Social Capital, Cooperation, and Human Capital

Fali Huang*

Department of Economics

The University of Pennsylvania

November 6, 2002

Abstract

In this paper I study the formation of social capital and its effects in a game theoretic setting. I formalize the concept of social trust and show that appropriate social trust enables strangers to cooperate in a one-period prisoner's dilemma. The relationship between several widely used forms of social capital is characterized. The analysis also sheds lights on the strong externality of the social component of human capital among people and suggests an important link between human capital and social capital. Social trust is determined in equilibrium by the aggregate choices of optimizing individual players. Multiple equilibria are possible, which implies that social capital levels may be history dependent. Those people with highest investment costs play a crucial role in determining whether there exists under-investment in social trust. The model suggests several ways to improve long run social trust. It provides new insights in the complex relationship between formal institutions and social capital. It also shows the importance of families, schools, and mass media in affecting the formation of social trust in the society. The paper provides some plausible explanations for many stylized facts in the empirical literature on social capital.

*I am grateful to George Mailath, Andrew Postlewaite, and Rafael Rob for their support. I also thank Hanming Fang, Mokoto Hanazono, Volker Nocke, Dan Silverman, and Huanxing Yang for comments and suggestions. All remaining errors are mine.

1 Introduction

As Arrow (1972, p.357) has observed, "Virtually every commercial transaction has within itself an element of trust, certainly any transaction conducted over a period of time. It can be plausibly argued that much of the economic backwardness in the world can be explained by the lack of mutual confidence." This generally aggreed view is supported by recent empirical studies. Knack and Keefer (1997) present evidence that average trusting level $TRUST$ and civic cooperation norm $CIVIC$ based on the World Value Survey are significantly associated with economic growth. La Porta et al. (1997) find that the effects of $TRUST$ on performance of various organizations in a society are both statistically significant and quantitatively large.

On the theoretical front the term 'social capital' has recently been coined to conceptualize the cooperative ability of a society in promoting social welfare (Coleman 1988, Putnam 1993, 1995). As Coleman has emphasized, social capital's value to a society parallels human capital's value to the individual. Putnam (1993) refers social capital to "features of social organization such as networks, norms and social trust that facilitate coordination and cooperation for mutual benefit."

Though many empirical investigations about social capital exist in the literature, theoretical research is still in its infancy. There are at least two obstacles to formal analysis of social capital. One is that social capital, though intuitively appealing, is largely a 'buzzword' that is difficult to pin down conceptually. The other obstacle is that social capital often represents group level characteristics, but the group, such as a loosely organized community, is not an optimizing unit in general to invest in social capital.

In this paper, we show that an appropriately defined social trust concept, together with a game theoretic model, overcomes these two barriers and allows us to study the formation of social capital at a societal level. A discussion of intuition and justifications for our approach follows.

In our view the various forms of social capital at a societal level should not be lumped together as conceptually homogenous species. Instead, social trust is the common element unifying them under the name of social capital. That is, social trust captures the essence of social capital. We show in this paper that the formalized concept of social trust can be used to characterize the relationships among a variety of frequently used examples of social capital.

A game theoretic setting is required since social capital "exists in the relations among persons" (Coleman 1988). Furthermore, note that the concept of social *trust* is vacuous without the discrepancy between social and individual returns, since otherwise rational people can always be 'trusted' to choose their optimal actions.[1] The prisoner's dilemma thus seems to be the very context where social trust matters. In particular, we will focus on a one-period prisoner's dilemma game.[2]

The experimental literature demonstrates that some people rationally choose to cooperate in one-period public goods games because they may get utility from the very act of behaving cooperatively (Palfrey and Prisbrey 1997, Andreoni and Croson 2002). In this paper we label this taste for cooperation as a person's *cooperative tendency*. The *distribution* of cooperative tendency among players in a group is interpreted as *social trust* in the group. In this way social trust as a group level characteristic is naturally linked to the individual members of the group. These two definitions also characterize the relationships between several often-used trust concepts like trustworthy and trusting level.

When people with different cooperative tendencies randomly match with each other in a one-period prisoners' dilemma, those with high enough cooperative tendencies will cooperate, and the cooperation level in a group with higher social trust will be higher. In this way both individual and total outputs are increased. The exact quantitative effects of the same social trust on outputs, however, depend on detailed specifications of the game, including game payoffs, information structure, and duration. As a result the same social trust level can correspond to different cooperation levels across games, a fundamental reason for the non-trivial discrepancies among the various forms of social capital.

Then cooperative tendency is endogenized to study the formation of social trust within a society. Specifically, it is treated as a component of human capital which is distinct from cognitive ability. Both cooperative tendency and cognitive ability can increase a player's payoff over a life time, and they can be complements or substitutes to each other depending on some parameters. We develop a human capital investment game where, taking as given the expected social trust in the society, each player chooses his cooperative tendency and cognitive ability to maximize lifetime utility. In this game the equilibrium social trust is very likely to be inefficient because investment in cooperative tendency has strong positive externality on social outputs. Multiple

[1] See Hardin (2001) for more evidence supporting this usage of 'trust'.

[2] This is consistent with the general usage of *social* trust among strangers rather than aquaintances, friends, and family members.

equilibrium is possible in all cases, which implies that long-run social trust levels may be history dependent.

The model suggests that families, schools, and early intervention programs may be important in affecting the formation of social trust in the society, since they can reduce investment costs by nurturing cooperative tendency in a person's formative years. As the efficiency of information flow goes up and people more easily learn each other's cooperative tendencies, social trust also increases. The model provides new insights into the complex relationship between formal institutions and social capital. It shows that although people's cooperative tendency levels may be crowded out by formal institutions, the proportion of cooperative people in the society would increase.

The paper provides some plausible explanations for many stylized facts in the empirical literature on social capital. For example, it shows how social trust improves individual and social outputs in prisoner's dilemmas. It quantifies the discrepancies among several widely used empirical measures of social capital – $TRUST$, $CIVIC$, and organization membership $MEMBER$,[3] and provides an explanation for the fact that the trust indicator based on survey questions is not strongly related to, and sometimes even contradicts, the trust measure in public goods experiments. The behavioral pattern predicted by the paper is also consistent with the experimental evidence that cooperation levels in public goods games differ across subjects and games (Glaeser et al. 2000b). The paper also clarifies the working mechanisms of several forces that have been empirically linked with the recent decline of social capital in the US.

The current paper sheds new lights on the relationship between social trust and repeated games. It shows that appropriate social trust generates cooperation even in a one-period prisoner's dilemma, which identifies the unique role of social trust in generating cooperation.[4] The repeated games, however, can increase the effects of *existing* social trust on individual and total outputs under incomplete information, and may provide incentives for people to invest in *future* social trust if the repeated

[3] In their studies of social capital, Knack and Keefer (1997) use $TRUST$, $CIVIC$; La Porta et al (1997) use $TRUST$; and Putnam (1993) and Glaeser et al. (2000a) use $MEMBER$.

[4] Note that on the other hand, repeated games can generate cooperation without social trust. When we look at each stage game isolated from the repeated process, it seems there exists social trust among players that enable them to cooperate. However, in many cases the true motivation for cooperation in repeated games is not social trust but rational calculation of the rewards and punishments associated with repeated interactions among players.

interactions help to reveal people's types.

By treating cooperative tendency as a component of human capital, this paper is related to the human capital literature where several recent works have shown that incentive-enhancing preferences are important in determining individual earnings (Heckman 2000, Bowles et al. 2001).[5] Furthermore, we show that these preferences may have strong positive externality on aggregate welfare and suggest that investment inefficiency may be more severe than in the case of conventional human capital.

The paper complements the work of Rob and Zemsky (2002) which studies social capital in a firm's environment. Taking the initial stock of social trust among employees as exogenously given, they show firms could use incentive structures to affect employees' preferences and thus foster social capital at the firm level. The current paper, taking a life-time perspective, endogenizes players' heterogenous predisposition to cooperate in the context of the whole society.

A closely related paper is by Glaeser et al. (2000a), where an individual capital investment model is used to study social capital formation. The current paper differs from theirs in several important aspects. First, they do not distinguish between various forms of social capital. Second, they treat social capital at group level as the simple sum of 'individual social capital', and thus ignore the important externality among players. Third, individual players in their model make investment decisions in isolation from each other rather than in a game theoretic setting.

The paper is organized as follows. In the next section social trust and cooperative tendency are formally defined in the one-period prisoner's dilemma context, and their effects on individual and aggregate outputs are analyzed. They are endogenized in section three, where players invest in human capital to maximize their life time utility, taking as given the expected social trust level. The comparative statics and their empirical implication in improving social trust are also discussed. The final section presents conclusions.

[5] For example, in National Employers Survey 1997, the most important criterion used by employers in their decision to hire employees is "attitude", scoring 4.6 out of maximum 5. In comparison, the "score on tests given by employer" and "academic performance" are both at 2.5.

2 The Effects of Social Trust on Outputs

2.1 The Basic Set Up

2.1.1 Payoff Assumptions in Prisoners' Dilemma

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

		player j	
		C	D
player i	C	(g,g)	$(-l,g+d)$
	D	$(g+d,-l)$	$(0,0)$

where C is cooperate or exert efforts, D is defect or not exert effort, $i,j \in [0,1]$; and $g,l,d,> 0$ represent outputs produced by the players. The letters g,l are set to represent *g*ain and *l*oss from making cooperative efforts respectively, and d is for extra gain from *d*efecting (by not making efforts) when the other player makes efforts.

To make the game interesting for our purpose, we make two assumptions about the levels of the payoffs:

$$\begin{aligned} d &< l, & (1) \\ g+d-l &> 0. & (2) \end{aligned}$$

These two assumptions are standard in the literature (Kreps et al 1982, Rotemberg 1994, Bar-Gill and Fershtman 2000). The rationale behind them is as follows. The first assumption means that a player behaving cooperatively lowers his partner's cost of acting similarly. When a player i plays C, his partner j gets g if playing C and $g+d$ if playing D. The difference of the two payoffs, d, is the marginal cost or net loss of j playing C when i also plays C. Using a similar argument we get that when i defects, player j's marginal cost of playing C is l. So d and l represent marginal costs of making cooperative efforts under two different situations: in one the partner cooperates, in the other he does not. The assumption $d < l$ thus implies that a player's cooperative behavior has positive externality on his partner's incentive to behave cooperatively.

The second assumption, meaning that cooperative behavior always improves aggregate output, implies two conditions. The obvious one is that both players exerting effort yields a higher payoff than only one player exerting effort. This is true iff $2g > g + d - l$, a condition already satisfied by our first assumption. The more demanding condition implied by the second assumption is that exerting effort unilaterally as in (C, D) or (D, C) is better than both defecting (D, D), which requires $g + d - l > 0$. This is reasonable since efforts made by even one player should have higher productivity than no efforts at all.

2.1.2 Cooperative Tendency and Social Trust

We assume that player i incurs a disutility $\alpha_i \in R^+$ when not exerting effort, for $i \in [0, 1]$.[6] Specifically, the utility of player i matched with a partner j is

$$u_i(a_i, a_j) = m_i(a_i, a_j) - \alpha_i \chi_D(a_i),$$

where $a_i, a_j \in \{C, D\}$ are the actions of player i and j; $m_i(a_i, a_j)$ is the material payoff for player i; and $\chi_D(a_i)$ is an index function such that

$$\chi_D(a_i) = \begin{cases} 1 & \text{if } a_i = D \\ 0 & \text{if } a_i = C. \end{cases}$$

This kind of preference is characterized by *warm-glow* motivation: people often derive utility from the very act of cooperating, independent of the exact utility their cooperative behavior delivers to others. Palfrey and Prisbrey (1997) show that the warm-glow effect is highly significant in inducing cooperation in public good experiments. As we will show below the higher the α_i, the more likely in general player i is to behave cooperatively. In this sense α_i measures player i's utility of warm-glow or his taste for cooperation. We thus define α_i as the *cooperative tendency* of player i.

Note that cooperative tendency is a stable characteristic of a person and an internal discipline against defecting. It acts as a life-time commitment enabling its owner to cooperate in situations where cooperation is otherwise impossible.[7] Coop-

[6] Note that the disutility of defecting is a natural component of the payoff players get from the game. There are several reasons why we single it out from other payoffs in the paper. First, it is in general not directly observable as material payoffs are. Second, it is not determined by a specific game but systematically associated with a single player across different games. Third, we would endogenize it in the paper to study the formation of social trust.

[7] In this sense it has a similar role as tit-for-tat in Kreps et al (1982), but we would argue that cooperative tendency is a better analytical tool than tit-for-tat. First, players with positive cooperative tendencies are still rational in that they always aim to maximize their utility, while tit-for-tat

erative tendency and trust are closely related in that players with higher cooperative tendency are more *trustworthy* in real life.[8]

The distribution of cooperative tendency α_i in the population, denoted by $F(\cdot)$, is defined as *social trust* among players. This definition captures the intrinsic relationship between cooperative tendency (trustworthiness) at an individual level and social trust at an aggregate level, enabling us to clarify the relationships between several trust concepts. It is also a useful analytical tool, as will become clear below, for studying various forms of social capital and their innate relationships.

2.2 The Effects of Social Trust

In this section we study the effects of social trust on total and individual outputs in various games. Specifically we would prove the following proposition:[9]

Proposition 1 *Social trust could induce cooperation even in a one-period prisoners' dilemma. It strictly increases both total outputs and (at least weakly) the expected individual outputs for all players. The exact effects of social trust on outputs depend on game specifications.*

2.2.1 One-Period Complete Information Game

Suppose players' cooperative tendencies are observed publicly. They are randomly matched with each other to play a one-period game where the material outputs are the same as in the above prisoners' dilemma. With the introduction of α in players' payoff function, the game is different for players of different cooperative tendencies, even though the material outputs g, l,and d are the same as in the above game.

players rigidly and blindly follow a certain behavior rule regardless of its associated utility. Second, cooperative tendency is more general and flexible than tit-for-tat. In certain situations, say finitely repeated games, players with appropriate cooperative tendencies may act like they are of the tit-for-tat type. Thus tit-for-tat is only one specific manifestation of the cooperative tendency preference. Third, tit-for-tat preference cannot reasonably explain cooperation in one-period priosoners' dilemma, a phenomenon at the heart of social trust.

[8] We believe our definition of cooperative tendency captures the essence of trustworthiness, however, since in daily life the latter term has various connotations, we choose to use *cooperative tendency* to reduce possible confusion.

[9] Technical details about the effects on outputs are in the appendix.

Take for example the game between a player Ann with cooperative tendency α_A and a player Mike with α_M:

		Mike	
		C	D
Ann	C	(g, g)	$(-l\ , g + d - \alpha_M)$
	D	$(g + d - \alpha_A, -l)$	$(-\alpha_A, -\alpha_M)$

In this game when Mike plays C Ann would choose to play C if $g \geq g + d - \alpha_A$ holds, which is true iff $\alpha_A \geq d$. When Mike plays D Ann would also play D iff $\alpha_A \leq l$ or play C iff $\alpha_A > l$.[10] Since the game is symmetric, Mike has the same best response function.

It is clear that d and l are the two thresholds dividing the parameter space of cooperative tendency into three ranges: $[0, d), [d, l], (l, \infty)$.[11] A player will never cooperate in this game if her cooperative tendency is in the lowest range $[0, d)$. She would be discreet in her best response if she falls into the middle range $[d, l]$: cooperate if her partner does so, defect if her partner defects.[12] In other words, she behaves reciprocally in this range. She will always cooperate if her α is higher than l, regardless of what her partner does.

Thus in a given game under complete information, players could be categorized into three corresponding behavioral types: the *selfish* type who always defects, the *reciprocal* type who makes in-kind responses to what her partner does, and the *selfless* type who always cooperates.[13] The latter two types are *non-selfish.*

Let π_R denote the proportion of the reciprocal type, π_S the selfless type, then the remaining $1 - (\pi_R + \pi_S)$ is the proportion of the selfish type players. The social trust in this case could be characterized by (π_R, π_S). By definition $\pi_R \equiv \Pr(\alpha_i \in [d, l])$,

[10] Note that we assume, without loss of generality, that when a player is indifferent between C and D, she always chooses the same action as that of her partner.

[11] This is not coincidence. Note that d is the lowest cost of behaving cooperatively in the game while l is the highest. A player in her best response would always make her marginal cost of an extra unit of cooperative effort equal to her cooperative tendency.

[12] In a continuous version the only difference is that a player with $\alpha \in [d, l]$ would make more cooperative efforts if her cooperative tendency is higher while keeping the marginal cost of extra effort equal to her α.

[13] Many experimental studies have found that between 40 and 66 percent of subjects exhibit reciprocal behaviors, while between 20 and 30 act completely selfish. Selfless type players are relatively rare (Fehr and Gachter 2000). Note that the assumption $d < l$ is crucial for players to demonstrate reciprocity in the game.

$\pi_S \equiv \Pr(\alpha_i > l)$, where $\alpha_i \sim F(\cdot)$. This implies that in games with different defecting benefits (d, l), the same social trust would have different manifestations.

The pure strategy Nash equilibria for the one-period complete information games are described in the following claim:

Claim 1 *In the above one-shot games under complete information, (C, C) and (D, D) are the two pure strategy Nash equilibria when both players are reciprocal. The unique pure strategy Nash equilibrium is (D, D) when both players are selfish; (C, C) when both of them are non-selfish and at least one is selfless; (C, D) when the two players are (selfless, selfish).*

Proof. When both Ann and Mike are reciprocal, which means $\alpha_A, \alpha_M \in [d, l]$, both would play C if the other one plays C since $\alpha_A, \alpha_M \geq d$. Similarly, both will play D if the other plays D because $\alpha_A, \alpha_M \leq l$ by assumption. So (C, C) and (D, D) are the two Nash equilibria if the two players are both reciprocal. If Ann is selfless while Mike is reciprocal, Ann's 'always cooperate' strategy would induce Mike to cooperate reciprocally. Thus the only Nash equilibrium is (C, C).

When both players are selfish, they always defect regardless of circumstance, so (D, D) is the only Nash equilibrium. On the contrary, two selfless players would always cooperate, and the unique Nash equilibrium is (C, C). If Ann is selfless and Mike is selfish, both players will still play their dominant strategy C and D respectively, meaning that the only Nash equilibrium is (C, D). ■

Remark 1 *Between the two Nash equilibria (C, C) and (D, D) for two reciprocal players, each can unilaterally avoid (D, D) by always playing C when the partner is known to be reciprocal. So individual utility maximization would essentially eliminate (D, D) and leaves the Pareto dominant equilibrium (C, C) as the only one ever played between two reciprocal players. Without much loss of generality, we will focus our attention on (C, C) in the subsequent discussion.*

The expected material outputs for these three types of players are: $\pi_S(g + d)$ for selfish players, $(\pi_R + \pi_S)g$ for reciprocal players, and $(\pi_R + \pi_S)(g + l) - l$ for selfless players.[14] The individual outputs of all three types strictly increase with social trust (π_R, π_S). The selfless players get the highest marginal benefit from social

[14] In this game only non-selfish players produce output by making efforts, while selfish players unfairly get benefits from the unilateral efforts of selfless players.

trust, though in terms of output levels they fare less well than the reciprocal types in that they can not avoid being taken advantage of by selfish players. The total output for the population, $Q_1^r \equiv (\pi_R + \pi_S)^2 g + \pi_S(1 - \pi_R - \pi_S)(g + d - l)$, also strictly increases with social trust (π_R, π_S).

An alternative matching system is assortative matching by type, where a selfish player matches only with another selfish player and vise versa. Now all non-selfish players get g, while all selfish ones get zero. The total output in this case, $Q_1^a \equiv g(\pi_R + \pi_S)$, is higher than Q_1^r. It strictly increases with social trust (π_R, π_S), while the individual outputs only weakly increase with it. The intuition is that, for an individual player's income, social trust matters only through his partner's cooperative tendency. For aggregate output, however, the total number of cooperative players also plays an important role.

2.2.2 One-Period Incomplete Information Game

Under incomplete information players' cooperative tendencies are private information. The social trust $F(\cdot)$ is common knowledge and assumed to be continuous. Now players can no longer use different strategies corresponding to their partners' types. Whether players choose to cooperate or not depends only on their own cooperative tendency and the publicly known social trust.[15]

Claim 2 *In the one-period game under incomplete information, the Bayesian Nash equilibrium is "all players with $\alpha_i \geq \pi d + (1 - \pi)l$ play C, others play D," where π is the proportion of cooperative players in the game, uniquely determined by the equation $\pi + F(\pi d + (1 - \pi)l) = 1$.*

Proof. In the game a player i's probability of matching with a cooperative partner is π, and with a defecting partner $1 - \pi$. By playing C she would get g if her partner is also cooperative, $-l$ if her partner is defecting. So her expected payoff from playing C is $\pi g - (1 - \pi)l$. By playing D her expected utility is $\pi(g + d - \alpha_i) - (1 - \pi)\alpha_i$. Thus a player will play C iff it brings her higher utility, i.e. $\pi g - (1 - \pi)l \geq \pi(g + d - \alpha_i) - (1 - \pi)\alpha_i$. This condition holds iff $\alpha_i \geq \pi d + (1 - \pi)l$.

To guarantee that the belief π is consistent with players' strategies, it must be true that $\pi = \Pr(\alpha_i \geq \pi d + (1 - \pi)l) \equiv 1 - F(\pi d + (1 - \pi)l)$. The RHS is continuous

[15] If the trust in a group is revealed gradually, say after each round of the repeated one-shot games, our model may be used to explain some results documented in the public good experiments (see the survey by Andreoni and Croson 2002).

in π on the closed interval $[0,1]$. It increases with π since $\frac{\partial RHS}{\partial \pi} = (l-d)DF \geq 0$. Furthermore, $RHS(\pi = 0) = 1 - F(l) > 0$ and $RHS(\pi = 1) = 1 - F(d) < 1$. So π is uniquely determined. ■

Remark 2 *Note that $\pi d + (1-\pi)l > d$ when $\pi < 1$. So the minimum cooperative tendency to induce cooperative behavior is now higher, or equivalently the proportion of cooperative players is smaller, under incomplete information than that under complete information.*

Under incomplete information the social trust could be characterized by π, the proportion of players who cooperate in the game.[16] We still categorize the players with $\alpha_i < \pi d + (1-\pi)l$ the selfish type and those with $\alpha_i \geq \pi d + (1-\pi)l$ non-selfish. Again the manifestation of the same social trust differs across games and is determined by (d,l) in each game.

The expected output for a selfish player, $G_1^M \equiv \pi(g+d)$, is higher than that of a non-selfish player, $G_1^A \equiv \pi g - (1-\pi)l$.[17] The individual outputs for both types, however, strictly increase with social trust. And non-selfish players get higher marginal benefit from social trust than that of selfish ones. The total output in this game, $Q_1^I = (l-d)\pi^2 + (g+d-l)\pi$, also strictly increases with social trust.

Note that when there is no player with cooperative tendency in the range $[d, \pi d + (1-\pi)l)$, the proportions of non-selfish players under both incomplete and complete information are the same, i.e. $\pi = \pi_R + \pi_S$. In this case $Q_1^r < Q_1^I < Q_1^a$. The first inequality holds because under incomplete information the reciprocal players have to cooperate even when they are matched with selfish players, while by assumption (C,D) generates higher output than (D,D). The implication is that lack of information increases total outputs at the cost of reciprocal players. The second inequality means that random matching is inferior to assortative matching in terms of total output, since (C,C) is more productive than (C,D).

[16] In a continuous version, however, the level of cooperative tendencies higher than $\pi d + (1-\pi)l$ and lower than l would also matter. This applies to similar situations and will not be mentioned everytime.

[17] Here again, though it is still true that only non-selfish players produce output by exerting effort, the selfish players get the most out of this cooperation. As will be clear soon, we need repeated games to induce selfish players to exert effort.

2.2.3 T-period incomplete information game

The above analysis has shown that in one-shot games social trust improves total outputs by enabling non-selfish players to make efforts in an otherwise prisoners' dilemma environment. Now we will show that in repeated games social trust can elicit cooperative behavior even from selfish players. In particular, a sequential equilibrium in a finite T-period game is characterized and used to illustrate the important interaction between social trust and reputation effect.[18]

Suppose players are randomly paired to play the above stage game for finite $T \geq 2$ periods. Each and every pair lasts for all the T periods after they are matched.[19] We still assume that social trust is common knowledge, but a specific player's cooperative tendency is not publicly observed under incomplete information. After each period the actions taken during the period are known to the players. Let $\beta \in [0, 1]$ represent the time discount factor for all players. Again we define players with $\alpha_i \geq \pi d + (1-\pi)l$ non-selfish, and others selfish. Among non-selfish players those with $\alpha_i \geq l$ are still the selfless.

Claim 3 *In the T-period game described above, the following strategy profile and belief system is a sequential equilibrium if* $\beta \geq \frac{d}{(g+d)(\pi-\pi_S)}$ *and* $\pi - \pi_S \geq \frac{d}{g+d}$*, where* π *is the solution to the equation* $\pi + F(\pi d + (1-\pi)l) = 1$*, and* $\pi_S \equiv 1 - F(l)$*. The strategy profile is: (1) Selfless players always play C; all other non-selfish players take the strategy "play C first; play C if* (C, C) *is played in the previous period, play D otherwise." (2) All selfish players' strategy is "play C first; at any period* $1 < t < T-1$*, play C if* (C, C) *is played in the previous period, play D otherwise; play D at period T." The belief system is: (1) In the first period and every period following the history which only* (C, C) *has been played, every player assigns probability* π *to his partner being non-selfish. (2) In all the following periods after the first time* (C, D)

[18] A finite-period game is used because it is well known that even selfish players can cooperate among themselves in infinitely repeated games. Actually cooperation can also be achieved in a finitely repeated prisoners' dilemma game if one player plays tit-for-tat (Kreps et al. 1982), and indeed a similar theorem could be reproduced in our setting. So one characterized equilibrium can suffice our purpose.

[19] This assumption is for simplicity of exposition and not essential to the result. In fact, the same result holds if players can exchange partners, as long as all players' actions in the history are observed by their possible partners. In equilibrium, there is no difference regardless of whether partners are changed or not, since everybody acts the same until the very last period.

is observed, the player who has played D is believed to be selfish, the player who has played C is still believed to be non-selfish with probability π.

Proof. In appendix. ■

The expected material payoff of a non-selfish player in this sequential equilibrium is again, because of the incomplete information, smaller than that of a selfish player. However, selfish players now have incentives to cooperate until the last period because of the reputation effect, which greatly increases the effects of social trust on both individual and total outputs. Indeed, the total output in the above sequential equilibrium may be higher than the maximal output under complete information.

This example shows that repeated games do *not creat* trust among current players, rather they increase the effects of *existing* social trust on generating cooperation and improving outputs.

2.3 Social Trust and Social Capital

Since the game could be interpreted as the representative prisoners' dilemma game a person encounters in real life, the qualitative results generated above could be applied not only to the environment of a certain game, but also to an organization, a community, or even the society as a whole. Now we will use the results obtained in the above section to explain the relationships among various social capital and social trust concepts used in the literature. We will show that our definition of social trust is quite useful in analyzing social capital.

2.3.1 Trusting Level, Individual and Average Trustworthiness

In a prisoners' dilemma game γ with corresponding payoffs g_γ, d_γ, l_γ, let π_γ denote the proportion of non-selfish players in the population. When players are randomly matched to play the game γ under incomplete information, π_γ is the probability that a generic partner is non-selfish. In this sense π_γ measures the trustworthiness of an average person in the prisoners' dilemma γ. It indicates to what extent a rational player should trust other people to behave cooperatively in the game. Indeed, the *average trustworthiness* of a group member in game γ, π_γ, is exactly the *trusting level* of all group members when social trust is public knowledge.[20]

[20] In reality people usually do not know the true degree of social trust among relevant players, and their trusting level is equal to their expected / perceived average trustworthiness of other players.

On the other hand the cooperative tendency of a player measures his own trustworthiness, which is quite stable over time. In contrast his trusting level of others differs across games and groups of players.[21] At the individual level a player's trustworthiness and her trusting level can be very different. For example, a player with low α_i may behave selfishly in the game making himself untrustworthy, but he may highly trust others to cooperate if π_γ is high. On the contrary a non-selfish player could be trusted to cooperate even though her trusting level (equal to π_γ) is low.

This difference, however, does not carry over to the aggregate level since $\pi_\gamma = \Pr(\alpha_i \geq \pi_\gamma d_\gamma + (1 - \pi_\gamma) l_\gamma)$ according to Claim 2. This condition means that only when there are many trustworthy players would the trustworthiness of an average player be high.

2.3.2 Variations of Social Capital: $TRUST$, $CIVIC$, $MEMBER$

If the representative game in a country is game γ, then the widely used trust indicator, $TRUST$, corresponds to π_γ in our model. $TRUST$ is equal to the percentage of respondents in each nation replying "most people can be trusted" in the World Values Survey when they are asked the trust question: "Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?"

Note that since π_γ is the proportion of non-selfish players in the population, it is equal to the proportion of people who indeed are matched with a non-selfish player in game γ. If players randomly matched with each other to play game γ are asked the same trust question, what would be the percentage of players replying "most people can be trusted"? It would be π_γ, since exactly π_γ percentage of players meet a partner that can be trusted. That is $TRUST = \pi_\gamma$.

Another trust measure, $CIVIC$, roughly corresponds to people's average cooperative tendency. The World Values Survey asks people whether they think, on a scale of 1-10, defecting behaviors in five prisoners' dilemma situations "can always be justified, never be justified or something in between."[22] These situations can be characterized by complete information games with cooperative partners, and people's answers

[21] See evidence in Glaeser et al. (2000b).

[22] The five prisoners' dilemmas are: A) "claiming government benefits which you are not entitled to," B) "avoiding a fare on public transport," C) "cheating on taxes if you have the chances," D) "keeping money that you have found," and E) "failing to report damage you've done accidentally to a parked vehicle."

should reflect their cooperative tendencies. Suppose people say 10 when their $\alpha_i \geq d_q$ in each question q, and say $\max(1, \frac{10\alpha_i}{d_q})$ if otherwise, where $q = 1, 2, ..., 5$. Then $CIVIC = 10\sum_{q=1}^{5}[\pi_q + \int \max(1, (\frac{\alpha_i}{d_q}|\alpha_i < d_q))di]$, where $\pi_q = \Pr(\alpha_i|\alpha_i \geq d_q)di$.[23]

Now we discuss what is measured by organization membership, $MEMBER$. Suppose joining an organization m corresponds to a prisoners' dilemma game γ_m associated with d_m, where $d_m < d_{m+1}$, and $m = 1, ..., M$. Let $\pi_m = \Pr(\alpha_i \geq d_m)$, then the proportion $(\pi_m - \pi_{m+1})$ of players would join the organizations with index up to m. Therefore the average number of memberships in the population is equal to $MEMBER = \sum_{m=1}^{M-1} m(\pi_m - \pi_{m+1})$.

So the three social capital measures, $TRUST$, $CIVIC$, $MEMBER$, actually represent the same social trust $F(\cdot)$ in different game contexts. This is the underlying reason why there are discrepancies among them. The relationship between social trust and these forms of social capital is like that between human capital and its returns in different jobs. An empirical implication is that when using these measures of social capital to study the levels, effects, and formation mechanisms of social trust, we should take the underlying game contexts into consideration.

2.3.3 Trust Measures in Public Goods Experiments

The experimental literature also provides various measures of social trust. If we denote a public goods experiment as a game γ_P with (d_P, l_P), then the number of cooperative subjects is $\pi_P = \Pr(\alpha_s \geq \pi_P d_P + (1-\pi_P)l_P)$ and the average cooperation level roughly corresponds to the average cooperative tendency of non-selfish players: $\int(\alpha_s|\alpha_s \geq \pi_P d_P + (1-\pi_P)l_P)ds$, where $s \in \{1, 2, ..., S_P\}$ is the index of the S_P subjects in the experiment γ_P. If the distribution of cooperative tendency of the subjects in the experiment γ_P is a random sample drawn from the distribution $\{\alpha_i, i \in [0, 1]\}$ of the whole population, and its (d_P, l_P) is equal to (d_γ, l_γ) of the representative game γ in the society, then π_P is an unbiased estimate of $TRUST$ in game γ. However, this is usually not the case since a) most subjects are college students who in many countries may not be a representative group, b) typically the public game γ_P is designed by researchers to be exactly the same across countries, but it is very probable that the representative game γ in each country is different.

[23] If for some question q the net gain of defecting d_q is so high that $\pi_q = 0$, then $CIVIC_q = \frac{10}{d_q}\int \alpha_i di$, where $CIVIC_q$ is the average response to question q. In the other extreme, if d_q is so low that $\pi_q = 1$, then $CIVIC_q = 10$.

Indeed, there are discrepancies and even contradictions between the trust measures in the public goods experiments and $TRUST$.[24] For example, UK subjects "free-rode to a much greater extent" than Italians in a public goods experiment (Burlando and Hey 1997). Specifically about 60 percent of players in UK sessions completely free-rode versus 42 percent in Italy. However, the $TRUST$ in UK (44.4) is much higher than that of Italy (26.3) (Knack and Keefer 1997).[25] Similarly Weimann (1994) shows that US subjects free-rode more than German ones, though the $TRUST$ in the US (45.4) is also much higher than that in Germany (29.8).

3 Social Trust Formation

We have shown that non-selfish players (thus social trust) are valuable to the society not only because they can cooperate between themselves but also because they can elicit cooperative efforts from self-interested people. To improve social welfare every society desires more social trust. This need seems quite urgent in the US where the social trust has recently been on the decline (Knack and Keefer 1997, Putnam 1993, 1995).[26] The question that follows is how social trust is generated in the society.

So far we've assumed that the distribution of non-selfish players, hence the social trust, is determined exogenously like a kind of natural endowment. However, we have some reason to believe that behaving cooperatively to strangers is a trait acquired early in life and nurtured over time.[27] Parents and teachers may deliberately teach children to be more cooperative, acting as role models and choosing appropriate home and school inputs.[28] For adults Putnam (1993, 1995) argues that organization

[24] A possible explanation of this phenomenon is provided in section 3.5.

[25] The $TRUST$ measures in Knack and Keefer (1997) are based on the 1990-1991 survey and differ from those in 1997. The raw data has been weighted to correct for an oversample of citydwellers and the better-educated.

[26] The trust indicator from the General Social Survey is about 55-60 in the late 1950s and early 1960s. It falls to the mid- and upper-30s in the 1990s.

[27] Little about the social psychological process of (trust) cooperative tendency formation is known. It is suggested, however, that the social learning model may be an appropriate proxy (see 'Trust in Society' edited by Karen S. Cook 2001)

[28] Indeed, parents do choose certain desirable traits to invest in children. For example, 77.2% of parents in the General Social Survey from 1986 to 1998 think that "help others when they need help" is one of the three most important traits that their children should learn, while 96.8% rank it among the top four. And empirical evidence has shown that children's cognitive and social development are affected by home inputs (Huang 2002).

membership can instill in their members "habits of cooperation, solidarity, and pubic-spiritedness," which implies that being involved in associations might increase one's cooperative tendency.

If social trust is a choice variable, how much is the social optimal level? Would it be created in equilibrium by people who intrinsically care only about their own material payoffs?[29] If so, what are the possible social trust levels in various situations? How and in what ways could we improve social trust? These issues will be addressed in this section.

3.1 Basic Setup with Human Capital

3.1.1 Cooperative Tendency: a Component of Human Capital

Cooperative tendency is an inalienable trait that may benefit its owner over a long time. If we agree that human capital is the knowledge and skills invested in a person that yield returns to him/her in many periods (OECD 2001), cooperative tendency should also belong to human capital.[30] It is, however, distinct from cognitive ability in that it does not directly affect production functions. The way cooperative tendency affects players' output is through social interaction with other players.[31] Specifically, players equipped with adequate cooperative tendencies can cooperate with others and produce more outputs.

If we explicitly define conventional human capital, denoted by h, as the *productive ability* composed by the knowledge and skills that directly enter a specific production function, then α could be characterized as the *cooperative ability* of a player that can increase the opportunities to use his producing ability h. The combination of these two types of abilities, (h, α), determines a person's overall productivity when other things (say social trust) are given.[32] This motivates us to denote a player's human

[29] See Rotemberg (1994) for arguments about why material payoffs should be the criterion for determining welfare. We do not actually need the discipline of maximizing material payoff because it coincides with the life time utility maximization criterion in the model.

[30] In the same spirit, other personal characteristics such as working attitude, self-discipline, motivation, and time preference are treated as components of human capital in Becker (1996), Bowles and Gintis (1998), Heckman (2000), and Bowles et al. (2001).

[31] In this sense, cooperative tendency has some similarity to the concept of social asset proposed by Mailath and Postlewaite (2001).

[32] An analogy might be helpful in seeing the relationship between the two components. A person, as an optimizing unit, is like a firm. The profitability of a firm depends not only on its technology

capital by its two components (h, α).

These components should be discussed simultaneously because they are correlated with each other in human capital accumulation and goods production processes. On the one hand, productive ability and cooperative tendency are competing for the scarce resources in the society. For example, a child could spend time alone in studying mathematics or in a group socializing with other children. A company may have to choose between funding to train its employees in productive skills or to encourage socialization among them. The allocation of time and resources depends on the perceived importance of productive ability relative to cooperative ability. On the other hand, productive ability and cooperative tendency may also be complementary. A cooperative person could more easily get help from others to improve his cognitive ability or could have better opportunities to produce more outputs. Groups with cooperative members may achieve more than those with frequent in-fighting.

Again, the distribution of cooperative tendency among the population is social trust, which represents the *cooperative infrastructure* of the society. Similarly the distribution of productive ability in the society is its *intellectual infrastructure*, the importance of which in economic growth has been shown by Lucas (1988) among others. The interaction of these two dimensions of aggregate human capital is only briefly discussed in the paper, in order to maintain our focus on cooperative tendency and social trust.

3.1.2 Human Capital Version of the Stage Game

Now we let the material payoffs in the prisoner's dilemma game explicitly depend upon players' human capital. The underlying rationale is that players with different productive abilities usually play different games. For example, prisoner's dilemmas that senior managers have to deal with are typically different from those faced by front-line workers.

We also assume that a player's material payoff is determined solely by his own producing ability h, though players have to cooperate with each other in order to produce more than the default amount (which is normalized to zero). This assumption abstracts from complementarity among productive skills, making it clear that all positive total outputs completely represent the effects of cooperative tendency and

but also on its management. It is not difficult to see the parallels between cognitive ability and technology, social ability and management, a person's earning and a firm's profit, etc..

social trust.

Let h^i and h^j denote the productive abilities of player i and player j respectively. The human capital version of the prisoner's dilemma between two players i and j, denoted by game γ_h, is

	C	D
C	$g(h^i), \quad g(h^j)$	$-l(h^i), \quad g(h^j) + d(h^j)$
D	$g(h^i) + d(h^i), \quad -l(h^j)$	$0, \quad 0$

The production functions $g(\cdot), d(\cdot), l(\cdot)$, are (weakly) increasing and concave in h. Corresponding to assumptions (1) and (**??**), $d(h) < l(h)$ and $g(h) + d(h) - l(h) > 0$ are assumed for all h. Following the same algebra as in the second section, we find that for player i to be able to cooperate under complete information, his cooperative tendency α_i should be at least as high as $d(h^i)$. Note that the minimum cooperative tendency for a non-selfish player now depends on his productive ability. The rationale is that a player with higher cognitive ability usually has defecting chances associated with larger values.

The relationship between h and α under incomplete information is derived similarly. Let Π denote the expected proportion of non-selfish players in the population, which is taken as exogenously given.[33] Player i's cooperative tendency α_i must satisfy $\alpha_i \geq \Pi d(h^i) + (1 - \Pi) l(h^i)$ to enable him to cooperate under incomplete information. Thus the minimum cooperative tendency to enable non-selfish player i to make cooperative efforts under incomplete information is

$$\underline{\alpha}(h^i, \Pi) \equiv \Pi d(h^i) + (1 - \Pi) l(h^i).$$

Note that $\partial \underline{\alpha}(h^i, \Pi)/\partial \Pi = d(h^i) - l(h^i) < 0$. That is, the higher the expected proportion of non-selfish players in the population, the lower cooperative tendency needed for each individual player to make cooperative effort. Another partial derivative is $\partial \underline{\alpha}(h^i, \Pi)/\partial h^i \geq 0$. This means that among non-selfish players, those with higher productive ability h also need a higher minimum level of cooperative tendency α, since they are faced with greater temptation to defect.

[33] Note that in real life Π represents people's belief of future social trust. It could be affected by mass media, people's life philosophies, group morale, etc.. In equilibrium and in the long run, however, Π must be equal to the realized social trust.

3.2 Human Capital Investment Model

3.2.1 Timing and Information Structure of the Game

Each player lives three periods. The first period is the *human capital investment stage*, during which each player chooses how much human capital (h^i, α^i) to invest to maximize his life-time utility,[34] taking as given the expected proportion of non-selfish players $\Pi \in [0, 1]$ in the population.[35] The following two periods belong to the *production stage*, where players interact with each other to produce outputs.

In the second period all players are strangers whose types are private information. They randomly match with each other playing the above stage game γ_h under incomplete information. In the final period, with probability $1 - p \in [0, 1]$, players have no chance to meet with each other and all get zero payoff. With probability p players meet again and their cooperative tendencies are fully revealed.[36] In this case a new game γ'_h is played where the row player i's payoff from (C, C) is $G(h_i)$, and from (D, C) is $G(h_i) + D(h_i)$. Without loss of generality we will focus on the game where $D(h_i) = \underline{\alpha}(h^i, \Pi)$ holds.[37] We also assume that the assortative matching rule applies under complete information.

Note that in the final period, because of complete information and assortative matching, only genuinely non-selfish players with at least $\underline{\alpha}(h^i, \Pi)$ could form a cooperative relationship and produce $G(h_i)$, all others get zero. This feature rules out

[34] Or we could say the parents of the player choose the human capital for him to maximize his lifetime utility. This explanation could avoid the problem of changing preferences. Another way of justification is assuming a stable meta-utility function which has human capital α as an argument, as Becker (1996) has suggested.

[35] Further invesment in cooperative tendency is abstrated away in the paper, since a person's human capital is largely dertermined before adulthood. However, to the extent that it could be changed later on, the same setup still applies, with the only difference being that the initial human capital is already given.

[36] Here we assume that the second period is informative enough to allow players' types to be fully revealed. A more general assumption is that people know each other's cooperative tendency from their last period's interaction with probability $q \in [0, 1]$. However, as long as q is big enough to deter any mimicing, we will get the same results as in the simple model where q is assumed to be one.

[37] This condtion is actually general enough for us to study the formation of a wide range of cooperative tendencies, since $\underline{\alpha}(h, \Pi)$ is different across players. In essence our model captures the following general situation: there are a series of prisoners' dilemma games γ with different defecting benenfit d_γ; players with different cooperative tendencies could find appropriate complete information games $\{\gamma : d_\gamma \leq \alpha_i\}$ to play once their types are revealed.

the possibility that in equilibrium players with $\alpha < \underline{\alpha}(h^i, \Pi)$ may want to mimic non-selfish players in the second period. Knowing this, all players would play honestly in the second period. That is, all players with at least $\underline{\alpha}(h^i, \Pi)$ cooperate, and all others defect.

In the production stage, a non-selfish player i gets total (time discounted) payoff $\beta[\Pi g(h^i) - (1-\Pi)l(h^i)] + \beta^2 pG(h^i)$, while a selfish player j gets $\beta\Pi[g(h^j) + d(h^j)]$. The basic logic driving the model is that non-selfish players have lower payoff than selfish players in the incomplete information game, but they could get future benefits under complete information; in contrast, selfish players gain immediately from their opportunism, barring themselves from access to future cooperation opportunities. Players would take these trade-offs into consideration when deciding whether to become non-selfish or not.

This two-step information-revealing assumption is aimed at capturing the essence of how people deal with prisoners' dilemmas in real life situations. People usually exert a lot of effort trying to determine each other's true type before making any (long-term) cooperation commitment in which the lion's share of the production is conducted. For example, strangers would not become friends until they have had some cooperative interactions. Before getting married couples date each other or live together for some time to see whether they will be able to cooperate for their lifetimes. In the labor market a complicated recruiting process (requiring credentials, reference letters, interviews, tests, internships, etc.) is conducted before an employment relationship is established. As the efficiency of these type-revealing processes goes up, cooperative people benefit more from their cooperative tendency.

As these examples show, p measures the efficiency of information flowing in the society. It is generally determined by social, economic, and technological structures in a society and can be different over time and across countries. For example, advancement in information technology greatly increases the availability of information: a car buyer's credit type, a firm's track record in product quality, a community's general safety. Many social institutions can substitute for family or kinship ties to facilitate the information flow.[38] On the other hand, increased mobility in modern society brings together people with different ethnic, cultural, language, geographical,

[38] Actually the relationship between social network and social trust is two-way: more social networks improve information efficiency thus increase future social trust; while more non-selfish players would cooperate and build more social networks. However, the informaton efficiency p is also determined by other slow-moving forces so that it is considered as exogenously given.

and institutional backgrounds. These differences may reduce the efficiency of information exchange and lead to less cooperation. Indeed, subjects paired with a partner of a different race or nationality are less cooperative in the public goods experiments (Glaeser et al. 2000b).

3.2.2 Human Capital Investment

Taking as given their expected payoffs in production stage as a function of their own human capital and the social trust, players make human capital investment decisions at time zero to maximize their expected lifetime utility. Players have different investment costs which depend on human capital and players' index. Specifically, the cost function is $c(h, \alpha, i)$, where $h, \alpha \in R^+$, $i \in [0, 1]$. We assume that investing in either kind of human capital incurs positive costs, and the cost is higher to players with higher index. That is, $c(0, 0, i) = 0$, $c_h > 0$, $c_\alpha > 0$, $c_i > 0$. The cost function is convex with respect to both α and $h : c_{hh} \geq 0$, $c_{\alpha\alpha} \geq 0$. Some further restrictions on the cost function will be discussed in due course.

Recall that the gain is invariant with α once it is at or above the minimum level for cooperation, but the cost of increasing α is strictly positive. It is thus rational for players to invest only $\underline{\alpha}(h^i, \Pi)$ if they do choose a positive α. As a result each player i is faced with only two choices of cooperative tendency: either $\alpha = 0$ to remain selfish or $\underline{\alpha}(h^i, \Pi)$ to become non-selfish.

Accordingly a player i's expected life-time utility function $V(h, \alpha, i)$ takes two types of values: $V_A^i(h) \equiv V(h, \underline{\alpha}(h, \Pi), i)$ if he becomes non-selfish, $V_M^i(h) \equiv V(h, 0, i)$ if he remains selfish. They are, respectively,[39]

$$\begin{aligned} V_A^i(h) &= \beta[\Pi g(h) - (1 - \Pi)l(h)] + \beta^2 pG(h) - c(h, \underline{\alpha}(h, \Pi), i), \\ V_M^i(h) &= \beta\Pi[g(h) + d(h)] - c(h, 0, i). \end{aligned}$$

The existence and comparative statics of the optimal human capital investments for both types of players are summarized in the following claim:

Claim 4 *i) If boundary conditions* $\lim_{h \to 0} c_h(h, 0, i) = 0$ *and* $\lim_{h \to 0} g'(h) > 0$ *hold, there exists a unique optimal solution* $h_M^i \equiv h_M(\Pi, p, \beta, i, k)$ *that maximizes* $V_M^i(h)$,

[39] Note that in equilibrium the value function for a non-selfish player is the same whether or not the psychological effect of α is taken into account. The reason is that a non-selfish player always cooperates in equilibrium so that the guilty feeling of non-cooperation represented by α is never realized and thus does not appear in the utility function.

where k represents all other parameters. Similarly, under boundary conditions $\lim_{h\to 0}\frac{\partial c(h,\alpha(h,\Pi))}{\partial h} = 0$ *and* $\lim_{h\to 0}[\beta pG'(h) + \Pi(g'(h) + l'(h)) - l'(h)] > 0$*, together with a sufficient condition*

$$l''(h) = 0, \ \ and \ \frac{\partial^2 c(h, \alpha(h, \Pi), \Pi)}{\partial h^2} \geq 0, \tag{A2}$$

$h_A^i \equiv h_A(\Pi, p, \beta, T, i, k)$ *is the unique solution maximizing* $V_A^i(h)$.

ii) h_M^i *increases with* Π *and* β *but is not affected by* p. h_A^i *increases with* p *and* β. *It increases with* Π *if the following condition holds.*

$$c_{h\alpha}(h, \alpha(h, \Pi), i) \geq 0, d'(h) - l'(h) \leq 0 \tag{A3}$$

The claim says that player i chooses different levels of productive abilities h_M^i and h_A^i corresponding to cooperative tendency 0 and $\alpha(h_A^i, \Pi), i)$ respectively. Under quite general conditions both h_M^i and h_A^i are uniquely determined as a function of Π, p, β, i and parameters in production functions $G, g, d,$and l. The comparative statics suggest that conventional human capital increases with the expected social trust Π and patience β for both types. Non-selfish players' productive ability h_A^i also increases with information efficiency p, while h_M does not depend on p.

The ranking of productive abilities across players is determined by their relative marginal costs of investing in both components of human capital. In order to focus on cooperative tendency, we impose a futher restriction on the cost function:

$$c_{hi}(h, \alpha, i) = 0, \ c_{\alpha i}(h, \alpha(h), i) > 0. \tag{A4}$$

This assumption means that the marginal cost of investing in h is the same for all players, while the marginal cost in α increases with a player's index. Accordingly, all players would choose the same level of productive skill h_M if they were to become selfish. That is, without any consideration about cooperative tendency, all players would have the same conventional human capital h_M. In other words, the investment in cooperative tendency now affects people's conventional human capital choices and creates a new margin for difference among them. The details are specified by the following proposition.

Proposition 2 *i) Under assumption (A4),* h_A^i *decreases with* i, *while* $h_M^i = h_M$, *for all* $i \in [0, 1]$.

ii) Furthermore, there exists a unique $\bar{i}(\Pi, p) \in [0,1]$ *such that* $h_A^i \geq h_M$ *for all* $i \in [0, \bar{i}]$, *while* $h_A^i < h_M$ *for all* $i \in (\bar{i}, 1]$ *under condition*

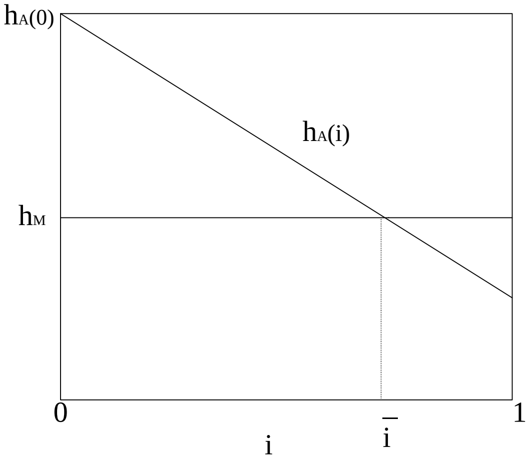
$$B(1, \Pi) \geq A(\Pi, p) \geq B(0, \Pi), \tag{A5}$$

where $A(\Pi, p) \equiv \beta^2 p G'(h_M) - \beta[\Pi d'(h_M) + (1-\Pi) l'(h_M)]$ *and* $B(i, \Pi) \equiv \partial(c(h_M, \alpha(h_M, \Pi), i) - c_h(h_M, 0, i))/\partial h_M$.

iii) $\bar{i}(\Pi, p)$ *increases with* p *and* Π.

Proof. See the Appendix. ■

The first result is not surprising given assumption *(*A4*)*. If players choose to invest in cooperative tendency, those with lower investment cost will choose higher productive abilities. Furthermore, the productive abilities h_A^i of players with sufficiently low cost in α are higher than h_M if assumption *(*A5*)* is true. Therefore productive and cooperative abilities complement each other for low cost players. However, they may be substitutes for high cost players whose h_A^i is lower than h_M. See the following figure for illustration.



Relation Between h_A^i and h_M

Player $\bar{\imath}$ is a threshold player who has the same productive ability independent of his type choice, i.e. $h_A^{\bar{\imath}} = h_M$. The index of this threshold player increases with p and Π, implying that as information structure becomes more efficient and the expected social trust is higher, the proportion of people whose h and α are complementary is larger. Note that as p goes up, the line of h_A^i shifts up while h_M remains the same, this is why $\bar{\imath}$ increases with p.[40] Since both h_A^i and h_M increases with Π, the fact that $\bar{\imath}$ increases with Π implies that h_A^i increases more.

Since h_A^i and h_M are the optimal solutions, $V_A^i(h_A^i)$ and $V_M^i(h_M)$ are the maximized value functions respectively. Let $V_d(i,\Pi)$ denote the net gain of being non-selfish versus selfish, where

$$\begin{aligned}V_d(i,\Pi) &\equiv V_A^i(h_A^i) - V_M^i(h_M)\\ &= \beta[\Pi g(h_A^i) - (1-\Pi)l(h_A^i) - \Pi(g(h_M)+d(h_M))] + \beta^2 pG(h_A^i)\\ &\quad -[c(h_A^i,\alpha(h_A^i,\Pi),i) - c(h_M,0,i)].\end{aligned}$$

Player i will choose to be non-selfish if and only if $V_d(i,\Pi) \geq 0$. The comparative statics of $V_d(i,\Pi)$ is summarized by the following lemma.

Lemma 1 $\frac{\partial V_d(i,\Pi)}{\partial i} < 0$ *under assumption (A3).* $\frac{\partial V_d(i,\Pi)}{\partial \Pi} > 0$ *under assumptions (A3), (A4), and (A5).*

Proof. In the appendix. ■

The lemma says that $V_d(i,\Pi)$ decreases with i and increases with Π under certain conditions. That is, players with lower investing costs would be more likely to become non-selfish, and a higher expected social trust Π would provide more incentives for all players to do so. The intuition is quite clear. The marginal effect of Π on the net gain of being non-selfish has two aspects: improving the chance of meeting a non-selfish player and reducing the cost of investing in cooperative tendency. All players benefit from the first channel (though at different levels), but only the non-selfish players benefit from the cost reduction. Thus the net gain of being non-selfish strictly increases with Π. $V_d(i,\Pi)$ decreases with players' identity because investing cost of cooperative tendency increases with index. As a result, when there are π altruistic players, their index ranges from 0 to π.

3.3 The Optimal Social Trust

In this human capital investment game, how much social trust would a social planner choose in order to maximize social welfare and total outputs? Suppose the social

[40] Furthermore since more players want to become non-selfish as p goes up ($\partial V_d(i,\Pi)/\partial p > 0$), the overall stock of human capital also increases with p.

planner's objective function is the sum of all players' life-time utility:

$$\max_{\pi} V(\pi) = \int_{i=0}^{i^S} V_A^i(h_A^i)di + \int_{i=i^S}^{1} V_M^i(h_M)di,$$

where i^S is the highest index among all non-selfish players. Let π denote the proportion of non-selfish players, then $\pi = \Pr(i \leq i^S) = i^S$, since we assume that i is uniformly distributed on the interval $[0, 1]$.[41]

Proposition 3 *The social return to investment in cooperative tendency is larger than a player's individual return.*

Proof. Take first order derivative with respect to π, we get

$$\frac{dV}{d\pi} = \underbrace{\int_{i=0}^{i^S} \frac{\partial V_A^i(h_A^i)}{\partial \pi}di + \int_{i=i^S}^{1} \frac{\partial V_M^i(h_M)}{\partial \pi}di}_{\text{externality to others due to } \pi \text{ increase}} + \underbrace{\left[V_A^{i^S}(h_A^{i^S}) - V_M^{i^S}(h_M)\right]}_{\substack{\equiv V_d(i^S,\pi) \\ \text{individual netgain for player } i^S}} .$$

$\frac{dV}{d\pi}$ is the marginal effect of π on social welfare. It is composed of two parts: the individual net gain for player i^S becoming non-selfish, and the externality on all other players due to the marginal increase of π. Now we need to prove that the externality is postive. Indeed, both non-selfish and selfish players benefit from the social trust increase caused by another player becoming non-selfish since

$$\begin{aligned} \frac{\partial V_A^i(h_A^i)}{\partial \pi} &= [\beta(g(h_A^i) + l(h_A^i)) + c_\alpha(h_A^i, \alpha(h_A^i, \Pi), i)(l(h_A^i) - d(h_A^i))] > 0, \\ \frac{\partial V_M^i(h_M)}{\partial \pi} &= \beta(g(h_M) + d(h_M)) > 0. \end{aligned}$$

Note that the externality is strictly positive for all $\pi > 0$. So the social return to any player being non-selfish is always strictly bigger than his individual return. ■

This means that the socially optimal proportion of non-selfish players is always strictly larger than the equilibrium proportion, except in the case where the equilibrium social trust already reaches the maximum level $\pi = 1$. In other words, in equilibrium the investment in social trust is generally inefficient, or it is efficient only when all people are non-selfish. The exact level of socia trust chosen by the social planner, however, depends on the parameters of the production and cost functions. For example, if the investing costs for some players are extremely high, say higher than the positive externalities received by all other players, then the benevolent social planner would allow them to remain selfish. Otherwise, it is probable that the social planner would order everybody to become non-selfish.

[41] The same result holds if players are distributed according to a general distribution function.

3.4 The Social Trust Equilibria

In this section we will study the existence and properties of Nash Equilibrium (NE thereafter) at the human capital investment stage. Every NE can be characterized by a pair $(\Pi = e, \pi = e)$ where π is the actual proportion of non-selfish players and $e \in [0, 1]$. The reason is quite straightforward. Given all other players' strategies that are summarized by the expected number of non-selfish players Π, player i chooses to invest in α if and only if his $V_d(i, \Pi)$ is non-negative. The actual investment choices of all players are the realized social trust at the end of time zero, which in this model can be summarized by π. No player would want to deviate from his choice when the expected social trust is exactly realized, i.e., when $\Pi = \pi$.

We partition the parameter space according to properties of the net gain function $V_d(i, \Pi)$ and characterize the corresponding equilibria. In the first case the net gain of investing in α has rather evenly distributed values across players. The second case is the opposite, where the gap of net gains between the lowest and highest indexed players is very large. In the other two cases, the investing cost in α is either quite high or quite low for all players. For illustrative purpose a graph is shown in each case, where the best response function has monotone slope. In the first two cases, the linear best response functions and the corresponding NE are also calculated.

To further investigate the properties of these NE, we consider them as the steady states in a dynamic process. Suppose there are countable infinite generations of players denoted by the integers $1, 2, ..., N, ...$ Each generation has a continuum of players with measure one and lives for three periods playing the same game as above. Every following generation $N + 1$ takes the realized proportion of non-selfish players in its immediately previous generation N, denoted by π_N, as its expected proportion of non-selfish players, i.e. $\Pi_{N+1} = \pi_N, \forall N = 1, 2, ...$, and the initial $\Pi_{N=1} \in [0, 1]$ is assumed exogenously given. In other words, these generations are identical ex ante except their Πs. In each case we will show which equilibrium is a stable steady state with respect to small perturbations of Π.

The process of finding NE is the same for all cases, so we will show details for only the first case. Note that the 'no social trust' equilibrium $(\Pi = 0, \pi = 0)$ always exists regardless of the underlying fundamentals. The proof is simple. When $\Pi = 0$ every player in the population is expected to be selfish. Since mutual cooperation needs at least two non-selfish players, there is no gain from being the only non-selfish player. Thus the realized number of non-selfish players is also 0.

3.4.1 Case I: $V_d(0, \Pi_0) = 0, V_d(1, \Pi_1) = 0$.

Suppose there exist $\Pi_0, \Pi_1, \in (0, 1)$ such that

$$\begin{aligned} V_d(0, \Pi_0) &= 0, & (3) \\ V_d(1, \Pi_1) &= 0. & (4) \end{aligned}$$

The equation (3) means that player with index 0 is indifferent to being selfish or not when the expected proportion of non-selfish players is Π_0. Since $V_d(i, \Pi)$ increases with Π by lemma 1, player 0 will choose to invest in α if $\Pi \geq \Pi_0$ is true; he will remain selfish if $\Pi < \Pi_0$. Similarly, the player with index 1 will become non-selfish for $\Pi \geq \Pi_1$ and remain selfish otherwise.

Lemma 2 *i) Given conditions (3) and (4), the best response function of the population, denoted by $B(\Pi)$, is*

$$B(\Pi) \equiv \{ \begin{array}{cc} 0 & \textit{if } \Pi \in [0, \Pi_0] \\ i^*(\Pi) & \textit{if } \Pi \in [\Pi_0, \Pi_1] \\ 1 & \textit{if } \Pi \in [\Pi_1, 1] \end{array} ,$$

where $i^*(\Pi)$ is deterimined by $V_d(i^*(\Pi), \Pi) = 0$ for any $\Pi \in [\Pi_0, \Pi_1]$.

ii) $B(\Pi)$ is continuous and non-decreasing with Π on $[0, 1]$. It strictly increases with Π on the closed interval $[\Pi_0, \Pi_1]$.

iii) $\partial B(\Pi; p, \beta)/\partial p \geq 0; \partial B(\Pi; p, \beta)/\partial \beta \geq 0$

Proof. First we show that $\Pi_1 > \Pi_0$. Since $V_d(i, \Pi)$ is decreasing with player's index i, it is straightforward that $V_d(1, \Pi_0) < V_d(0, \Pi_0) = 0$. But by condition (4) we get $V_d(1, \Pi_1) = 0$. Combining these two conditions leads to $V_d(1, \Pi_1) > V_d(1, \Pi_0)$. Since $V_d(i, \Pi)$ increases with Π, we know $\Pi_1 > \Pi_0$.

Now we can use Π_0 and Π_1 to partition the interval $[0, 1]$ into three segments: $[0, \Pi_0]$, $[\Pi_0, \Pi_1]$, $[\Pi_1, 1]$. For any given $\Pi \in [0, \Pi_0]$, nobody would want to be non-selfish, since Π is so low that the net gain of being non-selfish is negative for all $i > 0$. As a result we get $\pi = 0$ for all $\Pi \in [0, \Pi_0]$. The opposite is true for $\Pi \in [\Pi_1, 1]$. Now Π is so high that even the player the with highest cost would want to invest in α, implying that $\pi = 1$ for all $\Pi \in [\Pi_1, 1]$.

The situation with $\Pi \in [\Pi_0, \Pi_1]$ needs more discussion. For all $\Pi \geq \Pi_0$ we have $V_d(0, \Pi) > 0$. Similarly we know $V_d(1, \Pi) < 0$ for all $\Pi \leq \Pi_1$. So the following conditions hold:

$$V_d(0,\Pi) > 0, V_d(1,\Pi) < 0,\ \forall \Pi \in [\Pi_0, \Pi_1].$$

By $\frac{\partial V_d(i,\Pi)}{\partial i} < 0$, we know that $V_d(i,\Pi)$ is continuous and strictly decreasing in $i \in [0,1]$ given any Π. These conditions guarantee that for each $\Pi \in [\Pi_0, \Pi_1]$, there exists a unique $i^* \equiv i^*(\Pi) \in [0,1]$ such that

$$V_d(i^*(\Pi), \Pi) = 0.$$

Players $i \leq i^*(\Pi)$ would choose to become non-selfish because their net gain $V_d(i,\Pi)$ is positive. So the real proportion of non-selfish player in the population, denoted by π, is $\pi \equiv \Pr(i \leq i^*(\Pi)$. Since i is uniformly distributed on $[0,1]$, we have $\pi = i^*(\Pi)$.

It is straightforward to verify that $B(\Pi)$ is continuous and non-decreasing in Π on $[0,1]$. More specifically $B(\Pi)$ strictly increases in Π on $[\Pi_0, \Pi_1]$, since

$$\frac{\partial i^*(\Pi)}{\partial \Pi} = -\frac{\partial V_d(i^*,\Pi)/\partial \Pi}{\partial V_d(i^*,\Pi)/\partial i^*} > 0.$$

Proofs of the last results are in the appendix. ■

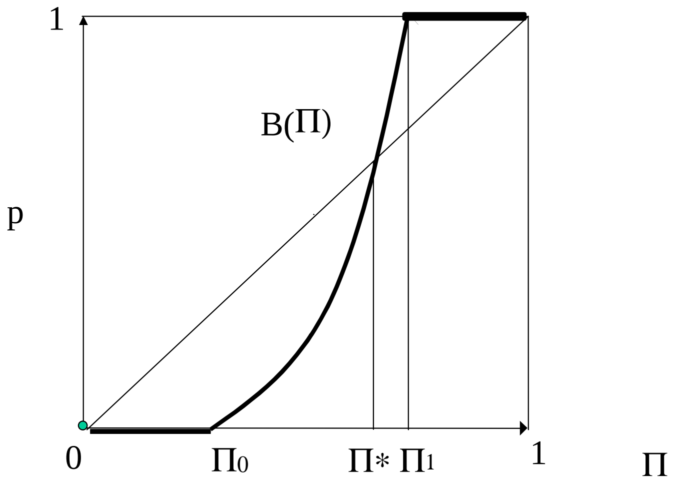
Since $B(\Pi)$ is continuous and strictly increasing in Π on $[\Pi_0, \Pi_1]$, and $B(\Pi_0) = 0$, $B(\Pi_1) = 1$, there must exist at least one fixed point $\Pi^* \in [\Pi_0, \Pi_1]$ such that $(\Pi = \Pi^*, \pi = \Pi^*)$ is a *NE*. The exact number of the fixed points on $[\Pi_0, \Pi_1]$ depends on the curvature of $i^*(\Pi)$ or $\frac{\partial^2 i^*(\Pi)}{\partial \Pi^2}$:

$$\frac{\partial^2 i^*(\Pi)}{\partial \Pi^2} = \frac{\frac{\partial^2 V_d(i^*,\Pi)}{\partial i^* \partial \Pi}\frac{\partial V_d(i^*,\Pi)}{\partial \Pi} + (-\frac{\partial V_d(i^*,\Pi)}{\partial i^*})\frac{\partial^2 V_d(i^*,\Pi)}{\partial \Pi^2}}{(\partial V_d(i^*,\Pi)/\partial i^*)^2}.$$

Since all other terms are positive, the sign of $\frac{\partial^2 i^*(\Pi)}{\partial \Pi^2}$ depends on $\frac{\partial^2 V_d(i^*,\Pi)}{\partial \Pi^2}$. If $\frac{\partial^2 V_d(i^*,\Pi)}{\partial \Pi^2} \geq 0$ then $\frac{\partial^2 i^*(\Pi)}{\partial \Pi^2} > 0$, so that the slope of $B(\Pi)$ strictly increases on the interval $[\Pi_0, \Pi_1]$. Otherwise it is possible that $\frac{\partial^2 i^*(\Pi)}{\partial \Pi^2} \leq 0$ or that it might not be monotone.

Remark 3 *Hereafter we focus our discussion on situations where the best response function has monotone slopes with respect to π.*

When $B(\Pi)$ has monotone slopes on the interval $[\Pi_0, \Pi_1]$, the NE $(\Pi = \Pi^*, \pi = \Pi^*)$ is unique. It is easy to check that $(\Pi = 0, \pi = 0)$ and $(\Pi = 1, \pi = 1)$ are the other two NE. See the following figure for illustration.



Case I

If the initial expectation at time $N = 1$ is such that $\Pi_{N=1} < \Pi^*$, then ultimately this economy will fall into the no-trust trap $(\Pi = 0, \pi = 0)$ where no cooperation happens. On the contrary, if $\Pi_{N=1} > \Pi^*$, the economy will gradually reach the social optimal level $(\Pi = 1, \pi = 1)$ where everybody cooperates. These two corner NEs are stable with respect to small perturbations. The interior NE $(\Pi = \Pi^*, \pi = \Pi^*)$ only happens when $\Pi_1 = \Pi^*$. It is unstable since any small perturbation will lead the economy to the two corner NEs. Thus we have proved the following proposition.

Proposition 4 *Under conditions (3) and (4), there are three Nash equilibria:* $(0, 0)$, (Π^*, Π^*), *and* $(1, 1)$, *where* $\Pi^* \in [\Pi_0, \Pi_1] \subset (0, 1)$. *Among them* $(0, 0)$ *and* $(1, 1)$ *are stable.*

This case shows that the initial conditions or random historical events can be of great importance to economic growth. It can be used to account for dramatically different performances between two otherwise identical communities or organizations. Suppose, for example, that both western and eastern regions of the same country were faced with the same social and economic structures. If for some exogenous reason $\Pi_{N=1} = \Pi^* + \frac{\varepsilon}{2}$ in the west, while in the east $\Pi_{N=1} = \Pi^* - \frac{\varepsilon}{2}$, then overtime this almost negligible ε difference in initial beliefs could explode into two stable equilibria: everyone cooperates in the west, but no one does so in the east.

The intuition is as follows. Players' investment costs in this case are quite evenly distributed in a narrow range, so that the interaction among players becomes relatively more important in affecting players' decision. If they believe enough people (over a threshold Π^*) will be non-selfish then everybody prefers to be non-selfish, otherwise all will remain selfish. In other words, nobody is different enough in their investing costs to avoid being swept away by others' choices.

Here is an example of the linear best response function in this case:

Example 1 *Suppose $i^*(\Pi) = a\Pi - b$, where a, b are constants to be characterized below. It turns out that to satisfy the conditions (3) and (4), it must be true that $a = \frac{1}{\Pi_1 - \Pi_0}$ and $b = \frac{\Pi_0}{\Pi_1 - \Pi_0}$. Thus we get $i^*(\Pi) = \frac{\Pi - \Pi_0}{\Pi_1 - \Pi_0}$. Note that the slope a is bigger than 1. Let Π_l^* be the solution to $i^*(\Pi_l^*) = \Pi_l^*$, then $\Pi_l^* = \frac{\Pi_0}{1 + \Pi_0 - \Pi_1}$. It is trivial to check that $\Pi_l^* \in [\Pi_0, \Pi_1]$.*

3.4.2 Case II: $lim_{\Pi \to 0^+} V_d(0, \Pi) > 0, V_d(1, 1) < 0$.

We know if $\Pi = 0$ then $\pi = 0$, but the best response function may not be continuous at $\Pi = 0$. For example, the players with very low costs would want to invest in α if other players, however few, are expected to do so. In other words, once the expected proportion of non-selfish players is positive, regardless of how close it might be to zero, the players with lowest costs will choose to be non-selfish. This means the following condition holds

$$lim_{\Pi \to 0^+} V_d(0, \Pi) > 0. \tag{5}$$

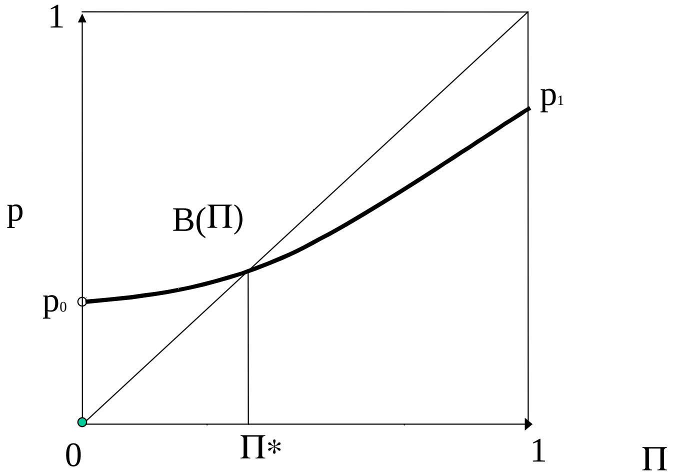
Let π_0 be defined by $\pi_0 \equiv lim_{\Pi \to 0^+} i^*(\Pi)$ and $lim_{\Pi \to 0^+} V_d(i^*(\Pi), 0) = 0$. Then this condition is equivalent to $\pi_0 > 0$. On the other hand, some player's cost could be so high that he would not invest in α even when everyone except himself is expected to be non-selfish. That is

$$V_d(1, 1) < 0. \tag{6}$$

Let π_1 be defined by $\pi_1 \equiv i^*(1)$ and $V_d(i^*(1), 1) = 0$, then the above condition is the same as $\pi_1 < 1$.

Proposition 5 *Under conditions (5) and (6), there exist two NE: $(0, 0)$ and (Π^*, Π^*), where $\Pi^* \in (\pi_0, \pi_1)$. Only (Π^*, Π^*) is stable.*

Proof. In the Appendix. ■



Case II

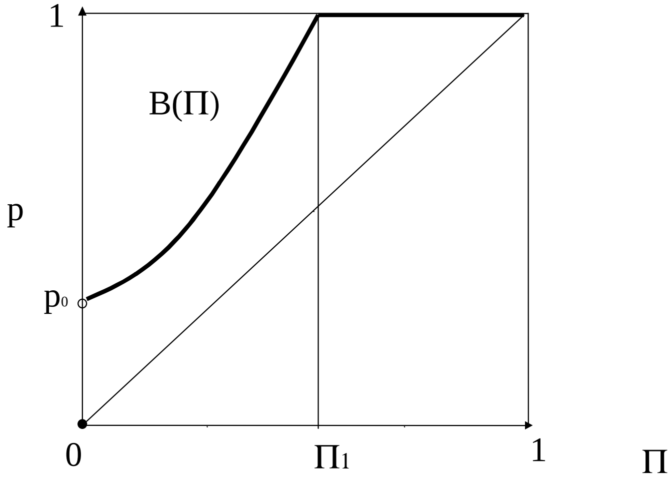
Here the interior NE $(\Pi = \Pi^*, \pi = \Pi^*)$ is the only focal point of the history and is stable to perturbations. Contrary to the previous case, the long run social trust level in this case is immune to random events. Wherever it starts (as long as $\Pi_{N=1} \geq 0$), the society will definitely settle down at $(\Pi = \Pi^*, \pi = \Pi^*)$. However, social optimum is not achievable in equilibrium. See the above figure for illustration.

The intuition here is also clear. Players have very different investment costs in this case. Those with extremely low costs would become non-selfish even when there are few others willing to do so. At the same time, players with the highest costs would always remain selfish. These two groups of players have very stable behavior patterns, leaving little room for dramatic changes though beliefs or other transitory influences.

Example 2 *Suppose* $i^*(\Pi) = c\Pi + d$, *where* c, d *are some constants determined by the underlying parameters. Solving* $i^*(\Pi_l^*) = \Pi_l^*$, *we get* $\Pi_l^* = \frac{d}{1-c}$. *According to (5) and (6),* $\lim_{\varepsilon \to 0^+} i^*(\varepsilon) = \lim_{\varepsilon \to 0^+}(c\varepsilon + d) = d > 0$, $i^*(1) = c + d < 1$. *A similar result as in Case I is* $i^*(\Pi)/\partial\Pi > 0$, *which implies that* $c > 0$. *In summary, we get* $c, d > 0$ *and* $c + d < 1$. *Thus the slope of* $i^*(\Pi)$ *is less than* 1, *and* $\Pi_l^* = \frac{d}{1-c}$ *lies in the open interval* $(0, 1)$.

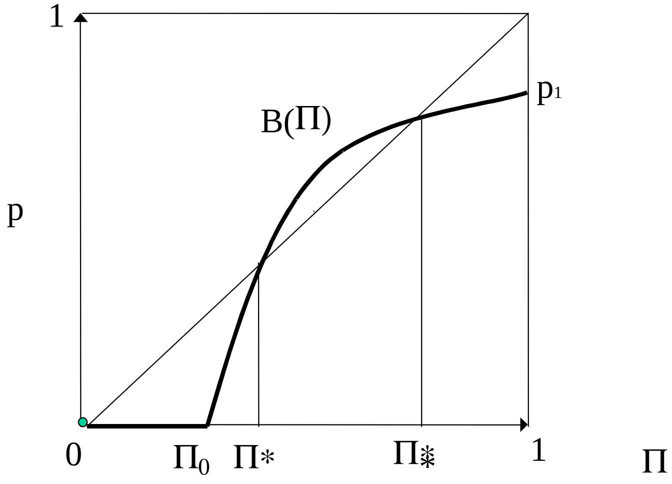
3.4.3 Case III & Case IV

The joint result of (4) and (5) is the third case, where $\Pi_1 \in [0, 1]$. This differs from the second case only in that even the highest cost players would consider being non-selfish when they believe enough people are doing so. See the next figure for illustration.



Case III

The fourth case is defined by the combination of conditions (3) and (6), where $\Pi_0 \in (0, 1]$. In contrast with the third case, the investing cost here is relatively high for all players. See the figure below for illustration.



Case IV

Proposition 6 [42] *Under conditions (4) and (5),* $(1, 1)$ *always exists and is stable. It can be the only stable equilibrium when* $\partial^2 B(\Pi)/\partial\Pi^2 \leq 0$, *or when* $\partial^2 B(\Pi)/\partial\Pi^2 > 0$ *and* $\Pi_1 \leq \pi_0$ *both hold. Otherwise it is possible to have an even number of NE at interior points, among which the odd-numbered ones are stable.*

[42] The proof is similar to the other two cases and thus omitted.

ii) Under conditions (3) and (6), $(0,0)$ *always exists and is stable. It can be the only stable equilibrium when* $\partial^2 B(\Pi)/\partial\Pi^2 \geq 0$, *or when* $\partial^2 B(\Pi)/\partial\Pi^2 > 0$ *and* $\Pi_0 \leq \pi_1$ *both hold. Otherwise it is possible to have an even number of NE at interior points, among which the even-numbered ones are stable.*

These last two cases could account for why firms spend a lot of resources in selecting employees with right attitudes. For example, a firm in a situation like case III could easily achieve optimal cooperation among employees, while in case IV cooperation level is never optimal and might not exist at all. Even when firms are otherwise identical, their employees' investing costs in cooperative tendency would singly make a big difference in the firms' performances. A similar result is derived by Rob and Zemsky (2001).

The proposition below summarizes some common results of the above four cases:

Proposition 7 *i) The social optimal trust level is achievable in stable NE when condition (4) is satisfied but never so when (6) holds.*

ii) 'No trust' equilibrium always exists. It is stable under condition (3).

iii) Multiple equilibrium is possible in all cases.

The proposition implies that to achieve the social optimal trust level, it is crucial to reduce the investing cost of those high-cost players, while in order to generate *positive* social trust in the long run, it is very important to reduce the investing cost of the low-cost players.

3.5 Several Ways to Increase Social Trust

Now we will discuss the possible ways, together with their empirical implications, suggested by the model to increase social trust levels in the long run.

Efficient Information Structure. As Lemma 3 shows more efficient information structure in the society (a higher p) shifts up the best response function $B(\Pi)$ in all cases. The reason is that other things being given, players prefer to become non-selfish if they have more opportunities to form cooperative matches. The result is that social trust π^* in stable equilibrium increases with p. Empirically this implies that social trust is higher in a society/community where information flow is smoother.

As discussed above there are many forces affecting information efficiency in a society, one example being mass media. Temple and Johnson (1998) show that across

29 countries the measurement of social trust is positively correlated with both daily newspaper circulation (0.73) and the number of radios per capita (0.53). One of their major conclusions is that "an assessment of mass communications, given the absence of other good measures, is probably the best way of capturing variation in social trust across developing countries."

In developed countries the information structure seems to differ somewhat. For example, high mobility rate in modern society increases background variation in one's aquaintances, making it difficult to infer a person's cooperative type. Since we do not learn much about people's cooperative tendency through casual daily encounters, we are less likely to form cooperative relationships. Accordingly, the benefit of investing in cooperative tendency is smaller, possibly reducing the proportion of trustworthy people. This could be a force contributing to the steady decline of trust indicators in the US.

Longer Tenures. As the duration of cooperative relationships goes up, non-selfish players' gain from complete information games becomes bigger. An earlier version of the paper shows that longer tenure shifts the best response function upward and leads to a higher stable social trust equilibrium. This suggests that in communities and organizations where members have longer tenures together, social trust is higher. For example, under the following circumstances: (a) when the divorce rate is lower and families are more stable; (b) when the average turnover rate within a group is lower; (c) when people are encouraged to become home-owners rather than tenants, social trust is generally higher. Shorter tenure might be another reason to account for the decline of social trust in the US.

Higher Time Preference. The positive effects of a higher time preference β on social trust and individual productive ability are not surprising. The empirical implication is obvious: more patient players tend to generate more social trust and to invest in higher human capital.

Optimistic Expectation. It is obvious that in all cases the equilibrium social trust π^* (weakly) increases with the initial expected social trust $\Pi_{N=1}$. This implies that, in the long run, it never hurts to have an optimistic Π, and sometimes, as when there are multiple equilibria, it makes a world of difference. Of course the more realistic the Π is (in terms of its closeness to the equilibrium level π^*), the lower the adjustment costs for players. One empirical implication is that, in general, we should observe a positive correlation between an optimistic population and a higher level of social trust.

The important role of optimistic expectation is clearly recognized in the real life. Examples are abundant: mass media's tendency to highlight positive aspects of society and present a bright picture of the future; morale boosting materials, talks, and activities; efforts exerted on cultivating optimistic personalities. People's expectations about future social trust are generally quite stable over time at steady sate. However, they can also be abruptly changed by some historical event, which switches the social trust level from one equilibrium to another. For example, Putnam (1995) suggests that some political scandals have contributed to the recent decline of social trust in the US.

Inside and Outside Discipline. In all cases a lower cost of investing in cooperative ability α also shifts up the best response function $B(\Pi)$ and increases π^*. The same is true for lower levels of $l(h)$ and $d(h)$ that reduce the costs of behaving cooperatively. These two channels for improving social trust can be categorized respectively as the *inside discipline* and *outside discipline* schemes.

Inside discipline schemes aim at cultivating cooperative tendency in people. Examples are encouraging informal socialization among people; increasing the coverage of cooperation stories in the mass media; and teaching children how to cooperate in daily life through role models, examples, books, games, and special activities. In this respect, families and formal education systems are crucial to social trust formation. They not only teach children productive knowledge and skills, represented by h, but also affect their cooperative tendency and other important characteristics. Early intervention programs like Headstart are also important in improving social trust, since they are effective in helping children from poor family backgrounds to behave cooperatively (Heckman 1999).

Examples of outside discipline schemes include the legal system, incentive schemes in organizations, contracts, game rules, and social norms. If these outside institutions are perfect, the defecting benefits or cooperation costs represented by d and l should be zero. In this sense the levels of d and l measure the efficiency of outside institutions in promoting cooperation.

While the relationship between these two schemes is a rather complex one, our social trust formation model can provide new insights to it.[43] Specifically, there is substitution between formal institutions and cooperative tendency but complemen-

[43] A thorough analysis of the interdependence of formal institutions and inside discipline goes beyond the scope of the paper, since the determination of the formal institutions should also be endogenized.

tation between formal institutions and the proportion of cooperative people. The reason is as follows.

When outside disciplinary institutions are efficient, defecting benefits l and d are low. The required cooperative tendencies to achieve cooperation are, therefore, also low. In other words, the outside discipline crowds out innate discipline.[44] On the other hand, the equilibrium proportion of non-selfish players π is higher, since more people can afford the investment costs of lower cooperative tendencies. If the cost of designing and enforcing formal disciplinary mechanisms is very high, people have to invest in higher cooperative tendencies to achieve cooperation. As a result fewer people are non-selfish, but those who are have higher cooperative tendencies.

An interesting implication is that, though the cooperation level is generally higher in a society where outside disciplinary mechanisms are more effective, people may not be able to cooperate in games with very high defecting benefits. The opposite is true for a society with less effecitive outside institutions. As a result, social trust measures in different games could have contradictory rankings across countries (see the example in section 2.3.3).

Much attention in economics literature has been focused on outside instiutions, while the importance of using inside discipline to improve social trust is not well appreciated. This bias is also prominent in many real life situations. For example, the current policies regarding education and job training "...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"(Heckman 2000). In the business sector firms often report shortage in appropriate working habits and poor attitudes, suggesting an underinvestment in these qualities, either due to high investment cost and/or inadequate recognition of the importance of cooperative tendency to individual and total outputs.[45]

[44] The fact that extrinsic motivators crowd out intrinsic motivation is observed by social psychologists and tested by experimental economists (see Frey and Oberholtzer-Gee 1997). Also see Bar-Gill and Fershtman (2000) for a similar application in economics.

[45] Abundant evidence is provided by Cappelli (1997). An anecdote may further illuminate people's view on this. Jack Welch, the former CEO of GE, recalls in his autobiography that expenditures (on meeting places and tredmills), aimed at encouraging informal socialization among employees across different departments, was strongly objected to as a waste of money.

4 Conclusions

In this paper we have formally shown that cooperative tendency and social trust can, in the presence of prisoners' dilemma, elicit cooperation and thus improve individual and total outputs. They are endogenously generated in equilibrium in the human capital investment game where rational individual players simultaneously maximize their expected life-time earnings by choosing their optimal level and combination of human capital.

This paper contributes to the social capital and human capital literature in several ways. First, we formally define social trust as the distribution of cooperative tendency in a society, making it an operational concept for studying various forms of social capital. This helps pin down the undesirable flexibility of the social capital concept and offers a uniform conceptual tool to conduct further study of social capital. For instance, the relationship among several widely used empirical measures of social capital is analyzed, and the discrepancies among them are accounted for using this concept of social trust.

Second, we show that the exact quantitative effects on outputs of the same social trust vary with the detailed specifications of the games. This means we could design game rules to increase the effects of social trust. In empirical studies the game specifications should be taken into account to measure the level and effects of social trust.

Third, we treat cooperative tendency as another component of human capital and link its aggregate distribution to social trust. This allows us to look at the important effects of individual social abilities from the macro level and provide intellectual support to the related study in the human capital literature. For example, we now understand that social ability, by increasing players' opportunities to engage in productive cooperation, is not only important to individual earnings, but also crucial to aggregate output in the group through the social trust channel. This implies that the effects of social ability are not cancelled out at the macro level, rather they are multiplied, becoming more important in aggregation.

Fourth, we endogenize the formation of social trust as the aggregated result of the optimal human capital choices by individual players. The model shows that the equilibrium social trust is usually lower than the social optimal level, though the latter can be achieved in some situations. It is possible to have multiple equilibrium where historical conditions are important.

Finally, the model clarifies the mechanism of various forces in improving social trust. For example, it shows that a society's formal institutions, acting as outside discipline, may crowd out an individual's internal cooperative tendency, but they can increase the proportion of cooperative players in a representative prisoners' dilemma. The cost of investing in individual human capital, especially cooperative tendency, is found to have important effects in determining social trust level in the society. If it is true that cooperative tendency is a trait planted in early childhood and developed gradually over many years, then social trust might significantly improved through work in the family and in the formal education system.

The paper is, however, only a preliminary step in characterizing the formation mechanisms of social capital and the relationship between social capital and human capital. More research is needed. For example, elements such as information structure, formal institutions, and costs of investment in cooperative tendency might in the long run have dynamic interactions with social trust, even though they are treated as exogenously given in this paper. The effects of these forces on various forms of social capital and the relative importances of these effects also need to be measured with empirical studies. The efforts to understand social capital may ultimately generate profound change in society's resource allocation for the enhancement of our long term social and economic well-being.

References

1. Abreu, D. (1998), "On the Theory of Infinitely Repeated Games with Discounting," *Econometrica*, vol. 56(2): p383-96.

2. Andreoni, J. and R. Croson (2002), "Partners versus Strangers: Random Rematching in Public Goods Experiments," forthcoming in *Handbook of Experimental Economics Results.*

3. Arrow, K. (1972), "Gift and Exchanges," *Philosophy and Public Affairs*, I (1972), p343-62.

4. Bar-Gill, O., and C. Fershtman (2000), "The Limit of Public Policy: Endogenous Preferences," *working paper*, Aug. 2000.

5. Becker, G. (1996) *Accounting for Tastes*, Harvard University Press.

6. Bowles, S., and H. Gintis (1998) "The Determinants of Individual Earnings: Cognitive Skills, Personality, and Schooling," *working paper.*

7. Bowles, S., H. Gintis, and M. Osborne (2001) "The Determinants of Earnings: A Behavioral Approach," *Journal of Economics Literature*, Vol. XXXIX (Dec. 2001), pp 1137-1176.

8. Burlando, R., and J.D. Hey (1997), "Do Anglo-Saxons Free-ride More?" *Journal of Public Economics* 64 (1997) 41-60.

9. Coleman, J.S., "Social Capital in the Creation of Human Capital," *American Journal of Sociology* 94 (1988): S95-S120.

10. Fehr, E. and S. Gachter (2000), "Fairness and Retaliation: The Economics of Reciprocity," *Journal of Economic Perspectives* 14, 159-182.

11. Frey, B.S. and F. Oberholtzer-Gee (1997), "The Cost of Price Incentives: an Empirical Analysis of Motivation Crowding Out," *The American Economic Review* 87, 746-755.

12. Glaeser, E.L., David Laibson, and Bruce Sacerdote (2000a), "The Economic Approach to Social Capital," *NBER working paper* 7728.

13. Glaeser, E.L., David Laibson, J.A. Scheinkman, and C.L. Soutter (2000b), "Measuring Trust," *Quarterly Journal of Economics*, Aug. 2000.

14. Fudenberg, Drew, Maskin Eric (1986) "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information." *Econometrica*, vol. 54(3): p533-54.

15. Hardin, Russell (2001), "Conceptions and Explanations of Trust," in *Trust in Society,* edited by Karen S. Cook, New York: Russell Sage Foundation, 2001.

16. Heckman, James J. (1999) "Policies to Foster Human Capital," *NBER Working Paper* 7288, August 1999.

17. Huang, Fali (2002), "Estimations of Child Development Production Functions," *mimeo*, University of Pennsylvania.

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

19. Kreps, D., P. Milgrom, J. Roberts, and R. Wilson (1982) "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma." *Journal of Economic Theory,* vol. 27: p245-52.

20. La Porter, R., F. Lopez-de-Silanes, A. Shleifer, and R.W. Vishny, "Trust in Large Organizations," *American Economic Review*, May 1997.

21. Lucas, R.E. Jr. (1988), "On the Mechanics of Economic Development," *Journal of Monetary Economics* 22 (1988) 3-42.

22. Mailath, G. and A. Postlewaite (2001), "Social Assets," *CARESS working paper*, University of Pennsylvania.

23. OECD (2001), *The Well-being of Nations: the Role of Human and Social Capital*, Paris.

24. Putnam, R. D. (1993) (with R. Leonardi and R.Y. Nanetti), *Making Democracy Work*, Princeton, NJ: Princeton University Press, 1993.

25. Putnam, R. D. (1995), "Bowling Alone: America's Declining Social Capital," *Journal of Democracy*, Vol.6 (1995), pp. 65-78.

26. Rob, R., and P. Zemsky (2002), "Social Capital, Corporate Culture and the Incentive Intensity," *RAND Journal of Economics*, Vol. 33 No. 2, Summer 2002.

27. Rotemberg, J. J. (1994), "Human Relations in the Workplace." *Journal of Political Economy* 102 (August 1994): 684-718.

28. Temple, J. and P.A. Johnson (1998), "Social Capability and Economic Growth," *Quarterly Journal of Economics*, August,1998.

29. Weimann, J. (1994), 'Individual Behavior in a Free Riding Experiment," *Journal of Public Economics* 54, 185-200.

30. Welch, Jack (2001), *Straight from the Gut,* Warner Books, Inc., New York, NY.

Appendix:

- **Proof for Proposition 1: technical details.**

Proof. (1) The total output for the population is $Q_1^r = (\pi_R+\pi_S)^2 g + \pi_S(1-\pi_R-\pi_S)(g+d-l)$, which strictly increase with social trust (π_R, π_S) because

$$\frac{dQ_1^r}{d(\pi_R,\pi_S)} = (\frac{\partial Q_1^r}{\partial \pi_R}, \frac{\partial Q_1^r}{\partial \pi_S})' = \begin{pmatrix} 2\pi_R g + \pi_S(g+l-d) \\ \pi_R g + (g+d-l) + (\pi_R + 2\pi_S)(l-d) \end{pmatrix} > 0.$$

Note that $\frac{\partial Q_1^r}{\partial \pi_S} - \frac{\partial Q_1^r}{\partial \pi_R} = (1-\pi_R-\pi_S)(g+d-l) > 0$ iff $\pi_R + \pi_S < 1$.

(2) $Q_1^a < Q_1^r$.

$$\begin{aligned} Q_1^r - Q_1^a &= (\pi_R+\pi_S)g - (\pi_R+\pi_S)^2 g - \pi_S(1-\pi_R-\pi_S)(g+d-l) \\ &= (\pi_R+\pi_S)(1-(\pi_R+\pi_S))g - \pi_S(1-(\pi_R+\pi_S))(g+d-l) \\ &> \pi_S(1-\pi_R-\pi_S)(l-d) > 0 \end{aligned}$$

(3) $Q_1^r < Q_1^I | \pi = \pi_R + \pi_S$.

$$\begin{aligned} Q_1^r &= (\pi_R+\pi_S)^2 g + \pi_S(1-\pi_R-\pi_S)(g+d-l) \\ &< \pi^2 g + \pi(1-\pi)(g+d-l) \\ &= \pi^2(l-d) + \pi(g+d-l) \\ &= Q_1^I \end{aligned}$$

(4) $Q_1^I < Q_1^a | \pi = \pi_R + \pi_S$.

$$\begin{aligned} Q_1^I &= \pi^2(l-d) + \pi(g+d-l) \\ &= \pi g - \pi(1-\pi)(l-d) \\ &< \pi g = (\pi_R+\pi_S)g = Q_1^a \end{aligned}$$

QED. ■

- **Proof for Claim 3**:

Proof. We first prove that given the above specified belief system, non-selfish players cannot do better by deviation. (1) In period T following the history that only (C,C) has been played in all previous $T-1$ periods, a player assigns probability π to his partner being non-selfish. In this case the last period game is the same as the above incomplete information one-shot game, where non-selfish players would play C according to Claim 2. (2) At any period $t < T$ following the history in which only (C,C) has been played, given his partner playing C according to the specified equilibrium strategy, a non-selfish player strictly prefers to play C since C dominates D due to $\alpha > d$. So non-selfish players will not deviate from the equilibrium strategy specified above by the assumption that they have $\alpha_i \geq \pi d + (1-\pi)l$. If D is played by his partner, then a selfless player would still play his dominant strategy C, while other non-selfish players would play D since (D,D) is a NE.

Next we prove that selfish players cannot do better by deviation if $\beta \geq \frac{d}{(g+d)\pi}$ given the above specified belief system. At period T, playing D is selfish players' dominant strategy, so he will not deviate. Suppose he deviates at some period $t < T$ by playing D. But then his selfish type is revealed because of the belief system. According to the equilibrium strategies, if his partner is not selfless, (D, D) is played in the left periods after t; only when his partner is selfless, (C, D) is played. Denote the proportion of selfless players by $\pi_S \equiv \Pr(\alpha_i > l)$. The deviation payoff for a selfish player from period t until T is

$$(g+d)\beta^{t-1} + (g+d)(\beta^t + \beta^{t+1} + ... + \beta^{T-1})\pi_S.$$

By not deviating he can get

$$g\beta^{t-1} + g\beta^t + ... + g\beta^{T-2} + (g+d)\pi\beta^{T-1} = g\beta^{t-1}\frac{1-\beta^{T-t+1}}{1-\beta} + d\pi\beta^{T-1}.$$

The non-deviation condition at period t for a selfish player is

$$(g+d)\beta^{t-1} + (g+d)(\beta^t + \beta^{t+1} + ... + \beta^{T-1})\pi_S < g\beta^{t-1} + g\beta^t + ... + g\beta^{T-2} + (g+d)\pi\beta^{T-1}$$

$$\Rightarrow [g - (g+d)\pi_S]\frac{\beta(1-\beta^{T-t-1})}{1-\beta} + (g+d)(\pi - \pi_S)\beta^{T-t} > d$$

The LHS's partial derivation with respect to t is

$$\begin{aligned}
\frac{\partial LHS}{\partial t} &= [g - (g+d)\pi_S]\frac{\beta^{T-t}\ln\beta}{1-\beta} - (g+d)(\pi - \pi_S)\beta^{T-t}\ln\beta \\
&= \frac{\beta^{T-t}\ln\beta}{1-\beta}[g - (g+d)\pi_S - (g+d)(\pi - \pi_S)(1-\beta)] \\
&= \frac{\beta^{T-t}\ln\beta}{1-\beta}[g - (g+d)\pi_S\beta - (g+d)\pi(1-\beta)] \\
&= \frac{\beta^{T-t}\ln\beta}{1-\beta}[(g+d)(\frac{g}{g+d} - \pi) + (g+d)(\pi - \pi_S)\beta] \\
&= \frac{-\beta^{T-t}\ln\beta}{1-\beta}(g+d)[\pi - \frac{g}{g+d} - (\pi - \pi_S)\beta].
\end{aligned}$$

It is negative if

$$\beta \geq (\pi - \frac{g}{g+d})/(\pi - \pi_S). \tag{7}$$

That is, if players are patient enough, they would wait until later to deviate, since deviation becomes more attractive as time goes by. In other words, if a selfish players do not want to deviate at period $T-1$, then they will not deviate at any time earlier. Non-deviation at period $T-1$ means

$$\begin{aligned}
(g+d)\beta^{T-2} + (g+d)\beta^{T-1}\pi_S &< g\beta^{T-2} + (g+d)\pi\beta^{T-1} \\
&\Rightarrow d < (g+d)(\pi - \pi_S)\beta \\
&\Rightarrow \beta > \frac{d}{(g+d)(\pi - \pi_S)}.
\end{aligned} \tag{8}$$

It is easy to check that condition (7) is implied by condition (8) because $\pi < 1$. So that the condition (8) guarantees that selfish players will not want to deviate at any time. To make sure that there is such β, $\frac{d}{(g+d)(\pi-\pi_S)}$ must be smaller than 1, which implies that $\pi - \pi_S > \frac{d}{g+d}$.

We have proved that the above specified strategy profile is sequentially rational w.r.t. the belief system. Now we want to show that the belief system is fully consistent given the strategy profile. In the first period, the probability that a player is non-selfish is equal to the actual proportion of non-selfish players in the population π since the match is random. In any period after the history that only (C, C) is played, the initial belief is still maintained because the two types of players cannot be distinguished from each other by both playing C. If in some period (C, D) is observed after a series of (C, C), the player who plays D must be selfish since D is never a best response for a non-selfish player when his partner plays C. So the player who does not play D must update his belief about his partner's probability of being non-selfish from π to 0. While the probability of the player who plays C in (C, D) being non-selfish is still π because both types could do so according to the equilibrium strategy profile. ■

- **Proof for Claim 4.**

Proof. (1) The Existence of Unique Solutions h_A^i and h_M^i.
The objective functions are

$$\begin{aligned} V_A^i(h) &= \beta[\Pi g(h) - (1-\Pi)l(h)] + \beta^2 pG(h) - c(h, \underline{\alpha}(h, \Pi), i), \\ V_M^i(h) &= \beta\Pi[g(h) + d(h)] - c(h, 0, i). \end{aligned}$$

The FOC of V_M^i for an interior solution is,

$$[V_M^i(h)]' = \beta\Pi[g'(h) + d'(h)] - c_h(h, 0, i) = 0 \tag{9}$$

Since $g''(h) \leq 0$, $d''(h) \leq 0$, and $c_{hh}(h, \alpha, i) > 0$, we know that $[V_M^i(h)]'$ is a decreasing function of h. If we assume that

$$\lim_{h\to 0} c_h(h, 0, i) = 0, \lim_{h\to 0} g'(h) > 0, \tag{A1}$$

we get $\lim_{h\to 0} V_M^{i\prime}(h, 0) > 0$. So there is a unique solution $h_M^i = h_M(\Pi, p, \beta, T, i, k) \geq 0$ such that $V_M^{i\prime}(h_M^i) = 0$, where k represents all other parameters.

The FOC of V_A^i for an interior solution is,

$$[V_A^i(h)]' = \beta[\Pi g'(h) - (1-\Pi)l'(h)] + \beta^2 pG'(h) - \frac{\partial c(h, \alpha(h, \Pi))}{\partial h} = 0. \tag{10}$$

The second derivative of value function $V_A^i(h)$ w.r.t. to h is

$$[V_A^i(h)]'' = \beta[\Pi g''(h) - (1-\Pi)l''(h)] + \beta^2 pG''(h) - \frac{\partial^2 c(h, \alpha(h, \Pi))}{\partial h^2},$$

where

$$\begin{aligned}\frac{\partial c(h,\alpha(h,\Pi))}{\partial h} &= c_h(h,\alpha(h,\Pi),i)+c_\alpha(h,\alpha(h,\Pi),i)\alpha_h(h,\Pi),\\ \frac{\partial^2 c(h,\alpha(h,\Pi),\Pi)}{\partial h^2} &= c_{hh}+(c_{h\alpha}+c_{\alpha h})\alpha_h(h,\Pi)+c_{\alpha\alpha}(\alpha_h(h,\Pi))^2+c_\alpha\alpha_{hh}(h,\Pi).\end{aligned}$$

A sufficient condition to enable the second order condition to hold is

$$l''(h)=0,\ \text{and}\ \frac{\partial^2 c(h,\alpha(h,\Pi),\Pi)}{\partial h^2}\geq 0. \tag{A2}$$

To guarantee a non-negative solution, we have to assume that $[V_A^i(h=0)]'\geq 0$, which requires the boundary condition

$$\lim_{h\to 0}\frac{\partial c(h,\alpha(h,\Pi))}{\partial h}=0, \lim_{h\to 0}[\beta pG'(h)+\Pi(g'(h)+l'(h))-l'(h)]>0, \tag{A1'}$$

Under these two conditions, we can get a unique solution $h_A^i=h_A(\Pi,p,\beta,T,i,k)$ such that $V_A^{i\prime}(h_A^i)=0$.

(2) Comparative Statics for h_A^i and h_M w.r.t. Π for any $i\in[0,1]$.

By Implicit Theorem,

$$\begin{aligned}\frac{\partial h_M}{\partial \Pi} &= -\frac{\partial^2[V_M^i(h)]}{\partial\Pi\partial h}/\frac{\partial^2[V_M^i(h)]}{\partial h^2}=-\beta[g'(h)+d'(h)]/\frac{\partial^2[V_M^i(h)]}{\partial h^2}>0.\\ \frac{\partial h_A^i}{\partial \Pi} &= -\frac{\partial^2[V_A^i(h)]}{\partial\Pi\partial h}/\frac{\partial^2[V_A^i(h)]}{\partial h^2}=-[\beta(g'(h)+l'(h))-\frac{\partial c(h,\alpha(h,\Pi),i)}{\partial h\partial\Pi}]/\frac{\partial^2 V_A^i(h)}{\partial h^2}>0.\end{aligned}$$

if

$$\frac{\partial c(h,\alpha(h,\Pi),i)}{\partial h\partial\Pi}\leq 0, \tag{11}$$

where

$$\begin{aligned}\frac{\partial c(h,\alpha(h,\Pi),i)}{\partial h\partial\Pi} &= \frac{\partial[c_h(h,\alpha(h,\Pi),i)+c_\alpha(h,\alpha(h,\Pi),i)\alpha_h(h,\Pi)]}{\partial\Pi}\\ &= [c_{h\alpha}(h,\alpha(h,\Pi),i)+c_{\alpha\alpha}(h,\alpha(h,\Pi),i)\alpha_h(h,\Pi)](d(h)-l(h))\\ &\quad +c_\alpha(h,\alpha(h,\Pi),i)(d'(h)-l'(h)).\end{aligned}$$

A sufficient condition for (11) to hold is

$$c_{h\alpha}(h,\alpha(h,\Pi),i)\geq 0, d'(h)\leq l'(h). \tag{A4}$$

(3) Comparative Statics for h_A^i and h_M for any $i\in[0,1]$ w.r.t. p

$$\begin{aligned}\frac{\partial h_M}{\partial p} &= -\frac{\partial^2 V_M^i(h)}{\partial p\partial h}/\frac{\partial^2[V_M^i(h)]}{\partial h^2}=0,\\ \frac{\partial h_A^i}{\partial p} &= -\frac{\partial^2 V_A^i(h)}{\partial p\partial h}/\frac{\partial^2[V_A^i(h)]}{\partial h^2}=-\beta^2 G'(h)/\frac{\partial^2 V_A^i(h)}{\partial h^2}>0.\end{aligned}$$

(4) Comparative Statics for h_A^i and h_M for any $i \in [0,1]$ w.r.t. β

$$\begin{aligned}
\frac{\partial h_M}{\partial \beta} &= -\frac{\partial^2 V_M^i(h)}{\partial \beta \partial h} / \frac{\partial^2 V_M^i(h)}{\partial h^2} = -\Pi[g'(h) + d'(h)] / \frac{\partial^2 V_M^i(h)}{\partial h^2} \geq 0, \\
\frac{\partial h_A^i}{\partial \beta} &= -\frac{\partial^2 V_A^i(h)}{\partial \beta \partial h} / \frac{\partial^2 V_A^i(h)}{\partial h^2} = -\{[\Pi g'(h) - (1-\Pi) l'(h)] + G'(h) 2\beta p\} / \frac{\partial^2 V_A^i(h)}{\partial h^2} \geq 0,
\end{aligned}$$

since $\Pi(g'(h) + l'(h)) + 2\beta p G'(h) - l'(h) > 0$ at h_A^i by condition (10). ■

- **Proof for Proposition 2.**

Proof. (1) The Relation Between h_M^i and h_M^j, h_A^i and h_A^j for any $i, j, \in [0,1]$

Since $[V_M^i(h)]' = 0$ by condition (9), we can use the Implicit Function Theorem and get

$$\begin{aligned}
\frac{\partial h_M^i}{\partial i} &= -\frac{\partial [V_M^i(h)]'}{\partial i} / \frac{\partial [V_M^i(h)]'}{\partial h} \\
&= \frac{\partial c_h(h, 0, i)}{\partial i} / \frac{-\partial^2 V_M^i(h)}{\partial h^2} \gtreqless 0, \\
\text{iff } \partial c_{hi}(h, 0, i) &\lesseqgtr 0;
\end{aligned}$$

Similarly from $[V_A^i(h)]' = 0$ we get

$$\begin{aligned}
\frac{\partial h_A^i}{\partial i} &= -\frac{\partial [V_A^i(h)]'}{\partial i} / \frac{\partial [V_A^i(h)]'}{\partial h} \\
&= \frac{\partial^2 c^i(h, \alpha(h, \Pi), i)}{\partial h \partial i} / \frac{\partial^2 V_A^i(h)}{\partial h^2} \gtreqless 0, \\
\text{iff } \frac{\partial^2 c^i(h, \alpha(h, \Pi), i)}{\partial h \partial i} &= c_{hi}(h, \alpha(h, \Pi), i) + c_{\alpha i}(h, \alpha(h, \Pi), i) \alpha_h(h, \Pi) \lesseqgtr 0.
\end{aligned}$$

Under the following assumption,

$$c_{hi}(h, 0, i) = 0, \ c_{hi}(h_A^i, \alpha(h_A^i, \Pi), i) = 0, \ c_{\alpha i}(h, \alpha(h), i) > 0, \tag{A3}$$

we get that $h_M^i = h_M$ for any $i \in [0,1]$, and $h_A^i > h_A^j$ for any $i < j$, where $i, j, \in [0,1]$.

(2) The Relation Between h_A^i and h_M for any $i \in [0,1]$

We know that $[V_A^i(h_A^i)]' = 0$ and $[V_A^i(h_A^i)]'' < 0$. If we can show that $[V_A^i(h_M)]' \gtreqless 0$, then $h_A^i \gtreqless h_M$ is proved. By condition (9), $[V_M^i(h)]' = 0$. It implies that

$$-\beta \Pi [g'(h_M) + d'(h_M)] + c_h(h_M, 0, i) = 0.$$

Add this zero term to $[V_A^i(h_M)]'$, we get

$$\begin{aligned}
[V_A^i(h_M)]' &= \underbrace{\beta^2 p G'(h_M) - \beta[\Pi d'(h_M) + (1-\Pi) l'(h_M)]}_{A(\Pi)} \\
&\quad - \underbrace{[\partial c(h_M, \alpha(h_M, \Pi), i) / \partial h_M - c_h(h_M, 0, i)]}_{B(i,\Pi)}. \\
&\equiv A(\Pi, p) - B(i, \Pi).
\end{aligned} \tag{12}$$

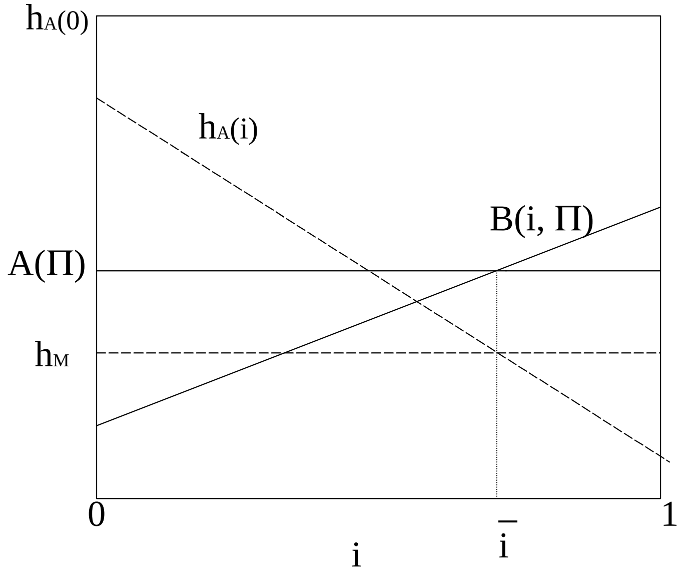


Figure 1:

The first term $A(\Pi, p)$ is the same for all players. The extra cost of investing in a positive α for player i, $B(i, \Pi)$, increases with players' index by assumption (A3):

$$\frac{\partial B(i,\Pi)}{\partial i} = c_{\alpha i}(h_M, \alpha(h_M, \Pi), i)\alpha_h(h_M, \Pi) > 0.$$

If the boundary condition $[V_A^0(h_M)]' \geq 0 \geq [V_A^1(h_M)]'$ holds given Π and p, i.e. if

$$B(1,\Pi) \geq A(\Pi, p) \geq B(0,\Pi), \tag{A5}$$

then there must exist a unique $\bar{i}(\Pi, p) \in [0,1]$ such that

$$[V_A^{\bar{i}}(h_M)]' = A(\Pi) - B(\bar{\imath}, \Pi) = 0. \tag{13}$$

See figure (1) below for illustration. This condition means that for player $\bar{i}$, his optimal choice $h_A^{\bar{i}}$ is equal to h_M. In other words, the benefit of being non-selfish is equal to the cost of investing in appropriate cooperative tendency so that his productive ability choice is not affected at all. Then for all $i \leq \bar{i}$, we have $[V_A^i(h_M)]' > 0 \Longleftrightarrow h_A^i \geq h_M$; for all $i > \bar{i}$, $[V_A^i(h_M)]' < 0 \Longleftrightarrow h_A^i < h_M$. If $A(\Pi, p) \geq B(0, \Pi)$, then $h_A^i \geq h_M$ for all i; on the other hand, if $A(\Pi, p) < B(0, \Pi)$, the opposite is true.

(3) Now we check the sign of $\frac{\partial \bar{i}}{\partial \Pi}$ based on equation (13).

$$\frac{\partial \bar{i}(\Pi, p)}{\partial \Pi} = -\frac{\partial V_A^i(h_M)/\partial h \partial \Pi}{\partial V_A^i(h_M)/\partial h \partial \bar{i}} = \frac{\partial V_A^i(h_M)/\partial h \partial \Pi}{c_{\alpha i}(h_M, \alpha(h_M, \Pi), i)\alpha_h(h_M, \Pi)} > 0$$

by condition (A4). Similarly, we get that

$$\frac{\partial \bar{i}(\Pi,p)}{\partial p} = -\frac{\partial V_A^i(h_M)/\partial h \partial p}{\partial V_A^i(h_M)/\partial h \partial \bar{i}} = \frac{\beta^2 G'(h_M)}{c_{\alpha i}(h_M, \alpha(h_M,\Pi), i)\alpha_h(h_M,\Pi)} > 0.$$

■

- **Proof of Lemma 1**

Proof. The derivative of $V_d(i,\Pi)$ with respect to i is

$$\begin{aligned}\frac{\partial V_d(i,\Pi)}{\partial i} &= \frac{\partial V_A^i(h_A^i) - \partial V_M^i(h_M)}{\partial i} = -\partial C(i,\Pi)/\partial i \\ &= -[c_i(h_A^i, \alpha(h_A^i,\Pi), i) - c_i(h_M, 0, i)] \\ &< 0,\end{aligned}$$

by assumption (A3). The second equality holds according to the Envelope Theorem, since both $V_A^i(h_A^i)$ and $V_M^i(h_M^i)$ are maximized value functions. The effects of Π and i on abilities h_A^i and h_M^i have already been taken into consideration through the maximization process, and thus have no further power on the difference between the two optimal value functions.

Now we prove $\frac{\partial V_d(i,\Pi)}{\partial \Pi} > 0$. The derivative of $V_d(i,\Pi)$ with respect to i is

$$\begin{aligned}\frac{\partial V_d(i,\Pi)}{\partial \Pi} &= \frac{\partial V_A^i(h_A^i) - \partial V_M^i(h_M)}{\partial \Pi} \\ &= \beta[g(h_A^i) + l(h_A^i) - g(h_M) - d(h_M)] + c_\alpha(h_A^i, \alpha(h_A^i,\Pi), i)[l(h_A^i) - d(h_A^i)].\end{aligned}$$

It is obvious that $\frac{\partial V_d(i,\Pi)}{\partial \Pi} > 0$ when $h_A^i \geq h_M$, which is true for low index players by condition (A5). If we can show that $\frac{\partial V_d(i,\Pi)}{\partial \Pi}$ reaches its infimum at the lowest index $i = 0$, then $\frac{\partial V_d(i,\Pi)}{\partial \Pi} > 0$ for all players. Indeed this is the case since $\frac{\partial^2 V_d(i,\Pi)}{\partial \Pi \partial i} > 0:$

$$\begin{aligned}\frac{\partial^2 V_d(i,\Pi)}{\partial \Pi \partial i} &\equiv \frac{\partial^2 V_d(i,\Pi)}{\partial i \partial \Pi} = \frac{-\partial[c_i(h_A^i, \alpha(h_A^i,\Pi), i) - c_i(h_M, 0, i)]}{\partial \Pi} \\ &= \underbrace{c_{i\alpha}(h_A^i, \alpha(h_A^i,\Pi), i)[l(h_A^i) - d(h_A^i) + \alpha_h \frac{\partial h_A^i}{\partial \Pi}]}_{\substack{>0 \\ \text{because } c_{i\alpha}(h,\alpha(h,\Pi),i) > 0,\ \frac{\partial h_A^i}{\partial \Pi} \geq 0}} \\ &\quad \underbrace{+ c_{ih}(h_A^i, \alpha(h_A^i,\Pi), i)\frac{\partial h_A^i}{\partial \Pi} - c_{ih}(h_M^i, 0, i)\frac{\partial h_M}{\partial \Pi}}_{\substack{=0 \\ \text{because } c_{ih}(h,0,i) = 0}}.\end{aligned}$$

The intuition behind $\frac{\partial^2 V_d(i,\Pi)}{\partial \Pi \partial i} > 0$ is that the high cost players get more benefit from the reduced $\alpha(h_A^i,\Pi)$ due to a higher Π. ■

- **Proof of Proposition 5.**

Proof. The arguments are very similar to the first case. Here only the different part is presented. Under these two conditions, for each $\Pi \in (0,1]$ we can find a player $i^* \in (0,1)$ such that $V_d(i^*, \Pi) = 0$. Accordingly the best response function $B(\Pi)$ is

$$B(\Pi) = \{ \begin{array}{cc} 0 & \text{if } \Pi = 0 \\ i^*(\Pi) & \text{if } \Pi \in (0,1] \end{array}$$

Let $\Pi^* \in (0,1)$ denote the solution to the equation $i^*(\Pi^*) = \Pi^*$. ■

- **Proof of Lemma 2 (3)**.

Proof. (1) $\partial B(\Pi; p)/\partial p > 0$.

$$\frac{\partial B(\Pi; p)}{\partial p} = -\frac{\partial V_d(i^*, \Pi)/\partial p}{\partial V_d(i^*, \Pi)/\partial i^*} = -\frac{\beta^2 G(h_A^i)}{\partial V_d(i^*, \Pi)/\partial i^*} > 0.$$

(2) $\partial B(\Pi; \beta)/\partial \beta > 0$.

$$\begin{aligned} \partial B(\Pi; \beta)/\partial \beta &= -\frac{\partial V_d(i^*, \Pi)/\partial \beta}{\partial V_d(i^*, \Pi)/\partial i^*} \\ &= -\frac{\Pi g(h_A^i) - (1-\Pi) l(h_A^i) + 2\beta p G(h_A^i) - \Pi[g(h_M) + d(h_M)}{\partial V_d(i^*, \Pi)/\partial i^*} > 0. \end{aligned}$$

This condition would be satisfied automatically to allow for $\pi > 0$ in equilibrium. Since investing in $\alpha > 0$ needs extra cost, the associated gain of being non-selfish should be at least bigger than zero, i.e.

$$\beta[\Pi g(h_A^i) - (1-\Pi) l(h_A^i)] + \beta^2 p G(h_A^i) - \beta \Pi[g(h_M) + d(h_M)] > 0,$$

which is smaller than

$$\Pi g(h_A^i) + 2\beta p G(h_A^i) - (1-\Pi) l(h_A^i) - \Pi[g(h_M) + d(h_M) > 0.$$

(3) $\partial F(i^*(\Pi, T))/\partial T > 0$

$$\begin{aligned} \partial B(\Pi; p, T, \beta))/\partial T &= -\frac{\partial V_d(i^*, \Pi)/\partial T}{\partial V_d(i^*, \Pi)/\partial i^*} \\ &= -\frac{\beta^2 g(h_A^i) \frac{-\beta^T \ln \beta}{1-\beta}}{\partial V_d(i^*, \Pi)/\partial i^*} > 0. \end{aligned}$$

This proof is presented here to illustrate the role of longer tenure in a model where $T \geq 1$. ■