the following thresholds for player  $i$ .<sup>9</sup>

$$T_1^j(0) = \frac{1}{\theta_1^j} \ln \frac{d_1}{\omega} \quad T_1^j(n_2) = \frac{1}{\theta_1^j} \ln \frac{d_1}{\omega(1 - \beta\gamma n_2)} \quad T_1^j(1) = +\infty \quad (3.5)$$

$$T_2^j(0) = +\infty \quad T_2^j(n_1) = \frac{1}{\theta_2^j} \ln \frac{d_2}{\beta\gamma\omega n_1} \quad T_2^j(1) = \frac{1}{\theta_2^j} \ln \frac{d_2}{\beta\gamma\omega} \quad (3.6)$$

The trust thresholds of an investor satisfy  $T_1^j(0) < T_1^j(n_2) \leq T_1^j(1)$ . The

---

<sup>9</sup>See footnotes 7 and 8. If  $1 - \beta\gamma n_2 < 0$ , then we set  $T_1^j(n_2) = +\infty$ .

trust threshold increases as a receiver cooperates more likely. She cooperates against a receiver at any distance if the receiver also cooperates with probability one. Without knowing a partner's type, an investor bears the risk of trusting a receiver (possibly untrustworthy). Investors trust receivers at greater distance as the partners are trustworthy more likely. The intermediate threshold  $T_1^j(n_2)$  rises if the ratio of repayment ( $\beta$ ) from receivers rises, or if the return of investment ( $\gamma$ ) rises, or if the fraction of good types ( $n_2$ ) in the population of receivers rises.

The receiver's trustworthiness thresholds satisfy the opposite inequalities,  $T_2^j(1) < T_2^j(n_1) \leq T_2^j(0)$ . The threshold decreases as the investor cooperates more likely. As we have noted, the receiver's net expected material gain  $\beta\gamma\omega\pi_1(y)$  from playing  $N$  rather than  $C$  increases as the investor trusts him more likely. In order to overcome the higher temptation to cheat, the receiver should be matched with an investor more nearby so that the norm of good conduct (trustworthiness) can be applied with the greater force. The receiver cooperates in a match with any distance if the partner does not invest. This follows from the investor-receiver relationship that the receiver's material payoff is constant at zero if the investor chooses  $N$ , while he enjoys noneconomic benefit of trustworthiness no matter how distant matches are. The intermediate threshold  $T_2^j(n_1)$  of the receiver depends on parameters in the opposite way to the case of an investor. Namely,  $T_2^j(n_1)$  rises if the ratio of repayment ( $\beta$ ) from receivers falls, or if the return of investment ( $\gamma$ ) falls, or if the fraction of good types ( $n_2$ ) in the population of receivers falls.

**Theorem 1.** There exists a unique equilibrium in the trust game with random matching. The equilibrium probabilities  $\pi_1(y)$  and  $\pi_2(y)$  of two players'

choosing  $C$  are characterized as follows.

1.  $\pi_1(y) = 1, \pi_2(y) = 1$  for  $0 \leq y \leq T_2^b(1)$
2.  $\pi_1(y) = 1, \pi_2(y) = n_2$  for  $T_2^b(1) < y \leq \min(T_1^b(n_2), T_2^g(1))$
3.  $\pi_1(y) = 1, \pi_2(y) = 0$  for  $T_2^g(1) < y \leq T_1^b(0)$
4.  $n_1 < \pi_1(y) < 1, 0 < \pi_2(y) < n_2$  for  $\max(T_1^b(0), T_2^g(1)) < y < \min(T_1^b(n_2), T_2^g(n_1))$
5.  $n_1 < \pi_1(y) < 1, n_2 < \pi_2(y) < 1$  for  $\max(T_1^b(n_2), T_2^g(1)) < y < T_2^b(n_1)$
6.  $\pi_1(y) = n_1, \pi_2(y) = n_2$  for  $\max(T_1^b(n_2), T_2^b(n_1)) < y \leq \min(T_1^g(n_2), T_2^g(n_1))$
7.  $\pi_1(y) = n_1, \pi_2(y) = 0$  for  $\max(T_1^b(0), T_2^g(n_1)) < y \leq T_1^g(0)$
8.  $0 < \pi_1(y) < n_1, 0 < \pi_2(y) < n_2$  for  $\max(T_1^g(0), T_2^g(n_1)) < y < T_1^g(n_2)$
9.  $0 < \pi_1(y) < n_1, n_2 < \pi_2(y) < 1$  for  $\max(T_1^g(n_2), T_2^b(n_1)) < y \leq S$ .

The theorem shows that the equilibrium configuration of the trust game is more complicated than that of the prisoner's dilemma considered by Tabellini (2008b).<sup>10</sup> In the prisoner's dilemma, the (Pareto superior) equilibrium has only three phases. With two thresholds,  $Y^0$  and  $Y^1$ , for a bad type and for a good type, both types cooperate in a match with distance  $y$  if  $0 \leq y \leq Y^0$ , and only good types cooperate if  $Y^0 < y \leq Y^1$  and no types cooperate if  $Y^1 < y \leq S$ . In contrast, the equilibrium of the trust game has nine phases.

The reason for the complexity of equilibrium configurations in the trust game is that investors and receivers have different strategic incentives. In the prisoner's dilemma, two players have the same incentives. Different incentives of investors and receivers lead to a cycle of their best responses when the noneconomic benefits of trust and trustworthiness are small. Investors are willing to trust receivers if they are trustworthy. On the other hand, receivers are tempted to cheat investors' trust. Anticipating this, it is optimal for in-

---

<sup>10</sup>The prisoner's dilemma game with random matching has multiple equilibria due to strategic complementarity. Tabellini (2008b) confines attention to the Pareto superior equilibrium that sustains the maximal cooperation.

vestors not to trust them. Then, receivers are motivated to be trustworthy if they enjoy noneconomic benefit to be trustworthy, no matter how small it is. Due to the cycle of best responses, a mixed-strategy equilibrium prevails in the trust game if noneconomic benefits are not very large. Actually, it always holds that some type of at least one player may cooperate with a positive probability.

The equilibrium behavior of investors and receivers are summarized as follows. First, when social distance is sufficiently small, investors trust receivers, who reciprocate their trust (case 1). As in Tabellini (2008b), this property is the direct result of the assumption that the closer their matches are, the higher noneconomic benefits individuals enjoy from trust and trustworthiness. This case corresponds to a local community where every individual strongly internalizes the norm of trust and trustworthiness. Second, the probability of investors' cooperation (trusting) is monotonically decreasing in social distance, as in the prisoner's dilemma. Good types of investors cooperate until the distance  $T_1^g(0)$  (cases 1-7), and beyond that they play mixed strategies. Bad types of investors cooperate over a smaller range until the distance  $T_1^b(0)$  (cases 1-3), and they play mixed strategies until  $T_2^b(n_1)$  (cases 4 and 5), and beyond that they do not cooperate. Third, the probability of receivers' cooperation (trustworthiness) is not monotonic in social distance. There are four phases: monotonically decreasing (cases 1-3), monotonically increasing (cases 4 and 5), monotonically decreasing (cases 6 and 7), and monotonically increasing (cases 8 and 9). Fourth, as social distance becomes large beyond the local community, bad types of receivers are the first who behave non-cooperatively (case 2). When distance becomes greater, good types of receivers do not cooperate, either (case 3). In this case, all types of investors trust untrustworthy receivers since their noneconomic benefits of trusting motivate them to do so. Fifth, when social distance is intermediate (cases 4 and 5), bad types of investors are motivated not to trust their partners. As a result, a mixed equilibrium prevails. Finally, if social distance is very large (cases 8 and 9),

a mixed equilibrium prevails again. Bad types of investors never trust their partners, and only good types of them may cooperate.

To make our analysis transparent, we place some restrictions on model parameters. In Theorem 1, we rule out two cases 3 and 7 that no types of receivers act in a trustworthy manner. Namely, we assume that  $T_1^b(0) < T_2^g(1)$  and  $T_1^g(0) < T_2^g(n_1)$  for every  $0 < n_1 < 1$ . Since  $T_2^g(1) < T_2^g(n_1)$  for every  $n_1$ , these conditions are reduced to  $T_1^g(0) < T_2^g(1)$ , which is rewritten as

$$(A1) \quad \frac{\ln d^1 - \ln \omega}{\ln d^2 - \ln \beta\gamma\omega} < \frac{\theta_1^g}{\theta_2^g}.$$

Under (A1), the minimum threshold for a good investor is smaller than that for a good receiver. We also assume that  $T_1^j(n_2) = \infty$  for every  $j = g, b$ ,<sup>11</sup> equivalently:

$$(A2) \quad n_2 > \frac{1}{\beta\gamma}.$$

When only the good receivers are trustworthy, every type  $j$  of the investor earns material payoff  $\beta\gamma\omega n_2$  if she trusts a receiver, and  $\omega$  otherwise. Thus, under (A2), the optimal choice of an investor is to trust her partner, regardless of a distance. Intuitively, if the fraction of good receivers is sufficiently large, then every type of an investor is willing to trust a receiver, no matter how far they are connected. (A2) rules out cases 5, 6 and 9 in Theorem 1.

The next proposition is a corollary of Theorem 1. For simplicity of analysis, we will assume (A1) and (A2) in what follows.<sup>12</sup>

**Proposition 1.** Under (A1) and (A2), a unique equilibrium of the trust game

---

<sup>11</sup>See footnote 7.

<sup>12</sup>The result of the paper holds true without these assumptions.

with random matching is characterized as follows.

1.  $\pi_1(y) = 1, \pi_2(y) = 1$  for  $0 \leq y \leq T_2^b(1)$
2.  $\pi_1(y) = 1, \pi_2(y) = n_2$  for  $T_2^b(1) < y \leq T_2^g(1)$
3.  $n_1 < \pi_1(y) < 1, 0 < \pi_2(y) < n_2$  for  $T_2^g(1) < y < T_2^g(n_1)$
4.  $0 < \pi_1(y) < n_1, 0 < \pi_2(y) < n_2$  for  $T_2^g(n_1) < y \leq S$

The equilibrium in the proposition has a simple structure. A bad investor cooperates until the threshold  $T_2^g(1)$  and a good investor does further until  $T_2^g(n_1)$ . A bad receiver cooperates until the threshold  $T_2^b(1)$ , and he never cooperates beyond the threshold. A good receiver cooperates in matches with longer distances until  $T_2^g(1)$ , and he may cooperate with a positive probability in every match beyond the threshold. If a match becomes sufficiently far, only good types of investors and receivers may cooperate with positive probability. Interestingly, the equilibrium configuration in Proposition 1 depends only on trustworthiness thresholds for receivers.

Recent experimental studies of trust games support the prediction of Theorem 1 and Proposition 1 that trust and trustworthiness can be maintained more likely among socially closer partners. Glaeser et al. (2000) report experimental evidence that when individuals are socially closer (in the number of friends in common and the duration of their acquaintanceship), both levels of trust and trustworthiness rise. Specifically, trustworthiness declines when partners are of different races or nationality, even after controlling social connection. In their experiments, eleven out of the twelve times with no money returned, partners were of different races. Employing an economically relevant measure of social distance, Etang et al. (2011) conduct trust game experiments in two villages in rural Cameroon. They found that significantly more money was sent when players lived in the same village, although levels of trust and trustworthiness towards the other village were still high. Gender, education and membership



of rotating credit groups also influenced transfers. Many experiment studies in social psychology show that even the ad hoc categorization (minimal-group condition) of individuals affect their cooperative behavior, in general. In cross-national experiments of the trust game, Buchan et al. (2002) show that the level of trust is higher among “neighbors” than among “strangers” where the group identity of neighborhood is experimentally manipulated.

## 4 Cultural Transmission

In this section, we consider how the proportion  $n_i$  of good investors and good receivers change through cultural transmission of the values of trust and trustworthiness. Following recent theoretical studies of cultural transmission (Bisin and Verdier 2001 and Tabellini 2008b), we assume that parents rationally choose values to transmit to their children. Parents evaluate their children’s actions with their own preferences. Under this assumption of “imperfect empathy” (Bisin and Verdier 2001), parents attempt to socialize their children to their own preference trait. Dohmen et al. (2012) provide empirical evidence for transmission of risk and trust attitudes from parents to children.

The trust game in the last section is played over infinitely many periods. Each period has two generations, kids and parents, in two populations of investors and receivers. In the first period of their life, kids are educated by their parents. The education increases the probability that a kid becomes a good type. After the education, kids play the trust game, observing their own types. In the second period of their life, each player becomes a parent of a single kid and devotes effort to education. Education is costly.

To analyse the dynamics of cultural transmission, we first consider the parents’ educational choice where their kids behave according to the equilibrium of Proposition 1.

For each player  $k = 1, 2$  and each type  $i = g, b$ , let  $\sigma_k^i(y)$  be the probability that a kid of type  $i$  cooperates in a match with distance  $y$ , and  $\pi_k(y)$  be the

probability that a kid cooperates at distance  $y$ . In the population of player 1 (investors), consider a parent of type  $i$  who has a kid of type  $j$  in period  $t = 1, 2, \dots$ . Let  $V_{1,t}^{ij}$  denote the parent's evaluation of the kid's expected utility. We assume that  $V_{1,t}^{ij}$  is given by

$$V_{1,t}^{ij} = \int_0^S \{u_1^j(\sigma_1^j(z), \pi_2(z)) + d^1 \sigma_1^j(z) e^{-\theta^i z}\} g(z) dz \quad (4.7)$$

where  $u_1^j(\sigma_1^j(z), \pi_2(z))$  is the material payoff that a kid of type  $j$  receives in a match with distance  $z$  when a partner may cooperate with probability  $\pi_2(z)$ . The parent's evaluation  $V_{1,t}^{ij}$  implies that the parent cares about the kid's material payoff in the same way as the kid, but that the parent values the good conduct (trust) in her own way, possibly different from the kid's. For player 2 (receivers), the type  $i$  parent's evaluation,  $V_{2,t}^{ij}$ , of the type  $j$  kid's expected utility is defined in the same way as (4.7).

Types of kids depend on the effort of their parents in every population  $i (= 1, 2)$  as follows. If the parents devote a positive effort  $f_i$ , then their kids turn out to be good types with probability  $\delta_i + f_i$  and bad types with probability  $1 - \delta_i - f_i$ . Note that the probability that a kid becomes a particular type does not depend on the types of their parents. It depends only on the natural rate  $\delta_i$  of transmission and on the effort level  $f_i$  of the parents. This formulation of cultural transmission is adapted from Tabellini (2008b). The rule of cultural transmission implies that parents are motivated to educate their kids if they evaluate higher a good kid than a bad one.

The next proposition shows the parent's educational choice.

**Proposition 2.** For every population  $i = 1, 2$ , a good parent is motivated to devote a positive effort to educating a kid. A bad parent exerts no effort.

Proposition 2 reproduces the result of Tabellini (2008b, Lemma 2) in our context of the trust game. Good parents evaluate good kids higher than bad

ones, and thus they are motivated to educate their kids. On the other hand, bad parents evaluate bad kids higher than good ones, and thus they exert no effort to educating their kids. While the proposition compares the expected evaluations of parents for having kids of different types over the whole range  $[0, S]$ , it actually shows that parents weakly prefer having kids of the same types as theirs in every match in  $[0, S]$  and do strictly so in some matches.

The intuition for the result can be explained as follows. From their own value judgements, parents evaluate the equilibrium outcome of the trust game in Proposition 1. Parents are indifferent between having different types of kids if their actions are identical (cases 1 and 2 for investors and case 1 for receivers). Consider first the population of investors. When social distance is in a middle range (case 3), a good kid cooperates, and a bad kid plays a mixed strategy. Thus, a good kid prefers to cooperate rather than playing the mixed strategy for the bad kid. Since a good parent has the same value as the good kid, she prefers to have a good kid. On the other hand, a bad kid is indifferent between trusting and not trusting her partner. Since a bad parent has the same value as the bad kid, she is indifferent between having a good or bad kid. When social distance is large (case 4), a good kid plays a mixed strategy and a bad kid does not trust her partner. The good kid is indifferent between trusting and not trusting the partner. Since a good parent has the same value as the good kid, she is indifferent to a type of her kid. On the other hand, the bad kid prefers not to trust the partner rather than playing the mixed strategy for the good kid. Since a bad parent has the same value as the bad kid, she prefers to have a bad kid rather than a good kid who plays a mixed strategy.

Consider next a parent in the population of receivers. When social distance is in a middle range (case 2), a good kid sends a positive money and a bad kid does not. Since a good parent has the same value as the good kid, she prefers to have a good kid. By the same reason, a bad parent prefers to have a bad kid. When social distance becomes larger (cases 3 and 4), a good kid plays a

mixed strategy and a bad kid prefers not to behave in a trustworthy manner. The good kid is indifferent between sending and not sending a positive money. Since a good parent has the same value as the good kid, he is indifferent to a type of his kid. The bad kid prefers not to send a positive money rather than playing the mixed strategy for the good kid. Since a bad parent has the same value as the bad kid, he prefers to have a bad kid rather than a good kid who plays a mixed strategy.

We now consider the good parent's optimal effort level for education. Following Tabellini (2008b), we assume that the parent incurs quadratic costs  $(1/2\varphi)f^2$  where  $f > 0$  is an educational effort. In every population  $i = 1, 2$ , let  $V_i^{gj}$  be the good parent's evaluation of the type  $j$  kid's expected utility defined by (4.7). The parent chooses an effort  $f_i$  to maximize her or his expected evaluation

$$(\delta_i + f_i)V_i^{gg} + (1 - \delta_i - f_i)V_i^{gb} - (1/2\varphi)f_i^2,$$

given the effort chosen by all other parents. The first-order optimality condition is given by

$$f_i/\varphi = V_i^{gg} - V_i^{gb}. \quad (4.8)$$

For  $i = 1$ , it follows from Proposition 1 that

$$\begin{aligned} f_1/\varphi &= \int_{T_2^g(1)}^{T_2^g(n_1)} \{\beta\gamma\omega\pi_2(z) - \omega + d^1 e^{-\theta_1^g z}\} g(z) dz \\ &= d^1 \int_{T_2^g(1)}^{T_2^g(n_1)} (e^{-\theta_1^g z} - e^{-\theta_1^b z}) g(z) dz. \end{aligned} \quad (4.9)$$

In the range  $[T_2^g(n_1), S]$ , a good kid plays a mixed strategy, and thus she is indifferent between choosing  $C$  and  $N$ . Since a bad kid chooses  $N$ , this implies that a good parent is indifferent to a type of her kid. In the range  $[T_2^g(1), T_2^g(n_1)]$ , the mixed strategy equilibrium condition for a bad kid implies that  $\beta\gamma\omega\pi_2(z) - \omega = -d^1 e^{-\theta_1^b z}$ . The parent's optimal effort is determined by the difference of her evaluations between having good and bad kids in the

intermediate range  $[T_2^g(1), T_2^g(n_1)]$  where both types of kids behave differently. It follows from (4.9) that the parent's optimal effort  $f_1$  is a function of the fraction  $n_1$  of good kids, and we denote it by  $f_1(n_1)$ .<sup>13</sup>

Since  $T_2^g(n_1)$  is a decreasing function of  $n_1$ , the optimal effort  $f_1(n_1)$  is decreasing in  $n_1$ . This means that the trust game does not have the strategic complementarity with respect to parents' education discussed in Tabellini (2008b). This property of the trust game is due to the interaction between different populations of investors and receivers, which is missing in the prisoner's dilemma. If the ratio  $n_1$  of good kids increases in the population of investors, then receivers have more incentive to cheating, and thus their trustworthiness threshold  $T_2^g(n_1)$  decreases. In turn, the range  $[T_2^g(1), T_2^g(n_1)]$  becomes smaller and it makes investor parents to appreciate less the trusting behavior of their kids. As a result, they devote less effort  $f_1(n_1)$  to education.

Similarly, for  $i = 2$  we have

$$\begin{aligned} f_2/\varphi &= \int_{T_2^b(1)}^{T_2^g(1)} \{(1 - \beta)\gamma\omega + d_2e^{-\theta_2^g z} - \gamma\omega\}g(z)dz \\ &= \int_{T_2^b(1)}^{T_2^g(1)} (d_2e^{-\theta_2^g z} - \beta\gamma\omega)g(z)dz. \end{aligned} \quad (4.10)$$

The receiver parent's optimal effort  $f_2$  is determined by the difference of his evaluations between having good and bad kids in the intermediate range  $[T_2^b(1), T_2^g(1)]$  where both types of kids behave differently. It does not depend on the fraction  $n_i$  of good kids in either population under (A1) and (A2).

Given the optimal efforts (4.9) and (4.10) of parents, a kid of a good parent turns out to be a good type with probability  $\delta_i + f_i(n_i)$ ,<sup>14</sup> while that of a bad parent becomes a good type with probability  $\delta_i$ . By the Law of Large Numbers,

---

<sup>13</sup>In general, the parent's optimal effort level may depend on fractions  $n_1, n_2$  of good players in both populations in the equilibrium characterized by Theorem 1.

<sup>14</sup>Cost parameter  $\varphi$  is assumed to be sufficiently small so that  $0 < \delta_i + f_i(n_i) < 1$ .

the fraction of good kids in each population  $i = 1, 2$  in period  $t$  is given as

$$n_t = (\delta_i + f_i(n_t))n_{t-1} + \delta_i(1 - n_{t-1}) = \delta_i + n_{t-1}f_i(n_t). \quad (4.11)$$

We now obtain the following proposition.

**Proposition 3.** For sufficiently small  $\varphi$ , the cultural transmission (4.11) in population  $i (= 1, 2)$  has a unique stationary point  $n_i^*$  which is globally stable. The stationary point satisfies  $n_i^* = \frac{\delta_i}{1 - f_i(n_i^*)}$ .

Good parents devote positive efforts to educating their kids. Starting from an initial situation where there are few good individuals in the populations of investors and receivers, the fractions of good types monotonically increase and converge to the stationary points. Higher efforts of parents increase the stationary levels of trust and trustworthiness.<sup>15</sup>

## 5 Trust and Institution

Tabellini (2008b) investigates whether good values of cooperation are crowded in or out by better external enforcement, based on his model of the prisoner's dilemma with random matching. The quality of external enforcements is formulated as the probability of cheating being detected. Comparing two different, local and legal, enforcements, which are applied to neighboring and distant matches, respectively, he argues that better legal enforcement contributes the diffusion of good values, but that better local enforcement destroys good values. The reason for this result is that good players cooperate over a larger range of matches under better legal enforcement, and thus parents devote more efforts

---

<sup>15</sup>Replacing imperfect empathy with full altruism, Tabellini (2008b) considers the cultural transmission by parents who care about kids' welfare evaluated by the kids themselves. He shows that all (good and bad) parents devote positive efforts to educating their kids since the expected utility of a good kid is always higher than that of a bad kid. It can be shown that the same result holds in the trust game.

to education. In contrast, when local enforcement is strengthened, the threshold of cooperation for bad players increases, and thus the range of matches where bad and good players behave differently shrinks. As a result, parents put less efforts to education. In this section, we examine whether or not these insights of Tabellini (2008b) can be extended to trust and trustworthiness.

Recently, many empirical studies of trust show that trust is composed of social preferences and beliefs in reciprocal behavior (Ashraf et al. 2006, Fehr 2009, Ben-Ner and Halldorsson 2010, and Sapienza et al. 2013, for example). Better enforcements positively affect people's belief in others' reciprocal behavior. But, how about social preferences? If external intervention through monetary rewards or punishments undermines intrinsic motivation as the motivation crowding theory suggests,<sup>16</sup> better enforcement may hurt trust and trustworthiness, and may have a long run effect on social, economic, and political outcomes through cultural transmission.

We formalize external enforcements in the trust game as follows. If an investor cooperates and a receiver cheats in a match of distance  $y$ , then cheating may be detected with probability  $p(y)$ . In the case of detection, the receiver is enforced to implement a contract, that is, to repay the  $\beta$  amount of the return  $\gamma\omega$  to the investor. If the cheating is not detected, the two players receive payoffs in Table 1. Note that enforcement is relevant only for a receiver. The quality of external enforcement is captured by the detection probability  $p(y)$ . For an action pair  $(C, N)$ , the investor obtains expected payoff  $\beta\gamma p(y)\omega$ , and the receiver does  $(1 - \beta p(y))\gamma\omega$ . The payoff matrix under enforcement is given in Table 2 where  $p = p(y)$ . We assume that  $\beta\gamma p < 1$  so that the trust games in Tables 1 and 2 have the same strategic properties.

---

<sup>16</sup>Frey and Jegen (2001) provide an excellent survey on empirical evidences supporting the existence of motivation crowding effects in a wide variety of social, economic and political situations.

	$C$	$N$
$C$	$\beta\gamma\omega, (1-\beta)\gamma\omega$	$\beta\gamma p\omega, (1-\beta p)\gamma\omega$
$N$	$\omega, 0$	$\omega, 0$

Table 2 Material payoffs under enforcement

An investor of type  $j$  is indifferent between playing  $C$  and  $N$  if

$$\beta\gamma\omega\{\pi_2(y) + (1 - \pi_2(y))p(y)\} + d_1e^{-\theta_1^j y} = \omega \quad (5.12)$$

where  $\pi_2(y)$  is the probability that a receiver cooperates. A receiver of type  $j$  is indifferent between playing  $C$  or  $N$  if

$$(1 - \beta)\gamma\omega\pi_1(y) + d_2e^{-\theta_2^j y} = (1 - \beta p(y))\gamma\omega\pi_1(y) \quad (5.13)$$

where  $\pi_1(y)$  is the probability that an investor cooperates in a match with distance  $y$ .

From (5.12) and (5.13), the three kinds of thresholds (3.5) and (3.6) for two players are changed as follows. Let  $p = p(y)$ .

$$T_{1p}^j(0) = \frac{1}{\theta_1^j} \ln \frac{d_1}{\omega(1 - \beta\gamma p)}, T_{1p}^j(n_2) = \frac{1}{\theta_1^j} \ln \frac{d_1}{\omega(1 - \beta\gamma n_2^p)}, T_{1p}^j(1) = \infty \quad (5.14)$$

where  $n_2^p = n_2 + (1 - n_2)p$ ,<sup>17</sup> and

$$T_{2p}^j(0) = \infty, T_{2p}^j(n_1) = \frac{1}{\theta_2^j} \ln \frac{d_2}{\beta(1-p)\gamma\omega n_1}, T_{2p}^j(1) = \frac{1}{\theta_2^j} \ln \frac{d_2}{\beta(1-p)\gamma\omega}. \quad (5.15)$$

Comparing the two players' thresholds without enforcement in (3.5) and (3.6) and those with enforcement in (5.15) and (5.16), we can see that the introduction of enforcement can expand the ranges of cooperation for them, regardless

<sup>17</sup>If  $1 - \beta\gamma n_2^p < 0$ , then we set  $T_{1p}^j(n_2) = +\infty$ .



of their types.

The equilibrium analysis in the last section holds true under enforcement, replacing the thresholds of two players with those defined by (5.13) and (5.14). In what follows, we consider the same type of equilibrium as in Proposition 1. Note that (A2) implies  $1 - \beta\gamma n_2^p < 0$  and thus  $T_{1p}^j(n_2) = \infty$  in (5.14). (A1) is easily modified under enforcement.

Given the equilibrium configurations in Proposition 1, we consider three levels of enforcements. Let  $y^1$  belong to a right-hand neighborhood of  $T_2^b(1)$ , and  $y^2$  to a right-hand neighborhood of  $T_2^g(1)$ . Three level enforcements  $p^0$ ,  $p^1$ , and  $p^2$  are applied to matches in the ranges  $[0, y^1]$ ,  $(y^1, y^2]$ , and  $(y^2, S]$ , respectively. The parameter  $p^0$  corresponds to local enforcement which is relevant within families, small groups, and communities. The parameter  $p^2$  corresponds to legal or formal enforcement that is more relevant beyond communities where local enforcement is ineffective. The parameter  $p^1$  captures an intermediate enforcement.

The following proposition shows how the optimal efforts of parents are affected by different kinds of enforcement institutions.

**Proposition 4.** The educational effort of an investor parent is constant, monotonically decreases, and monotonically increases as local, intermediate, and legal enforcement increase, respectively. That of a receiver parent monotonically decreases, increases, and is constant as local, intermediate, and legal enforcement increase, respectively.

We obtain two implications of Proposition 4. First, enforcements which are applied to transactions in matches between socially closer partners are more likely to hurt trust and trustworthiness. Those which are applied to more distant matches are more likely to crowd in these values. Second, different institutions influence trust and trustworthiness in different ways. In the population of investors, a good parent is less willing to educate her kid when an

intermediate enforcement increases, and is more willing when legal enforcement increases. In the population of receivers, a good parent is less willing to educate his kid when local enforcement increases, and is more willing when an intermediate enforcement increases.

The intuition for these results can be explained as follows. The introduction of enforcement increases all trustworthiness thresholds for receivers,  $T_2^b(1), T_2^g(1), T_2^g(n_1)$ , which determine the equilibrium behavior of kids. By the definition of the three types of enforcement, the smallest threshold  $T_2^b(1, p^0)$  depends on  $p^0$ , and the intermediate one  $T_2^g(1, p^1)$  does on  $p^1$ , and the largest one  $T_2^g(n_1, p^2)$  does on  $p^2$  (assuming that  $p^0$  and  $p^1$  are not very large). All the thresholds are monotonically increasing functions of the corresponding levels of enforcement. In the population of investors, a good kid trusts a partner in all matches over the range  $[T_2^g(1, p^1), T_2^g(n_1, p^2)]$ , and a bad kid does not. Thus, a good parent evaluates a good kid more than a bad kid over this range. Since the threshold  $T_2^g(1, p^1)$  is increasing in  $p^1$ , the range shrinks and the expected evaluation of the parent for having a good kid decreases if  $p^1$  is strengthened. As a result, the parent is less willing to educate her kid. On the other hand, the increase of legal enforcement  $p^2$  expands the range of trusting, and thus the parent becomes more willing to educate her kid. The same logic can be applied to the population of receivers although the enforcement level enhancing the parent's educational effort is different from the case of investors.

Proposition 4 shows that the insight of Tabellini (2008b) into the relationship between enforcement and the value of cooperation in the prisoner's dilemma can be basically extended to the case of trust and trustworthiness. Better enforcement such as law applied to distant transactions contributes to the diffusion of trust and trustworthiness, and a strong local enforcement applied to a smaller group may hurt them. We, however, should remark that value-enhancing enforcements for trust and trustworthiness are possibly different. For example, according to Proposition 4, the improvement of an intermediate enforcement enhances trustworthiness and deteriorates trust.

Bohnet and Baytelman (2007) provide empirical evidence supporting the results of crowding-out discussed above. With subjects of senior executives, they compare the amounts of money sent and returned in an anonymous one-shot trust game with those under four institutional constraints (pre-play communication, post-play communication, post-play punishment, and repetition). They find that all settings with tighter institutional constraints compared to the one-shot trust game decrease intrinsically motivated trust. In contrast, the institutional constraints do not affect trustworthiness. This finding supports our theoretical result that trust and trustworthiness are crowded out under different types of institution. Frey and Oberholzer-Gee (1997) present a case study of the siting of a noxious facility (repositories for nuclear wastes) in Switzerland, and show that financial compensation reduces the willingness to host the facility due to motivation-crowding out.

Finally, Proposition 4 is consistent with empirical studies on cross-national differences of trust in the literature. Yamagishi and his associates (Yamagishi and Yamagishi 1994, Yamagishi et al. 1998) provide empirical evidence that American people are more trusting other people in general than Japanese people. They argue that this cross-national difference prevails since networks of committed relations play a more prominent role in Japanese society than in American society.<sup>18</sup> General trust is irrelevant under commitment formation and assurance. Proposition 4 provides a theoretical support to their arguments. It is conceivable that Japanese society has a higher level of local enforcement which are applied to closely connected individuals such as families, firms, and communities. Thus, caused by the crowding-out effect shown in Proposition 4, the value of trust has been diffused less in Japanese society than in American society.

---

<sup>18</sup>Yamagishi et al. (1998) write: “strong and stable relations (such as family ties and group ties) promote a sense of security within such relations but endanger trust that extends beyond these relations.”

## 6 Concluding Remarks

We have considered the cultural transmission of trust and trustworthiness when players are intrinsically motivated by these good values. Cultural transmission takes place through socialization by parents' education. Since parents with imperfect empathy evaluate the welfares of their children by their own preferences, only good parents with higher levels of generalized trust evaluate good children more than bad ones, and thus they make positive efforts to educate their children. In a cultural transmission process from parents to children, the ratios of good players globally converge to a stationary point when educational costs are sufficiently small. The strengthening of a more local enforcement which apply to transactions between socially closer partners deteriorates the good values of trust and trustworthiness, and that of a legal enforcement which applies to transactions with far distance enhances them. The strengthened local enforcement reduces the range of transactions where good and bad players behave differently, and as a result, it discourages parents' education. It has a long-run effect on trust and trustworthiness. Institution, however, influences trust and trustworthiness in different ways. A better "intermediate" enforcement crowds out trust, and crowds in trustworthiness.

Finally, institution is exogenously given in our analysis. We have focused on a one-way relation from institution to behavior under the assumption that institution affects the intrinsic motivation of trust and trustworthiness. Tabellini (2008b) extends his analysis to an issue of endogenous institution, considering the opposite relation from behavior to institution. Viewing that the quality of distant matches is determined by a political process, he argues that politics and culture interact with mutually reinforcing effects. Further analysis of interactions between trust and politics remains for future research.

## Appendix

**Proof of Theorem 1.** For every  $i = 1, 2$  and  $j = g, b$ , let  $\sigma_i^j(y)$  be the equilibrium probability that player  $i$  of type  $j$  chooses  $C$  in a match at distance  $y$ . Also, let  $\pi_i(y)$  be the equilibrium probability that player  $i$  chooses  $C$  in a match at distance  $y$ . It holds that  $\pi_i(y) = n_i \sigma_i^g(y) + (1 - n_i) \sigma_i^b(y)$ .

The best response of player 1 of type  $j = g, b$  is given as

$$\begin{aligned} \sigma_1^j(y) &= 1 \quad \text{if } y < T_1^j(\pi_2(y)) \\ 0 \leq \sigma_1^j(y) &\leq 1 \quad \text{if } y = T_1^j(\pi_2(y)) \\ \sigma_1^j(y) &= 0 \quad \text{if } y > T_1^j(\pi_2(y)) \end{aligned} \tag{A.1}$$

where  $T_1^j(\pi_2(y))$  is defined by (3.2) and (3.5). The best response of player 2 of type  $j = g, b$  is given as

$$\begin{aligned} \sigma_2^j(y) &= 1 \quad \text{if } y < T_2^j(\pi_1(y)) \\ 0 \leq \sigma_2^j(y) &\leq 1 \quad \text{if } y = T_2^j(\pi_1(y)) \\ \sigma_2^j(y) &= 0 \quad \text{if } y > T_2^j(\pi_1(y)) \end{aligned} \tag{A.2}$$

where  $T_2^j(\pi_1(y))$  is defined by (3.4) and (3.6). Keeping  $\pi_2(y)$  fixed, we have  $T_1^b(\pi_2(y)) < T_1^g(\pi_2(y))$ . This implies that if  $0 < \sigma_1^b(y)$ , then  $\sigma_1^g(y) = 1$ . Thus, there are three possible cases,  $\pi_1(y) = 0, n_1, 1$ , in a pure-strategy equilibrium, and two possible cases,  $0 < \pi_1(y) < n_1, n_1 < \pi_1(y) < 1$ , in a mixed-strategy equilibrium. By the same reason, there are three possible cases,  $\pi_2(y) = 0, n_2, 1$ , in a pure-strategy equilibrium, and two possible cases,  $0 < \pi_2(y) < n_2, n_2 < \pi_2(y) < 1$ , in a mixed-strategy equilibrium.

We will check the equilibrium conditions in all 13 ( $=3 \times 3 + 2 \times 2$ ) possible cases.

case 1.  $\pi_1(y) = 1, \pi_2(y) = 1$  :

Since  $T_1^j(1) = +\infty$  for every  $j = g, b$ ,  $\pi_2(y) = 1$  and (A.1) imply  $\sigma_1^j(y) = 1$  and thus  $\pi_1(y) = 1$  for every  $y$ . By (A.2),  $\sigma_2^j(y) = 1$  is a best response to

$\pi_1(y) = 1$  if  $y \leq T_2^j(1)$ . Thus,  $(\pi_1(y) = 1, \pi_2(y) = 1)$  is an equilibrium if  $y \leq T_2^b(1)$ .

case 2.  $\pi_1(y) = n_1, \pi_2(y) = 1$  :

By the same reason as in case 1,  $\pi_2(y) = 1$  must imply that  $\pi_1(y) = 1$ . A contradiction. Thus,  $(\pi_1(y) = n_1, \pi_2(y) = 1)$  is not an equilibrium for any  $y$ .

case 3.  $\pi_1(y) = 0, \pi_2(y) = 1$  :

By the same reason as in case 2,  $(\pi_1(y) = 0, \pi_2(y) = 1)$  is not an equilibrium for any  $y$ .

case 4.  $\pi_1(y) = 1, \pi_2(y) = n_2$  :

By (A.1),  $\sigma_1^j(y) = 1$  is a best response to  $\pi_2(y) = n_2$  if  $y \leq T_1^j(n_2)$ . By (A.2),  $\sigma_2^j(y) = 1$  is a best response to  $\pi_1(y) = 1$  if  $y \leq T_2^j(1)$ . Thus,  $(\pi_1(y) = 1, \pi_2(y) = n_2)$  is an equilibrium if  $T_2^b(1) < y \leq \min(T_1^b(n_2), T_2^g(1))$ .

case 5.  $\pi_1(y) = 1, \pi_2(y) = 0$  :

By (A.1),  $\sigma_1^j(y) = 1$  is a best response to  $\pi_2(y) = 0$  if  $y \leq T_1^j(0)$ . By (A.2),  $\sigma_2^j(y) = 1$  is a best response to  $\pi_1(y) = 1$  if  $y \leq T_2^j(1)$ . Thus,  $(\pi_1(y) = 1, \pi_2(y) = 0)$  is an equilibrium if  $T_2^g(1) < y \leq T_1^b(0)$ .

case 6.  $\pi_1(y) = n_1, \pi_2(y) = n_2$  :

By (A.2),  $\sigma_1^j(y) = 1$  is a best response to  $\pi_2(y) = n_2$  if  $y \leq T_1^j(n_2)$ . By (A.2),  $\sigma_2^j(y) = 1$  is a best response to  $\pi_1(y) = n_1$  if  $y \leq T_2^j(n_1)$ . Thus,  $(\pi_1(y) = n_1, \pi_2(y) = n_2)$  is an equilibrium if  $\max(T_1^b(n_2), T_2^b(n_1)) < y \leq \min(T_1^g(n_2), T_2^g(n_1))$ .

case 7.  $\pi_1(y) = 0, \pi_2(y) = n_2$  :

Since  $T_2^j(0) = +\infty$  for every  $j = g, b$ ,  $\pi_1(y) = 0$  and (A.2) imply  $\sigma_2^j(y) = 1$  and thus  $\pi_2(y) = 1$  for every  $y$ . A contradiction. Thus,  $(\pi_1(y) = 0, \pi_2(y) = n_2)$  is not an equilibrium for any  $y$ .

case 8.  $\pi_1(y) = n_1, \pi_2(y) = 0$  :

By (A.1),  $\sigma_1^j(y) = 1$  is a best response to  $\pi_2(y) = 0$  if  $y \leq T_1^j(0)$ . By (A.2),  $\sigma_2^j(y) = 1$  is a best response to  $\pi_1(y) = n_1$  if  $y \leq T_2^j(n_1)$ . Thus,  $(\pi_1(y) = n_1, \pi_2(y) = 0)$  is an equilibrium if  $\max(T_1^b(0), T_2^g(n_1)) < y \leq T_1^g(0)$ .

case 9.  $\pi_1(y) = 0, \pi_2(y) = 0$  :

Since  $T_2^j(0) = +\infty$  for every  $j = g, b$ ,  $\pi_1(y) = 0$  and (A.2) imply  $\sigma_2^j(y) = 1$  and thus  $\pi_2(y) = 1$  for every  $y$ . A contradiction. Thus,  $(\pi_1(y) = 0, \pi_2(y) = 0)$  is not an equilibrium for any  $y$ .

case 10.  $n_1 < \pi_1(y) < 1, 0 < \pi_2(y) < n_2$  :

It holds that  $0 < \sigma_1^b(y) < 1$  and  $0 < \sigma_2^g(y) < 1$ . By (A.1) and (A.2), we must have

$$y = T_1^b(\pi_2(y)) = \frac{1}{\theta_1^b} \ln \frac{d_1}{\omega(1 - \beta\gamma\pi_2(y))} \quad (\text{A.3})$$

$$y = T_2^g(\pi_1(y)) = \frac{1}{\theta_2^g} \ln \frac{d_2}{\beta\gamma\omega\pi_1(y)}. \quad (\text{A.4})$$

(A.4) solves

$$\pi_1(y) = \frac{d_2}{\beta\gamma\omega} e^{-\theta_2^g y}.$$

Thus,  $n_1 < \pi_1(y) < 1$  is equivalent to that  $T_2^g(1) < y < T_2^g(n_1)$ . (A.3) solves

$$\pi_2(y) = \frac{1}{\beta\gamma} \left(1 - \frac{d_1}{\omega} e^{-\theta_1^b y}\right).$$

Thus,  $0 < \pi_2(y) < n_2$  is equivalent to that  $T_1^b(0) < y < T_1^b(n_2)$ . In sum,  $(n_1 < \pi_1(y) < 1, 0 < \pi_2(y) < n_2)$  is a mixed equilibrium if  $\max(T_1^b(0), T_2^g(1)) < y < \min(T_1^b(n_2), T_2^g(n_1))$ .

case 11.  $n_1 < \pi_1(y) < 1, n_2 < \pi_2(y) < 1$  :

It holds that  $0 < \sigma_1^b(y) < 1$  and  $0 < \sigma_2^b(y) < 1$ . By (A.1) and (A.2), we must have

$$y = T_1^b(\pi_2(y)) = \frac{1}{\theta_1^b} \ln \frac{d_1}{\omega(1 - \beta\gamma\pi_2(y))} \quad (\text{A.5})$$

$$y = T_2^b(\pi_1(y)) = \frac{1}{\theta_2^b} \ln \frac{d_2}{\beta\gamma\omega\pi_1(y)}. \quad (\text{A.6})$$

(A.6) solves

$$\pi_1(y) = \frac{d_2}{\beta\gamma\omega} e^{-\theta_2^b y}.$$

Thus,  $n_1 < \pi_1(y) < 1$  is equivalent to that  $T_2^b(1) < y < T_2^b(n_1)$ . (A.6) solves

$$\pi_2(y) = \frac{1}{\beta\gamma} \left(1 - \frac{d_1}{\omega} e^{-\theta_1^b y}\right).$$

Thus,  $n_2 < \pi_2(y) < 1$  is equivalent to that  $T_1^b(n_2) < y$ . Note that  $\pi_2(y) < 1$  holds for any  $y$  since  $\beta\gamma > 1$ . In sum,  $(n_1 < \pi_1(y) < 1, n_2 < \pi_2(y) < 1)$  is a mixed equilibrium if  $\max(T_1^b(n_2), T_2^g(1)) < y < T_2^b(n_1)$ .

case 12.  $0 < \pi_1(y) < n_1, 0 < \pi_2(y) < n_2$  :

It holds that  $0 < \sigma_1^g(y) < 1$  and  $0 < \sigma_2^g(y) < 1$ . By (A.1) and (A.2), we must have

$$y = T_1^g(\pi_2(y)) = \frac{1}{\theta_1^g} \ln \frac{d_1}{\omega(1 - \beta\gamma\pi_2(y))} \quad (\text{A.7})$$

$$y = T_2^g(\pi_1(y)) = \frac{1}{\theta_2^g} \ln \frac{d_2}{\beta\gamma\omega\pi_1(y)}. \quad (\text{A.8})$$

(A.8) solves

$$\pi_1(y) = \frac{d_2}{\beta\gamma\omega} e^{-\theta_2^g y}.$$

Thus,  $0 < \pi_1(y) < n_1$  is equivalent to that  $T_2^g(n_1) < y$ . Note that  $0 < \pi_1(y)$  holds for any  $y$ . (A.7) solves

$$\pi_2(y) = \frac{1}{\beta\gamma} \left(1 - \frac{d_1}{\omega} e^{-\theta_1^g y}\right).$$

Thus,  $0 < \pi_2(y) < n_2$  is equivalent to that  $T_1^g(0) < y < T_1^g(n_2)$ . In sum,  $(0 < \pi_1(y) < n_1, 0 < \pi_2(y) < n_2)$  is a mixed equilibrium if  $\max(T_1^g(0), T_2^g(n_1)) < y < T_1^g(n_2)$ .

case 13.  $0 < \pi_1(y) < n_1, n_2 < \pi_2(y) < 1$  :

It holds that  $0 < \sigma_1^g(y) < 1$  and  $0 < \sigma_2^b(y) < 1$ . By (A.1) and (A.2), we must have

$$y = T_1^g(\pi_2(y)) = \frac{1}{\theta_1^g} \ln \frac{d_1}{\omega(1 - \beta\gamma\pi_2(y))} \quad (\text{A.9})$$

$$y = T_2^b(\pi_1(y)) = \frac{1}{\theta_2^b} \ln \frac{d_2}{\beta\gamma\omega\pi_1(y)}. \quad (\text{A.10})$$



(A.10) solves

$$\pi_1(y) = \frac{d_2}{\beta\gamma\omega} e^{-\theta_2^b y}.$$

Thus,  $0 < \pi_1(y) < n_1$  is equivalent to that  $T_2^b(n_1) < y$ . Note that  $0 < \pi_1(y)$  holds for any  $y$ . (A.9) solves

$$\pi_2(y) = \frac{1}{\beta\gamma} \left(1 - \frac{d_1}{\omega} e^{-\theta_1^g y}\right).$$

Thus,  $n_2 < \pi_2(y) < 1$  is equivalent to that  $T_1^g(n_2) < y$ . Note that  $\pi_2(y) < 1$  holds for any  $y$  since  $\beta\gamma > 1$ . In sum,  $(0 < \pi_1(y) < n_1, n_2 < \pi_2(y) < 1)$  is a mixed equilibrium if  $\max(T_1^g(n_2), T_2^b(n_1)) < y$ . Q.E.D.

### Proof of Proposition 2.

Consider first the population of investors. In a match with distance  $y$ , let  $\sigma_1^j(y)$  be the probability that a kid of type  $j (= g, b)$  may cooperate, and  $\pi_2(y)$  the probability that a partner may cooperate. Let  $v_1^{kj}(\sigma_1^j(y), \pi_2(y))$  be the expected utility that the type  $j$  kid receives in a match with distance  $y$  where noneconomic benefit of trust is evaluated by the value of a parent of type  $k (= g, b)$ . Namely, we have

$$v_1^{kj}(1, \pi_2(y)) = \beta\gamma\omega\pi_2(y) + d_1 e^{-\theta_1^k y}, \quad v_1^{kj}(0, \pi_2(y)) = \omega.$$

Let  $V_1^{ij}$  be the evaluation of a type  $i$  parent for having a type  $j$  kid (omitting time variable  $t$ ). By (4.7), it holds that

$$V_1^{ij} = \int_0^S v_1^{ij}(\sigma_1^j(y), \pi_2(y)) g(y) dy.$$

Consider a good parent. It follows from Proposition 1 that

$$\begin{aligned} V_1^{gg} - V_1^{gb} &= \int_{T_2^g(1)}^{T_2^g(n_1)} \{v_1^{gg}(1, \pi_2(z)) - v_1^{gb}(\sigma_1^b(z), \pi_2(z))\} g(z) dz \\ &+ \int_{T_2^g(n_1)}^S \{v_1^{gg}(\sigma_1^g(z), \pi_2(z)) - v_1^{gb}(0, \pi_2(z))\} g(z) dz. \quad (\text{A.11}) \end{aligned}$$

Note that the parent is indifferent between having a good kid and a bad kid over the range  $[0, T_2^g(1)]$  since kids' actions are the same.

We will examine the sign of the right hand side in (A.11) term by term. Consider the first term. In equilibrium, a good kid chooses  $\sigma_1^g(z) = 1$ , and thus  $v_1^{gg}(1, \pi_2(z)) > \omega$ . Since a bad kid plays a mixed strategy  $0 < \sigma_1^b(z) < 1$ , it holds that

$$\begin{aligned} v_1^{gb}(\sigma_1^b(z), \pi_2(z)) &= \sigma_1^b(z)v_1^{gg}(1, \pi_2(z)) + (1 - \sigma_1^b(z))\omega \\ &< v_1^{gg}(1, \pi_2(z)). \end{aligned}$$

Therefore, the first term is strictly positive. Consider the second term. A good kid plays a mixed strategy  $0 < \sigma_1^g(z) < 1$ . Since she is indifferent between choosing 0 and 1, it holds that  $v_1^{gg}(\sigma_1^g(z), \pi_2(z)) = \omega$ . Since a bad kid chooses  $\sigma_1^b(z) = 0$ ,  $v_1^{gb}(0, \pi_2(z)) = \omega$ . Therefore, the second term is zero. In total, we have  $V_1^{gg} > V_1^{gb}$ .

For a bad parent, we have

$$\begin{aligned} V_1^{bb} - V_1^{bg} &= \int_{T_2^g(1)}^{T_2^g(n_1)} \{v_1^{bb}(\sigma_1^b(z), \pi_2(z)) - v_1^{bg}(1, \pi_2(z))\}g(z)dz \\ &+ \int_{T_2^g(n_1)}^S \{v_1^{bb}(0, \pi_2(z)) - v_1^{bg}(\sigma_1^g(z), \pi_2(z))\}g(z)dz. \quad (\text{A.12}) \end{aligned}$$

Consider the first term of (A.12). In equilibrium, a bad kid plays a mixed strategy  $0 < \sigma_1^b(z) < 1$ , and thus is indifferent between choosing 0 and 1. This implies that  $v_1^{bb}(\sigma_1^b(z), \pi_2(z)) = v_1^{bb}(1, \pi_2(z)) = v_1^{bb}(0, \pi_2(z)) = \omega$ . By definition, we have  $v_1^{bb}(1, \pi_2(z)) = v_1^{bg}(1, \pi_2(z))$ . Thus, the first term is equal to zero. Consider the second term of (A.12). A good kid plays a mixed strategy  $0 < \sigma_1^g(z) < 1$ , and thus is indifferent between choosing 0 and 1. This implies that

$$v_1^{gg}(\sigma_1^g(z), \pi_2(z)) = \beta\gamma\omega\pi_2(z) + d_1e^{-\theta_1^g z} = \omega.$$

Since  $\theta_1^g < \theta_1^b$ , it holds that

$$\begin{aligned} v_1^{bg}(\sigma_1^g(z), \pi_2(z)) &< \sigma_1^g(z)(\beta\gamma\omega\pi_2(z) + d_1e^{-\theta_1^g z}) + (1 - \sigma_1^g(z))\omega \\ &= \omega \\ &= v_1^{bb}(0, \pi_2(z)). \end{aligned}$$

Therefore, the second term is strictly positive. In total, we have  $V_1^{bb} > V_1^{bg}$ .

Consider next the population of receivers. We use similar notations to the case of investors. Let  $v_2^{kj}(\pi_1(y), \sigma_2^j(y))$  be the expected utility that a kid of type  $j (= g, b)$  receives in a match with distance  $y$  where noneconomic benefit of trustworthiness is evaluated by the value of a parent of type  $k (= g, b)$ . Namely, we have

$$v_2^{kj}(\pi_1(y), 1) = (1 - \beta)\gamma\omega\pi_1(y) + d_2e^{-\theta_2^k y}, \quad v_2^{kj}(\pi_1(y), 0) = \gamma\omega\pi_1(y).$$

For a good parent, it follows from Proposition 1 that

$$\begin{aligned} V_2^{gg} - V_2^{gb} &= \int_{T_2^b(1)}^{T_2^g(1)} \{v_2^{gg}(1, 1) - v_2^{gb}(1, 0)\}g(z)dz \\ &+ \int_{T_2^g(1)}^S \{v_2^{gg}(\pi_1(z), \sigma_2^g(z)) - v_2^{gb}(\pi_1(z), 0)\}g(z)dz. \quad (\text{A.13}) \end{aligned}$$

We will examine the sign of the right hand side in (A.13) term by term. Consider the first term. In equilibrium, a good kid chooses  $\sigma_2^g(z) = 1$ . This implies that

$$v_2^{gg}(1, 1) = (1 - \beta)\gamma\omega + d_2e^{-\theta_2^g y} > v_2^{gg}(1, 0) = \gamma\omega.$$

Since  $v_2^{gb}(1, 0) = \gamma\omega$  by definition, it holds that  $v_2^{gg}(1, 1) > v_2^{gb}(1, 0)$ . Thus, the first term is strictly positive. Consider the second term. A good kid plays a mixed strategy  $0 < \sigma_2^g(z) < 1$  and thus he is indifferent between choosing 0 and 1. This implies that  $v_2^{gg}(\pi_1(z), \sigma_2^g(z)) = v_2^{gg}(\pi_1(z), 1) = v_2^{gg}(\pi_1(z), 0)$ . Clearly,  $v_2^{gg}(\pi_1(z), 0) = v_2^{gb}(\pi_1(z), 0)$ . Therefore, the second term is zero. In

total, we have  $V_2^{gg} > V_2^{gb}$ .

For a bad parent, we have

$$\begin{aligned} V_2^{bb} - V_2^{bg} &= \int_{T_2^b(1)}^{T_2^g(1)} \{v_2^{bb}(1, 0) - v_2^{bg}(1, 1)\}g(z)dz \\ &+ \int_{T_2^g(1)}^S \{v_2^{bb}(\pi_1(z), 0) - v_2^{bg}(\pi_1(z), \sigma_2^g(z))\}g(z)dz. \quad (\text{A.14}) \end{aligned}$$

Consider the first term of (A.14). In equilibrium, a bad kid chooses  $\sigma_2^b(z) = 0$  and thus  $v_2^{bb}(1, 0) > v_2^{bb}(1, 1)$ . By definition, we have  $v_2^{bb}(1, 1) = v_2^{bg}(1, 1)$ . Thus, the first term is strictly positive. Consider the second term of (A.14). A good type kid plays a mixed strategy  $0 < \sigma_2^g(z) < 1$ , and thus he is indifferent between choosing 0 and 1. This implies that

$$v_2^{gg}(\pi_1(z), \sigma_2^g(z)) = (1 - \beta)\gamma\omega\pi_1(z) + d_2e^{-\theta_2^g z} = \gamma\omega\pi_1(z).$$

Since  $\theta_2^g < \theta_2^b$ , it holds that

$$\begin{aligned} v_2^{bg}(\pi_1(z), \sigma_2^g(z)) &< \sigma_2^g(z)\{(1 - \beta)\gamma\omega\pi_1(z) + d_2e^{-\theta_2^g z}\} + (1 - \sigma_2^g(z))\gamma\omega\pi_1(z) \\ &= \gamma\omega\pi_1(z) \\ &= v_2^{bb}(\pi_1(z), 0). \end{aligned}$$

Therefore, the second term is strictly positive. In total, we have  $V_2^{bb} > V_2^{bg}$ . Q.E.D.

### Proof of Proposition 3.

Consider the dynamics (4.11). Since  $\delta_i + f_i(n_t)$  is a probability for a kid to become a ‘‘good’’ type, it must hold that

$$\delta_i + f_i(n) \leq 1 \quad \text{for all } 0 < n < 1. \quad (\text{A.15})$$

It is well known in the theory of dynamical systems that (4.11) has a unique

stationary point which is globally stable if  $0 < \frac{dn_t}{dn_{t-1}} < 1$  for every  $n_{t-1}$ . By (4.11), we have

$$\frac{dn_t}{dn_{t-1}} = f_i(n_t) + n_{t-1}f'_i(n_t)\frac{dn_t}{dn_{t-1}}.$$

This gives

$$\frac{dn_t}{dn_{t-1}} = \frac{f_i(n_t)}{1 - n_{t-1}f'_i(n_t)}. \quad (\text{A.16})$$

By (A.16),  $0 < \frac{dn_t}{dn_{t-1}} < 1$  is equivalent to that

$$f_i(n_t) + n_{t-1}f'_i(n_t) < 1. \quad (\text{A.17})$$

(A.15) and (A.17) hold for sufficiently small  $\varphi$ . Putting  $n_{t-1} = n_t = n^*$  in (4.11), we obtain the stationary point  $n_i^* = \frac{\delta_i}{1 - f_i(n_i^*)}$  of (4.11). Q.E.D.

#### Proof of Proposition 4.

Consider the equilibrium in Proposition 1 under enforcement. Players' trust and trustworthiness thresholds  $T_{ip}^j(\pi_k)$  are given by (5.13) and (5.14) where  $\pi_k = 0, n_k, 1$  is the probability that a partner may cooperate. Emphasizing that  $T_{ip}^j(\pi_k)$  is a function of enforcement level  $p$ , we denote it as  $T_i^j(\pi_k, p)$ . It can be shown that  $\frac{\partial}{\partial p}T_i^j(\pi_k, p) > 0$  for every  $i$  and  $j$ . The three level enforcements  $p^0, p^1$ , and  $p^2$  are applied to matches in the ranges  $[0, y^1]$ ,  $(y^1, y^2]$ , and  $(y^2, S]$ , respectively, where  $y^1$  belongs to a right-hand neighborhood of  $T_2^b(1)$  and  $y^2$  to a right-hand neighborhood of  $T_2^g(1)$ .

Similarly to (4.9), the optimal effort of an investor parent is given by

$$\begin{aligned} f_1/\varphi &= \int_{T_2^g(1, p^1)}^{T_2^g(n_1, p^2)} \{\beta\gamma\omega\{\pi_2(z) + (1 - \pi_2(z))p\} - \omega + d^1e^{-\theta_1^g z}\}g(z)dz \\ &= d^1 \int_{T_2^g(1, p^1)}^{T_2^g(n_1, p^2)} (e^{-\theta_1^g z} - e^{-\theta_1^b z})g(z)dz. \end{aligned} \quad (\text{A.18})$$

Note that we take values  $p^1$  and  $p^2$  so that  $T_2^g(1, p^1) < y^2 < T_2^g(n_1, p^2)$ . The investor parent's effort does not depend on  $p^0$ . With  $\theta_1^g < \theta_1^b$ , it can be shown that  $\frac{\partial f_1}{\partial p^2} > 0 > \frac{\partial f_1}{\partial p^1}$  since  $T_2^g(1, p^1)$  and  $T_2^g(n_1, p^2)$  are increasing functions of  $p^1$

and  $p^2$ , respectively. This proves the first part of the proposition.

Similarly to (4.10), the optimal effort of a receiver parent is given by

$$\begin{aligned} f_2/\varphi &= \int_{T_2^b(1,p^0)}^{T_2^g(1,p^1)} \{(1-\beta)\gamma\omega + d_2e^{-\theta_2^g z} - (1-\beta p)\gamma\omega\}g(z)dz \\ &= \int_{T_2^b(1,p^0)}^{T_2^g(1,p^1)} \{d_2e^{-\theta_2^g z} - \beta(1-p(z))\gamma\omega\}g(z)dz. \end{aligned} \quad (\text{A.19})$$

We take values  $p^0$  and  $p^1$  so that  $T_2^g(1,p^0) < y^1 < T_2^g(1,p^1) < y^2$ . The receiver parent's effort  $f_2$  does not depend on  $p^2$ . Since  $T_2^b(1,p^0)$  and  $T_2^g(1,p^1)$  are increasing functions of  $p^0$  and  $p^1$ , respectively, it can be shown without much difficulty that  $\frac{\partial f_1}{\partial p^1} > 0 > \frac{\partial f_1}{\partial p^0}$ .<sup>19</sup> This proves the second part of the proposition. Q.E.D.

## References

- Alesina, A. and La Ferrara, E. (2002), "Who Trusts Others?", *Journal of Public Economics*, 85, 207–234.
- Ashraf, N., Bohnet, I. and Piankov, N. (2006), "Decomposing Trust and Trustworthiness", *Experimental Economics*, 9, 193–208.
- Ben-Ner, A. and Halldorsson, F. (2010), "Trusting and Trustworthiness: What are They, How to Measure Them, and What Affects Them", *Journal of Economic Psychology*, 31, 64–79.
- Berg, J. Dickhaut, J. and McCabe, K. (1995), "Trust, Reciprocity, and Social History", *Games and Economic Behavior*, 10, 122–142.
- Bisin, A. and Verdier, T. (2001), "The Economics of Cultural Transmission and the Dynamics of Preferences", *Journal of Economic Theory*, 97, 298–319.

---

<sup>19</sup>Note that the integrand is increasing in  $p(z)$  where  $p(z) = p^0$  on  $[T_2^b(1,p^0), y^1]$  and  $p(z) = p^1$  on  $[y^1, T_2^g(1,p^1)]$ . We use the derivative formula of  $F(a) = \int_0^a f(x,a)dx$  for a differential function  $f(x,a)$ :  $F'(a) = f(a,a) + \int_0^a \frac{\partial}{\partial a} f(x,a)dx$ .

- Bohnet, I. and Baytelman, Y. (2007), “Institutions and Trust”, *Rationality and Society*, 19, 99–135.
- Bohnet, I. and Zeckhauser, R. (2004), “Trust, Risk and Betrayal”, *Journal of Economic Behavior and Organization*, 55, 467–484.
- Buchan, N., Croson, R. and Dawes, R. (2002), “Swift Neighbors and Persistent Strangers: A Cross-Cultural Investigation of Trust and Reciprocity in Social Exchange”, *American Journal of Sociology*, 108, 168–206.
- Camerer, C. (2003), *Behavioral Game Theory*. Princeton, NJ: Princeton University Press.
- Cesarini, D., Dawes, C. T., Fowler, J. H., Johannesson, M., Lichtenstein, P. and Wallace, B. (2008), “Heritability of Cooperative Behavior in the Trust Game”, *Proceedings of the National Academy of Sciences*, 105, 3721–3726.
- Coleman, J.S. (1990), *Foundations of Social Theory*. Cambridge, MA: Harvard University Press.
- Cox, J. C. (2004), “How to Identify Trust and Reciprocity”, *Games and Economic Behavior*, 46, 260–281.
- Dixit, A. K. (2004), *Lawlessness and Economics—Alternative Models of Governance*. Princeton, NJ: Princeton University Press.
- Dohmen, T., Falk, A., Huffman, D. and Sunde, U. (2012), “The Intergenerational Transmission of Risk and Trust Attitudes”, *Review of Economic Studies*, 79, 645–677.
- Ermisch, J. and Gambetta, D. (2010), “Do Strong Family Ties Inhibit Trust?”, *Journal of Economic Behavior and Organization*, 75, 365–376.
- Etang, A., Fielding, D. and Knowles, S. (2011), “Does Trust Extend Beyond the Village? Experimental Trust and Social Distance in Cameroon”, *Experimental Economics*, 14, 15–35.
- Fehr, E. (2009), “On the Economics and Biology of Trust,” *Journal of the European Economic Association* 7, 235–266.

- Frey, B. S. and Oberholzer-Gee, F. (1997), “The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-Out,” *American Economic Review* 87, 746–755.
- Frey, B. S. and Jegen, R. (2001), “Motivation Crowding Theory,” *Journal of Economic Surveys* 15, 589–611.
- Glaeser, E.L., Laibson, D.I., Scheinkman, J.A. and Soutter, C.L. (2000), “Measuring Trust”, *Quarterly Journal of Economics*, 115, 811–846.
- 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. (2006), “Does Culture Affect Economic Outcomes?”, *Journal of Economic Perspectives*, 20, 23–48.
- Guiso, L., Sapienza, P. and Zingales, L. (2008), “Social Capital as Good Culture”, *Journal of the European Economic Association*, 6, 295–320.
- Guiso, L., Sapienza, P. and Zingales, L. (2009), “Cultural Biases in Economic Exchange”, *Quarterly Journal of Economics*, 124, 1095–1131.
- Hoffman, E., McCabe, K., and Smith, V. (1996), “Social Distance and Other-Regarding Behavior in Dictator Games”, *American Economic Review*, 86, 653–660.
- Kiyonari, T., Yamagishi, T., Cook, K. and Cheshire, C. (2006), “Does Trust Beget Trustworthiness? Trust and Trustworthiness in Two Games and Two Cultures: A Research Note”, *Social Psychology Quarterly*, 69, 270–283.
- Knack, S. and Keefer, P. (1997), “Does Social Capital Have an Economic Payoff? A Cross-Country Investigation”, *Quarterly Journal of Economics*, 112, 1251–1288.
- Kosfeld, M., Heinrichs, M., Zak, P. J., 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. W. (1997), “Trust in Large Organizations”, *American Economic Review*, 87, 333–338.
- Sapienza, P., Toldra-Simats, A., and Zingales, L. (2013), “Understanding Trust”, *The Economic Journal*, 123, 1313–1332.
- Tabellini, G. (2008a), “Institutions and Culture”, *Journal of the European Economic Association*, 6, 255–294.
- Tabellini, G. (2008b), “The Scope of Cooperation: Values and Incentives”, *Quarterly Journal of Economics*, 123, 905–950.
- Yamagishi, T., Cook, K. S. and Watabe, M. (1998), “Uncertainty, Trust, and Commitment Formation in the United States and Japan”, *American Journal of Sociology*, 104, 165–194.
- Yamagishi, T. and Yamagishi, M. (1994), “Trust and Commitment in the United States and Japan”, *Motivation and Emotion*, 18, 129–166.