Building Social Trust: A Human Capital Approach

Fali Huang
September 2007

Building Social Trust: A Human Capital Approach

FALI HUANG[*]
School of Economics
Singapore Management University

September 15, 2007

ABSTRACT

Much evidence suggests individuals differ in their predisposition to cooperate, which is essentially a component of human capital. This paper examines the role of individual cooperative tendencies and their interactions with institutions in generating social trust; it also endogenizes cooperative tendencies using a human capital investment model. Multiple equilibria and inefficiencies exist due to positive externalities. An innovative finding is that, when institutions are more effective in punishing defecting behaviors, more people invest in cooperative tendencies and hence the endogenous social trust is higher, though the equilibrium cooperative tendencies are lower. This paper provides a plausible explanation for many empirical and experimental results. (JEL: Z13, J24)

1 Introduction

The importance of social trust in economy was suggested long ago by Arrow [1972]. In recent years social trust has attracted the attention of many economists as well as other social scientists.[1] For example, several empirical studies show that the average trust level in a society is significantly associated with economic growth (KNACK AND KEEFER [1997]) and has large positive effects on the performance of various organizations (LA PORTA et al. [1997]). At the same time, many experimental economists have documented substantial amounts of trust in various games (see PALFREY AND PRISBREY [1997], BERG, DICKHAUT, AND MCCABE [1995], and GLAESER et al. [2000] among others). The formal analysis

---

[1]Social trust is considered an important form of social capital, which represents the cooperative infrastructure of a society (COLEMAN [1988], PUTNAM [1993], [1995]).

of social trust, however, is lagging behind, and answers to many basic questions about social trust are still elusive (James [2002]). For example, what is the relationship between trustworthiness, trust, and social trust? Where does social trust come from? Is its level efficient? How is it related to human capital? This paper attempts to answer these questions using a human capital investment approach in a game theoretic setting.

To induce voluntary cooperation in prisoner's dilemmas, the conventional way in economics is to embed a dilemma in a bigger game with repeated interactions, where the potential gains from future dealings motivate people to cooperate in a seemingly one-shot prisoner's dilemma. In real life, however, this kind of "shadow of future" (Axelrod [1984]) generated by expected future dealings with other players is often too vague to have strong enough disciplines against defecting in the game or situation studied. For example, when a person finds a lost wallet containing some cash on the sidewalk in New York City, the gains from future interactions with the wallet owner are not likely to be large enough to prompt one to return the wallet; yet among hundreds of such lost wallets more than half of them were returned with the cash intact (Frank [1992]). In these occasions shrewd calculative thinking is rendered less helpful, whereas personal character cultivated from socialization and moral training may offer a more reliable guide for actions.

It is a mundane observation that some people are simply more trustworthy than others: they are more inclined to help others at their own costs and less likely to shirk their responsibilities. Abundant evidence from experiments suggests people, in general, differ in predisposition to cooperate (Palfrey and Prisbrey [1997], Andreoni and Croson [2008]). Extensive research in industrial and organization psychology demonstrates that an individual's conscientiousness, a reliable and consistent dimension of personality, relates strongly to job performance across different types of jobs (Judge and Ilies [2002]). Studies of moral development demonstrate that such pro-social attitudes develop a functional autonomy over time and become distinct from short-term calculations of self-interest (Staub [1978]). The daily life usage of trust, as reflected in the dictionary definitions, also focuses on the trusted person's essential integrity and character, rather than on whether she has external incentives to refrain from taking advantage of others. For example, Webster's New Collegiate Dictionary [1979, p. 1246] defines trust as "assured reliance on the character, ability, strength, or truth of someone or something."

All of these facts and studies point to a distinct way for players to reach cooperation in one-period prisoner's dilemmas: some players have acquired a certain type of trait or skill that enables them to resist short-run opportunistic temptations and to cooperate for mutual gains (Frank [1987], Kandel and Lazear [1992], Rotemberg [1994], Kreps [1997]). Such a stable personal trait, which is called in this paper the *cooperative tendency* of a player, can be interpreted as a credible commitment to do general good. It is essentially

a component of human capital that is costly to cultivate but yields a stream of returns in the future. When more people have higher cooperative tendencies in a society, they are more trustworthy and hence social trust is higher. How many people choose to inculcate cooperative tendencies and at what levels are likely to be affected by the costs and returns of cultivation, which may vary across society and over time.

These insights are formalized in this paper. When we fix a one-period prisoner's dilemma, players with cooperative tendencies below a certain threshold behave like a *selfish* type who always defects; those with high enough cooperative tendencies, in contrast, behave as a *selfless* type who always cooperates; while those with middle level cooperative tendencies behave as a *reciprocal* type who makes in-kind responses to her partner's action.[2] The latter two types, jointly called the *cooperative type*, resemble the irrational types typically assumed in reputation literature, where the honest players (TIROLE [1996] and DIXIT [2003]) correspond to the selfless type here, the tit-for-tat (KREPS et al. [1982]) and the reciprocal players (FEHR AND GACHTER [2000]) correspond to the reciprocal type here. These various ad hoc types are now naturally unified and rationalized in this paper in that their behaviors, though seemingly irrational and mechanic, are actually rational responses of players with different levels of cooperative tendencies. And players with higher cooperative tendencies are more likely to behave as a cooperative type than others. Furthermore, the same player with a fixed cooperative tendency may exhibit all three behavior types across different games, where her probability of behaving as a cooperative type is higher when the defecting benefits are lower.

Now we can define the trust-related concepts. The *trustworthiness* of a player in a game is the probability that she would cooperate in it, which obviously depends on her cooperative tendency, the payoffs of defecting in the game, and the institutional background. How much *trust* a player has in her partner is equal to the latter's perceived trustworthiness. *Social trust* in a group is equal to the perceived trustworthiness of a typical member or the average trustworthiness of all members, which is often characterized by the proportion of cooperative players in the group. The level of social trust is thus determined by the distribution of cooperative tendency in the group and the specific game features, which is why it varies across players and games (GLAESER et al. [2000]). This formalization of trust based on individual cooperative tendencies seems fruitful, since it not only unifies various ad hoc behavioral types widely adopted in reputation literature, but also systematically accounts for related empirical and experimental results.

The next step is to study how cooperative tendencies are acquired by individuals. The cultivation process is modeled in this paper as a human capital investment decision where

---

[2] Many experimental studies have found that between 40% and 66% of subjects exhibit cooperative or reciprocal behaviors, while between 20% and 30% act completely selfish (FEHR AND GACHTER [2000]).

parents choose the optimal level of cooperative tendency for their child to maximize her lifetime income minus the inculcation cost. We find that the individual returns of investing in cooperative tendencies are always lower than the social returns, since an individual's investment changes others' beliefs about future social trust and thus produces positive externalities. As a result, there exist multiple equilibria where the social trust levels are typically not optimal. When the net returns are quite similar among individuals, a negligible difference in initial beliefs may lead the economy to either 'no trust' or 'full trust' stable equilibrium; this may account for why otherwise similar communities may end up with very different levels of social trust (PUTNAM [1993]). When individual returns are diverse enough, the economy may have a unique stable equilibrium, where more people invest in cooperative tendencies if the information structure is better (so that one's cooperative tendency can be assessed more accurately), or if disciplinary institutions that punish defecting behaviors are more effective; the individual cooperative tendencies in the equilibrium, however, are lower since the relative temptation of defecting is lower.

This paper belongs to the burgeoning literature of social trust. For an important and complex concept such as social trust, there are naturally many different angles to investigate, each providing some new insights. For example, one may analyze social capital formation using the method of physical capital accumulation (GLAESER, LAIBSON, AND SACERDOTE [2002]), or study how incentive structures in a firm affect social trust among employees (ROB AND ZEMSKY [2002]). This paper explores a new perspective, namely the human capital approach, in understanding social trust and its formation. In addition to many findings consistent with existing empirical and experimental evidence, it yields new results that advance the social trust literature. For example, this paper illustrates the importance of strategic interdependence among individual decisions in cultivating cooperative tendencies, which sheds new light on cross-sectional differences in social trust. Though the crowding in and crowding out effects of legal institutions on intrinsic motivation are already known (HUCK [1998], BAR-GILL AND FERSHTMAN [2005], BOHNET, FREY, AND HUCK [2001], GUTH AND OCKENFELS [2005]), this paper's finding that a more effective legal system may induce *more* people to invest in *lower* cooperative tendencies is brand new; this result is also consistent with the evidence that nations with better legal institutions tend to have higher social trust levels based on the World Value Surveys (KNACK AND KEEFER [1997]).

Related literature studies the endogeneity of moral preferences using the indirect evolutionary approach, which combines individual rational decision-making guided by a given preference with the evolutionary approach of preference determination (e.g. GÜTH, KLIEMT, AND PELEG [2000], BRENNAN, GÜTH, AND KLIEMT [2003], GÜTH AND OCKENFELS [2005]). Specifically, the proportion of people with preferences containing a certain intrinsic motivational parameter increases in the population automatically if their relative material payoffs

are higher than those without it. The human capital investment approach in the current paper, however, treats the preference formation process itself as a rational choice, too, albeit not by the individual herself whose preference is to be determined, but by her parents aiming to maximize her lifetime material payoffs minus the inculcation cost. It implies that, even when players with cooperative preferences earn higher incomes than those without, their equilibrium proportion may not increase to one due to positive inculcation costs, and it may even drop to zero if it is too costly to inculcate such preferences. Such a prediction differs from the costless evolution approach of preferences. Nonetheless, the two approaches do share a fundamental element that the prevalence of cooperative preferences in a population positively correlates with the associated material gains. In terms of methodology, one can view the rational investment approach as a generalized model, with the indirect evolutionary approach as a special case with zero inculcation cost.[3]

This paper is organized as follows. In the next section cooperative tendency and social trust are formally defined and investigated in various games. A simple human capital investment model is developed in Section 3, where an individual's cooperative tendency is chosen (by her parents) to maximize her lifetime income given the others' investment decisions. The final section presents conclusions. All proofs are in the Appendix.

2 Cooperative Tendencies and Social Trust

2.1 The Basic Setup

There is a continuum of agents indexed by $i \in [0, 1]$. Agents are randomly paired to play the following one-shot game,

<div align="center">
player i
</div>

		C	D
player j	C	(g, g)	$(-l\ , g + d - \alpha_j)$
	D	$(g + d - \alpha_i, -l)$	$(-\alpha_i, -\alpha_j)$

where C is cooperate or exert effort, D is defect or do not exert effort, and g, l, and d are payoffs from material outputs. We assume

$$d \quad < \quad l, \tag{1}$$

$$g + d - l \quad > \quad 0, \tag{2}$$

which are quite standard in the relevant literature.[4]

---

[3] I thank an anonymous referee for suggesting this point.

[4] See, for example, KREPS et al [1982], ROTEMBERG [1994], BAR-GILL AND FERSHTMAN [2005] and DIXIT [2003].

Besides the game-specific payoffs specified by (g, l, d), each player i also incurs an idiosyncratic psychological cost $\alpha_i \in R^+$ when she defects; α_i acts as an internal penalty against defecting and hence measures player i's *cooperative tendency*.[5] Players have heterogenous cooperative tendencies such that $\alpha_i \sim F(\cdot)$, where $F(\cdot)$ is a continuous and strictly increasing function. When $\alpha_i = \alpha_j = 0$, the game is a typical prisoner's dilemma. To avoid confusion and be consistent with the standard usage in the literature, we call a game a prisoner's dilemma if it is so for players with zero cooperative tendency.

When individual cooperative tendencies are publicly observed, it is straightforward to get the following result: a player always defects (as a selfish type) if $\alpha_i < d$, always cooperates (as a selfless type) if $\alpha_i \geq l$, and makes in-kind responses to her partner's action (as a reciprocal type) if $\alpha_i \in [d, l)$. The trustworthiness of a player in the game is the probability that she would cooperate in it. Selfish players have zero trustworthiness since they never cooperate; by contrast, selfless players are completely trustworthy. A reciprocal player exhibits zero trustworthiness if her partner is (perceived) selfish, and full trustworthiness if otherwise. How much trust a player has in her partner is equal to the latter's (perceived) trustworthiness. So no one trusts a selfish partner, whereas everyone trusts a selfless one. A cooperative player will trust a reciprocal partner, but a selfish player will not. Social trust in this game is equal to the average trustworthiness of all players, which is $1 - F(l) + (1 - F(d))(F(l) - F(d))$ when the matching is random among players, and $1 - F(d)$ when cooperative players match among themselves.

2.2 Trust Among Strangers: Cooperative Tendencies and Disciplinary Institutions

Suppose cooperative tendencies are private information and players randomly match to play the above game for one period. There are disciplinary institutions such as the legal system or monitoring schemes in organizations, which detect the defecting behaviors with probability $q \in [0, 1]$, and assign zero material return to players caught defecting and also to their partners. When one player defects and the other cooperates, this punishes the defector since it confiscates her return $g + d$ from the game and gives l to the cooperator to compensate her loss, where $l < g + d$ under assumption (2); when both players defect, it does not change anything since both get zero return anyway. Let π denote the proportion of cooperative players in equilibrium.

Proposition 1 *In the one-period incomplete information game, every player cooperates in the Bayesian Nash equilibrium when $q \geq d/(g + d)$; when $q < d/(g + d)$, the Bayesian Nash*

---

[5] Modeling a cooperative tendency as an intrinsic *benefit* of cooperation does not make any difference for the results. A similar motivational parameter is used by, among others, ROB AND ZEMSKY [2002], BRENNAN, GÜTH, AND KLIEMT [2003] and GÜTH AND OCKENFELS [2005].

equilibrium is "players with $\alpha_i \geq \underline{\alpha}$ cooperate and others defect," where

$$\underline{\alpha} \equiv \pi d + (1 - \pi)l - q(\pi(g + d - l) + l), \tag{3}$$

and $\pi \equiv \Pr(\alpha_i \geq \underline{\alpha})$ is unique and stable with $\partial\pi/\partial d < 0$, $\partial\pi/\partial l < 0$, and $\partial\pi/\partial q > 0$.

This proposition shows that when the punishment imposed by the disciplinary institutions is large enough, that is, when $q \geq d/(g + d)$, the extrinsic incentives alone can induce cooperation from all players, regardless of their individual cooperative tendencies. But such a situation is often very costly to reach since detecting defecting behaviors and enforcing the punishment scheme uses valuable resources. Substantial costs may be saved if the general public has certain levels of cooperative tendencies. For an extreme example, when all players have a cooperative tendency $\alpha \geq d$, they will choose to cooperate out of intrinsic incentives alone and hence there is no need for any disciplinary institutions. For a more general example, suppose π proportion of players have a cooperative tendency α while the others have zero cooperative tendency, where $\alpha \in (0, \pi d + (1 - \pi)l)$. To achieve a social trust level of π, the society needs to establish institutions with the minimum degree of punishment

$$q(\alpha, \pi) = \frac{\pi d + (1 - \pi)l - \alpha}{\pi(g + d) + (1 - \pi)l},$$

which is derived from (3) by setting $\underline{\alpha} = \alpha$; it is straightforward to see that $q(\alpha, \pi)$ decreases in both α and π. In other words, the extrinsic incentives imposed by institutions and the intrinsic incentives represented by players' cooperative tendencies are substitutes to each other in determining the social trust level.

Note that $\underline{\alpha}$ denotes the minimum cooperative tendency for a player to behave cooperatively in this game. Proposition 1 implies that under incomplete information, a player with $\alpha_i < \underline{\alpha}$ behaves as a selfish type, $\alpha_i \in [\underline{\alpha}, (1 - q)l)$ reciprocal, and $\alpha_i \geq (1 - q)l$ selfless (who cooperates even when $\pi = 0$). The social trust, the expected trustworthiness of a typical member in this game, is thus characterized by $\pi = 1 - F(\underline{\alpha})$, the proportion of cooperative players. It is higher when the effectiveness of disciplinary institutions (q) is higher, when more players have higher cooperative tendencies, when the benefit of cooperation g is higher, and when the costs of cooperation d and l are lower. So social trust varies across players, games, and institutional backgrounds.

Such a social trust concept coincides with the commonly used trust measure in experiments, the proportion of players who cooperate in a prisoner's dilemma. It is also consistent with the widely used indicator $TRUST_C$, which measures the percentage of respondents in a community C replying "most people can be trusted" (to the question in the World Values Surveys "Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?"), if individuals who have met a trustworthy

partner agree that most people can be trusted, or those who say so are likely to cooperate themselves (GLAESER et al. [2000]): In the one-period incomplete information game studied above, exactly π proportion of players are matched with a trustworthy partner, and π proportion of players cooperate in the game. In other words, if the players who have played the above game are asked the same trust question as in the World Values Survey, the two trust measures are likely to coincide. Our finding that trust may differ across players, games, and countries can thus account for the commonly observed discrepancies between survey- and experiment-based measures of social trust (PALFREY AND PRISBREY [1997], GLAESER et al. [2000], BURLANDO AND HEY [1997], WEIMANN [1994]), since the distribution of cooperative tendencies, game features, and the effectiveness of disciplinary institutions are likely to vary across games and real life situations.

2.3 Trust Among Acquaintances: Cooperative Tendencies and Repeated Interactions

When the legal system is not available or too expensive to operate, repeated social interactions may serve as an informal disciplinary institution to discourage defecting. Suppose players are randomly paired to play the above stage game for more than one period; at the end of each period a player's action is observed by her partner, and there is a positive probability $\beta \le 1$ that they will play the game for another period. Each pair lasts at most T periods, where T is a finite integer.[6] The probability β that the game continues measures the strength of the social network such as a residential community, a club, or other social organizations where people may interact with each other for more than one occasion.

Suppose the minimum cooperative tendency of the players is $\alpha_l \in [0, d)$. A sequential equilibrium in this game is characterized below to illustrate the interactions between cooperative tendencies and repeated interactions in determining the social trust level. In this equilibrium all players cooperate until the last period, when they behave according to Proposition 1; let $\pi_0 = 1 - F(\pi_0 d + (1 - \pi_0)l)$ denote the proportion of cooperative players determined by (3) with $q = 0$, and $\pi_R = F(l) - F(\pi_0 d + (1 - \pi_0)l)$ denote the proportion of reciprocal players.

Proposition 2 *In the T-period game, the following strategy profile plus belief system is a sequential equilibrium when*

$$\alpha_l + \beta \pi_R (g + d) > d. \tag{4}$$

The strategy profile is: (i) Players with $\alpha \ge l$ are selfless who always play C. (ii) Players with $\alpha \in [\pi_0 d + (1 - \pi_0)l, l)$ are reciprocal, who play C in the first period, play C if (C, C) is played in the previous period, and play D otherwise. (iii) All the other players are of

---

[6]An infinitely repeated game brings similar insights and hence is not discussed here.

selfish type, who mimic reciprocal players until the last period when they play D. The belief system is: (i) In the first period and every period following the history in which only (C,C) has been played, every player believes that with probability π_0 her partner is a cooperative type. (ii) In all the following periods after the first time D is observed, the player who has played D is believed to be a selfish type; the player who has played C is still believed to be a cooperative type with probability π_0.

Condition (4) suggests that cooperation is more likely when the minimum cooperative tendency α_l in the community is higher, when there are more reciprocal players (so that π_R is larger), and when the possibility of future encounters, β, is higher. It points to two different sources of cooperation: One is trust based on players' intrinsic cooperative tendencies, the other is the reputation effects as represented by $\beta \pi_R (g + d)$, which are obviously substitutes to each other.

The role of a player's intrinsic cooperative tendency in achieving cooperation is often conveniently hidden under the assumption of irrational types, which actually are rational responses of players with certain levels of cooperative tendencies. For example, tit-for-tat players in KREPS et al. [1982] and the reciprocal ones in FEHR AND GACHTER [2000] act similarly as our reciprocal type players with $\alpha \in [\pi_0 d + (1 - \pi_0)l, l)$, and the honest types in TIROLE [1996] and DIXIT [2003] act like selfless players with $\alpha \geq l$. These cooperation-enhancing behavioral types can be unified and systematically analyzed in our framework using a single concept of cooperative tendency.

The external rewards and punishments contingent on past behaviors may make cooperation appealing to players with low or even zero cooperative tendencies. Such a reputation effect, however, can *not* be generated by repeated interactions alone in finitely repeated games; note that without enough reciprocal players (i.e. if condition (4) $\pi_R \geq (d - \alpha_l)/\beta(g + d)$ does not hold), the reputation effect vanishes immediately and hence selfish players would never cooperate. The effects of cooperative tendencies, however, can be greatly amplified by repeated interactions, where a few cooperative players may act as a powerful lever to induce many selfish players to cooperate.

3 The Formation of Cooperative Tendencies

In this section, a player's cooperative tendency is endogenized as an equilibrium result of parental investment in children. The disciplinary institutions are the same as in Section 2.2, which detect any defecting behaviors with probability $q \in [0, 1]$, and then assign zero material return to both players.

3.1 The Basic Model

Each player lives for two periods. The first period is the investment stage where each player's cooperative tendency is chosen (by her parents) to maximize her life-time income minus the investment cost, taking as given the expected proportion of cooperative players $\Pi \in [0, 1]$ in the population.[7] Investing in a cooperative tendency α incurs positive costs. For example, parents have to repeatedly make effort in teaching children to share toys and be considerate. This task is easier when parents are more skillful and when the child is more obedient. The cost function is $c(\alpha, i)$, where $c(0, i) = 0$, $c_\alpha > 0$, $c_i > 0$, $c_{\alpha\alpha} \geq 0$, and $c_{i\alpha} \geq 0$. That is, the inculcation cost strictly increases and is convex in the cooperative tendency α; to capture player heterogeneity in the investment cost, it also strictly increases in player index i.[8]

The second period is the production stage. With probability $1 - p$, players' cooperative tendencies are private information and they randomly match each other to play the one-period prisoner's dilemma characterized by (g, d, l). Recall from Proposition 1 that players with $\alpha \geq \underline{\alpha}(\Pi)$ cooperate in this game while the others defect, where

$$\underline{\alpha}(\Pi) \equiv \Pi d + (1 - \Pi)l - q(\Pi(g + d - l) + l) \tag{5}$$

is adapted from (3). With probability p, players' cooperative tendencies are observed, and only those who have high enough cooperative tendencies $\alpha \geq D$ are allowed to play a one-period prisoner's dilemma characterized by (G, D, L), where $G \geq g$ and $D \geq \underline{\alpha}(\Pi)$.

Such a set-up is chosen to capture real life situations where in some instances we have to engage with strangers in a one-period prisoner's dilemma without knowing their individual trustworthiness, while in other instances players' cooperative tendencies are revealed so that we can choose to deal only with those who are known to be trustworthy.[9] The fundamental message conveyed by the model is that the pros and cons of cultivating a cooperative tendency are not limited in a single game, but balanced throughout the whole lifetime: The loss of being cheated by a selfish player in one game may be compensated by the benefit

---

[7]We assume a person's cooperative tendency is fixed throughout adulthood, which is consistent with the fact that her trustworthiness may change across games, partners, and with updated information (ALESINA AND LA FERRARA [2002]). This was proved in Section 2.

[8]It is natural to conjecture that the inculcation cost may be lower when the parents' cooperative tendency is higher due to social learning and imitation at home. Such an influence is already accommodated in the cost function $c(\alpha, i)$ where the index i can be regarded as negatively correlated with the parents' cooperative tendency, and hence will not affect any Nash equilibrium results. Such parental influence may, however, generate different dynamics in an overlapping-generation framework; see Huang (2006) for an explicit treatment.

[9]A player's cooperative tendency may be correctly detected by others through various ways. For example, many subtle physical or emotional signals may enable us to distinguish a genuinely trustworthy person from a pretending one since it is often difficult to completely control these signals (FRANK [1987], [1988]). Information about a player's past behaviors may also reveal her innate cooperative tendency (GREIF [1989]). The effectiveness of such type-revealing processes is represented by p in the model.

of cooperating with a trustworthy partner in another game that may not be related to the former one.

Lemma 1 $\alpha_i = D$ if $\alpha_i > 0$ for any $i \in [0, 1]$.

This lemma shows that if a player ever invests, her cooperative tendency will be equal to D, which just enables her to cooperate in the complete information game (G, D, L). We assume $D = \underline{\alpha}(\Pi)$ without much loss of generality. When player i becomes a cooperative type, her expected lifetime income minus the investment cost is

$$\widehat{V}(i, \Pi; D) = \beta pG + \beta(1 - p)[\Pi g - (1 - \Pi)(1 - q)l] - c(\underline{\alpha}(\Pi), i); \tag{6}$$

if she remains a selfish type, no investment cost is incurred, and her expected lifetime income is

$$\widehat{V}(i, \Pi; 0) = \beta(1 - p)(1 - q)\Pi(g + d). \tag{7}$$

Let $V(i, \Pi)$ represent the net return of investing in $\underline{\alpha}(\Pi)$ versus remaining selfish:

$$V(i, \Pi) \equiv \widehat{V}(i, \Pi; D) - \widehat{V}(i, \Pi; 0) = \beta pG - \beta(1 - p)\underline{\alpha}(\Pi) - c(\underline{\alpha}(\Pi), i),$$

where the first term is the expected gain of being cooperative, the second term is the expected loss, and the last one is the investment cost. Players will choose to invest in $\underline{\alpha}(\Pi)$ if and only if $V(i, \Pi) \geq 0$.

Lemma 2 $\partial V(i, \Pi)/\partial i < 0$, $\partial V(i, \Pi)/\partial \Pi > 0$.

The intuition for this lemma is quite clear. $V(i, \Pi)$ decreases with player index i because the investing cost increases with it. A marginal increase of Π not only improves the chance of meeting a cooperative player, but also reduces the threshold cooperative tendency $\underline{\alpha}(\Pi)$ and hence the investment cost. Since cooperative players benefit more from both channels, the net return $V(i, \Pi)$ strictly increases with Π.

3.2 The Equilibrium

Every Nash Equilibrium (*NE* thereafter) at the investment stage is characterized by a pair (Π, π) where $\Pi = \pi$, i.e., the expected proportion of cooperative players is equal to the actual one. Note that 'no social trust' equilibrium $(\Pi = 0, \pi = 0)$ always exists since there is no gain for being the only cooperative player. And the equilibrium investment in cooperative tendency is generally inefficient because an individual's investment decision brings positive externalities to all players.[10] We partition the parameter space into four cases

---

[10]Actually the under-investment in appropriate working habits and attitudes has already been felt by many firms in the U.S., where the current human capital policies focus on cognitive skills to the exclusion of social skills, self-discipline and a variety of non-cognitive skills that are known to determine success in life (CAPPELLI [1995], HECKMAN [2000]).

and characterize the corresponding equilibria when the slopes of the best response functions are monotone. We also check whether these *NE*s are stable to small perturbations of Π.[11]

The Benchmark: The Diverse Cost Case

In the benchmark case, players have quite diverse costs: Some have costs so low that they would invest in cooperative tendencies no matter how few players are expected to do so, while others have such high costs that they would not invest even if everybody else does so. This case is characterized by the following conditions

$$lim_{\Pi \to 0^+} V(0, \Pi) \quad > \quad 0, \tag{8}$$

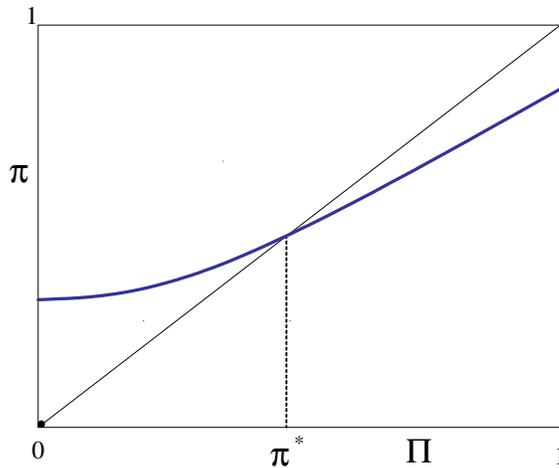$$V(1, 1) \quad < \quad 0. \tag{9}$$

Figure 1: Equilibrium Social Trust in the Diverse Cost Case

Proposition 3 *When players have diverse costs as specified by (8) and (9), there exist two NEs $(0,0)$ and (π^*, π^*), where only (π^*, π^*) is stable; π^* is uniquely determined by $V(i^*(\pi^*), \pi^*) = 0$, where $\partial \pi^*/\partial p > 0$ and $\partial \pi^*/\partial q > 0$; in contrast, $\partial \underline{\alpha}(\pi^*)/\partial p < 0$ and $\partial \underline{\alpha}(\pi^*)/\partial q < 0$.*

This proposition suggests that when the idiosyncratic differences in the investment cost outweigh the externalities, players are less affected by other people's choices and hence the resulting interior equilibrium π^* is unique and stable. See Figure 1 for illustration. In the stable equilibrium, more people invest in cooperative tendencies when it is easier to

---

[11]An NE can be reached as a steady state in a dynamic process with infinite generations, where the expected proportion of cooperative players in each following generation is equal to the realized proportion in the current one, that is, $\Pi_{N+1} = \pi_N$ for $N = 1, 2, ...$ where Π_1 is exogenously given.
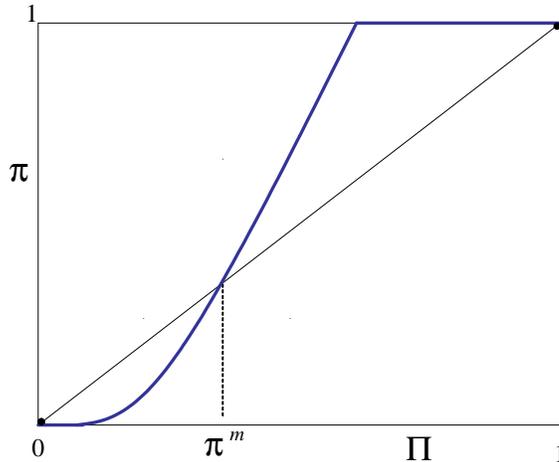
Figure 2: Equilibrium Social Trust in the Medium Cost Case

detect a partner's cooperative tendency (p is higher) and when the punishment imposed by disciplinary institutions is larger (q is higher). In contrast, the equilibrium cooperative tendency $\underline{\alpha}(\pi^*)$ that is necessary to resist the temptation of defecting is lower when the institutions are more effective (higher p and q). These results also hold for any interior stable equilibrium in the other cases discussed below.

Cases with Similar Costs

In contrast to the benchmark case, players may have similar investment costs, which can be either too high, too low, or in the medium range compared to the returns. In the *Medium Cost Case*, the net returns of investing in cooperative tendencies are quite similar across players, not too high or too low. This case is characterized by

$$lim_{\Pi \to 0^+} V(0, \Pi) \leq 0, \tag{10}$$

$$V(1, 1) \geq 0, \tag{11}$$

which are exactly the opposite to (8) and (9) that define the benchmark case. They imply that there exist Π_0 and Π_1 in $(0, 1)$ such that no players invest when $\Pi \leq \Pi_0$, and all do so when $\Pi \geq \Pi_1$. Following similar arguments as in the benchmark, we get the following proposition.

Proposition 4 *When players have medium level costs as specified by (10) and (11), there are three NEs: $(0,0)$, (π^m, π^m), and $(1,1)$, where $\pi^m \in [\Pi_0, \Pi_1] \subset (0,1)$. Among them $(0,0)$ and $(1,1)$ are stable.*

This case is illustrated by Figure 2. The interior NE (π^m, π^m) is unstable, happening only when the initial belief is exactly π^m. A negligible ε difference in initial beliefs may

13

lead to two polar stable equilibria: If the initial belief is 0.5ε lower than π^m, the economy will ultimately fall into the 'no-trust' trap $(0,0)$. On the contrary, if the initial belief is 0.5ε higher than π^m, the economy will gradually reach the 'full-trust' state $(1,1)$. The intuition is that, when more than π^m players are expected to invest in cooperative tendencies, the associated positive externalities outweigh idiosyncratic cost differences and make net returns positive for everybody, while the opposite is true when people believe that less than π^m will do so. In other words, when the inculcation costs are at the medium level and similar across players, an individual's decision is heavily affected by others' choices.

In the *Low Cost Case*, even the highest indexed players invest in cooperative tendencies when they believe enough people are doing so. It is characterized by conditions (8) and (11). Here the full trust *NE* $(1,1)$ always exists and is stable; it is either the only equilibrium, or there exist two other *NE*s at interior points where the one with lower π is stable. The *High Cost Case* is defined by (10) and (9), where the no-trust *NE* $(0,0)$ is stable and unique. The proof is omitted since it is similar to the first two cases. See Figure 3 for illustration. The following proposition summarizes some common results of these four cases.
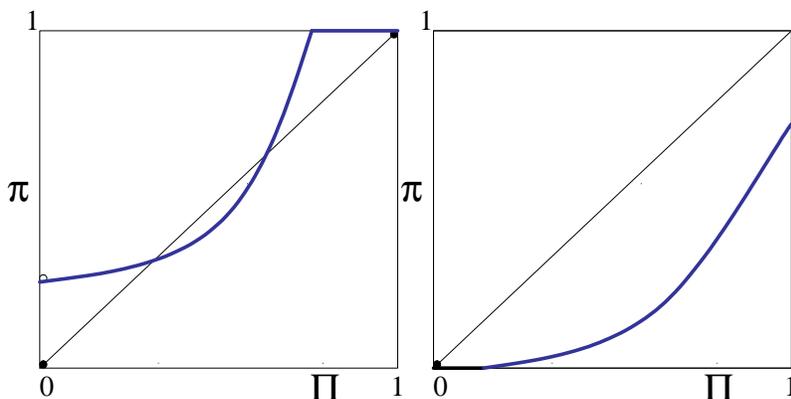

Figure 3: Equilibrium Social Trust in the Low and High Cost Cases

Proposition 5 *In all the stable interior NEs, social trust strictly increases in p and q, while the individual cooperative tendencies strictly decrease in p and q.*

The Relationship between Extrinsic and Intrinsic Incentives. A cooperative tendency is an *intrinsic discipline* against defecting, which is cultivated primarily at home and school during one's childhood. *Extrinsic incentives* against defecting are represented by p and q, which are determined by the information structure and the economic governance institutions, including the legal system and informal governance methods. The more

effective these institutions are in punishing defecting behaviors and encouraging cooperation, the higher p and q, which leads to higher social trust in equilibrium; the resulting cooperative tendency in equilibrium $\underline{\alpha}(\pi^*)$, however, is lower. That is, *higher extrinsic incentives induce more people to become cooperative, but with lower cooperative tendencies.* In contrast, when the governance institutions are less effective, people have to invest in higher cooperative tendencies to achieve cooperation; so fewer people are cooperative, but the cooperative ones are able to withstand larger temptations. The relationship between the intrinsic and extrinsic incentives is thus more complex than either being substitutes in promoting trust, or simply crowding in or crowding out each other.

These results imply that, in a country with more effective disciplinary institutions, survey-based trust indicator $TRUST$ is higher, whereas the experiment-based social trust measure may be lower if subjects are faced with quite high defecting benefits due to the lower p and q in the experiments than in real life. It may shed light on the contradictory social trust ranking across countries. For example, $TRUST$ in the UK (44.4) is much higher than Italy (26.3) (KNACK AND KEEFER [1997]), but UK subjects "free-rode to a much greater extent" than Italians in a public goods experiment (BURLANDO AND HEY [1997]). A similar comparison is the US ($TRUST = 45.4$) vs. Germany ($TRUST = 29.8$) based on survey results, whereas US subjects free-rode more than Germans in experiments (WEIMANN [1994]).

The disciplinary institutions are assumed exogenously given throughout the paper in order to focus on the role of individual cooperative tendencies and their endogenous formation. The endogeneity of these institutions, however, is readily acknowledged, and some direct implications of our results on the optimal design of the institutions is briefly discussed below, though a full-fledged analysis is best left for future research.[12]

Let's consider a society in the Medium Cost Case that has reached the full-trust equilibrium $\pi^*(q) = 1$, where q is the initial quality of disciplinary institutions. If the social planner is not aware of the endogeneity of individual cooperative tendencies, his best ex post choice of q should be zero, which is obviously not optimal, since it discourages investment in cooperative tendencies and may actually bring the society to the no-trust equilibrium in the next generation.[13] Similar things could also happen to the Low Cost Case where a lower ex post q is likely to reduce the social trust level significantly. The impact of endogenous cooperative tendencies on the optimal design of institutions is less dramatic in the Diverse Cost Case, though the optimal quality is still higher than the naive best choice due to $\pi^{*\prime}(q) > 0$ by Proposition 3. The only scenario where such endogeneity issues may not

---

[12] Some initial attempts are already made in this direction (see, for example, HUCK [1997], [1998], and GÜTH AND OCKENFELS [2005]), though most of them adopt an evolutionary approach for preference determination.

[13] A lower q shifts the best response curve downward in Figure 2 and hence makes the no-trust equilibrium more likely to happen.

make any difference on the optimal choice of q is the High Cost Case where nobody ever invests in cooperative tendencies.

It is also important to note the negative effects of high quality disciplinary institutions on the equilibrium levels of individual cooperative tendencies. The consequences may be quite serious if the effectiveness of cultivating cooperative tendencies in children depends on parents' cooperative tendencies; for example, the society may slip into the following vicious cycle: strict disciplinary institutions lead to lower cooperative tendencies, which then drive up the cultivation cost, and hence fewer people find it beneficial to inculcate cooperative tendencies in the next generation; as a result stricter and more costly disciplinary institutions are required to maintain the same level of cooperation as before. In other words, a society may find itself depleting its stock of intrinsic motivation and forced to rely exceedingly on extrinsic incentives. So the optimal design of disciplinary institutions should be combined with that of educational institutions, which, by directly reducing the cultivation costs, encourage more people to invest in higher cooperative tendencies and hence mitigate the negative effects of the former on the level of individual cooperative tendencies.

4 Conclusions

Social trust is an important social phenomenon, which has been extensively studied by social scientists (see for example COOK [2001] and HARDIN [2002]). Empirical studies in the economics literature have shown that it facilitates economic performance at various levels. This paper formalizes trust-related concepts and studies the formation of individual cooperative tendencies in society using a model of human capital investment. It provides a plausible explanation for many empirical and experimental results. It also generates fresh insights and policy implications about social trust, especially on its relationship with human capital and economic governance institutions.

This paper finds that more effective governance institutions encourage more people to become cooperative (hence higher social trust), though they may lead to lower individual cooperative tendencies. Just as criminal rates could be reduced either by more policing or by helping poor children acquire pro-social attitudes through Headstart or similar programs (GARCES, THOMAS, AND CURRIE [2002]), cooperation can be promoted by establishing governance institutions or cultivating cooperative tendencies in children both at home and school. How to achieve an optimal combination of these two sources seems an intriguing topic for future research. For instance, the effectiveness of legal institutions and monitoring schemes in firms and organizations may evolve interactively with the distribution of cooperative tendency in society and result in different combinations of extrinsic and intrinsic incentives across societies and over time. HUANG [2006], for example, uses a similar version of the human capital approach developed in this paper to study the dynamic interactions

between monitoring and trust in a principal-agent setting with overlapping generations.

The interactions between social trust, formal institutions, and other forms of social capital such as social networks and norms may also be studied in future research. One can imagine that when repeated dealings among players are frequent enough due to technological reasons (such as in an agricultural society), it is beneficial to establish sophisticated signaling and type-revealing arrangements so that a person's type is more accurately observed by others; these informal social arrangements may necessarily decline when interaction with strangers becomes more prevalent and profitable (as in an industrial economy), because impersonal disciplinary institutions taking advantage of economies of scale are more effective in curbing defecting behaviors in a large and mobile society. Discovering the dynamic relationship between them may contribute in important ways to our understanding of how substantial cooperation can be achieved over time and across societies.

Appendix

Proof for Proposition 1.

In this game the probability of a player matching with a cooperative partner is believed to be π. By playing C, player i gets g if her partner is cooperative, $-l$ if her partner defects. So her expected payoff of playing C is

$$V_C = \pi g + (1 - \pi)(1 - q)(-l).$$

Similarly, her expected utility of playing D is

$$V_D = (1 - q)[\pi(g + d - \alpha_i) + (1 - \pi)(-\alpha_i)] + q(-\alpha_i).$$

Thus she will play C iff $V_C \geq V_D$, which is simplified to

$$\alpha_i \geq \pi d + (1 - \pi)l - q(\pi(g + d - l) + l) \equiv \underline{\alpha}.$$

Note that when $\underline{\alpha} \leq 0$, all players cooperate so that $\pi = 1$, which then implies $q \geq d/(g+d)$.

For the belief π to be consistent with players' strategies, it must be true that

$$\pi = 1 - F(\pi d + (1 - \pi)l - q(\pi(g + d - l) + l)).$$

The RHS is continuous in π on the closed interval $[0, 1]$, where $RHS(\pi = 0) = 1 - F(l - ql) > 0$ and $RHS(\pi = 1) = 1 - F(d - q(g + d)) < 1$. Hence the existence of π. Note $\partial RHS/\partial \pi = (l - d + q(g + d - l))\partial F > 0$ since $\partial F > 0$ and $l - d + q(g + d - l) > 0$ by assumptions (1) and (2). So the RHS crosses the 45^0 line from above only once, and its

slope at the crossing point π must be smaller than one, that is, $(l-d+q(g+d-l))\partial F|_\pi < 1$ holds, and hence π is stable. By the Implicit Function Theorem we have

$$\frac{\partial \pi}{\partial d} = -\frac{\pi(1-q)\partial F|_\pi}{1-(l-d+q(g+d-l))\partial F|_\pi} < 0,$$

$$\frac{\partial \pi}{\partial l} = -\frac{(1-\pi)(1-q)\partial F|_\pi}{1-(l-d+q(g+d-l))\partial F|_\pi} < 0,$$

$$\frac{\partial \pi}{\partial q} = \frac{(\pi(g+d-l)+l)\partial F|_\pi}{1-(l-d+q(g+d))\partial F|_\pi} > 0.$$

Proof for Proposition 2.

Given the belief system, selfless and reciprocal players would not deviate by the same arguments as in Proposition 1 with $q = 0$. At period T, playing D is the dominant strategy for selfish players with $\alpha < \pi d + (1-\pi)l$, so they will not deviate. If she deviates in some period $t < T$ by playing D, her selfish type is revealed. According to the equilibrium strategies, (D, D) would be played in all future periods unless her partner is selfless, in which case (C, D) is played. So a selfish player who defects at period $t < T-1$ gets the deviation payoff $V_{D,t} = (g+d-\alpha) + (\pi_S(g+d) - \alpha)(\beta + \beta^2 + ... + \beta^{T-t})$; if she defects one period later, her payoff is $V_{D,t+1} = g + (g+d-\alpha)\beta + (\pi_S(g+d)-\alpha)(\beta^2 + \beta^3 + ... + \beta^{T-t})$, where $\pi_S = 1 - F(l)$ is the proportion of selfless players. She will deviate later if $V_{D,t} < V_{D,t+1}$, which holds when

$$\alpha + \beta(1-\pi_S)(g+d) > d. \tag{A1}$$

A player will not defect at period $T-1$ when $(g+d-\alpha) + \pi_S(g+d)\beta - \alpha\beta < g + (\pi_S + \pi_R)(g+d)\beta - \alpha\beta$ holds, which is equivalent to

$$\alpha + \beta\pi_R(g+d) > d.$$

When this condition holds, (A1) is guaranteed since $\pi_R \leq 1 - \pi_S$.

That is, earlier defecting is less attractive when a player's cooperative tendency α is higher, when there are more reciprocal players, and when the possibility β of repeated interactions is larger. So if a player with the minimum cooperative tendency α_l does not defect at period $T-1$, she will not defect at any time earlier, and neither will all the other players. Non-deviation for all players at period $T-1$ thus leads to (3), which guarantees that no players deviate at any time. Since it is obvious that the belief system is fully consistent with the strategy profile, the formal proof is omitted.

Proof for Lemma 1.

Players with $\alpha \geq D$ cooperate in both games and get an income $pG + (1-p)[\Pi g - (1-\Pi)(1-q)l]$. Since it does not depend on α and investing in α is costly, it is optimal to choose the lowest possible level D. Players with $\alpha < \underline{\alpha}(\Pi)$ always defect and get an income $(1-p)(1-q)\Pi(g+d)$; again it is independent of α, so it is optimal to set $\alpha = 0$

to save the investment cost. A player with $\alpha \in [\underline{\alpha}(\Pi), D)$ cooperates in the incomplete information game, but not in the complete information game, and hence she gets an income $(1-p)[\Pi g - (1-\Pi)(1-q)l]$, which is lower than that with either $\alpha \geq D$ or $\alpha < \underline{\alpha}(\Pi)$.

Proof for Lemma 2.

$\partial V(i, \Pi)/\partial i = -c_i < 0$; $\partial V(i, \Pi)/\partial \Pi = (\beta(1-p) + c_\alpha)(-\underline{\alpha}'(\Pi)) > 0$, where

$$\underline{\alpha}'(\Pi) = -(l-d)(1-q) - qg < 0. \tag{A2}$$

Proof for Proposition 3.

Since $V(i, \Pi)$ strictly decreases in i, conditions (8) and (9) imply that there exists a unique $i^*(\Pi)$ such that $V(i^*(\Pi), \Pi) = 0$ for any $\Pi > 0$, where

$$V(i^*(\Pi), \Pi) = \beta pG - \beta(1-p)\underline{\alpha}(\Pi) - c(\underline{\alpha}(\Pi), i^*(\Pi)) = 0. \tag{A3}$$

This implies that for any $\Pi > 0$, players with a lower index than $i^*(\Pi)$ will choose to become cooperative, while the others will not; the proportion of cooperative players is thus $\pi = \Pr(i \leq i^*(\Pi)) = i^*(\Pi)$ since by definition i is uniformly distributed. The best response function of the population is thus

$$\pi \equiv B(\Pi) = \left\{ \begin{array}{cc} i^*(\Pi) & \text{for } \Pi \in (0, 1] \\ 0 & \text{for } \Pi = 0 \end{array} \right. .$$

It is straightforward to see that $i^*(0) = 0$. So $B(\Pi)$ is continuous in Π on $[0, 1]$. And it strictly increases in Π since

$$\frac{\partial i^*(\Pi)}{\partial \Pi} = -\frac{\frac{\partial V(i^*, \Pi)}{\partial \Pi}}{\frac{\partial V(i^*, \Pi)}{\partial i}} = \frac{(\beta(1-p) + c_\alpha)[(l-d)(1-q) + qg]}{c_i} > 0.$$

Its curvature is undetermined in general, though a sufficient condition for strict convexity is when $c_{\alpha\alpha}$ is small enough:

$$\frac{\partial^2 i^*(\Pi)}{\partial \Pi^2} = [(l-d)(1-q) + qg]\frac{\partial(\beta(1-p) + c_\alpha)/c_i}{\partial \Pi} \text{ where}$$

$$\frac{\partial(\beta(1-p) + c_\alpha)/c_i}{\partial \Pi} = \frac{(c_{\alpha\alpha}\underline{\alpha}'(\Pi) + c_{i\alpha}\frac{\partial i^*(\Pi)}{\partial \Pi})c_i - (\beta(1-p) + c_\alpha)(c_{i\alpha}\underline{\alpha}'(\Pi) + c_{ii}\frac{\partial i^*(\Pi)}{\partial \Pi})}{c_i^2}$$

$$= \frac{-\underline{\alpha}'(\Pi)}{c_i}[(2\beta(1-p) + c_\alpha)\frac{c_{i\alpha}}{c_i} + (\beta(1-p) + c_\alpha)^2\frac{-c_{ii}}{c_i^2} - c_{\alpha\alpha}].$$

Let $\pi^* \in (0, 1)$ denote the solution to $B(\pi^*) = \pi^*$, then π^* is unique because $B(\Pi \to 0) > 0$, $B(\Pi = 1) < 1$, and $B(\Pi)$ strictly increases in $\Pi \in (0, 1]$. It is stable since the slope of $B(\Pi)$ is smaller than one when crossing the 45^0 line.

To prove $\partial \pi^*/\partial p > 0$, we show that p shifts up the best response function $B(\Pi)$ for each $\Pi \in (0, 1]$ and increases π_0, which is determined by

$$lim_{\Pi \to 0}V(i = \pi_0, \Pi) = \beta pG - \beta(1-p)(1-q)l - c((1-q)l, \pi_0) = 0. \tag{A4}$$

19

Accordingly, the intersection of $B(\Pi)$ with the 45^0 line, π^*, must also increase with p. Based on the equation (A3) we get

$$\frac{\partial B(\Pi)}{\partial p} = \frac{\partial i^*(\Pi)}{\partial p} = \frac{\partial V(i^*,\Pi)/\partial p}{\partial V(i^*,\Pi)/\partial i^*} = \frac{\beta[G + \underline{\alpha}(\Pi)]}{-\partial V(i^*,\Pi)/\partial i^*} > 0,$$

Based on the equation (A4) we get

$$\frac{\partial \pi_0}{\partial p} = -\frac{\beta(G + (1-q)l)}{-c_i} > 0.$$

Similarly, from (A3) and (A4) we have

$$\frac{\partial i^*(\Pi)}{\partial q} = -\frac{\partial V(i^*,\Pi)/\partial q}{\partial V(i^*,\Pi)/\partial i^*} = \frac{(\beta(1-p) + c_\alpha)(\pi(g+d-l)+l)}{-\partial V(i^*,\Pi)/\partial i^*} > 0,$$

$$\frac{\partial \pi_0}{\partial q} = -\frac{(\beta(1-p) + c_\alpha)l}{-c_i} > 0.$$

The comparative statics of $\underline{\alpha}(\pi^*)$ are as follows.

$$\frac{\partial \underline{\alpha}(\pi^*)}{\partial p} = \underline{\alpha}'(\pi^*)\frac{\partial \pi^*}{\partial p} < 0,$$

$$\frac{\partial \underline{\alpha}(\pi^*)}{\partial q} = -\pi^*(g+d-l) - l + \underline{\alpha}'(\pi^*)\frac{\partial \pi^*}{\partial q} < 0.$$

Proof for Proposition 4

Note that Π_0 and Π_1 are determined respectively by $V(0,\Pi_0) = 0$ and $V(1,\Pi_1) = 0$. It is easy to show $\Pi_0 < \Pi_1$. For any $\Pi \in [\Pi_0,\Pi_1]$, there exists a unique $i^*(\Pi)$ where $V(i^*(\Pi),\Pi) = 0$. By Lemma 2 the inequality $V(1,\Pi_0) < V(0,\Pi_0)$ holds, while $V(0,\Pi_0) = 0 = V(1,\Pi_1)$ by definition; so we have $V(1,\Pi_0) < V(1,\Pi_1)$, which implies $\Pi_0 < \Pi_1$ since $V(i,\Pi)$ strictly increases in Π.

Conditions (10) and (11) imply that $V(0,\Pi) \geq 0$ and $V(1,\Pi) \leq 0$ for any $\Pi \in [\Pi_0,\Pi_1]$. Since $V(i,\Pi)$ strictly decreases in i, there exists a unique $i^* \equiv i^*(\Pi)$ such that $V(i^*(\Pi),\Pi) = 0$. From Proposition 3 we know $i^*(\Pi)$ is strictly increasing and convex in Π. It is straightforward to see that $i^*(\Pi_0) = 0$ and $i^*(\Pi_1) = 1$. The proportion of players who invest in the cooperative tendency is thus

$$\pi \equiv B(\Pi) = \begin{cases} 0 & \Pi \leq \Pi_0 \\ i^*(\Pi) & \Pi \in (\Pi_0,\Pi_1) \\ 1 & \Pi \geq \Pi_1 \end{cases}$$

Following similar arguments as in Proposition 3, $B(\Pi)$ is continuous, strictly increasing and convex in Π on $[\Pi_0,\Pi_1]$, $B(\Pi_0) = 0$ and $B(\Pi_1) = 1$; so there must exist a unique fixed point $\pi^m \in [\Pi_0,\Pi_1]$ such that $B(\pi^m) = \pi^m$, and hence the *NE* (π^m,π^m) exists, though it is not stable since the slope of $B(\Pi)$ is bigger than one when crossing the 45^0 line. It is easy to check that $(0,0)$ and $(1,1)$ are the other two *NE*s and both are stable.

References

ALESINA, A., AND E. LA FERRARA [2002], "Who Trusts Others?" *Journal of Public Economics*, 85, 207-34.

ANDREONI, J., AND R. CROSON [2008], "Partners versus Strangers: Random Rematching in Public Goods Experiments," in: C. Plott and V. Smith (eds.), *Handbook of Experimental Economics Results*, Amsterdam: North-Holland.

ARROW, K. [1972], "Gift and Exchanges," *Philosophy and Public Affairs*, I, 343-62.

AXELROD, R. [1984], *The Evolution of Cooperation*, Basic Books: New York.

BAR-GILL, O., AND C. FERSHTMAN [2005], "Public Policy with Endogenous Preferences," *Journal of Public Economic Theory*, 7(5), 841–857.

BERG, J., J. DICKHAUT AND K. MCCABE [1995], "Trust, Reciprocity, and Social History," *Games and Economic Behavior*, 10(1), 122-142.

BOHNET, I., B. FREY, AND S. HUCK [2001], "More Order with Less Law: On Contract Enforcement, Trust, and Crowding," *American Political Science Review*, 95(1), 131-144.

BRENNAN, G., W. GÜTH, AND H. KLIEMT [2003], "Trust in the Shadow of the Courts," *Journal of Institutional and Theoretical Economics*, 159, 16–36.

BURLANDO, R., AND J. D. HEY [1997], "Do Anglo-Saxons Free-ride More?" *Journal of Public Economics,* 64, 41-60.

CAPPELLI, P. [1995], "Is the 'Skill Gap' Really About Attitudes?" *California Management Review*, 37(4), 108-124.

COLEMAN, J. S. [1988], "Social Capital in the Creation of Human Capital," *American Journal of Sociology,* 94, S95-S120.

COOK, K. S. [2001], *Trust in Society,* eds., New York: Russell Sage Foundation.

DIXIT, A. [2003], "On Modes of Economic Governance," *Econometrica*, 71(2), 449-481.

FEHR, E., AND S. GACHTER [2000], "Fairness and Retaliation: The Economics of Reciprocity," *Journal of Economic Perspectives,* 14, 159-182.

FRANK, R. H. [1987], "If Homo Economicus Could Choose His Own Utility Function, Would He Want One With a Conscience?" *American Economic Review,* 77(4), 593-604.

– – [1988], *Passions Within Reason: The Strategic Role of Emotions*, New York: W.W. Norton.

– – [1992], "Melding Sociology and Economics: James Coleman's Foundations of Social Theory," *Journal of Economic Literature,* XXX, 147-170.

GARCES, E., D. THOMAS, AND J. CURRIE [2002], "Longer-Term Effects of Head Start," *American Economic Review*, 92(4), 999-1012.

GLAESER, E. L., D. LAIBSON, AND B. SACERDOTE [2002], "The Economic Approach to Social Capital," *The Economic Journal*, 112(483), 437-458.

– –, D. Laibson, J. A. Scheinkman, and C. L. Soutter [2000], "Measuring Trust," *Quarterly Journal of Economics*, 115(3), 811-846.

Greif, A. [1989], "Reputation and Coalitions in Medieval Trade: Evidence on the Maghribi Traders," *Journal of Economic History*, 49, 857–882.

Güth, W., H. Kliemt, and B. Peleg [2000], "Coevolution of Preferences and Information in Simple Games of Trust," *German Economic Review*, 1, 83–110.

– –, and A. Ockenfels [2005], "The Coevolution of Morality and Legal Institutions: An indirect evolutionary approach," *Journal of Institutional Economics,* 1(2), 155-174.

Hardin, R. [2002], *Trust and Trustworthiness*, New York: Russell Sage Foundation.

Heckman, J. J. [2000], "Policies to Foster Human Capital," *Research in Economics*, 54, 3-56.

Huang, F. [2006], "To Trust or to Monitor: A Dynamic Analysis," Working Paper, School of Economics, Singapore Management University, Singapore.

Huck, S. [1997], "Preferences and institutions: An Evolutionary Perspective," *Journal of Institutional and Theoretical Economics*, 153, 771-779.

– – [1998], "Trust, Treason, and Trials: An Example of How the Evolution of Preferences Can Be Driven by Legal Institutions," *Journal of Law, Economics, and Organization,* 14, 44-60.

James, H. S. [2002], "The Trust Paradox: A Survey of Economic Inquiries into the Nature of Trust and Trustworthiness," *Journal of Economic Behavior and Organization*, 47(3), 291-307.

Judge, T. A., and R. Ilies [2002], "Relationship of personality to performance motivation: A meta-analytic review," *Journal of Applied Psychology,* 87(4), 797-807.

Kandel, E., and Lazear, E. P. [1992], "Peer pressure and partnerships," *Journal of Political Economy,* 100(4), 801–817.

Knack, S., and P. Keefer [1997], "Does Social Capital Have an Economic Payoff? A Cross-Country Investigation," *Quarterly Journal of Economics*, CXII, 1251-1288.

Kreps, D. [1997], "Intrinsic Motivation and Extrinsic Incentives," *American Economic Review*, 87(2), 359-364.

– –, P. Milgrom, J. Roberts, and R. Wilson [1982], "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma." *Journal of Economic Theory,* 27, 245-52.

La Porta, R., F. Lopez-de-Silanes, A. Shleifer, and R.W. Vishny [1997], "Trust in Large Organizations," *American Economic Review*, 87(2), 333-38.

Palfrey, T. R., and J. E. Prisbrey [1997], "Anomalous Behavior in Public Goods Experiments: How Much and Why?" *American Economic Review*, 87(5), 829-846.

Putnam, R. D. with R. Leonardi and R.Y. Nanetti [1993], *Making Democracy Work: Civic Traditions in Modern Italy*, Princeton, NJ: Princeton University Press.

– – [1995], "Bowling Alone: America's Declining Social Capital," *Journal of Democracy,* 6, 65-78.

ROB, R., AND P. ZEMSKY [2002], "Social Capital, Corporate Culture and the Incentive Intensity," *RAND Journal of Economics,* 33(2), 243-257.

ROTEMBERG, J. J. [1994], "Human Relations in the Workplace," *Journal of Political Economy,* 102, 684-718.

STAUB, E. [1978], *Positive Social Behavior and Morality,* New York : Academic Press.

TIROLE, J. [1996], "A Theory of Collective Reputations," *Review of Economic Studies,* 63, 1-22.

WEIMANN, J. [1994], "Individual Behavior in a Free Riding Experiment," *Journal of Public Economics,* 54, 185-200.