To Trust or to Monitor: A Dynamic Analysis

Fali Huang[*]
School of Economics
Singapore Management University

September 30, 2008

Abstract

In a principal–agent framework, principals can mitigate moral hazard problems not only through extrinsic incentives such as monitoring, but also through agents' intrinsic trustworthiness. Their relative usage, however, changes over time and varies across societies. This paper attempts to explain this phenomenon by endogenizing agent trustworthiness as a response to potential returns. It finds that agents acquire lower trustworthiness when monitoring becomes relatively cheaper over time, which may actually drive up the overall governance cost in society. Across societies, those giving employees lower weights in choosing governance methods tend to have higher monitoring intensities and lower trust.

Keywords Monitoring · Trustworthiness · Trust · Screening · Economic Governance

JEL Classification Numbers D2 · J5 · L2 · M5 · Z13.

1 Introduction

All societies have to deal with moral hazard problems. But each society resolves such problems in different ways; some rely more on trust, while others depend on heavy use of governance and monitoring rules. In the late medieval period, for example, agency relations among Maghribi traders were characterized by the prevalence of trust: "Despite the many opportunities for agents to cheat, only a handful of documents contain allegations of misconduct ..." (Greif 1993). This is, however, not the case in Italy, "where allegations of misconduct are well-reflected in the historical records" (Greif 1993). In current times, labor-management relations in Japan depend on a high level of trust, while "[t]he twentieth-century American system of industrial labor relations, with its periodic massive layoffs,

[*] 90 Stamford Road, Singapore 178903. Email: flhuang@smu.edu.sg. The author gratefully acknowledges the financial support of SMU Research Grant 03-C208-SMU-018.

book-length contracts, and bureaucratic, rule-bound personal interactions, would seem the very model of low-trust social relations" (Fukuyama 1995, p. 218). In a sample of fifteen developed economies, the supervision intensities in the UK, US and Canada are the highest, with an average over two times as high as that of the rest (Gordon 1994); and in most of these countries, monitoring intensities in the manufacturing sector followed an upward trend from the 1970s to 1990s (Vernon 2001). Why do societies differ in their usage of trust and monitoring? How does it change over time? These questions are explored in this paper.

From a society's point of view, the substitution between intrinsic trustworthiness and extrinsic governance in mitigating moral hazard problems is quite obvious. If the technology is so advanced that it costs very little to achieve perfect monitoring, the society may deem it unnecessary to invest in individual trustworthiness. In contrast, if there is a sufficiently large supply of trustworthy individuals and there are easy ways to recognize them, the society may not need to improve monitoring technologies. Since both monitoring and cultivating trustworthy people are costly, most societies fall somewhere between these two extremes,[1] and their exact locations along the spectrum depend on their relative costs of inculcation, screening, and monitoring; for instance, societies with higher screening costs may adopt more monitoring than others. If the cost of monitoring decreases faster than the costs of bringing up and screening trustworthy agents, which is plausible, at least recently, given that monitoring technologies are easier to standardize and improve upon than technologies of cultivating and screening trustworthiness,[2] monitoring intensities are likely to increase over time while the average trustworthiness tends to decline.

The distribution of agent trustworthiness is endogenized in this paper through agents' skill investment choices, where the relative cost of monitoring affects the returns of being trustworthy. As a result, the average trustworthiness declines when monitoring becomes cheaper because parents anticipate that trustworthiness will be less rewarded. Cheaper monitoring technologies, however, may in the end drive up the overall governance cost by crowding out too much trustworthiness and thus forcing society to rely excessively on

[1] Considerable resources are involved in setting up schools and religious institutions, not to mention the time and resources spent by parents, to inculcate moral values in a society's people (Shavell 2002). Meanwhile, monitoring is also costly: "more than 70,000 U.S. companies spent more than $500 million on surveillance software between 1990 and 1992, and that by 1990 more than 10 million workers were under electronic surveillance" (Kipnis 1996). As a result, both intrinsic and extrinsic incentives are commonly used; see, for example, Baron and Kreps (1999) and Nagin et al. (2002).

[2] Such an imbalance in knowledge accumulation is well-observed. "The history of the West shows asymmetric progress, with advances in technological knowledge steadily progressing whereas 'progress' in institutions ... seems to be much less pronounced and monotonic. ... Unravelling the mysteries of nature has turned out to be much easier than unraveling the complexities of human interaction." (Mokyr 2003).

extrinsic incentives.[3]

The costs of monitoring, screening, and cultivating trustworthy agents, however, are not only affected by exogenous technical features, but, more importantly, also by the incentive structure that shapes the relevant resource allocation decisions on cost reduction. In fact, this paper shows that principals and agents have conflicting interests in such matters: agents are better off with higher trustworthiness and higher monitoring costs; in contrast, principals always prefer cheaper monitoring methods and do not necessarily benefit from hiring more trustworthy agents. A natural implication is that principals have strong incentives to reduce monitoring costs, but much less to reduce the costs of screening and cultivating agent trustworthiness, while the opposite is true for the agents.

Since principals have quite different incentives from agents in the choice of governance modes, which side has more weight in resource allocation becomes very important in shaping the relative cost of trust and monitoring and hence their actual usage. This yields the following cross-sectional variation: societies giving lower weights to the welfare of workers when choosing governance modes rely more on extrinsic governance and less on trust. It is indeed supported by empirical evidence: the collective labor power is negatively correlated with the supervision intensity among developed economies; specifically, the US, UK, and Canada have the lowest labor powers and the highest supervision intensities, while the opposite is true in Japan, Germany and Denmark (Esping-Anderson 1990, Gordon 1994, Rubery and Grimshaw 2003, Botero et al. 2004). A discrete version of this result is that an individualistic society tends to rely more on monitoring than a group-oriented one, if agents in individualistic societies enjoy lower weights in resource-allocating decisions; this is consistent with the distinct management styles in the US versus Japan,[4] and in the two medieval trader groups mentioned above.[5]

[3] The experiences of American firms seem to be consistent with this result. The intensively monitored workplace and the "conflict-ridden state of labor-management relations in many American industries" are held partly to blame for "the low productivity and poor quality of American work" (Mills 1994). Faced with intense competitive pressure from foreign firms, various high performance work practices relying more on cooperation efforts from employees started to be adopted from the 1980s (Appelbaum and Batt 1994, Cappelli 1995, Cappelli and Neumark 2001). But such a transforming process is slow and difficult to sustain due to "insufficient trust" (Commission on the Future of Worker-Management Relations 1994). The consequences of low trust have motivated lively discussions among public and social scientists. See, among others, Putnam (1995), Cook (2001), James (2002), and Durlauf and Fafchamps (2005).

[4] In a survey cited by then Secretary of Labor Ann McLaughlin in 1988, "only 9 percent of American workers felt they would benefit from their companies' increased productivity compared to 93 percent of Japanese workers interviewed in a similar survey" (Mills 1994). Not coincidentally, public education in Japan "does not shy away from teaching children proper 'moral' behavior, and moral education continues in the worker training programs sponsored by Japanese corporations." (Fukuyama 1995, p. 188-89)

[5] The "social structure of the Maghribi traders' group was 'horizontal,' as traders functioned as agents

The conflict of interest between principals and agents with regard to agent trustworthiness may seem puzzling at first sight. If hiring trustworthy agents reduces governance costs, then principals should necessarily benefit from it; this is also the general impression one gets from discussions on social trust. But there are two problems in such an argument: it ignores the competition among principals and the endogeneity of agent trustworthiness. As a standard result of competition, the rent from hiring more trustworthy agents disappears in equilibrium; the logic is similar in spirit to Becker's (1962) insight on firms' reluctance to invest in the general training of employees, anticipating that competitive firms will steal them away with higher wages.[6] Principals gain only when the bottom agents are more trustworthy, but then these agents have no incentives to make a costly investment in trustworthiness only to bring free windfall to principals. In other words, any rent captured by a principal is at risk of being bid away by ex post labor market competition and by ex ante investment of agents. The existence of labor market frictions may enable principals to capture some rent; such a rent, however, is not only limited in value, since it is bounded above by the degree of frictions, but also temporary in possession because it is rooted in the shifting sand of endogenous agent trustworthiness. In sharp contrast, the reduction of governance cost due to cheaper monitoring methods is immune to both hazards.

In summary, the main contribution of this paper is to provide a novel explanation for how the levels of intrinsic trustworthiness and monitoring intensities differ across societies and evolve over time. Specifically, it delivers three main implications that can be tested empirically. (1) Across societies, those giving workers lower weights in resource allocation choices tend to have higher monitoring intensities and lower trustworthiness levels. (2) As the monitoring cost is likely to fall relatively faster than screening and cultivating costs, monitoring intensities tend to increase over time while the average trustworthiness tends to decline, the more so in societies where workers have less power in making decisions. (3) Cheaper monitoring technologies may induce excessive monitoring that crowds out trustworthiness and drives up the governance cost.

In this paper, trustworthiness is essentially a trait or skill, endogenously invested by parents in one's childhood, that enables one to resist short-run opportunistic temptations.[7]

and merchants at the same time," while agency relations were organized "vertically" among the Italian traders in that "merchants and agents constitute two distinct subgroups" (Greif 1993). The Maghribi traders maintained close social ties to reduce the costs of training and screening trustworthy agents, while the Genoese traders adopted new technologies and institutions to reduce monitoring costs.

[6] In contrast, firms may be willing to invest in corporate culture (Rob and Zemsky 2002, Kreps 1997) and employee identity (Akerlof and Kranton 2005), in the same way as they are willing to invest in firm-specific human capital.

[7] A more systematic treatment of trust-related concepts based on personality traits is in Huang (2007). For other bases of trust such as social norms or altruism in the context of a principal-agent model, see

Though it often brings desirable results in terms of higher welfare, such results are typically not the goal in mind when trustworthy behaviors are exhibited. This interpretation of trustworthiness is consistent with its daily life usage as reflected in the dictionary definitions of trust, which focus on the trusted person's essential integrity and character, rather than on whether he or she has external incentives to refrain from taking advantage of others. Furthermore, the extensive experimental evidence produced over the past four decades on human behavior in social dilemmas "demonstrates that internalized trust is a common phenomenon; that it is at least in part learned rather than innate; and that different individuals vary in their inclinations toward trust." (Stout and Blair 2001)

This paper is thus related to studies of endogenous social preferences and ethical behaviors. Frank (1987) explores whether an individual wants to choose his own utility function that allows others-regarding elements. Güth and Ockenfels (2005) study the endogeneity of moral preferences using the indirect evolutionary approach, which combines individual rational decision making with the evolutionary approach of preference determination. Kaplow and Shavell (2007) examine how a social planner would inculcate guilt and virtue in individuals to foster social welfare. Using a similar approach of human capital investment as in the current paper, Huang (2007) studies the formation of social trust in the context of prisoner's dilemmas.

Another stream of literature investigates how intrinsic motivation can be crowded in or out by extrinsic incentives, such as by high-power incentive schemes (Kreps 1997, Rob and Zemsky 2002, Sliwka 2007), by public policies (Bar-Gill and Fershtman 2005), by legal institutions (Huck 1998, Bohnet, Bruno and Huck 2001, Huang 2007), and by explicit monitoring (Frey 1993). The current paper contributes to this literature by endogenizing both the intrinsic motivation and the monitoring intensities so that their dynamic interactions are studied, instead of the usual one-way crowding-out effects. More importantly, while most of these studies focus on instant or mechanical feedbacks of extrinsic incentives on agents' trustworthy behaviors, the current paper emphasizes their long-run effects on agent predisposition through rational human capital investment. The relative importance of these two channels of feedback remains to be assessed by empirical work, though available evidence suggests that a person's traits and skills are more difficult to change at older ages (Cunha et al. 2006, Cunha and Heckman 2007).

This paper proceeds as follows. In Section 2, a simple principal–agent model with monitoring and public observation of agent trustworthiness is introduced, and the intergenerational dynamics are analyzed where individual trustworthiness is endogenized through parental investment. This basic model is then extended to costly screening of agent trust-

Rotemberg (1994) and Casadesus-Masanell (2004).

worthiness in Section 3. The final section presents conclusions.

2 The Basic Model

2.1 A Principal–Agent Model with Monitoring

A principal hires an agent to complete a project. The outcome is stochastic: if the agent makes the appropriate effort, the outcome is $h > 0$ with probability $p \in [0, 1]$ and 0 with probability $1 - p$; if the agent shirks, the probability of getting h is $q \in [0, 1)$, where $q < p$, and that of getting 0 is $1 - q$. The cost of making effort is e, while shirking involves no cost. $hp - e > hq$ is assumed to be true so that making effort e is the social optimal choice.

There is a continuum of measure one of agents, who are heterogeneous in predisposition to cooperate. An agent has a degree of *trustworthiness* $\alpha \geq 0$ that measures the amount of guilt he feels if he shirks, whether caught or not by the principal.[8] The cumulative distribution function of trustworthiness among agents is $F(\cdot)$ on $\Re^+$. It characterizes the quality of workforce in this economy. Agents are risk neutral, and there is a liability constraint such that a negative payment is not feasible for agents.[9]

Principals are identical and of measure one. The reservation utility of agents and the alternative return for principals are normalized to zero. To reduce shirking, a principal may screen job candidates and monitor the agent on the job. In this basic model an agent's trustworthiness is publicly observed. A more general case is studied in the next section, where a principal can obtain a noisy signal of α through a screening process.

The monitoring intensity is denoted by $m \in [0, 1]$, which equals the probability that an agent who shirks gets caught by the principal. The total monitoring cost is mk, where k measures the unit cost of using monitoring technologies such as video cameras in the workplace. Monitoring is usually imperfect because effort is difficult to measure; for example, video cameras can record whether an agent is working, but they do not always enable the principal to tell whether the agent is making a conscientious effort or just daydreaming while working.

The payment to an agent has two components: one is the basic wage $b \geq 0$ that is independent of the agent's performance, and the other is the incentive payment $w \geq 0$,

[8] Presumably, α indicates an agent's cooperative tendency, which may lead to different levels of trustworthiness in different situations (Huang 2007). In this paper α is directly called an agent's trustworthiness because the game is fixed so that there is a one-to-one relationship between the two. Modeling α as an intrinsic benefit of cooperation does not affect the results.

[9] Risk averse agents were assumed in an earlier version of the paper, which yields similar results.

which will be forsaken if shirking is detected by the principal.[10] The utility of an agent with α is thus $w + b - e$ when he makes effort, and $(1 - m)w + b - \alpha$ if he shirks.

The time line of the game with publicly observed agent trustworthiness is as follows. Principals announce their incentive packages (m, w, b) as functions of the agent's perceived trustworthiness α. Agents then match with principals. After the matching is finished, agents get the basic wage b and choose whether to make the effort or shirk. Principals then monitor agents with intensity m, pay w if no shirking is found, and pay nothing if otherwise.[11]

2.2 The Competitive Equilibrium

The competitive equilibrium is reached in this game when there is no further changing of partners and, once in a match, nobody wants to deviate from their decisions. For positive levels of monitoring to happen and to simplify analysis, the following condition is assumed:

$$k < e < 0.5h(p - q). \tag{1}$$

We solve the game backwards.

Lemma 1 *In any given match, the optimal incentive package (m^*, w^*, b^*) includes $w^* = e$, $b^* = 0$,*

$$m^* = (e - \alpha)/e$$

for $\alpha \leq e$, and $m^* = 0$ for $\alpha > e$. Given this incentive scheme, all agents make the effort; the optimal profit of a principal is

$$Q^* = hp - e - k(e - \alpha)/e$$

if $\alpha \leq e$, and $hp - e$ if $\alpha > e$. The governance cost for $\alpha \leq e$ is

$$M^*(\alpha, k) \equiv hp - e - Q^* = k(e - \alpha)/e,$$

which decreases in α and increases in k, and zero for $\alpha > e$.

[10] Given that the paper's main focus is the interactions between trust and monitoring, and that an outcome-contingent wage serves the same purpose as monitoring in deterring shirking, the qualitative results will not be affected by allowing the incentive wage to vary across outcomes. In general, wages that are not contingent on outcomes may also be adopted when outcomes are not verifiable by the agent, or due to multi-tasking concerns (Holmstrom and Milgrom 1991).

[11] The qualitative results would not change if alternative combinations of schemes were used. For example, whatever repeated interactions can do to mitigate the moral hazard problem is either type-revealing or imposing extra extrinsic incentives, both of which are already represented by screening and monitoring in a one-period relationship.

Proof. Given the incentive package (w, m, b), an agent does not shirk if $w + b - e \geq (1 - m)w + b - \alpha$. This is simplified to the following no-shirking condition

$$\underbrace{mw}_{\text{extrinsic incentive}} + \underbrace{\alpha}_{\text{intrinsic incentive}} \geq \underbrace{e,}_{\text{cost of effort}} \tag{2}$$

where an agent won't shirk if the sum of extrinsic and intrinsic incentives is larger than the cost of effort.

For agents with $\alpha \leq e$, (2) implies that the minimum monitoring level required to deter shirking is $(e-\alpha)/w$. So if a positive monitoring level is ever chosen, the principal's objective function is

$$\max_{w,b} hp - w - \frac{k(e-\alpha)}{w} - b,$$

subject to the participation constraint $w + b - e \geq 0$. Note that $b^* = 0$ must hold, since if not, the profit can always be increased by reducing b and increasing w to reduce the monitoring intensity $(e - \alpha)/w$. The Lagrangian is thus

$$L = hp - w - \frac{k(e-\alpha)}{w} + \lambda(w - e).$$

The Kuhn-Tucker conditions are

$$\lambda - 1 + \frac{k(e-\alpha)}{w^2} = 0,$$
$$w - e \geq 0,\ \lambda \geq 0,\ \lambda(w - e) = 0.$$

If $\lambda = 0$, then $w = \sqrt{k(e-\alpha)} > e$ should hold, which cannot be true given that $k < e$ holds under assumption (1). If $\lambda > 0$, then $w^* = e$, which leads to $m^* = (e - \alpha)/e$. Given this incentive scheme, all agents make the effort, and the principal's optimal profit is

$$Q^* \equiv hp - w^* - m^*k - b^* = hp - e - k(e - \alpha)/e.$$

For agents with $\alpha > e$, their intrinsic incentive α alone is high enough to prevent shirking, so the principal would set $m^* = 0$ and $b^* + w^* = e$, where $b^* = 0$ and $w^* = e$ are assumed to be consistent with the case of $\alpha \leq e$. The profit is thus $hp - e$, which is the highest possible social surplus that can be achieved by a principal–agent couple in this economy. The gap between $hp - e$ and profit Q^* indicates the level of governance cost, which is $M^*(\alpha, k) \equiv hp - e - Q^* = k(e - \alpha)/e$ for $\alpha \leq e$ and 0 for $\alpha > e$; it coincides with the total monitoring cost m^*k. ■

Since an agent with a higher α requires a lower governance expenditure $M^*(\alpha, k)$ and thus brings a higher profit Q^*, all principals prefer to hire him; but then competition among principals would bid up a rent $r(\alpha, k)$ for the agent so that a principal's profit would become

$Q^* - r(\alpha, k)$. In contrast, an agent with the lowest trustworthiness in the economy, $\underline{\alpha} \geq 0$, gets a zero rent because there is no competitive bidding for him. So his principal earns a profit

$$Q^*_{\underline{\alpha}} = hp - e - k(e - \underline{\alpha})/e,$$

which increases in $\underline{\alpha}$ and decreases in k.

Given that principals are identical with a mass equal to that of agents, all of them would end up earning the same profit in equilibrium. This implies that a principal hiring an agent with $\alpha > \underline{\alpha}$ must pay a rent $r^*(\alpha, k)$ to her agent such that

$$r^*(\alpha, k, \underline{\alpha}) = Q^* - Q^*_{\underline{\alpha}} = k(\alpha - \underline{\alpha})/e. \tag{3}$$

The total compensation for an agent is thus equal to

$$I^*(\alpha, k, \underline{\alpha}) \equiv w^* + r^*(\alpha, k, \underline{\alpha}) = e + k(\alpha - \underline{\alpha})/e, \tag{4}$$

which increases in α and k but decreases in $\underline{\alpha}$. The income of the bottom agents with $\underline{\alpha}$ is e, independent of their own trustworthiness $\underline{\alpha}$. When $\underline{\alpha} = 0$, principals' profit becomes

$$Q^*_0 = hp - e - k, \tag{5}$$

which does not depend on agent trustworthiness at all. In contrast, when $\underline{\alpha} > 0$, principals benefit from agent trustworthiness by capturing a partial rent $Q^*_{\underline{\alpha}} - Q^*_0 = \underline{\alpha}/e > 0$.

Once all principals earn an identical profit $Q^*_{\underline{\alpha}}$, nobody wants to change agents anymore.[12] Similarly, agents do not gain from changing principals either. The agent income $I^*(\alpha, k, 0) = e + k\alpha/e$, principals' profit Q^*_0, and the governance cost $M^*(\alpha, k)$ are illustrated in Figure 1. The relevant results are summarized in the following proposition.

Proposition 1 *In the competitive equilibrium of the basic model, each principal pays a rent $r^*(\alpha, k, \underline{\alpha})$ in (3) to her agent with trustworthiness α in addition to the incentive pay $w^* = e$, so an agent's total income is $I^*(\alpha, k, \underline{\alpha})$ in (4). All principals make an identical profit $Q^*_{\underline{\alpha}}$, which is decreasing in k, increasing in $\underline{\alpha}$, but independent of α.*

The underlying intuition of this proposition is as follows. Given an agent's trustworthiness, principals adjust their monitoring intensities and incentive pays accordingly to save governance costs. Because of perfect competition between principals, the cost saved is transferred to agents as a rent, leaving principals with a profit that they would have made

[12] The profit of a principal is actually $\max\{Q^*_{\underline{\alpha}}, hq\}$, where hq arises when the principal pays the reservation wage to the agent and does not monitor him. So monitoring is chosen if and only if $Q^*_{\underline{\alpha}} \geq hq$; since $Q^*_{\underline{\alpha}} \geq Q^*_0$, this condition is true when $Q^*_0 \geq hq$, which holds by assumption (1).

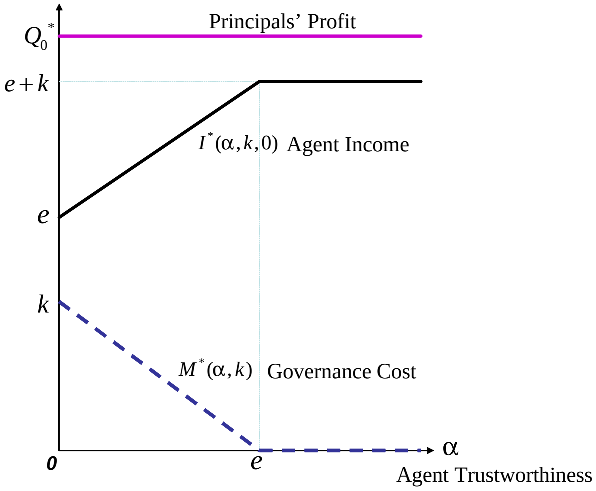

Figure 1: Principals' Profit, Agent Income and Governance Cost

when hiring the least trustworthy agents in the population. In other words, principals do not benefit from hiring agents with higher trustworthiness in equilibrium, though they do gain when the bottom agents are more trustworthy, since $Q^*_{\underline{\alpha}}$ increases in $\underline{\alpha}$. In this sense, the bottom-level trustworthiness $\underline{\alpha}$ serves as a public good for all principals.

2.3 Inter-Generational Dynamics: Endogenous Trustworthiness

The distribution of α in society is endogenized in this part. Suppose principals and agents live for one period, each raising a child to replace their role. The underlying technologies remain the same over generations, and all agent children are endowed with an identical productive ability to that of their parents. Their intrinsic trustworthiness is zero at birth, which can be improved by parental investment during childhood to maximize a child's lifetime income minus the investment cost.

The sequence of events is as follows. In the beginning of generation $n = 1, ..., \infty$, the distribution of α_n is realized. Then the above stage game is played, where the competitive equilibrium derived in Proposition 1 prevails. At the same time, the agent α_n inculcates trustworthiness α_{n+1} in his child, expecting him to get an equilibrium income $I^*(\alpha_{n+1}, k, \underline{\alpha_{n+1}})$ when the child becomes an adult, where $\underline{\alpha_{n+1}}$ denotes the lowest trustworthiness in generation $n + 1$. The inculcation cost $C(\alpha_{n+1}; \alpha_n)$ is increasing and convex in α_{n+1}, while it decreases in parental trustworthiness α_n; that is, $C_1 > 0$, $C_{11} > 0$, $C_2 < 0$, $C_{12} < 0$. And it costs nothing to retain the initial zero trustworthiness so that $C(0; \cdot) = 0$. Then generation $n + 1$ replaces the old one, and the sequence of events goes on.

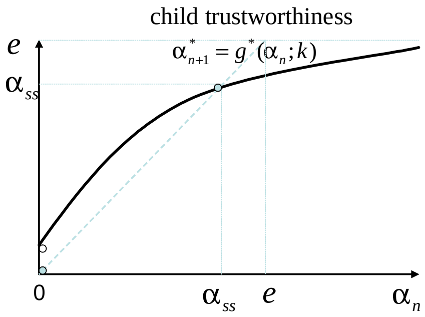

Figure 2: Endogenous Trustworthiness

The objective function of a parent in generation n is $max\{R(\alpha_n), e\}$ where

$$R(\alpha_n) \equiv \max_{\alpha_{n+1}} I^*(\alpha_{n+1}, k, \underline{\alpha_{n+1}}) - C(\alpha_{n+1}; \alpha_n). \tag{6}$$

Note that $R(\alpha_n) = e$ when $\alpha_{n+1} = 0$, that is, agent parents can always get a net return of e by not investing in their children's trustworthiness.

Given that the income of the bottom agents, e, is independent of their trustworthiness $\underline{\alpha_{n+1}}$ while investing in any positive trustworthiness is costly, $\underline{\alpha_{n+1}} = 0$ must be true. This implies that the lowest trustworthiness of agents in generation n would never be positive.[13] As a result, principals would always earn a profit Q^*_0 that is independent of α_{n+1}, and thus cannot capture any rent generated by agent trustworthiness from the second generation onwards. Similarly, no agent would have $\alpha_{n+1} > e$, since doing so yields the same income as having $\alpha_{n+1} = e$ but incurs larger investment costs. Thus we have proved the following proposition.

Proposition 2 *In any generation $n \geq 2$, the lowest trustworthiness is always 0 and the highest is not larger than e; as a result, the profit of all principals is Q^*_0 in (5).*

This proposition suggests that, if principals capture a partial rent $Q^*_{\underline{\alpha}} - Q^*_0$ in the above static model due to $\underline{\alpha} > 0$, then it has to be transferred back to agents in this dynamic model of endogenous trustworthiness. That is, principals do not benefit from agent trustworthiness once it is costly to cultivate.

The optimal solution to problem (6) with $\underline{\alpha_{n+1}} = 0$ and its comparative statics are stated in the following proposition and illustrated in Figure 2.

[13] We ignore the perverse case where α can be negative; even when it is allowed, there exists a lower bound for α, below which principals would choose not to monitor and give zero wage.

Proposition 3 *(i) There exists a unique optimal solution $\alpha^*_{n+1} \equiv g(\alpha_n; k)$ to problem (6), where $g(\alpha_n; k)$ strictly increases in α_n and k. (ii) There exists at least one stable steady state $\alpha_{ss} = g(\alpha_{ss}; k)$ when k is not too small. (iii) In all the stable steady states, α_{ss} strictly increases in k, and, contrary to the short-run result in Lemma 1, the governance cost $M^*(\alpha_{ss}, k)$ decreases in k when the elasticity of α_{ss} over k is high enough.*

Proof. In the Appendix. ■

This proposition suggests that, when trustworthiness is endogenously determined, cheaper monitoring technologies may *increase* the governance cost, which is in stark contrast to the short-run view in Lemma 1. The intuition is as follows. If monitoring is cheaper in the next generation, the lifetime return of trustworthiness is lower, so agents will invest less in it; but when the levels of trustworthiness are lower, principals have to monitor agents more intensively. When the effect of a higher monitoring intensity outweighs that of a lower unit monitoring cost, the total governance cost goes up; this happens when the elasticity of trustworthiness over k is large enough.[14] A specific case is provided by the following example.

Example. Suppose the cost function is

$$c(\alpha_{n+1}; \alpha_n) = [e^a - (e - \alpha_{n+1})^a](1 + \alpha_n)^{-b},$$

where $0 < a < 1$ and $b > 0$.[15] Then the objective function is

$$\max_{\alpha_{n+1}} I^*(\alpha_{n+1}, k, 0) - c(\alpha_{n+1}; \alpha_n) = e + k\alpha_{n+1}/e - [e^a - (e - \alpha_{n+1})^a](1 + \alpha_n)^{-b}.$$

The unique optimal child trustworthiness is

$$\alpha^*_{n+1} \equiv g(\alpha_n; k) = e - k^{\frac{-1}{1-a}}(ae(1 + \alpha_n)^{-b})^{\frac{1}{1-a}},$$

which is strictly increasing in k and α_n, and strictly concave in α_n. In any generation $n + 1$, the governance cost is

$$M^*(\alpha^*_{n+1}, k) = k(e - \alpha^*_{n+1})/e = k^{\frac{-a}{1-a}}(ae^a(1 + \alpha_n)^{-b})^{\frac{1}{1-a}},$$

which strictly decreases in k when α^*_{n+1} is endogenously determined.

[14] The mechanism is in some sense similar to the familiar phenomenon that, when the demand is elastic, a lower price may lead to a much higher demand and thus a higher total expenditure.

[15] When $a \geq 1$, the marginal benefit of investing in trustworthiness is ever-increasing before e is reached so that $\alpha^*_{n+1} = e$ for all agents.

3 Costly Screening

3.1 A Principal–Agent Model with Monitoring and Screening

An extension to the basic model is studied in this section. An agent's trustworthiness α is not publicly observed. A principal may pay a screening cost $S > 0$ to observe a signal $z = \alpha + \varepsilon$, where ε is a random variable with cdf $V(\cdot)$ and support $(-E, E)$. Screening enables principals to reduce governance cost by hiring more trustworthy agents and using more suitable incentive packages. Screening, however, is costly so that it may not be worthwhile to hire agents with low signals. Let z_l denote the threshold signal, below which an agent is not hired by a screening principal. Since the maximum reduction of the governance cost is k, a necessary condition for a positive mass of screened agents is $S < k$, which is assumed.

The time line of this screening game is similar to that in the basic model. Principals first decide whether to screen or not. Those who choose to screen announce their selection criterion z_l and the incentive package (m_s, w_s, b_s) as functions of z, hire the first agent with $z \geq z_l$ and reject others. Principals who do not screen would hire whoever comes first and adopt a single incentive package (m_r, w_r, b_r), since all agents look the same to them. Agents then decide where to apply for jobs; in fact, all agents would go to screening principals first, since searching involves zero cost for agents. If an agent is screened but turned down by a principal, it is publicly observed, though the signal z is not; then this agent can only work for a non-screening principal, since a screening principal is better off by screening a fresh agent than him. As a result, no agents are ever screened more than once. After matching is finished, agents get the basic wage and rent, if any, and then choose whether to make the effort or shirk. Principals monitor agents, pay the incentive wages if shirking is not detected, and pay nothing if detected. The competitive equilibrium is reached when all principals stick to their screening choices, there is no partner-changing, and, once in a match, nobody wants to deviate from their decisions.

Again we solve the game backwards. Given the incentive package (m_s, w_s, b_s) offered by a screening principal, the probability that the agent with a signal z will make the effort once hired is

$$\Pr(\alpha \geq e - m_s w_s) = \Pr(\varepsilon \leq z - e + m_s w_s) = V(z - e + m_s w_s),$$

as implied by the non-shirking condition (2). So with probability $V(z - e + m_s w_s)$ the agent makes the effort and produces an expected output hp, while with probability $1 - V(z - e + m_s w_s)$ he shirks and produces a lower expected output hq. He is caught with probability m_s and loses the incentive wage w_s. A screening principal's expected profit from hiring an agent with signal z is thus

$$Q_s = V(z-e+m_s w_s)(hp-w_s)+(1-V(z-e+m_s w_s))(hq-(1-m_s)w_s)-m_s k-b_s-S. \quad (7)$$

Note that the screening cost S can be interpreted as the expected cost of a successful hire, and hence does not depend on the actual number of job candidates screened.

A non-screening principal does not incur any screening cost and thus observes no signals of an agent's trustworthiness; she knows, however, that an agent not hired by a screening principal must have a signal $z < z_l$. Let $\Phi(\cdot)$ denote the cdf of the distribution of z, which is determined by the distributions of α and ε. Then given the incentive package (m_r, w_r, b_r) offered by a non-screening principal, the probability that an agent shirks is

$$\Pr(\alpha < e - m_r w_r | z < z_l) = \frac{F(e - m_r w_r)}{\Phi(z_l)},$$

as implied by (2). A non-screening principal's expected profit from hiring an agent is thus

$$Q_r = (1 - \frac{F(e - m_r w_r)}{\Phi(z_l)})(hp - w_r) + \frac{F(e - m_r w_r)}{\Phi(z_l)}(hq - (1 - m_r)w_r) - m_r k - b_r. \quad (8)$$

When the labor market clears, the proportion of agents working for non-screening principals must be equal to that rejected by screening ones, which is $\Phi(z_l)$; this is also the proportion of principals who choose not to screen. Competition among screening principals equalizes their profits, which, in equilibrium, would also be the same as the optimal profit Q_r^* made by all non-screening principals due to competition pressure from them. That is, just as in the basic model, all principals earn the same expected profit Q_r^* in equilibrium, regardless of their screening choices and their agents' trustworthiness. This is formally proved in the following proposition.

Proposition 4 *The optimal incentive packages are $w_s^* = w_r^* = e$, $b_s^* = b_r^* = 0$, while m_s^* and m_r^* maximize Q_s and Q_r, respectively. In the competitive equilibrium of this screening game: (i) There exists a unique signal z_l^* such that agents with $z \geq z_l^*$ work for screening principals and others for non-screening principals, where $\partial z_l^* / \partial S > 0$ and $\partial z_l^* / \partial k < 0$. (ii) All principals make the same profit Q_r^*, where $\partial Q_r^* / \partial S > 0$, $\partial Q_r^* / \partial k < 0$, and $Q_r^* > Q_{\underline{\alpha}}^*$; the two types of governance costs $M_s^*(z, k) \equiv hp - e - Q_s^*$ and $M_r^*(z_l^*, k) \equiv hp - e - Q_r^*$ both increase in k and decrease in z and z_l^*, respectively. (iii) An agent with trustworthiness α earns an expected income*

$$I_s^*(\alpha, k, S, Q_r^*) = e + \int_{z_l^* - \alpha}^{E} r_s^*(\alpha + \varepsilon, k, S, Q_r^*) dV(\varepsilon), \quad (9)$$

where $r_s^*(z, k, S, Q_r^*) = Q_s^* - Q_r^*$ is the rent received by an agent with $z \geq z_l^*$. $I_s^*(\alpha, k, S, Q_r^*)$ is increasing in k and α, decreasing in S and Q_r^*, and concave in α.

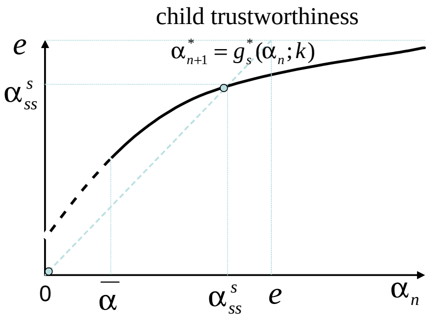

Figure 3: Endogenous Trustworthiness with Costly Screening

Proof. In the Appendix. ■

Note that the screening cost increases the profit of principals but reduces agent incomes in comparison to the basic model. The expected income of an agent is lower not only because it is the agent that ultimately pays the screening cost S, but also because principals earn a higher profit Q^*_r than before. The intuition is that the positive screening cost is a form of market friction that reduces competition among principals and hence enables them to capture some rent generated by agents. Again, as in the basic model, principals do not gain from hiring an agent with higher trustworthiness.

3.2 Endogenous Trustworthiness with Screening

The endogenization of α with costly screening is similar to that in the basic model, except that in each generation the stage game with costly screening is played, and the competitive equilibrium derived in Proposition 4 prevails.

The objective function of a parent in generation $n = 1, 2, ..., +\infty$ is

$$R_s(\alpha_n; Q^*_{r,n+1}) \equiv \max_{\alpha_{n+1}} I^*_s(\alpha_{n+1}; k, S, Q^*_{r,n+1}) - C(\alpha_{n+1}; \alpha_n),$$

taking as given $Q^*_{r,n+1}$, the equilibrium profit of principals in generation $n+1$. Since $R_s(\alpha_n; Q^*_0)$ strictly increases in α_n, there must exist a unique trustworthiness level $\overline{\alpha}$ such that

$$R_s(\overline{\alpha}; Q^*_0) = e. \tag{10}$$

To simplify analysis, we consider only the case with $\alpha^*_{n+1}(\overline{\alpha}) \geq \overline{\alpha}$ and $E \leq 0.5\overline{\alpha}$.[16]

[16] The alternative case with $\alpha^*_{n+1}(\overline{\alpha}) < \overline{\alpha}$ can be similarly analyzed with minor adjustment, which can be shown in Figure 3 by shifting down the α^*_{n+1} curve. When E is big, the problem becomes too complicated to analyze.

Proposition 5 *For any generation $n \geq 2$, the following beliefs and strategies constitute a Perfect Bayesian Nash Equilibrium: (i) Every non-screening principal believes that their agent has zero trustworthiness regardless of his signal and offers an incentive package $(m^*, w^*, b^*)|_{\alpha=0}$ as in Lemma 1, while each screening principal believes that $\alpha = z - \varepsilon$ for any $z \geq \overline{\alpha} - E$ and $\alpha = 0$ if otherwise, hires the first agent with $z \geq \overline{\alpha} - E$, and offers (m_s^*, w_s^*, b_s^*) as in Proposition 4. (ii) All principals get the same profit Q_0^* as in the basic model; agents with $\alpha_{n+1} = 0$ work for non-screening principals and get an income e, while those with $\alpha_{n+1} > 0$ work for screening principals and get an expected income $I_s^*(\alpha_{n+1}; k, S, Q_0^*)$. ($iii$) Agents who are descendants of families with $\alpha_n < \overline{\alpha}$ would have $\alpha_{n+1}^* = 0$, while those with $\alpha_n \geq \overline{\alpha}$ would choose $\alpha_{n+1}^* > 0$ to get*

$$R_s(\alpha_n; Q_0^*) \equiv \max_{\alpha_{n+1}} I_s^*(\alpha_{n+1}; k, S, Q_0^*) - C(\alpha_{n+1}; \alpha_n). \tag{11}$$

Proof. Given the belief of non-screening principals that their agents have zero trustworthiness, the optimal incentive package is $(m^*, w^*, b^*)|_{\alpha=0}$ as proved in Lemma 1. Then their profit is Q_0^* in (5) and the income of their agents is e. Due to competition, screening principals also make Q_0^*, and the extra gain $Q_r^* - Q_0^*$ captured in the static model with screening is transferred back to agents. So principals do not benefit from agent trustworthiness once it becomes endogenous, the same as in the basic model.

An agent with $\alpha_{n+1} = 0$ can only work for a non-screening principal because his highest possible signal E is smaller than the threshold $\overline{\alpha} - E$ because $E \leq 0.5\overline{\alpha}$ by assumption. Thus e is the net return of no investment in a child. Since $R_s(\alpha_n; Q_0^*)$ strictly increases in α_n and $R_s(\overline{\alpha}; Q_0^*) = e$ holds by definition, families with $\alpha_n < \overline{\alpha}$ would have $\alpha_{n+1}^* = 0$ for any generation $n \geq 1$, while only those with $\alpha_n \geq \overline{\alpha}$ would ever invest in their children. Then agents with $\alpha_{n+1}^* > 0$ are always hired by screening principals and gets an expected income $I_s^*(\alpha_{n+1}; k, S, Q_0^*)$, since their lowest signals, $\alpha_{n+1}^*(\overline{\alpha}) - E$, are higher than the threshold $\overline{\alpha} - E$ given that $\alpha_{n+1}^*(\overline{\alpha}) \geq \overline{\alpha}$. Thus all agents working for non-screening principals have zero trustworthiness, which is consistent with the belief stated in (i). ■

The optimal trustworthiness of descendants in families with $\alpha_1 \geq \overline{\alpha}$ and the relevant comparative statics are stated in the following proposition, which is similar to Proposition 3 in the basic model. See Figure 3 for illustration.

Proposition 6 (i) *There exists a unique optimal solution $\alpha_{n+1}^* \equiv g_s(\alpha_n; k)$ to problem* (11)*, where $g_s(\alpha_n; k)$ strictly increases in α_n and k. (ii) When k is not too low, there exists at least one stable steady state with $\alpha_{ss}^s = g_s(\alpha_{ss}^s; k) > 0$ for all agents from families with $\alpha_1 \geq \overline{\alpha}$ and* 0 *for others; that is, a proportion $\pi \equiv 1 - F(\overline{\alpha})$ of agents have α_{ss}^s, where π increases in k but decreases in S. (iii) In all the stable steady states, α_{ss}^s strictly* *increases in k, and, contrary to the short-run result in Proposition 4, the governance cost $M_s^*(z_{ss}, k) = hp - e - Q_s^*(z_{ss}; k)$ decreases in k when the elasticity of α_{ss}^s over k is high enough.*

Proof. In the Appendix. ■

The intuition is similar to that presented previously, except for the new insight that cheaper monitoring not only reduces the amount of investment in each child's trustworthiness, but also induces fewer agents to invest in it, the more so when the screening cost is higher.

A social planner, when deciding on how to allocate resources in reducing monitoring, screening, and inculcation costs, would take into consideration the dynamic crowding-out and crowding-in effects of monitoring technologies on agent trustworthiness. Individual principals, however, do not necessarily internalize the negative externalities they impose on agent incomes when allocating resources to reduce monitoring costs. In other words, principals tend to over-invest in monitoring technologies. The reason is that the long-run profit of principals, Q_0^*, increases when monitoring is cheaper, but it does not change when inculcation or screening costs are lower; in contrast, the incomes of agents decrease in the former case, but increase in the latter. So principals gain but agents lose when the unit cost of monitoring is lower; agents benefit from lower inculcation and screening costs, whereas principals are indifferent. This conflict of interests between principals and agents seems fundamental in determining the basic incentive structure of a society's resource allocation choices between reducing monitoring costs versus reducing inculcation and screening costs, and hence may shape the long-term trends and cross-sectional variations of trust and monitoring intensity.

4 Conclusions

This paper analyzes the dynamic relationship between trust and monitoring in reducing moral hazard problems in a principal–agent setting. Agent trustworthiness and monitoring intensities are both determined by fundamental forces in society such as the costs of monitoring and screening agents and the cost of inculcating trustworthiness; their long-term trends and cross-sectional variation are thus shaped by how these relevant costs change. While acknowledging the influence of exogenous technical features on the cost-reduction process, we argue that an important role is also played by the inherent conflict of interests between principals and agents in equilibrium: principals benefit from lower monitoring costs, but not necessarily from lower screening and inculcation costs, whereas the opposite is true for agents. When monitoring becomes relatively cheaper, agent trustworthiness declines and monitoring intensities increase over time; they may do so at faster rates in societies where the interests of agents are given a lower weight in the choice of monitoring schemes. The overall governance costs, however, may be driven up by cheaper monitoring technologies, which crowd out intrinsic incentives and induce society to rely too much on extrinsic ones. These results are indeed consistent with preliminary empirical evidence, though more rigorous tests are needed in future research.

The main insights of this paper also apply to situations with general market frictions that give principals certain monopsony powers. The following results can be readily obtained with similar arguments as in the text. Principals may capture a rent from agent trustworthiness when labor market frictions exist; the rent, however, is limited by the degree of frictions and, more importantly, it again disappears once trustworthiness becomes endogenous. Since labor market frictions increase principals' profits but reduce agent incomes, principals have less incentive to eliminate them, while the opposite is again true for agents.

This paper can be extended in various directions to get a more thorough understanding of the relevant issues. For example, the resource allocation decisions on improving various governance modes can be explicitly modeled in a bargaining or political economy environment. The screening process can be fleshed out and repeated interactions between principals and agents may be added to better address potential problems associated with screening in a diverse and mobile society. The identical production ability of agents assumed in this paper can also be relaxed to study the trade-off or complementarity between investment in cognitive and non-cognitive skills from the perspective of aggregate welfare. For instance, if the difficulty in monitoring increases when higher cognitive abilities are involved, then the thoroughness of screening and the combination of governance modes should vary across jobs in some systematic way.

Appendix

Proof of Proposition 3. The objective function is

$$\max_{\alpha_{n+1,i}} e + k\alpha_{n+1}/e - C(\alpha_{n+1}, \alpha_n).$$

The first order condition for an interior solution in $(0, e)$ is

$$k/e - C_1(\alpha^*_{n+1}, \alpha_n) = 0. \tag{12}$$

It yields the unique optimal choice $\alpha^*_{n+1} \equiv g(\alpha_n; k)$ in each generation n since the second order condition $-C_{11} < 0$ always holds. α^*_{n+1} increases in α_n and k because

$$\frac{\partial \alpha^*_{n+1}}{\partial \alpha_n} = \frac{-C_{12}}{C_{11}} > 0, \quad \frac{\partial \alpha^*_{n+1}}{\partial k} = \frac{1}{eC_{11}} > 0.$$

Note that the left-hand side of (12) is strictly increasing in k; this implies that all parents except those with the lowest level of α_n will invest in a positive α^*_{n+1} when $k/e - C_1(0,0) \geq 0$ holds. Furthermore, parents with $\alpha_n > e$ will invest in $\alpha^*_{n+1} \leq e$, and $g(\alpha_n; k)$ is continuous and increasing in α_n. These three conditions suggest that there must exist at least one stable steady state $\alpha_{ss} \in (0, e)$ such that $g(\alpha_{ss}; k) = \alpha_{ss}$ holds. And α_{ss} is unique if $g(\alpha_n; k)$ is concave; this happens if

$$g''(\alpha_n; k) = \frac{-C_{11}C_{122} + C_{12}C_{112}}{C_{11}^2} \leq 0.$$

Note that $\partial \alpha_{ss}/\partial k > 0$ because a higher k shifts up the transition function $g(\alpha_n; k)$ due to $\partial \alpha^*_{n+1}/\partial k > 0$.

The governance cost at the steady state, $M(\alpha_{ss}, k) = k(e - \alpha_{ss})/e$, may actually increase when k is lower, since

$$\frac{\partial M(\alpha_{ss}, k)}{\partial k} = \underbrace{(1 - \frac{\alpha_{ss}}{e})}_{\substack{k\text{'s direct effect on} \\ \text{governance cost}}} \quad \underbrace{-(\frac{\partial \alpha_{ss}}{\partial k}\frac{k}{\alpha_{ss}})\frac{\alpha_{ss}}{e}}_{\substack{k\text{'s indirect effect via} \\ \text{agent trustworthiness}}} \quad < 0$$

holds if the elasticity of α_{ss} over k, $\frac{\partial \alpha_{ss}}{\partial k}\frac{k}{\alpha_{ss}}$, is high enough. In fact, as long as α^*_{n+1} is endogenous, the governance cost $M(\alpha^*_{n+1}, k)$ may be higher when monitoring is cheaper. Similar arguments suggest that, when there are multiple steady states, the trustworthiness levels and the corresponding governance costs in the stable states will exhibit the same properties with respect to k.

Proof of Proposition 4. Similar arguments as in the proof of Lemma 1 suggest that $b^*_s = b^*_r = 0$. After re-arranging terms, the profit function (7) becomes

$$Q_s = hp - w_s - m_s k - (1 - V(z - e + m_s w_s))(h(p - q) - m_s w_s) - S.$$

Maximizing Q_s subject to the participation constraint $w_s - e \geq 0$ leads to the first order conditions:

$$-1 + V'(z - e + m_s w_s)(h(p-q) - m_s w_s)m_s + (1 - V(z - e + m_s w_s))m_s + \lambda_s = 0,$$
$$-k + V'(z - e + m_s w_s)(h(p-q) - m_s w_s)w_s + (1 - V(z - e + m_s w_s))w_s = 0,$$
$$w_s - e \geq 0,\ \lambda_s \geq 0,\ \lambda_s(w_s - e) = 0.$$

After rearranging the first two conditions, we get $\lambda_s - 1 + km_s^*/w_s^* = 0$. If $\lambda_s = 0$, we have $w_s^* = km_s^*$, but then it is contradictory to $w_s^* > e$, since $km_s^* \leq k$ and $k < e$ by assumption (1). If $\lambda_s > 0$, we have $w_s^* = e$, and m_s^* is uniquely determined by

$$V'(z - e + m_s e)(h(p - q) - m_s e) + 1 - V(z - e + m_s e) - k/e = 0, \tag{13}$$

since the second order condition $SOC \equiv V''(h(p-q) - m_s e)e - 2V'e < 0$ holds given $V''(\cdot) \leq 0$ and $V'(\cdot) > 0$. Based on (13) we get

$$\frac{\partial m_s^*}{\partial k} = \frac{-1/e}{-SOC} < 0.$$

By the envelope theorem we get

$$\begin{aligned} \frac{\partial Q_s^*}{\partial k} &= -m_s^* < 0, \\ \frac{\partial Q_s^*}{\partial z} &= V'(z - e + m_s^* e)(h(p - q) - m_s^* e) > 0, \\ \frac{\partial^2 Q_s^*}{\partial z^2} &= V''(z - e + m_s^* e)(h(p - q) - m_s^* e) \leq 0. \end{aligned}$$

The analysis of profit function (8) is similar. After re-arranging terms, we get

$$Q_r = hp - w_r - m_r k - \frac{F(e - m_r w_r)}{\Phi(z_l)}(h(p - q) - m_r w_r) - b_r.$$

Maximizing Q_r subject to the participation constraint $w_r - e \geq 0$ leads to the first order conditions

$$\begin{gathered} -1 + \frac{F'(e - m_r w_r)m_r}{\Phi(z_l)}(h(p - q) - m_r w_r) + \frac{F(e - m_r w_r)m_r}{\Phi(z_l)} + \lambda_r = 0, \\ -k + \frac{F'(e - m_r w_r)w_r}{\Phi(z_l)}(h(p - q) - m_r w_r) + \frac{F(e - m_r w_r)w_r}{\Phi(z_l)} = 0, \\ w_r - e \geq 0,\ \lambda_r \geq 0,\ \lambda_r(w_r - e) = 0. \end{gathered}$$

Similar analysis as above leads to $w_r^* = e$, and m_r^* is uniquely determined by

$$\frac{F'(e - m_r e)}{\Phi(z_l)}(h(p - q) - m_r e) + \frac{F(e - m_r e)}{\Phi(z_l)} - k/e = 0. \tag{14}$$

By the envelope theorem we get

$$\begin{aligned} \frac{\partial Q_r^*(z_l)}{\partial k} &= -m_r^* < 0 \\ \frac{\partial Q_r^*(z_l)}{\partial z_l} &= \Phi(z_l)^{-2} F'(e - m_r^* e)(h(p - q) - m_r^* e) > 0. \end{aligned}$$

So the average governance costs $M_s^*(z, k) \equiv hp - e - Q_s^*$ and $M_r^*(z_l^*, k) \equiv hp - e - Q_r^*$ both increase in k, the same as in the basic model, and decrease in z and z_l^*, respectively.

In general, screening and hiring a more trustworthy agent brings a higher potential profit to justify the screening cost $S > 0$, while there is no benefit in screening a non-trustworthy agent. In the extreme case when α is perfectly observed by screening principals, $Q_s^*(e) > Q_r^*(e)$ and $Q_s^*(0) < Q_r^*(0)$ must be true; when the signal z is not too noisy to nullify the advantage of screening a more trustworthy agent, $Q_s^*(z_l) > Q_r^*(z_l)$ should still hold for high levels of z_l, while the opposite should hold true when z_l is low. And as both $Q_s^*(\cdot)$ and $Q_r^*(\cdot)$ are strictly increasing functions, there must exist a unique z_l such that

$$Q_s^*(z_l^*) = Q_r^*(z_l^*). \tag{15}$$

Based on (15) we get

$$\begin{aligned} \frac{\partial z_l^*}{\partial S} &= -\frac{-1}{\partial Q_s^*/\partial z_l - \partial Q_r^*/\partial z_l} > 0, \\ \frac{\partial z_l^*}{\partial k} &= -\frac{m_r^* - m_s^*}{\partial Q_s^*/\partial z_l - \partial Q_r^*/\partial z_l} < 0, \end{aligned}$$

where $\partial Q_s^*/\partial z_l - \partial Q_r^*/\partial z_l > 0$ holds because the non-screening principal, who has less information than the screening principals about her agent's trustworthiness, is less efficient in generating profits.

As in the basic model, competition between screening and non-screening principals implies that a rent

$$r_s^*(z, k, S, Q_r^*) = Q_s^*(z) - Q_r^*(z_l^*)$$

has to be passed to an agent who has a signal $z > z_l^*$ and works for a screening principal. So all principals earn an identical profit equal to $Q_r^*(z_l^*)$, which decreases in k and increases in S because $Q_r^{*\prime}(z_l^*)\partial z_l^*/\partial k + \partial Q_r^*/\partial k < 0$ and $Q_r^{*\prime}(z_l^*)\partial z_l^*/\partial S > 0$.

An agent with α gets an expected income

$$I_s^*(\alpha, k, S, Q_r^*) = e + \int_{z_l-\alpha}^{E} r_s^*(\alpha+\varepsilon, k, S) dV(\varepsilon);$$

it is strictly increasing in k and concave in α because

$$\begin{aligned} \partial r_s^*(z, k, S, Q_r^*)/\partial k &= m_r^* - m_s^* > 0, \\ \partial r_s^*(z, k, S, Q_r^*)/\partial z &= \partial Q_s^*(z)/\partial z > 0, \\ \partial r_s^*(z, k, S, Q_r^*)/\partial z &= \partial^2 Q_s^*(z)/\partial z^2 \leq 0. \end{aligned}$$

Note, however, $I_s^*(\alpha, k, S, Q_r^*)$ is lower when the screening cost S is higher, since $Q_s^*(z)$ decreases in S. In general, principals make a higher profit when screening is available than in the basic model, that is, $Q_r^* \geq Q_{\underline{\alpha}}^*$ always holds, since non-screening principals can always get $Q_{\underline{\alpha}}^*$ by treating their agents as all having the lowest trustworthiness level $\underline{\alpha}$.

Proof of Proposition 6. The objective function is

$$\max_{\alpha_{n+1,i}} I_s^*(\alpha_{n+1};k,S,Q_0^*) - C(\alpha_{n+1},\alpha_n).$$

The first order condition for an interior solution is

$$\partial I_s^*(\alpha_{n+1};k,S,Q_0^*)/\partial\alpha_{n+1} - C_1(\alpha_{n+1},\alpha_n) = 0. \tag{16}$$

It yields the unique optimal choice $\alpha_{n+1}^* \equiv g_s(\alpha_n;k)$ in each generation n because the second order condition $SOC_s < 0$ always holds due to concavity of $I_s^*(\alpha_{n+1};k,S,Q_0^*)$ and $C_{11} > 0$. α_{n+1}^* is increasing in both α_n and k because

$$\begin{aligned} \frac{\partial\alpha_{n+1}^*}{\partial\alpha_n} &= \frac{-C_{12}(\alpha_{n+1},\alpha_n)}{-SOC_s} > 0, \\ \frac{\partial\alpha_{n+1}^*}{\partial k} &= \frac{\partial^2 I_s^*(\alpha_{n+1};k,S,Q_0^*)/\partial\alpha_{n+1}\partial k}{-SOC_s} > 0, \end{aligned}$$

where $\partial^2 I_s^*(\alpha_{n+1};k,S,Q_0^*)/\partial\alpha_{n+1}\partial k > 0$ is true given that $\partial^2 r_s^*(z_{n+1},k,S,Q_0^*)/\partial k\partial z_{n+1}$ $= -\partial m_s^*/\partial z_{n+1} > 0$. So the left-hand side of (16) strictly increases in k. This implies that when k is not too small, there exists at least one positive steady state $\alpha_{ss}^s > 0$ such that $g_s(\alpha_{ss}^s,k) = \alpha_{ss}^s$ holds, and α_{ss}^s is unique if $g_s(\alpha_n;k)$ is concave in α_n. Note that $\partial\alpha_{ss}^s/\partial k > 0$ holds, since a higher k shifts up the transition function due to $\partial\alpha_{n+1}^*/\partial k > 0$.

The governance cost at the steady state $M_s^*(z_{ss},k) = hp - e - Q_s^*(z_{ss};k)$ may actually increase when k is lower, if the elasticity of α_{ss}^s over k is high enough:

$$\frac{\partial M_s^*(\alpha_{ss}^s,k)}{\partial k} = -\underbrace{\frac{\partial Q_s^*(z_{ss};k)}{\partial k}}_{\substack{(+)\\ \text{k's direct effect on}\\ \text{governance cost}}} \quad -\underbrace{\frac{\partial Q_s^*(z_{ss};k)}{\partial z_{ss}}\frac{\partial\alpha_{ss}^s}{\partial k}}_{\substack{(-)\\ \text{k's indirect effect via}\\ \text{agent trustworthiness}}} \quad < 0$$

holds when the indirect effect is big enough. In fact, as long as α_{n+1} is endogenous, the governance cost $M_s^*(\alpha_{n+1}^* + \varepsilon, k)$ may be higher when monitoring is cheaper. Similar arguments suggest that, when there are multiple steady states, the trustworthiness levels and the corresponding governance costs in the stable states will exhibit the same properties with respect to k.

Based on $\pi = 1 - F(\overline{\alpha})$ and (10) we get

$$\begin{aligned} \frac{\partial\pi}{\partial S} &= -F'(\overline{\alpha})\frac{\partial\overline{\alpha}}{\partial S} = F'(\overline{\alpha})\frac{\partial R_s(\overline{\alpha})/\partial S}{\partial R_s(\overline{\alpha})/\partial\overline{\alpha}} = F'(\overline{\alpha})\frac{-1}{-C_2} < 0, \\ \frac{\partial\pi}{\partial k} &= -F'(\overline{\alpha})\frac{\partial\overline{\alpha}}{\partial k} = F'(\overline{\alpha})\frac{\partial R_s(\overline{\alpha})/\partial k}{\partial R_s(\overline{\alpha})/\partial\overline{\alpha}} = F'(\overline{\alpha})\frac{\partial I_s^*(\alpha_{n+1},k,S,Q_0^*)/\partial_k}{-C_2} > 0. \end{aligned}$$

So the proportion of agents with a positive trustworthiness decreases in screening cost S and increases in monitoring cost k.

References

[1] Akerlof G.A., Kranton R.E., 2005. Identity and the Economics of Organizations. *Journal of Economic Perspectives* 19(1), 9-32.

[2] Appelbaum, E., Batt R., 1994. *The New American Workplace: Transforming Work Systems in the U.S.*. Ithaca, New York: Cornell ILR Press.

[3] Bar-Gill O., Fershtman C., 2005. Public Policy with Endogenous Preferences. *Journal of Public Economic Theory* 7(5), 841-857.

[4] Baron J., Kreps D., 1999. *Strategic Human Resources.* New York: John Wiley & Sons.

[5] Becker G.S., 1962. Investment in Human Capital: A Theoretical Analysis. *Journal of Political Economy* 70(5), Part 2: Investment in Human Beings, 9-49.

[6] Bohnet I., Frey B.S., and Huck S., 2001. More Order with Less Law: On Contract Enforcement, Trust, and Crowding. *American Political Science Review* 95(1), 131-144.

[7] Botero J., Djankov S., La Porta R., Lopez-de-Silanes F., and Shleifer A., 2004. The Regulation of Labor. *Quarterly Journal of Economics* 119(4), 1339-1382.

[8] Cappelli P., 1995. Is the 'Skills Gap' Really About Attitudes? *California Management Review* Reprint Series 37(4), 108-124.

[9] Cappelli P., Neumark D., 2001. Do 'High Performance' Work Practices Improve Establishment-Level Outcomes? *Industrial and Labor Relations Review*, 737-775.

[10] Casadesus-Masanell R., 2004. Trust in Agency. *Journal of Economics & Management Strategy* 13(3), 375-404.

[11] Commission on the Future of Worker-Management Relations, appointed by U.S. Secretaries of Labor and Commerce, 1994. Fact Finding Report.

[12] Cook K.S. (Eds.), 2001. *Trust in Society.* New York: Russell Sage Foundation.

[13] Cunha F., Heckman J., 2007. The Technology of Skill Formation, *American Economic Review* 97(2), 31-47.

[14] Cunha F., Heckman J., Lochner L., and Masterov D., 2006. Interpreting the Evidence on Life Cycle Skill Formation. In *Handbook of the Economics of Education*, edited by E. Hanushek and F. Welch. Amsterdam: North-Holland, 697–812.

[15] Durlauf S., Fafchamps M., 2005. Social Capital. In: Aghion P., Durlauf S. (Eds.), *Handbook of Economic Growth*, Amsterdam: North Holland.

[16] Esping-Anderson G., 1990. *The Three Worlds of Welfare Capitalism*. Oxford: Polity Press.

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

[18] Frey B.S., 1993. Does Monitoring Increase Work Effort? The Rivalry with Trust and Loyalty, *Economic Inquiry* 31(4), 663-70.

[19] Fukuyama F., 1995. *Trust: The Social Virtues and the Creation of Prosperity*. New York, Simon & Schuster.

[20] Gordon D.M., 1994. Bosses of Different Stripes: A Cross-National Perspective on Monitoring and Supervision. *The American Economic Review Papers and Proceedings* 84(2), 375-379.

[21] Greif A., 1993. Contract Enforceability and Economic Institutions in Early Trade: the Maghribi Traders' Coalition. *American Economic Review* 83(3), 525-48.

[22] Greif A., 1994. Cultural Beliefs and the Organization of Society: A Historical and Theoretical Reflection on Collectivist and Individualist Societies. *Journal of Political Economy* 102(5), 912-50.

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

[24] Holmstrom B., Milgrom P., 1991. Multi-task Principal–Agent Analyses. *Journal of Law, Economics, and Organization* 7, 24-52.

[25] Huang F., 2007. Building Social Trust: A Human Capital Approach. *Journal of Institutional and Theoretical Economics* 163(4), 552-573.

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

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

[28] Kaplow L., Shavell S., 2007. Moral Rules, the Moral Sentiments, and Behavior: Toward a Theory of an Optimal Moral System. *Journal of Political Economy* 115, 494-514.

[29] Kipnis D., 1996. Trust and Technology. In: Kramer R.M., Tyler T.R. (Eds.), *Trust in organizations: frontiers of theory and research.* Thousand Oaks, California: Sage Publications.

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

[31] Mills D.Q., 1994. *Labor-Management Relations.* New York : McGraw-Hill, 5th ed.

[32] Nagin D.S., Rebitzer J.B., Sanders S., and Taylor L.J., 2002. Monitoring, Motivation, and Management: The Determinants of Opportunistic Behavior in a Field Experiment. *American Economic Review* 92(4), 850-873.

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

[34] Putnam R.D., 1995. Bowling Alone: America's Declining Social Capital. *Journal of Democracy* 6, 65-78.

[35] Rob R., Zemsky P., 2002. Social Capital, Corporate Culture and the Incentive Intensity. *RAND Journal of Economics* 33(2), 243-257.

[36] Rotemberg J., 1994. Human Relations in the Workplace. *Journal of Political Economy* 102(4), 684-717.

[37] Rubery J., Grimshaw D., 2003. *The Organization of Employment: An International Perspective.* Palgrave MacMillan.

[38] Shavell S., 2002. Law versus Morality as Regulators of Conduct. *American Law and Economics Review* 4(2), 227-257.

[39] Sliwka D., 2007. Trust as a Signal of a Social Norm and the Hidden Costs of Incentive Schemes. *American Economics Review* 97(3), 999-1012.

[40] Stout L.A., Blair M.M., 2001. Trust, Trustworthiness, and the Behavioral Foundations of Corporate Law. *University of Pennsylvania Law Review.*

[41] Vernon G., 2001. Work Organization and Comparative Historical Statistics on the Extent of the Managerial Hierarchy. SKOPE Research Paper No. 18, University of Oxford.