**The London School of Economics and Political Science**

*Firms, Names, and the Organization of Financial Markets*

Tianxi Wang

A thesis submitted to the Department of Economics of the London School of Economics for the degree of Doctor of Philosophy, London, February 2009

# Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).



I warrant that this authorization does not, to the best of my belief, infringe the rights of any third party.

# Abstract

The thesis examines the nature of the organization, both as a whole and as a stage set up for the members to interact. Chapter One considers why and how an organization as a whole, represented by its name, holds reputation, like a natural person, even thought it, unlike the latter, has no fixed self, or "type" as called in economics. The chapter finds that having names hold reputations improves the economic efficiency; it also discovers two mechanisms that drive organizational reputation. Chapter Two considers the optimal allocation of ownership of physical capital. The effect of the allocation on control receives little attention in the literature and is the focus of the chapter. Control means here to affect the project choice of the agent, while incentive means the choice of ex ante human capital investment and ex post effort. The chapter finds that the principal ownership improves control, yet reduces incentive of the agent, compared to the agent ownership; thus the former, called "integration", happens iff the benefit of coordination outweighs the loss in incentive. Chapter Three provides a new angle of delineating the boundary of the firm, by the allocation of the liability to investors. In a Towsend economy, it examine all modes of financing, each defined by the according allocation of the liability; particularly, Financial intermediation (FI) is defined by the fact the monitor alone takes the liability. The real race is between FI and Conglomeration, where the entrepreneurs and the monitor form a conglomerate to take the liability. FI has "Number Advantage": when default is declared, the investors audit *one* bank asset under FI but *many* entrepreneur projects under conglomerate. Conglomeration has "Collateral Advantage": its collateral is the pool of the projects contains as a part the bank asset, the collateral of FI. Both FI and Conglomeration implement the benefit of diversification; indeed, under the perfect diversification, Conglomeration is as good as FI. The chapter thus challenges the view that the benefit of diversification drives Financial Intermediation (FI), a view first established by Diamond (1984) and well accepted by the literature.

# Acknowledgements

My full gratitude is devoted to John Moore, my supervisor. More than five years ago, September 2003, when I first arrived in England, my life was miserable. Living in Zone three southwest London, I was always sick after taking one hour of serpentinely moving bus; daunted by expensiveness of everything in the cafe, I was always hungry in evenings; with not even half comprehensible English, my room seems to have been the only place comfortable to me; and the worst, after continuously failing to find the gate of economics, which was the only end of my visiting this unacquainted land, I was losing confidence of myself. Meeting John and being accepted as his PhD student in June 2005 was the turning point of my life during the PhD stage. He encourages me to fly myself; he magically understands my English, now improved to be half comprehensible, and my messy brain; more importantly, he leads me to the gate of economics and delivers to me the professional perspective of observing, contemplating, and speculating the world, through the enormous time and patience engaged into discussing my work piece by piece – in one Thursday he even came from Edinburgh to LSE at 9 am peculiarly to help me practise the presentation. All these will be impressed in my heart forever.

I am indebted to my wife; her love appeases me when I am in anxiety and her endurance contains me when I am in unreasonably bad temperament, not mentioning what she is quietly doing for me in the plain everyday life. The stream of her contributions, to the thesis in particular and to my career in general, unclaimed and unnoticed, is by no means countable.

I would like to thank several professors for all their helps and advice. Professor Leonardo Felli gave me generous time of discussing my projects, and instructions in many other things. Professor Andrea Prat kindly arranged a desk for me in STICERD, which was providing wonderful conditions for research and life, and advised me for 2004-5 school year. Professor Michele Piccione benefited me with many nice feedbacks of the project of Chapter One. I am also more than thankful to Mr Madhav Aney, Mr Jan Bena, Mr Erlend Berg, Ms Tan Chankrajang, Mr Giovanni Ko, Dr Xuewen, Dr Liu, Dr Ricardo Sousa and Dr Jidong Zhou, for their warm friendship and extensive discussions of work and life.

# Table of Contents

**INTRODUCTION TO THE THESIS** 1

**CHAPTER ONE: THE REPUTATION OF AN ORGANIZATION AND ITS DYNAMICS** 6

INTRODUCTION 6
THE BASIC MODEL 10
THE COMPLETE MODEL 19
CONCLUSION 33

**CHAPTER TWO: CONTROL VERSUS INCENTIVE: THE OPTIMAL ALLOCATION OF PHYSICAL CAPITAL** 35

INTRODUCTION 35
THE MODEL 39
THE FOUR REGIME 43
THE COMPARISONS 50
CONCLUSION AND EMPIRICAL EVIDENCES 53

**CHAPTER THREE: THE ALLOCATION OF LIABILITY TO INVESTORS: WHY FINANCIAL INTERMEDIATION** 57

INTRODUCTION 57
THE MODEL 61
INDEPENDENT FINANCE 68
DIAMOND WORLD: B-MODEL 70
CONGLOMERATION: H-MODEL 78
CONCLUSION AND DISCUSSIONS 86

**CONCLUSION OF THE THESIS** 89

**APPENDICES** 91

**REFERENCES** 107

# Introduction

The thesis examines the nature of the organization, both as a whole (Chapter 1) and as a stage or framework set up for economic agents to interact (Chapters 2 and 3).

There is abundant economic research that take a firm in itself as an economic agent that, like a natural person, has utility of its own and is capable of making strategic decisions. Chapter 1 does not take that approach which is based on personification of organizations. Rather, its objective is to study why we personify organizations, in so many ways: we let them take liability; we compose histories of them, as if they have a constant identification; we do moral evaluations upon them, as if they can behave; and what is the most important from the economics' perspective, we let them hold reputations, the reason of which the chapter examines. Organizational reputation is established and evolves in a similar way to personal reputation, but there is a difference in nature between the two. A person's reputation is anchored by the person's physical or psychological characteristics, which constitute his "self" or "type" (as is called in economics), and are supposed to change little over his life. On the contrary, there is no such a type to anchor organizational reputation. The performance of an organization is decided by the aggregate quality of its member, but the members come and go from time to time and by no physical reason those coming are at the same level of quality as those gone. And the reputation of an organization often live long after the members who actually establish the reputation have gone.

The chapter examines the economic mechanisms that drive organizational reputation. It finds that organizational reputation is actually a genius device invented by human beings to improve the economic efficiency. The intuition is as follows. Reputation is important for us to handle information asymmetry problems. Personal reputation do help with the problems; for example, the adverse selection problem besetting a obscure youth is alleviated when a reputable senior writes a reference for him to certificate his quality. But personal reputation dies with the person who holds it. This fate of reputation is saved if it is held by an organization, which is inanimate and can technically live for ever; then reputation keep functioning long after the person who establishes it has gone, which improves efficiency.

More elaborately, organizational reputation works in the following way. The achievements of reputable seniors confer reputation not only on them personally, but also on the organization which they belong to. Nobody youths joins the reputable organization to signal their quality, which is not directly observable to others. This is equivalent to the way of the seniors writing references. In other words, organizational reputation is another channel, alternative to direct reference, through which a nobody borrows reputation from a somebody. This channel has an advantage over direct reference, namely that it does not require personal contact between the nobody and the somebody; indeed the former can benefit from the latter of hundreds years ago.

With the time passing by, the youths could remain being nobodies, or become somebodies themselves. Again, their performance not only confers personal reputation to themselves, but also contributes to the reputation of the organization. Therefore, the reputation of an organization is necessarily dynamic and evolves with the performance of the members. The chapter figures out the dynamics of its evolution in the social best equilibria.

While the organization is dealt with as a whole in Chapter One, Chapter Two and Chapter Three consider how it is structured or framed as the stage for the members to interact. Chapter Two considers the classic problem of the boundary of the firm, which is first studied by Coase (1937) and followed by a vast volume of literature. The chapter is, however, motivated by the observation that the literature seldom takes into account control or coordination side but overwhelmingly concentrates on incentive side. The chapter differentiates control problems from incentive problems. Both refer to situations where a principal wants an agent to make a preferable choice among ex ante uncontracible alternatives. Control problems differ from incentive problems in on-time negotiability. If on the time when the agent is deciding the choice the decision are negotiable (namely contractible) between the principal and the agent, it is a control problem, and if not, an incentive problem. Take an example from Milgrom and Roberts (1992), where a group of players are propelling a rowboat in a match. To have each player put his oar into the water at the same time is a control problem; the action is observable, and contractible. But to have him exert high effort to pull his oar in the water is an incentive problem; he could look but is actually not labored. For another example, it is a control problem to ensure G. W. Bush to or not to invade Iraq, but it is an incentive problem to ensure him spend more time considering serious stuff rather than having fun, as he claimed that he was working even in his Texas farm.

The model of the chapter consists of a principal (she), an agent (him) and a physical capital. With the capital, the agent can carry out two exclusive projects, one leading to the product of general interest and sold directly to the market, the other leading to the product specific to the principal's need but useless to others. Before the date of choosing the project to be done, the agent chooses the level of human capital investment and after the date, he chooses the level of effort doing the chosen project. The project choice is decided through bargaining between the principal and the agent and is thus control problem. In contrast, the choices of the levels of the human capital investment and the effort are privately decided by him, not subject to bargaining, and are, therefore, incentive problems. To give the agent incentive to make the investment and to exert the effort, he should be given the payoff rights of the capital, namely he owns what he produces. The chapter examines who should get ownership rights of the capital, which mean, following Grossman & Hart (1986) and Hart & Moore (1990), residual control rights, namely the rights to use the capital as he or she wishes when bargaining fails to reach any agreement. Following them again, integration is defined as the arrangement where the principal owns the capital and non-integration as the one where the agent owns it. The chapter shows that integration induces too much control, that is, the specific project is chosen even when it is not efficient, and non-integration induces too little control, that is, the specific project is not chosen even when it is efficient.

Under integration, the agent can be regarded as a division of a M-form organization: ownership of physical capital is centralized in the hands of the principal, but the agent has the payoff rights of the division, owning what he produces. The chapter therefore points out a rationale for M-form organizations, centralized ownership of physical capital to facilitate coordination, and payoff rights remained to divisions to give them incentive.

The perspective of Chapter Two is in the line of GHM, holding that ownership structure of physical capital delineates the border between the firm and the market. A new perspective is presented in Chapter Three, where the border is decided by the allocation of the liability, which is particularly relevant in considering the existence of financial intermediations (FIs).

Why does the fund not go directly from the investors to the entrepreneurs sometimes, but passed by financial intermediaries, which obviously adds one more level of agency problems? Could any difference be made by the plain fact that the fund changes hands one more time? A common sense, plausible, suggests that the intermediaries provide extra services, besides the intermediation of fund flow. For instance, she could see the quality of assets better than the investors and thus screen projects for them to invest; or she could observe the outcomes of the invested projects and fence the investors off being cheated of misreporting. This "extra-service" point of view, however, does not hit the exact target; the intermediaries may indeed provide valuable services, but the services could be provided separated from, rather than bound up with, the intermediation of fund flow. In other words, the intermediaries could provide the services only, with the fund flowing directly between the investors and the entrepreneurs, rather than both provide the services and intermediate between the two groups. To understand the existence of FIs, it needs exactly to answer why the latter arrangement is better than the former, which an extra-service point of view fails to understand.

If the valuable services are provided both under either arrangement, then how to define FI poses a challenge for the first place. A straightforward answer is that it is by the way the funds flow. According to this Chapter, however, this is just the appearance, not the essence; what matters is the allocation of the liability to the investors: if the provider of the services takes the liability, then the arrangement is FI, whereas if the entrepreneurs take the liability by some means, it is of direct finance. The way of fund flow is not more than a hallmark of the allocation of the liability.

The Chapter considers the economics of the allocation of the liability in an economy of Townsend (1979), where the output of an entrepreneur is verified to investors only through costly auditing, but observed by an expert at minor costs. The expert can thus provide the service of observing the outputs of entrepreneurs; the service is called monitoring, following Diamond (1984). Particularly, the chapter challenges the view that FI is driven by the benefit of diversification, a view first established by Diamond (1984) and then well accepted in the literature following him. Basically, Diamond (1984) shows that under sufficient diversification, the average costs of financing an entrepreneur under FI, where the monitor becomes the bank, is lower than those under Independent Finance (IF), where each entrepreneur is financed independently and separately, without monitoring provided. The chapter finds that the same benefit of diversification is implemented by an arrangement of direct finance, called Conglomeration. Under Conglomeration, the liability is taken by the conglomerate consisting of the entrepreneurs and the monitor, where each of entrepreneur-projects becomes a division and Ms X becomes the headquarter monitoring them, advising them of the overall performance of the conglomerate and of the contribution of each division to clear the overall liability of the conglomerate. Indeed, the chapter finds that to implement the benefit of diversification, what is needed is joint liability and monitoring, which are provided under both FI and Conglomeration.

# Chapter One: The Reputation of an Organization and Its Dynamics

## I.0 Introduction of the Chapter

What is the value of a name? It stands for the past glories, and because of them, it stands for the quality of the products (or services) currently provided under the name. However, often the glories were created by members long gone, and have nothing to do with current production. How can a name still stand for current quality in these cases? Moreover, organizational reputation is tradeable through mergers & acquisitions or trademark transactions. Unknown firms can become reputable by buying reputable names; for example, Tata acquires Jaguar and Lenovo acquires the PC subsidiary of IBM to become well known in the western world. It seems, therefore, that organizational reputation is backed by nothing intrinsic. On the other hand, it goes up after a success and down after a failure, as if it is backed by some intrinsic type and the posterior of the type being good is raised by success and lowered by failure. How do we explain these dynamics, if there is actually no such a type?

Some could still argue that behind names stands something intrinsic, such as Coca Cola's secret recipe. This chapter, however, shows that such things are not necessary for names to bear reputation. They can bear reputation in a way similar to how fiat money bears value. Furthermore, names bearing reputation helps mitigate the lemons problem (Akerlof (1970)) thereby improving social efficiency.

Consider an overlapping generation (OLG) economy. Young agents of each generation choose whether or not to produce a widget. They enter production with a name, either a new name formed at no cost or an existing name bought from a retiring agent; this name carries the history of performances of the previous owners (including the agent). Only good type youths produce a useful widget with a probability high enough to generate social surplus, and the production of bad types is a social waste. The type of a youth is his private information. If names of reputable histories are believed to signal or sort out good types, all youths want these names. This belief is rationalized, if good types outbid bad types in the competition for the names. This happens, because of following two mechanisms.

The first is the *value-adding* mechanism, which depends on the dynamics of name values. Consider the case where the type is the only private information. Good type agents succeed in producing a useful widget with a higher probability than bad ones. They outbid bad agents in competing for reputable names if the resale value of the names is higher after a success than after a failure. Therefore, the dynamics of name values decide how well names signal and sort out good types. In this chapter more sorting leads to higher efficiency. The first best where only good types produce is not achievable. The second best is implemented by a simple dynamics that involves only two states, one for new names and the other for reputable names. In the dynamics, a name is brought into the reputable state by a success and into the non-reputable state by a failure, which explains the dynamics mentioned in the first paragraph.

The second is the *commitment* mechanism. To illustrate it, another piece of private information is added: each agent privately receives a noisy signal on the quality of his widget after producing and before selling it. This post-production information cannot be transmitted through the names, which were bought before the signals arrive. It is transmitted through the price of widgets if agents are incentivized to price those widgets at the true value that they know are useless. The incentive is driven by the belief that if the useless widget of a name is overpriced, all the widgets subsequently produced under the name are also useless and overpriced. In other words, a *norm of setting honest prices* is imposed upon names.[1] Whenever a name overprices a widget, it breaks the norm, is never trusted again, and thus loses its resale value. The norm

[1] In real life, organizations are indeed subject to moral judgements. For example, *Financial Times* reported ten most ethically perceived brands in France, Germany, Spain, UK and US respectively (p. 24, Tuesday, Feb., 20, 2007).

forces an agent who sells a useless widget to choose between the name's resale value and the profit from setting a high dishonest price. Thus, buying names with high-enough resale values is equivalent to committing to pricing widgets honestly. Only good types are willing to make the commitment. They are thus sorted out by those names.

In the complete model, where agents have private information of both the type and the signal, the second best dynamics vary with the social surplus generated by good types. The smaller the surplus, the greater the number of successes new names need to accumulate to accomplish the top reputation. The surplus can be proxied by the profit margin and good types by high end firms. Then, the comparative statics predict an inverse relationship between the average profit margin of high end firms of an industry and the time span for new firms of this industry to fully establish reputation. For example, firms in the high-tech industry will build reputations more quickly than firms in the wine-making industry.

This chapter is closely related to the literature on corporate names, starting with Kreps (1990). In both papers, names refer to nothing intrinsic, and as a result, there exist babble equilibria where names are devoid of reputation.[2] In contrast, Tadelis (1999) offers a model where no babble equilibria exist, due to the assumption that change in name ownership is unobservable to customers.[3] This assumption, however, raises two issues. It backs names with something intrinsic, namely the type of the last period owners; and it does not capture big

[2]This and the requirement of infinite horizon to sustain a name's reputation are the two problems criticized by Tadelis (1999). Theoretically the first one is well justified. In real life, however, many great creations of human societies are driven by proper beliefs, such as fiat money, language (Crawford and Sobel (1982)), laws and authority (Mailath, Morris and Postlewaite (2001)). The second one is not a problem even theoretically. Names can physically live forever, which, this paper shows, is in fact the reason why having them bear reputation improves social efficiency.

[3]Following that paper, Tadelis (2002, 2003) incoporates moral hazard and Marvel and Ye (2008) analyze the welfare properties of allowing names to be traded. The same assumption is made by Mailath and Samuelson (2001).

corporative names well, ownership change of which is usually exposed by the media and hardly unobservable. Hakenes and Peitz (2007) and Deb (2008) establish that names can hold reputation with observable ownership change. In Hakenes and Peitz (2007), the reputation of a name is driven by neither the value-adding mechanism nor the commitment mechanism – they actually do not describe dynamics – but by customers switching. In Deb (2008), the reputation is driven by the value-adding mechanism; this mechanism is thus independently discovered by her paper and this chapter.[4] She describes some exogenously given dynamics, as Mailath and Samuelson (2001) and Tadelis (2002, 2003), but the efficiency of those dynamics is unclear. Compared to the existing literature, this chapter makes three contributions. First, it discovers the commitment mechanism. It is a surprise that the norm of setting honest prices makes a difference in the context of purely adverse selection of this chapter; in the literature, norms work in contexts of moral hazard, such as Klein and Laffler (1981) and Shapiro (1983). Second, this chapter derives the dynamics of organizational reputation in the second best equilibria. Third, it finds the comparative statics result that relates the second best dynamics to economic fundamentals.

In this chapter, names are similar to fiat money in the way of bearing value (Samuelson (1958) and Kiyotaki and Wright (1989)). Moreover, a name is essentially a record-keeping device, as is money by Kocherlakota (1998), although that paper uses a moral hazard framework, while this chapter develops an adverse selection model. The parallel goes further in equilibria where the value of a name depends only on the number of successes net of failures, not on the order of their occurrences, as if a unit of money were given because of a success and extracted because of a failure. Those equilibria, however, are not among the second best.

the chapter consists of two parts, the basic model and the complete model; the latter adds to the former another piece of private information. The basic model is intended to highlight

[4] A previous version of this paper, Wang (2007), was presented in the Econometric Society North American summer meeting at Duke University and then submitted for publication in that year. During the submission, I learnt of Deb's paper, which, following Tadelis (2002), involves both moral hazard and adverse selection.

the value-adding mechanism and the complete model delivers the commitment mechanism and the comparative statics. Subsection I.1 examines the basic model. Subsection I.2 examines the complete model. Then subsection I.3 concludes. Some proofs are relegated to the Appendix.

## I.1 The Basic Model

The basic model is a special case of the complete model and is interesting in itself. First we lay out the model.

### I.1.1 The Model

The time is from $-\infty$ to $+\infty$, with period $t$ starting at date $t$ and ending at date $t+1$. The economy has two goods, corn (endowed good and numeraire) and widget (produced good). Each period is populated with a mass 2 continuum of sellers and much more buyers. All agents are risk neutral. Sellers live for two periods, so that in each period mass one sellers are young and the other mass one old. The one-period discount rate for young sellers is $r < 1$. Only young sellers are active. Each of them chooses to produce either one widget at cost $c$ or nothing at all. Old sellers are idle. Sellers consume no widgets but corn only. Buyers are endowed with corn and consume both. How long they live does not matter since their only role is to compose the long demand side of the widget market in each period. A widget is either useful or useless. A useful widget is worth $\overline{v}$ for the buyer, while a useless one is worth $\underline{v}$. Sellers are of two types, good or bad. A good seller produces a useful widget with probability $\overline{q}$ and a bad one with probability $\underline{q} < \overline{q}$. Without loss of generality, let $\underline{v} = 0$ and $\overline{v} = v$; $\underline{q} = 0$ and $\overline{q} = q < 1$. The proportion of good sellers is $\gamma$ for each period.

**Assumption 1:** $\gamma qv < c < qv$.

Since $c < qv$, a good seller generates social surplus $\pi =: qv - c > 0$, whereas a bad seller generates $-c < 0$. Therefore, social efficiency is measured by the extent to which bad sellers are excluded from producing widgets. The question of how to exclude them becomes interesting because of the following information structure.

A seller's type, good or bad, is his private information. The quality of a widget, useful or useless, is not observable to the buyer when it is traded, but is revealed to all the agents of this and the next generations at the end of the period by word of mouth.

**Assumption 2:** Although the quality becomes publicly known at the end of the period, it is not contractible when the widget is traded.

This assumption implies that the price of a widget cannot be based on its quality. Otherwise, if it is priced at its value, $v$ or 0, bad sellers will never enter production and the question of how to exclude them becomes trivial.[5]

Suppose that after knowing his type but before engaging in production, a young seller obtains a name for his firm. He either forms a new name at no cost or buys an existing name from a retiring seller. Then the trading of names becomes the only inter-period link of the economy. All unsold names retire out of the economy with the owners. Because only young sellers hold names, it is common knowledge that ownership of names changes each period.

Each period goes through the following stages in order.

1. Young sellers are born, privately know their types, and decide whether to produce a widget.

2. The name market opens, where young sellers buy names from retiring sellers of the last generation.

[5] The same assumption is made by Tadelis (1999, 2002, 2003). This assumption is in the spirit of the Holmstrom (1999) career concern model, where at the end of a period a manager's performance is perfectly observed by the labor market, but at the beginning his wage contract cannot be based on it. The fundamental difference is that in Holmstrom (1999), the manager never dies and hence it is about personal reputation, whereas in this paper each seller retires after one period and it is about organizational reputation.

3. Widgets are produced, and sold in the widget market to buyers, who do not observe the quality of the widgets

4. This period ends and the next starts. The quality of all the widgets is publicly revealed. The name market opens again where this generation's retiring sellers sell their names to newborn young sellers.

If a seller buys his name at price $p_0$, sells his widget at $w$, and resells the name at $p_1$, then his overall return is $R = -p_0 + w - c + rp_1$, in which $w - c$ is the profit from the widget market and $-p_0 + rp_1$ is the capital gain from the name markets. The utility of a widget buyer is $\widetilde{v} - w$, where $\widetilde{v} = v$ or 0, depending on its quality. The reservation value of sellers who do not produce and of buyers who do not purchase are both 0.

A name could be used consecutively over several periods. The history of a name until date $t$ is defined as the sequence of the qualities of the widgets produced up to period $t-1$ under that name. This history is publicly known in period $t$. A name is characterized by its history; trading names is essentially trading histories. Let "$s$" denote success (a useful widget is produced), "$f$" failure (a useless widget is produced) and "$h$" a history. A name with history $h$ is called an $h$-name. $h$ is either empty (for new names), denoted by "$\phi$", or a sequence consisting of $s$ and $f$, such as "$s$", $sf$", "$sssf$" etc. Let "$s^n$" be the abbreviation of $n$ consecutive "$s$", and similarly for $f^n$. Denote by $H^n$ the set of all histories of length $n$. Then, $H^0 = \{\phi\}$, $H^1 = \{s, f\}$, and $H^2 = \{s^2, sf, fs, f^2\}$ etc. And let $H = \bigcup_{n \geq 0} H^n = \{\phi, s, f, s^2, sf, fs, f^2, ...\}$ be the set of all possible histories.

Names evolve with the performance of their owners. If an $h$-name is owned by a good seller, then with probability $q$, he succeeds, which transforms it into an $hs$-name in the next period, while with probability $1-q$ the failure transforms it into an $hf$-name. If an $h$-name is owned by a bad seller, it will definitely become an $hf$-name.

The equilibrium concept is competitive equilibrium, which consists of prices and decisions. The prices of widgets are denoted by $w_{ht}$ and the prices of names by $p_{ht}$, where subscript $h$ represents the histories of names and $t$ the dates of trading. Only young sellers have decisions to make. They first decide whether to produce and then which names to buy. Let $e_{Bt}, e_{Gt} \in [0, 1]$ denote the probability of a bad seller and a good seller entering production respectively and let $\lambda_{ht}$ denote the proportion of good sellers among the owners of all $h$-names in period $t$. Let $\rho_{ht}$ denote the mass of $h$-names in the date $t$ name market. The total value of the names in the market, which is the transfer from generation $t$ sellers to generation $t-1$, is $V_t = \sum\limits_{h \in H} \rho_{ht} p_{ht}$.

**Definition 1** $\{p_{ht}, w_{ht}; e_{Bt}, e_{Gt}, \lambda_{ht}\}_{h,t}$ *constitutes a competitive equilibrium if and only if*

*(i): Given the prices* $\{p_{ht}, w_{ht}\}_{h,t}$*, the optimal decisions of sellers at date* $t$ *are summarized by* $e_{Bt}, e_{Gt}$*, and* $\lambda_{ht}$*.*

*(ii): Given the decisions* $\{e_{Bt}, e_{Gt}, \lambda_{ht}\}_{h,t}$*,* $p_{ht}$ *clears the market of h-names at date t.*

*(iii): Given* $\{e_{Bt}, e_{Gt}, \lambda_{ht}\}_{h,t}$*,* $w_{ht}$ *clears the market of the widgets of h-names at date t.*

*(iv) (No Ponzi):* $\lim_{t \to \infty} r^t V_{T+t} = 0$ *for any T.*

All the conditions but No Ponzi are self evident. For No Ponzi, let $R_t$ be the total return of generation $t$ sellers and $\Pi_t$ their total profit from the widget market. Besides the profits, they pays $V_t$ in total to buy names and obtain $V_{t+1}$ in total from selling the names. Therefore, $R_t = \Pi_t - V_t + rV_{t+1}$. Then, if and only if the No Ponzi holds, we have

$$\sum_{t \geq 0} r^t R_{T+t} = \sum_{t \geq 0} r^t \Pi_{T+t} - V_T \tag{1}$$

The No Ponzi condition ensures that, taken as a whole all the sellers from generation $T$ onwards, their return must come from the "real values" they create in producing widgets. Thus sellers cannot earn arbitrarily large resources by simply buying and reselling names. No Ponzi condition is used here to prick asset bubbles, as in macroeconomics literature.

Only "*stationary equilibria*" are considered in this chapter, where $p_{ht}, w_{ht}, e_{Bt}, e_{Gt}$, and $\lambda_{ht}$ do not depend on $t$, but on $h$ only. In stationary Equilibria, names are classified into states; names in the same state have the same value and evolve into the same state after a success or a failure. The dynamics of names are then Markovian transformations over the states. In principle, each history could be a separate state, and there are infinite of histories. What happens in equilibrium, however, is much simpler, as will be shown. States are denoted by capitalized histories, such as $\Phi$, S, and $\mathrm{S}^2$, which respectively denote the states containing new names, $s$-names, and $s^2$-names.

Since new names are created at no cost, $p_\phi = 0$. Because buyers are on the long side of the widget market, competition drives them to obtain their reservation value, 0, in any equilibrium; this, in combination with them being risk neutral, implies that the market clearing price of a widget equals the expected value:

$$w_h = E(v|h) = q\lambda_h v \tag{2}$$

**Lemma 1** *In any equilibrium, bad sellers obtain 0 return.*

**Proof.** Suppose otherwise, in some equilibrium bad sellers get positive return from production. Then they all enter production in the equilibrium. The average price of the widgets, equal the expected value, is thus $\gamma qv$, while the average cost is $c$. By Assumption 1, $\gamma qv < c$. No Ponzi therefore implies that on average sellers obtain less than their reservation value, which is impossible in any equilibrium. ∎

As both buyers and bad sellers get 0 surplus, all social surplus goes to good sellers. Their return, therefore, measures the social efficiency.

Before we show how names bear reputation, we first consider as a benchmark what happens if names do not bear reputation, which sheds lights on why personal reputation is not sufficient and why we need organizational reputation.

### I.1.2 Benchmark: the Babble Equilibria

Suppose names are not believed to convey any information about the type of the current sellers. Then young sellers are not willing to pay for existing names, and the following proposition holds.

**Proposition 1** *: In any equilibrium where names do not bear reputation, the social surplus is 0.*

**Proof.** : In the equilibria where no widgets are produced, the social surplus is obviously 0. For the equilibria where widgets are produced and traded, first note that the following profile of prices and decisions forms an equilibrium. $p_h = 0$ and $w_h = c$ for all $h$; no sellers buy existing names (as discussed above), and the two types of sellers enter production in such a proportion that the expected value of the widgets is $c$. Given the prices, the return of both types of sellers is $w - c = 0$. Thus they are indifferent in entry and any proportion is justifiable. Given the entry decisions, the price of widgets is $c$ by (2). The price of all names is surely 0. Thus this is an equilibrium.

No other prices are possible in the equilibria. Names' price has to be 0. If the price of widgets $w < c$, no sellers want to produce. If $w > c$, then all sellers, particularly the bad ones, get positive return if entering production, which contradicts Lemma 1. Thus, $w = c$ in all the equilibria. Hence, the return of good sellers is always 0 and so is the social surplus. ■

In this benchmark, bad sellers enter production to the extent that all the social surplus generated by good sellers is wholly dissipated. At the end of each period, a mass $q$ of retiring good sellers has succeeded and established personal reputations. However, in the next period these people are retired and bring their personal reputations out of the economy. This explains why such inefficiency arises in the benchmark. Having names bear reputation improves efficiency, exactly because names can technically live forever, but persons cannot.

The next subsection provides a complete characterization of equilibrium payoffs and thus shows what can be done by having names bear reputation.

### I.1.3 The Characterization of Equilibrium Efficiency

A series of equilibria, ordered by $\lambda \in [\frac{c}{qv}, 1]$, are constructed below. The efficiency of the equilibria increases with $\lambda$; when $\lambda = 1$, we arrive at the highest efficiency, and when $\lambda = \frac{c}{qv}$, we go back to the babble equilibria, the lowest efficiency.

Two-State $\lambda$–Equilibrium (TSE-$\lambda$): In this equilibrium, names are of two states, $\Phi$ and $S$, for new names and reputable names respectively. A $\Phi$-name becomes an $S$-name after a success and remains a $\Phi$-name after a failure. An $S$-name remains an $S$-name after a success, and degenerates into (or is replaced by) a $\Phi$-name after a failure. The dynamic is illustrated as follows.

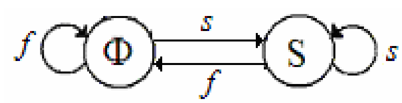


Figure 1: the Two-State Dynamics

The prices of names are $p_\Phi = 0$ and $p_S = \lambda qv - c$. The prices of widgets are $w_S = \lambda qv$ and $w_\Phi = c$. The decisions are summarized by $\lambda_S = \lambda$ and $\lambda_\Phi = \frac{c}{qv}$; $e_G = 1$ and $e_B = \frac{\gamma}{1-\gamma}\frac{\lambda(1-q)(1-\lambda_\Phi)+(1-\lambda)(q\lambda_\Phi+1-\lambda_\Phi)}{\lambda_\Phi}$.[6]

**Lemma 2** *The above profile of prices and decisions forms an equilibrium for any* $\lambda \in [\frac{c}{qv}, 1]$.

**Proof.** Let us verify that conditions (i)-(iv) are satisfied. Obviously, condition (iii), equivalent to (2), is satisfied, and so is No Ponzi (condition (iv)).

[6] To find $e_B$, first notice that in the steady state, the inflow of state $S$ equals the outflow, that is, $\rho_\Phi \lambda_\Phi q = \rho_S[\lambda(1-q)+1-\lambda]$. As all good sellers enter ($e_G = 1$), and they own either $\Phi$-names or $S$-names, $\gamma = \rho_\Phi \lambda_\Phi + \rho_s \lambda$. From these two equations, we find $\rho_\Phi$ and $\rho_S$. Then the total mass of bad sellers entering production is $\rho_\Phi(1-\lambda_\Phi) + \rho_s(1-\lambda)$, which divided by $1-\gamma$ gives $e_B$.

For condition (i), check that both types of sellers are indifferent in buying either state of names, and that good sellers get nonnegative return while bad sellers get 0. Hence any $\lambda_S, \lambda_\Phi$ and $e_B$ are optimal, and the optimal $e_G = 1$. If good sellers buy $S$-names, their return is $-p_S + w_S - c + r[qp_S + (1-q)p_\Phi] = rqp_S$. If they buy $\Phi$-names, the return is $-p_\Phi + w_\Phi - c + r[qp_S + (1-q)p_\Phi] = rqp_S$ again, where $p_S = \lambda qv - c \geq 0$. Hence good sellers are indifferent between buying $\Phi$- and $S$-names and prefer entering production. If bad sellers buy $S$-names, the return is $-p_S + w_S - c + rp_\Phi = 0$. If they buy $\Phi$-names, it is $-p_\Phi + w_\Phi - c + rp_\Phi = 0$. Hence, bad sellers are indifferent in buying any names, and in entry.

For condition (ii), given that good sellers buy both states of names on the equilibrium path, they must be indifferent in buying either state of names at the market clearing price of $S$-names. That is, $-p_S + w_S - c + rqp_S = w_\Phi - c + rqp_S \Rightarrow p_S = w_S - w_\Phi = \lambda qv - c$, as specified above. ■

In TSE-$\lambda$, the return of good sellers, which measures efficiency, is $rq(\lambda qv - c)$. It increases continuously with $\lambda$. Measured by the number of the states of names, the TSE are the simplest after the babble equilibria; the former involves two states, $\Phi$ and $S$, while the latter involves only one. By Proposition 1, the latter implements the lowest social efficiency. Therefore, it is surprising that, with only one more state added, the TSE already implement all levels of equilibrium efficiency with $\lambda \in [\frac{c}{qv}, 1]$. The following lemma helps prove this assertion.

**Lemma 3** $p_h \leq \frac{\pi}{1-r}$ *for any $h$ in equilibrium.*

**Proof.** See Appendix C. ■

Intuitively, if some $h$-names are sold at a price higher than $\frac{\pi}{1-r}$, for sellers to buy these names, the sum of the names' resale values $(q\lambda_h p_{hs} + (1 - q\lambda_h)p_{hf})$ must be even higher. The same consideration holds true for those names that evolve from the $h$-names (namely, $hs, hf, hs^2, hsf$ etc.), which pushes the sum total of the values of these subsequently evolved names higher and higher, and in the end breaks No Ponzi condition.

The main proposition of the basic model is stated here.

**Proposition 2** *The surplus of TSE-1, $rq\pi$, is the maximum social surplus among all equilibria. Therefore, the series of TSE-$\lambda$ implement all levels of equilibrium efficiency with $\lambda \in [\frac{c}{qv}, 1]$.*

**Proof.** Given any equilibrium, we are going to show that the equilibrium return of good sellers is not greater than $rq\pi$. For the equilibrium, $P = \sup\{p_h | h \in H\}$ is well defined by lemma 3. For any $\varepsilon$ such that $0 < \varepsilon < c$, there exist $h^*$-names such that $p_{h^*} > P - \varepsilon$. First, not all $h^*$-names are bought by bad sellers in equilibrium. Otherwise, the names sort out useless widgets, and $w_{h^*} = 0$. Buying the names, bad sellers obtain $-p_{h^*} + w_{h^*} - c + rp_{h^*f} \leq -p_{h^*} - c + rP < -P + \epsilon - c + rP < 0$. Thus they should not buy the names, a contradiction.

Thus $h^*$-names are bought by good sellers on the equilibrium path. The return of good sellers buying the names is $-p_{h^*} + w_{h^*} - c + r[qp_{h^*s} + (1-q)p_{h^*f}] \leq -p_{h^*} + w_{h^*} - c + r[qP + (1-q)p_{h^*f}] = rq(w_{h^*} - c) + rq(P - p_{h^*}) + (1-q)(-p_{h^*} + w_{h^*} - c + rp_{h^*f}) + q(1-r)(-p_{h^*} + w_{h^*} - c)$. Let us check the last sum term by term. For the first two terms, $w_{h^*} - c = \lambda_{h^*}qv - c \leq \pi$ by (2), and $P - p_{h^*} < \varepsilon$. As to the third and the fourth terms, consider what bad sellers get if they buy $h^*$-names. Their return is $-p_{h^*} + w_{h^*} - c + rp_{h^*f}$, which is nonpositive. It follows that the fourth term $-p_{h^*} + w_{h^*} - c \leq -rp_{h^*f} \leq 0$. Therefore, the return of good sellers buying $h^*$-names is no bigger than $rq\pi + rq\varepsilon$. This return is the equilibrium return of good sellers, since they buy $h^*$-names on the equilibrium path. The equilibrium return is thus no bigger than $rq\pi + rq\varepsilon$, for any $\varepsilon$ such that $0 < \varepsilon < c$. The proposition is proved by making $\varepsilon$ go to 0. ■

For an intuition, consider the extreme case where $r = 1$ and some $h^*$-names actually take the top value $P$. The return of good sellers buying these names consists of the profit, $w_{h^*} - c$, and the capital gain. As the $h^*$-names take the top value, the sellers obtain no capital gain from success. However, when they fail, the capital loss must be no less than the profit; otherwise, bad sellers earn positive return by buying the $h^*$-names. Therefore, at least $1 - q$ of the profits are offset by the expected capital loss and the return is thus no more than $q(w_{h^*} - c) = q(\lambda_{h^*}qv - c) \leq q(qv - c) = q\pi$.

By Proposition 2, we know the first best is not achievable: in the first best, only good sellers produce and hence the social surplus is $\pi > rq\pi$. Therefore names function as only an imperfect substitute for the contracts that would implement the first best when the quality of widgets is verifiable.

There are equilibrium dynamics that involve more than two states. An example is given in Appendix A, which involves four states. It implements the second best efficiency, however, if and only if it degenerates to the two-state dynamics above, that is, among the four states, two of them are equivalent and so are the other two.

Here we end the examination of the basic model. The next section considers the complete model that adds one element to the basic model, the post production signal. The basic model is thus a special case of the complete model where the signal is not informative at all. The complete model delivers two points that the basic model fails to deliver. It shows that organizational reputation can arise through a norm of setting honest prices for widgets. Also, it shows that in the second best equilibria, the smaller is $\pi$, the greater is the number of successes a name needs to accumulate in order to accomplish the top reputation. In contrast, in the basic model only one success is needed to accomplish the top reputation, independent of $\pi$.

## I.2 The Complete Model

Subsection 3.1 first presents the new element, the post-production signal. To utilize this extra information, the norm of honestly pricing widgets is introduced. This norm drives a new mechanism to sort out good types, the commitment mechanism.

### I.2.1 The Signal, the Norm, and the Commitment Mechanism

In the basic model, a seller has only private information of his type. In the complete model, besides the type, he privately receives a signal about the quality of his widget when it has been produced. The signal, denoted by $\widetilde{m}$, is either "n" ("nice") or "u" ("useless"), according to the following conditional distribution:

$$\Pr(\widetilde{m} = n|\widetilde{v} = v) = 1; \Pr(\widetilde{m} = n|\widetilde{v} = 0) = 1 - \tau \text{ and } \Pr(\widetilde{m} = u|\widetilde{v} = 0) = \tau < 1.$$

$\tau$ measures the informativeness of the signal. If $\tau = 0$, it is completely uninformative and we go back to the basic model. If $\tau = 1$, the seller knows precisely the quality of the widget, and there is no interesting stationary Markovian dynamics (see footnote 14). So $\tau < 1$ is assumed.

The timing is the same as in the basic model, except that at stage 3 (see page 7), after the widgets are produced and before they are sold, sellers receive the signals.

An additional assumption is introduced.

**Assumption 3:** $\frac{1-r}{(1-rq)rq} < \frac{qv-c}{qv}$.

The assumption complements Assumption 1 and says that the discount rate, $r$, is close to 1 enough. Its significance will be clear when Proposition 3 is proved in Subsection 3.2.

**The Norm of Setting Honest Prices:**

In order for the signals to make a difference, it is necessary to impose upon names a norm of setting honest prices for the widgets. The information of the signals is not transmitted by the names, which were bought before the signals arrive. If this information is to be utilized at all, it must be transmitted through the prices of the widgets. In particular, sellers must be incentivized to set price 0 for those widgets that they know are useless, even though they would like to sell at a higher price. After selling the widgets, sellers care only for the resale value of their names.

The incentive, therefore, must consist in the benefit of setting price 0 on the resale value. The benefit comes as follows.

Suppose buyers believe that a name that ever set a positive price for a useless widget keeps producing useless widgets and overpricing them. Then, no seller would want this name and the name's resale value would be destroyed. That is, names are subject to a *norm of setting honest prices*. Whenever a name sets a positive price for a useless widget, it breaks the norm, is regarded as *dishonest*, and is not trusted any more. This name then becomes useless and is destroyed. The norm will improve efficiency in this setting of purely adverse selection.

In this economy, it is not socially efficient to impose the norm upon all names and thus destroy any names that break the norm. If the resale value of a name is lower than the gain from setting a dishonest price, the threat of losing the value does not deter the seller from cheating. Imposing the norm in such a situation merely destroys the name, a resource of the economy, without helping transmit the signal's information. By this argument, in the socially best equilibria, we destroy because of dishonesty only those names for which the resale value is larger than the highest gain from cheating. This gain equals the highest widget price that buyers will ever accept: $\overline{w} \equiv E(\widetilde{v}|G, \widetilde{m} = n) = \frac{qv}{q+(1-q)(1-\tau)}$, where $G$ represents the condition that the widget is produced by a good seller. The present resale value of an $h$-name after producing a useless widget is $rp_{hf}$. Therefore, the norm is imposed upon only such $h$-names that $p_{hf} \geq \frac{\overline{w}}{r}$.

**The Commitment Mechanism**

In the basic model, organizational reputation is driven by the value-adding mechanism: good sellers succeed with a higher probability than bad ones; they outbid the latter for reputable names, because they can add value to the names with success. The norm introduces another sorting mechanism. Because of the norm, buying names of resale values $p_{hf} \geq \frac{\overline{w}}{r}$ is equivalent to committing ex ante to pricing the widgets honestly. These names thus sort out good sellers, because only good sellers are willing to make the commitment, whereas bad sellers are not. This mechanism is called the *commitment mechanism*. Accordingly, names such that $p_{hf} \geq \frac{\overline{w}}{r}$ are "*commitment names*", and names such that $p_{hf} < \frac{\overline{w}}{r}$ are "*non-commitment names*". The price of a non-commitment name's widget depends only on the name's history, as by (2), not on the post-production signal. In contrast, the price of a commitment name's widget depends both on the history and the post-production signal. The resale price of a commitment name, after a failure, is either $p_{hf}$ or 0 (the name being destroyed), contingent on the price of the widget.

The belief as to the pricing behavior drives the norm, which drives the commitment mechanism. The norm, however, makes a difference only if the post production signal is informative enough; therefore, that belief would make no difference in the basic model even if it were introduced there. This is proved in the next subsection. To prove it, we spell out first the constraints to which the dynamics of name values are subject in equilibrium. These constraints also help us study the equilibria with the highest efficiency.

#### I.2.2 The Dynamics and the Necessity of Post-Production Information

In any equilibrium, the dynamics of name values is decided by No Ponzi condition and following two incentive compatibility constraints. (E1) good sellers obtain the same return $R_G \geq 0$ from any names they buy on the equilibrium path; and (E2) bad sellers obtain 0 return on the equilibrium path and non-positive return off it. (E2) is proved in Lemma 1, which only depends on Assumption 1 and No Ponzi condition, and thus holds true in the complete model.

First, for both commitment names and non-commitment names,

$$p_{hf} \leq \frac{p_h + c}{r} \tag{3}$$

Otherwise, (E2) is violated; bad sellers would obtain a positive return, $-p_h - c + rp_{hf}$, by buying $h$-names and setting $w_h = 0$, which keeps the resale value, $p_{hf}$, in any case.

Then, we consider two types of names case by case. Consider non-commitment names first. If $h$-names are only bought by bad sellers, they become a sign of useless widgets. Thus $w_h = 0$

and the bad sellers obtain $-p_h - c + rp_{hf}$, which equals 0 by (E2). That is,

$$p_{hf} = \frac{p_h + c}{r} \tag{4}$$

If good sellers buy $h$-names, they obtain $R_G = -p_h + w_h - c + r(qp_{hs} + (1-q)p_{hf}) = (-p_h + w_h - c + rp_{hf}) + rq(p_{hs} - p_{hf})$. Two subcases arise. If $-p_h + \pi + rp_{hf} > 0$, bad sellers also buy the $h$-names in equilibrium.[7] By (E2), $-p_h + w_h - c + rp_{hf} = 0$. Thus $R_G = rq(p_{hs} - p_{hf})$. If $-p_h + \pi + rp_{hf} \le 0$, then $\lambda_h = 1$, which implies $w_h = qv$.[8] Consequently, $R_G = -(p_h - \pi - rp_{hf}) + rq(p_{hs} - p_{hf})$. Let $\Delta_h \equiv \max\{p_h - \pi - rp_{hf}, 0\}$. Then the two subcases are summarized altogether into

$$R_G = rq(p_{hs} - p_{hf}) - \Delta_h \tag{5}$$

Consider the case of $h$-names being commitment names. If bad sellers buy them, they honestly set price 0 for the widgets and (4) follows. Consider a good seller who buys such an $h$-name. With probability $(1-q)\tau$, he receives signal u, knows the uselessness of his widget, and accordingly sets price 0, to keep the resale value, $p_{hf}$. With probability $q + (1-q)(1-\tau)$, he receives signal n and sets price $\overline{w} = E(\widetilde{v}|G, \widetilde{m} = n)$ with probability 1.[9] However, with probability $\Pr(\widetilde{v} = 0|G, n) = \frac{(1-q)(1-\tau)}{q+(1-q)(1-\tau)}$ his widget is actually useless and hence price $\overline{w}$ is regarded as dishonesty, which leads the name to be destroyed; with probability $\frac{q}{q+(1-q)(1-\tau)}$, the widget is indeed useful and the name is resold at price $p_{hs}$. Therefore, $R_G = -p_h - c + (1 -$

[7] Otherwise, suppose no bad sellers buy the names. Then $w_h = qv$, by (2) (for non-commitment names). Therefore, if buying the names, bad sellers would obtain $-p_h + w_h - c + rp_{hf} = -p_h + \pi + rp_{hf} > 0$, contradictory to (E2).

[8] Otherwise, suppose $\lambda_h < 1$ and thus $w_h < qv$. Then, if buying the names, bad sellers obtain $-p_h + w_h - c + rp_{hf} < -p_h + \pi + rp_{hf} \le 0$, and hence they should not buy them, contradictory to the supposition that $\lambda_h < 1$.

[9] Suppose otherwise, concerned about keeping the resale value, he sets price 0 with some probability $\mu > 0$. Then conditional on price 0 and the $h$-names, the expected value of the widgets is positive: $E(\widetilde{v}|h, w = 0) \propto \mu \cdot \overline{w} > 0$. In equilibrium the price of these widgets equals the expected value and hence is positive, which contradicts the supposition that the price is 0.

$q)\tau[0 + rp_{hf}] + [q + (1-q)(1-\tau)][\overline{w} + \frac{q}{q+(1-q)(1-\tau)} rp_{hs}]$, which is simplified as

$$R_G = -p_h + \pi + rqp_{hs} + r(1-q)\tau p_{hf} \tag{6}$$

An equilibrium dynamics of name values is a function $p : H \longrightarrow R^+ \cup 0$, where $H$ is the set of all possible histories, such that

($a$): for new name, $p_\phi = 0$, and for any names, (3);

($b$): for non-commitment names, (4) if $h$-names are bought by bad sellers only and (5) otherwise;

($c$): for commitment names, (4) if $h$-names are bought by bad sellers ever and (6) if they are bought by good sellers;

($d$): No Ponzi.

As in the basic model, No Ponzi implies that $p_h < \frac{\pi}{1-r}$ by Lemma 3, the proof of which depends only on the two equilibrium conditions (E1) and (E2), not on any specific dynamic equations, and thus holds true in the complete model. Therefore, $P = \sup\{p_h | h \in H\}$ is well defined for any given equilibrium.

**Lemma 4** *(commitment mechanism) If $h$-names are commitment names and $p_h > rP - c$, then $h$-names sort out good sellers.*

**Proof.** If bad sellers ever buy the names, by (c), $p_{hf} = \frac{p_h + c}{r} > P$, contradictory to the definition of $P$. ■

By the lemma, it seems that only the commitment names within the top range sort out good sellers through commitment mechanism. However, Subsection 3.3 will show that in the equilibria with the highest efficiency, there are no other commitment names. Therefore, all commitment names sort out good sellers in these equilibria.

Any dynamics $\{p_h\}_{h\in H}$ that satisfies $(a)$, $(b)$, $(c)$ and $(d)$ above can be embedded into an equilibrium[10]. Hereinafter, we only consider the dynamics without fully spelling equilibria, unless necessary. The problem of finding the socially best equilibria is, then, to construct an equilibrium dynamics that bears the largest $R_G$. That is,

**Problem 1** $\max_{\{p_h\}_{h\in H}} R_G$, *s.t. (a), (b), (c), and (d).*

To obtain a lower bound of the maximum value, notice that any equilibrium of the basic model that involves only non-commitment names is also an equilibrium in the complete model, since non-commitment names satisfy the same constraints in both models. In particular, TSE-1, which implements $R_G = rq\pi$, is an equilibrium in the complete model. Therefore, the maximum is no smaller than $rq\pi$, which Proposition 2 states is the highest efficiency implemented in the basic model. We are looking for equilibria that implement $R_G > rq\pi$, that is, that strictly improve over the basic model. These equilibria are called "*Norm Equilibria*", since it is the norm that makes them possible.

In Norm Equilibria, $P \geq \frac{\overline{w}}{r}$; otherwise, there are no commitment names and the norm has no bite; moreover, the value of non-commitment names strictly increases with success.[11] That is,

$$p_{hs} > p_h \tag{7}$$

We are going to prove that Norm Equilibria exist if and only if the post production signal is informative enough. For that purpose, we establish an inequality that relates $P$ to $R_G$.

**Lemma 5** *In a Norm Equilibrium,* $P \leq \frac{\pi - R_G}{1 - rq - r(1-q)\tau}$.

---

[10] Actually this equilibrium is almost unique. Given $\{p_h\}_{h\in H}$, for non-commitment names, $w_h = p_h - rp_{hf} + c$ if $p_h - rp_{hf} < \pi$, and $qv$ otherwise, and $\lambda_h = \frac{w_h}{qv}$. For commitment names, $w_h = 0$ or $\overline{w}$; $\lambda_h = 1$ if $p_{hf} < \frac{p_h + c}{r}$, but undecided if $p_{hf} = \frac{p_h+c}{r}$. $e_B = \frac{1}{1-\gamma}\sum_h \rho_h(1-\lambda_h)$, and $e_G = \frac{1}{\gamma}\sum_h \rho_h \lambda_h$, where $\rho_h$, the steady state mass of $h$-names, is decided by the dynamics.

[11] By (5), $p_{hs} = p_{hf} + \frac{R_G}{rq} + \frac{\max\{p_h - \pi - rp_{hf}, 0\}}{rq}$. In Norm Equilibria, $\frac{R_G}{rq} > \pi$ by definition. Thus, $p_{hs} \geq \frac{R_G}{rq} > p_h$ if $p_h \leq \pi$. For $p_h > \pi$, notice that $p_{hf} + \frac{\max\{p_h - \pi - rp_{hf}, 0\}}{rq}$ is minimized at $p_{hf} = \frac{p_h - \pi}{r}$. Thus $p_{hs} > p_h - \pi + \frac{R_G}{rq} > p_h$.

**Proof.** See Appendix C. ■

For an intuition, consider the case where $P = p_h$ for some names and $r = 1$. These names are commitment names; otherwise, $p_{hs} > P$ by (7), a contradiction. Good sellers buying the names obtain the full surplus they create, $\pi$, since the names sort out good types through commitment mechanism. On the other hand, since the names are of the top value, they obtain no capital gain in any case; but they lose the names' value with probability $(1-q)(1-\tau)$, when they receive signal "n" for the useless widgets and unintentionally overprice them. Therefore, the return $R_G \leq \pi - (1-q)(1-\tau)P$. The intuition also helps us find the socially best equilibria.

The lemma paves the way to show the necessary and sufficient condition for Norm Equilibria to exist.

**Proposition 3** *Norm Equilibria exist if and only if* $\tau > \tau_c \equiv \frac{(qv - r\pi)(1-rq)}{r(1-q)[qv-(1-rq)\pi]}$. *Therefore, the post-production information is necessary for the norm to make a difference.*

**Proof.** The necessary part is proved here. By definition, $rq\pi < R_G$ in Norm Equilibria. Then $P < \frac{\pi - rq\pi}{1-rq-r(1-q)\tau}$ by Lemma 5. On the other hand, we saw $P \geq \frac{\overline{w}}{r}$. Therefore, Norm Equilibria exist only if $\frac{\overline{w}}{r} < \frac{\pi - rq\pi}{1-rq-r(1-q)\tau} \Leftrightarrow \tau > \tau_c$. The proof of the sufficiency is relegated in Appendix B. ■

Intuitively, the norm is introduced only to utilize the post-production information. Imposing the norm, however, incurs costs. With probability $(1-q)(1-\tau)$, the top range commitment names are destroyed, which is a social cost, because the good sellers who own the names unintentionally overprice the useless widgets. This cost is proportional to $1-\tau$, while the amount of the post-production information is measured by $\tau$. Therefore, imposing the norm brings about a net gain only if $\tau$ is beyond a threshold, $\tau_c$. $\tau_c < 1$ by Assumption 3.

By Lemma 5, $R_G < \pi$. Therefore, the first best where $R_G = \pi$ is not implementable in the complete model either. Subsequently, we proceed to examine the second best equilibria for the case where $\tau > \tau_c$, as by Proposition 3, if $\tau \leq \tau_c$, the second best is the same as in the basic model. The ultimate objective is to show that with $\pi$ decreasing, the dynamics that implement the second best become more and more complex, according to some measurement of complexity presented in the next subsection.

Hereinafter, to ease notations, we let $\beta \equiv (1-q)(1-\tau)$, the probability of good sellers unintentionally overpricing the widgets, and let $r = 1$[12].

### I.2.3 The Simplest Second Best Dynamics

After presenting the measurement of complexity, the subsection proceeds to reformulate Problem 1, which leads to a refinement of equilibria. The subsection then spells out two cases of simplest dynamics.

The complexity of a dynamics is measured by its length, denoted by $l$, which is defined as follows. If the minimum upper bound $P$ is never reached by any $h$-names, which means name values never stop growing, then define $l = \infty$. If some $h$-names take the top value $P$, then the length of the dynamics is defined as the smallest number of periods over which new names can reach the top position.[13] That is,

**Definition 2** *The **length** of a dynamic is $l = \min\{n|\ p_h = P \text{ for some } h \in H^n\}$, if the set is not empty; otherwise $l = \infty$. Moreover, if $l < \infty$, $h \in H^l$ such that $p_h = P$ are called the first top names.*

The length of the TSE of the basic model is 1. The improvement in efficiency comes with increase in complexity, as follows.

[12] More precisely, it is the case of $r$ close to but still less than 1. As everything below is continuous in $r$, we only consider what happens at $r = 1$.

[13] Norm Equilibria involve new names. As shown in the proof of Lemma 5, the top range names are commitment names and are bought by good sellers only. They are destroyed into new names with probability $\beta$.

**Lemma 6** *For the dynamics of any Norm Equilibrium, $l \geq 2$.*

**Proof.** It suffices to prove that $p_h < \overline{w}$ for any $h \in \{\phi, s, f\}$, since $\overline{w} \leq P$. $p_\phi = 0 < \overline{w}$. Apply (3) to $h = \phi$, $p_f \leq c < qv < \overline{w}$. Apply (5) to $\phi$ and rearrange, $p_s = p_f + \frac{R_G}{q} \leq c + \frac{R_G}{q}$. By Lemma 5, $R_G \leq \pi - \beta P \leq \pi - \beta\overline{w} = \pi + q\overline{w} - (\beta + q)\overline{w} = \pi + q\overline{w} - qv = q\overline{w} - c < q(\overline{w} - c)$. Therefore, $p_s \leq c + \frac{R_G}{q} < \overline{w}$. ■

The clue on how to find the second best dynamics is hinted in the intuition of Lemma 5. Start with examining the top names. Suppose $h$-names take the top value $P \geq \overline{w}$. We saw that these names are commitment names and that they sort out good sellers through the commitment mechanism (Lemma 4). Apply (6) to the names, $R_G = -P + \pi + qp_{hs} + (1-q)\tau p_{hf}$. To maximize $R_G$, we want to maximize $p_{hs}$ and $p_{hf}$. Therefore, $p_{hs} = p_{hf} = P$. That is, in the second best, the top names' value stops varying with performance: $p_{hs} = p_{hf} = P$, so long as the norm of setting honest prices is followed; it is destroyed only by unintended dishonesty (occurring with probability $\beta$).[14] Substitute $p_h = p_{hs} = p_{hf} = P$ into (6), then, in the second best,

$$R_G = -\beta P + \pi \tag{8}$$

Since $R_G$ is inversely related to $P$, Problem 1 becomes

**Problem 2** $\min_{\{p_h\}_{h \in H}} P$, *s.t. (a), (b), (c), and $P \geq \overline{w}$.*

The constraint that $P \geq \overline{w}$ ensures that there are commitment names. To solve the problem, we want the values of commitment names to be as small as possible, subject to the existence of such names. Therefore, the only commitment names are the top names in the second best equilibria; otherwise the top level commitment names could be cut off, which leaves some commitment names and lowers $P$. It follows that all commitment names sort out good sellers in the second best.

[14] It follows that if $\tau = 1$ and hence $\beta = 0$, the state of the top names will become an absorber of the dynamics. Then, there are no stationary Markovian Equilibria. That is why $\tau < 1$ is needed.

Ideally, $P = \overline{w}$. This could be implemented by some "strange" dynamics in which failure bears opposite meaning to names. To eliminate the equilibria with such dynamics, hereinafter, the following refinement is imposed.

**Refinement:** $p_{hf} \leq p_h$, for any $h$.

The refinement ensures that failure damages names always. An example of the "strange" equilibria is as follows. Suppose $c = \frac{\overline{w}}{2}$. Only bad sellers own new names. The names become $f$-names one period later with $p_f = c$ by (4). These $f$-names are again owned by bad sellers only and become $f^2$-names after one period with $p_{f^2} = 2c = \overline{w}$. These $f^2$-names are the first top names and owned by good sellers only. So $P = \overline{w}$. In this dynamics, only bad sellers own $f$-names, but only good sellers own $f^2$-names. That is, one failure stands for being totally bad, but two consecutive failures stand for being perfectly good. This U-turn poses a great difficulty to buyers in adjusting their belief as to the meaning of failure. This dynamics is eliminated by the refinement.

By the refinement, $p_{hf} \leq p_h < p_h + c$. Therefore, (3) is automatically satisfied, and (4) does not hold, which means that no non-commitment names are only bought by bad sellers and that no commitment names are ever bought by bad sellers. Constraint $(a)$ becomes $p_\phi = 0$. Constraint $(b)$, for non-commitment names, become

$$p_{hs} = p_{hf} + p_s + \frac{\max\{p_h - \pi - p_{hf}, 0\}}{q} \tag{9}$$

Here we apply $R_G = qp_s$, derived by substituting $p_f = 0$ into (5) for $h = \phi$. Substitute $R_G = qp_s$ into (8), and constraint $(c)$, for commitment names and thus the top names, become

$$qp_s = -\beta P + \pi \tag{10}$$

Incorporate the refinement and the above simplification of the constraints into Problem 2, and the problem becomes

**Problem 3** $\min_{\{p_h\}_{h\in H}} P$, *s.t.* $p_\phi = 0$; *(9) if* $p_h < P$; *(10);* $p_{hf} \leq p_h$; *and* $P \geq \overline{w}$.

By studying Problem 3, we proceed to construct two examples of the simplest second best dynamics in particular and then to prove the comparative statics on the second best dynamics in general.

By Lemma 6, the simplest Norm Equilibrium dynamics are of $l = 2$. Therefore, dynamics that are of $l = 2$ and implement the second best are among the simplest second best dynamics. Two cases of them are given in the following.

**Lemma 7** *If* $\pi \geq \frac{q+2\beta}{2-q}\overline{w}$, *for Problem 3, the constraint* $P \geq \overline{w}$ *is not binding and the solution dynamics are of* $l = 2$, *illustrated by figures 2 when* $\tau \leq \frac{2-q}{3-q}$ *or by figure 3 otherwise below. In the dynamics,* $p_{hs} \geq p_h \geq p_{hf}$ *for any* $h$ *and* $p_{hf} = 0$ *for* $p_h \leq \pi$. *Good sellers are sorted out by* $S$*-names through value-adding mechanism and* $S^2$*-names through commitment mechanism. Bad sellers enter only through new names and* $SF$*-names.*

**Proof.** See Appendix C. ■

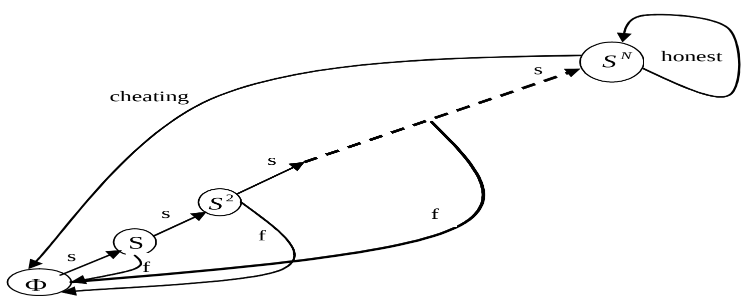


Figure 2: the second Best Dynamics if $\tau \leq \frac{2-q}{3-q}$

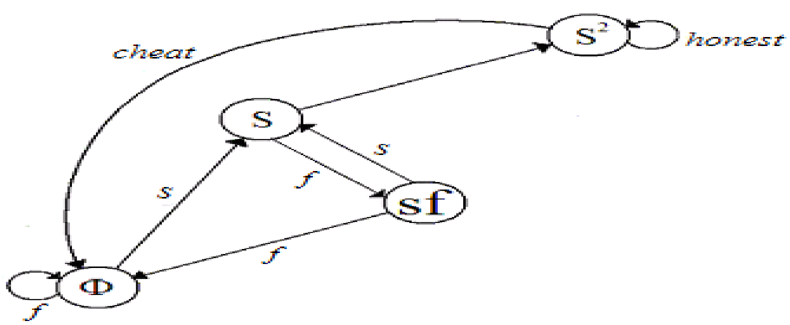

Figure 3: the second Best Dynamics if $\tau > \frac{2-q}{3-q}$

Both dynamics involve four states, $\Phi$, $SF$, $S$, and $S^2$, and $0 = p_\Phi < p_{SF} < p_S < p_{S^2} = P$. $S^2$-names are the only commitment names, they remain $S^2$-names after a success or a failure, so long as they follow the norm of setting honest prices; otherwise, they are destroyed into new names ($\Phi$-names). Names of other three states are non-commitment names. After a success, $\Phi$-names become $S$-names and $S$-names becomes $S^2$-names. After a failure, $\Phi$-names remain $\Phi$-names and $S$-names become $SF$-names. The evolution of $SF$-names gives the difference between the two dynamics. When $\tau \leq \frac{2-q}{3-q}$, they become $S$-names after a success and $\Phi$-names after a failure. When $\tau > \frac{2-q}{3-q}$, $SF$-names become $S^2$-names after a success and remain $SF$-names after a failure.

Lemma 7 summarizes what happens when $\pi$ is high enough. The next subsection shows that with $\pi$ smaller and smaller, the second best dynamics become longer and longer.

### I.2.4 Comparative Statics of $l$ with respect to $\pi$

To prove the comparative statics result, consider the condition under which some dynamics of $l = N$ could be a solution of Problem 3. Since $p_{hs} > p_h \geq p_{hf}$, the first top names are $s^N$-names if $l = N$. Consider then the maximum value of $p_{s^N}$ among all the dynamics of $l = N$. If this value is smaller than $\overline{w}$, then no dynamics of length $N$ could be a solution of Problem 3, because of no way to satisfy the constraint $P \geq \overline{w}$.

$p_{s^N}$ is maximized if and only if $p_{s^{n+1}} - p_{s^n}$ is maximized for each $n = 1, 2...N-1$. For these $n$, $s^n$-names are non-commitment names as they are not in $H^N$ and $l = N$. Apply (9) to these names, $p_{s^{n+1}} = p_{s^n f} + p_s + \frac{\max\{p_{s^n} - \pi - p_{s^n f}, 0\}}{q}$. Given $p_{s^n}$, $p_{s^{n+1}}$ is maximized by making $p_{s^n f}$ equal 0 or $p_{s^n}$ (by the refinement, $p_{s^n f} \leq p_{s^n}$). Therefore, to maximize $p_{s^N}$,

$$p_{s^{n+1}} = \max(p_{s^n}, \frac{p_{s^n} - \pi}{q}) + p_s, \text{ for } n = 1, 2...N-1 \tag{11}$$

And as before, since $p_{s^N} = P$, by (10),

$$-\beta p_{s^N} + \pi = q p_s \tag{12}$$

Out of these $N$ equations, we solve $p_s, p_{s^2}, \ldots$, and particularly, $p_{s^N} = P$. This $P$ depends on $\pi$ and $N$, and we denote it by $P(\pi, N)$. $P(\pi, N)$ gives the maximum value of $P$ among all the dynamics of $l = N$. Only if $P(\pi, N) \geq \overline{w}$, there are dynamics of $l = N$ that could be solutions of Problem 3. $P(\pi, N) = \overline{w}$ defines an implicit function $N(\pi)$. On the two functions, we have the following lemma:

**Lemma 8** $\frac{\partial P}{\partial N} > 0$ *and* $N'(\pi) < 0$.

**Proof.** See Appendix C. ■

By the lemma above, for any given $\pi$, $P(\pi, N) \geq \overline{w}$ if and only if $N \geq N(\pi)$. That is, $N(\pi)$ is the minimum length of the dynamics that can be solutions of Problem 3 at the given $\pi$. Moreover, the lemma says that the minimum length decreases with $\pi$. In combination, the lemma leads to the following Proposition.

**Proposition 4** *The smaller is $\pi$, the longer is the second best dynamics, and hence the greater is the number of successes new names have to accumulate to accomplish the top reputation.*

Intuitively, the value of a name increases with successes at a speed proportional to $\pi$: the increment in the value gives good sellers capital gain, which is proportional to the overall return; the return in itself is proportional to $\pi$. On the other hand, the threshold of the name value over which names accomplish the top reputation is fixed at $p = \overline{w}$. Therefore, the smaller $\pi$ is, the more successes new names have to accumulate to cross the threshold, and the longer the dynamics are. $\pi = qv - c$ is the surplus generated by a good seller. Empirically, the surplus could be proxied by the profit margin and good sellers proxied by high-end firms of an industry.

Then, the proposition lays down an inverse relationship between the average profit margin of high-end firms of an industry and the time span for new firms to fully establish reputation in this industry. If we believe that the high-end firms in the software industry averagely earn a higher profit margin than those in the wine industry, the comparative statics are consistent with the observation that it took a decade for Microsoft to build up its reputation, while it took a century for a wine firm to achieve some commensurate fame.

## I.3 Conclusion of Chapter One

This chapter presents an OLG model where names stand for nothing intrinsic. Names can still bear reputation, in the sense that past glories of a name stand for the quality of the product currently provided under the name, in the following way. If that is believed, all sellers want names with glorious past; this belief is rationalized if good types outbid bad types in the competition for these names. This happens by two mechanisms. One is the value-adding mechanism: good types are more capable of adding value to the names than the bad types, because they are more likely to succeed in producing high quality products. The other is the commitment mechanism. Highly reputable names are subject to the norm of pricing the products honestly. If the price is found beyond the true value, the names lose all the reputation. On the other hand, even when these names fail to produce high quality products, their reputation is not damaged by the failure if they honestly set the low price for the products. Buying these names are equivalent to commit to pricing the products honestly. Only good sellers are willing to make the commitment and thus sorted out by these names. It is a surprise that the norm of setting honest prices makes a difference in the context of purely adverse selection of the chapter.

the chapter spells out some cases of the dynamics of organizational reputation in the equilibria with the highest efficiency. The dynamics are Markovian transformation over several states, each defined by the value of the names of the state. The names of the top state sort out good sellers through the commit mechanism. They keep the reputation untouched, if and only if they follow the norm. The names of other states signal and sort good types through the value-adding mechanism. Their reputation increases after a success, decreases after a failure, and is totally destroyed by a failure when the reputation is already low enough.

Lastly, the chapters finds that in the equilibria with the highest efficiency, the smaller the surplus generated by the good type agents, the longer the dynamics, that is, the greater the number of successes which new names have to accumulate to accomplish the top reputation. This comparative statics result predicts an inverse relationship between the average profit margin of high-end firms of an industry and the time span over which firms can fully establish reputation in the industry.

# Chapter Two: Control versus Incentive – the Optimal Allocation of Physical Capital Ownership

## II.0 Introduction of the Chapter

Since Coase (1937) initiated the inquiry into the border between the firm and the market, it has motivated vast literature under the name "the theory of the firm". the chapter is motivated by the observation that this literature seldom takes into account control-coordination side but overwhelmingly concentrates on incentive side. However, in a widely-covered study into the industrial history of the US, Chandler (1977) finds that the giant corporations came into being only when transactions were better coordinated within the firm than in the market and that to the end of better coordination, it is necessary to put vast amounts of assets under centralized ownership. the chapter presents a theory on how centralized ownership of physical capital benefits coordination through enhancing control over human capital and on how this benefit is balanced by concomitant cost in incentive loss.

the chapter differentiates control problems from incentive problems. Both refer to situations where a principal wants an agent to make a preferable choice among ex ante uncontracible alternatives. Control problems differ from incentive problems in on-time negotiability. If on the time when the agent is deciding the choice the decision are negotiable (namely contractible) between the principal and the agent, it is a control problem, and if not, an incentive problem. Take an example from Milgrom and Roberts (1992), where a group of players are propelling a rowboat in a match. To have each player put his oar into the water at the same time is a control problem; the action is observable, and contractible. But to have him exert high effort to pull his oar in the water is an incentive problem; he could look but not actually be labored. For another example, it is a control problem to ensure G. W. Bush to or not to invade Iraq, but it is an incentive problem to ensure him spend more time considering serious stuff rather than having fun, as he claimed that he was walking even in his Texas farm.

The model of the chapter consists of a principal (she), an agent (him) and a physical capital. With the capital, the agent can carry out one of the two projects, on leading to the product of general interest and sold directly to the market, the other leading to the product specific to the principal's need but useless to others. The full details of either project are not clear before a particular date, and when the day arrives, she and he bargain on which project to be done; the two parties have equal bargaining power, in the sense that each has one half chance to make a take-it-or-leave-it offer (tioli) to the other. The project to be done is thus contractible at the date when it is decided. Before the date, the agent chooses the level of human capital investment, preparing for the future project. After the date, he chooses the level of effort doing the chosen project. Both choices are privately decided by him, not subject to bargaining. The value of both products increases with either incentive variable, but given the two incentive variables, the specific product is worth more than the general one, and the excess is the benefit of coordination. So the choices of human capital investment and effort are incentive problems; to ensure the specific project to be chosen is a control problem.

The last ingredient of the model is the friction of bargaining, without which the project choice follows ex post efficiency, independent of who owns the capital. In the chapter, the friction is due to information asymmetry at the contingencies. Under each contingency, the value of the general project is a fixed fraction of the specific project's value, across all levels of the incentive variables. At the date when the two parties negotiate on the project to be done, the agent knows the value of the fraction and the principal only knows its distribution.

The basic message of the chapter is that having the principal own the capital advantages control but disadvantages incentive. First, there is the trade-off between control and incentive. The agent obtains the full value of the general product, which is sold to the competitive market. On the contrary, the agent only reap half of the value of the specific product, as in Grossman and Hart (1986), because of the hold up problem. Therefore, the better the control, the higher the chance the specific project is chose, the worse the incentive. Since the specific project entails incentive loss, it is not always the second best, though given the levels of the incentive variables, it is worth more than the general project.

Then, move on to consider the effect of the allocation of ownership of the capital on the project choice, which is decided through bargaining between the principal and the agent. When the agent offers a tioli to the principal, which happens with probability one half, there is no efficiency loss, since the principal has no private information. The difference presents itself when she offers tioli.

First examine the arrangement where the principal owns the capital. The default choice is the specific project. If the agent nevertheless wants to go for the general project, he has to bargain with the principal, on the price to buy her assent. On the default option, she obtains half of the general project's value. Therefore, it is dominated for her to ask for any price no bigger than that half value, when she offers tioli. The agent accepts her asked price, so the general project is chosen, only if the value of the general project is no less than the sum of the price plus his default option value, half of the specific project's value. We saw the price is beyond the half of the specific project's value by a positive difference. Therefore, the general project is chosen only if its value is beyond the specific's by more than this difference. It follows that under those contingencies where the general project is worth more than the specific but by a smaller remainder than the difference, the general project is socially efficient but not chosen. That is, having the principal own the capital induces too much control.

Then examine the arrangement where the agent owns the capital. Now the default choice is the general project. If the principal wants him to work for her, to do the specific project, she has to bargain with him, on the price to buy his assent. On success, she will get half of the specific project's value. When offering tioli, therefore, she offers a price below the half value by a positive difference. If accepting the offer, the agent obtains the price plus half of the specific project's value, which altogether equal the specific project's value minus the positive difference.

The offer is accepted, so the specific project is chosen, therefore, only if the general project's value is below the specific project's by the difference. It follows that under those contingencies where the specific project is worth more than the general but by a smaller remainder than the difference, it is socially efficient but not chosen. That is, having the agent own the capital induces too little control.

Overall, the trade-off between too much control versus too little control decides who owns the capital in equilibrium. the chapter shows that the principal owns the capital if and only if the benefit of coordination is bigger enough.

The literature on the theory of the firm touches the control side only lightly. Given the volume of the literature, the chapter apologizes for not providing a complete survey but mainly relying on Gibbons (2004), who classifies the literature into four categories which are addressed below in order. The property rights theory (Grossman and Hart (1986) and Hart and Moore (1990) (GHM hereinafter)) is concerned with suboptimal provision of some investments. The levels of the investments, though observable after having sunk, are not decided by bargaining when the investments are being laid down; otherwise, the hold-up problems evaporate. This category of the literature, therefore, address only the incentive side, not the control side, even though Hart & Moore (1990) uses word "control". The incentive theory (Holmstrom & Milgrom (1991, 1994)) is concerned with the balance of the effort between multiple tasks. Similarly, the quasi-rent seeking theory (Baker & Hubbard (2000) etc.) is concerned with the effects of ownership of physical capital on the balance of the effort between rent seeking and the assignment for the principal. In both categories, how to distribute the effort is decided by the agent privately, unobservable to the principal, not by bargaining between the principal and the agent. The problem concerned is thus an incentive problem only. Actually, all the three categories discussed above are driven by incentive balance, either between tasks (the incentive theory and the rent-seeking theory) or between players (GHM).

The fourth and the last category of the theory of the firm, relational adaptation theory (Simon (1951) and Williamson (1971, 1975, 1991) etc.), pays attention to control problems. However, it does not have a formal model examining the trade-off between control and incentive; although Williamson (1975, 1991) argues, only informally, for the trade-off between adaptation and incentive, where adaptation, if decided through bargaining, could be on the control side, he does not microfound the trade-off on the information structure, as is done by this chapter to differentiate control from incentive. Moreover, that literature does not consider the allocation of ownership of physical capital, but is concerned with the comparison between authority or hierarchy and market.

Beyond the literature surveyed by Gibbons (2004), some papers also consider control problems. Hart and Holmstrom (2002) shares with this chapter the point that integration brings about too much coordination (control) while non-integration brings about too little. The two papers differ in the cost of integration; it is incentive loss of the agent in this chapter, but the loss of "the private benefits of managers and workers" in Hart and Holmstrom (2002). Rajan and Zingales (2002) considers the control problem of keeping the employees to work for the firm rather than to steal the critical resources. However, in that paper the trade-off is between growth of the firm and the risk of being expropriated.

The rest of the chapter is organized as follows. We first presents the model, then analyze all the possible contractual arrangements (called regime) one by one, then they are compared to find the equilibrium arrangement, and lastly we conclude and presents some empirical evidence that supports the theory of the chapter.

## II.1 The Model

The model consists of a principal (denoted by P), an agent (A) and a capital (K). K is indispensable for A to create value; on the other hand, A's human capital is indispensable for K to be utilized. The core question is how to use ownership of the physical capital to control A's human capital, the meaning of control being given as follows. Using K, A could engage into two exclusive projects. One is done to coordinate with P's integrated strategy and leads to a product that is specific to P's need and worth little to the market. The other project is independent of P's strategy and leads to a product that is of general interest and is to be sold to the market. The value of either product depends on the human capital investment A makes before the project is chosen and the effort level he exerts after the project choice. The specific project is denoted by "**cd**" and the general one by "**in**". Both players are risk neutral.

**Timing:**

There are five dates. At date 0, P and A decide the allocation of ownership and payoff right of K. Here ownership means, as GHM, residual control rights, namely that the owner walk away with K putting it in the alternative use when the bargaining fail to reach an agreement. And payoff rights mean the ownership of the final product. At date 1, A makes the human capital investment. The investment is specific to the capital, namely, if it helps nothing if A does not operate with K. At date 2, the state is realized, and P and A bargain on the project choice. At date 3, A chooses the effort level to do the project chosen at date 2. At date 4, the product of the project is yielded, and is traded if P has no payoff rights. An arrangement of the ownership and payoff rights of K is called a "*regime*". The timing is illustrated in figure 1 below.

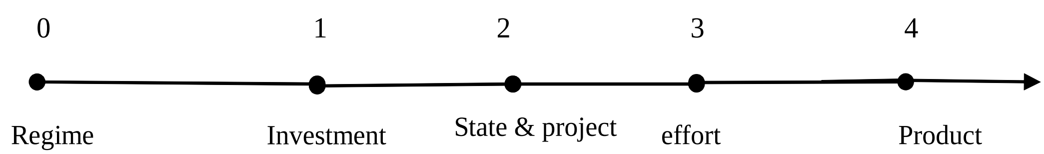


Figure 4: Timing Tree

The values of the two projects are as follows. If A invests $i \in [0, \infty)$ at date 1 and exerts effort level $e \in [0, \infty)$ at date 3, and the realized state is $s \in [0, 1]$, then the value of the specific product is $v_{cd}(i, e)$ and the value of the general product is $v_{in}(i, e; s) = sv_{cd}(i, e)$. As $s \in [0, 1]$, given the investment and effort level, the specific project is always worth more and the excess is the benefit of coordination. To capture the benefit, let $v_{cd}(i, e) = v(i, e) + B$, where $v(I, 0) = v(0, e) = 0$ and $B \geq 0$; thus $B$ has no incentive effect upon $i$ or $e$ and captures only the benefit of coordination. Denote by $c_i(i)$ the cost of the investment and by $c_e(e)$ the disutility of the effort. Assume, as usual, that the value functions are strictly increasing and concave and that the cost functions are strictly increasing and convex.

**Information:**

The project to be done is not contractible before date 2 and is contractible at the date (so that P and A bargain on it). The investment level, $i$, is not contractible and made privately by A at date 1, but observable at date 2; and the value of the product is never contractible but observable at date 4. Both assumptions are standard in the literature of incomplete contracting. The effort level $e$, is never observable to P, as in a typical moral hazard problem. Therefore, no contract conditional on the value of the final product is feasible to induce A to choose some levels of $i$ or $e$. He is incentivized only by obtaining the payoff rights. Including both the ex ante investment and the ex post effort seems redundant, and any one of them suffices to show the trade-off between incentive and control. Nevertheless, the purpose of introducing the effort is to show that the theory of the chapter does not rely on specific investment, as GHM does and the purpose of introducing the asset-specific investment ($i$) is to show that the theory is rick enough to incorporate GHM.

Assume $s$ uniformly distributes on $[0, 1]$ before date 2. Different from the literature, the realized state, $s$, is the private information of A at date 2. P may deduce $s$ from the observed value of the general product (if it is yielded) at date 4. Information asymmetry is a way to capture bargaining costs. Nevertheless, that is the only place digressing the standard set-ups.

**Assumption (Incomplete Contracting):** at date 0, P and A can do nothing but decide the allocation of the ownership and payoff rights of K.

To be sure, P and A could learn a lot from the implementation theory to design some clever mechanisms on how to choose the project at date 3 and how to trade the product at date 4. Thus the assumption is either justified by bounded rationality of both players (they may know nothing about game theory, not even say Maskin-Moore Theorem), or insisting on perfect rationality, by some way of Hart and Moore (1999).[15]

Anyway, by the assumption, the project choice is decided via bargaining between P and A at date 2, and if A has the payoff rights, the price of the specific product is decided via bargaining at date 4. Assume that both P and A has equal bargaining power, that is, each party has chance $\frac{1}{2}$ to make a take-it-or-leave-it (tioli) offer to the other.

So the project choice is negotiable when it is being decided, while the choices of the investment and the effort levels are not. Therefore, the former is on the control side and the latter on the incentive side. We call it "*loss of control*" (for P) if the specific project is not chosen at date 2. The specific project entails incentive loss since half of its product's value is appropriated by P. Therefore, it is not always socially efficient, even though given the levels of the incentive variables, it is worth more than the general project. We call it "*too much coordination*", if the specific project is chosen even when it is not efficient, and "*too little coordination*", if it is not chosen when it is efficient.

[15] I ever consider such a set-up. Besides the two relevant projects, each party could think out infinitely possible inefficient projects to abuse the other party, like extremely low cost but low value projects and extremely high cost but high value ones. And as Hart and Moore (1999) does, suppose renegotiation cannot be excluded. Then probably, as in that paper, ex ante null mechanism is the best mechanism. I cannot prove that strictly, but I can show that any mechanisms allocating decision rights (like P decide, or A decide, or P specifiy an extent within which A chooses etc.) do nothing better.

At date 0 P and A choose from the four following alternative regimes. The equilibrium regime will be the one that maximizes the total surplus because side payment is feasible.

**Regime 1**: A has the ownership and the payoff rights of K.

**Regime 2**: P owns K and A has the payoff rights of K.

**Regime 3**: A owns K and P has the payoff rights of K.

**Regime 4**: P has the both.

In regime 1, A is an independent contractor. In regime 2 A is a division of an M-form organization; the ownership of non-human capitals (K and P's other capitals) is centralized in the hands of P, but A has an independent account and owns what he produces. Regime 3 is actually an exclusive dealing arrangement, as A can only supply P. In regime 4, A is a salaried employee of P in the ordinary sense. It will be shown that regimes 3 and 4 are dominated by regime 2. Thus what matters is the allocation of ownership of K. Following Grossman and Hart (1986), regime 2 is called "Integration" and regime 1 "non-Integration". And to justify regimes 3 and 4 we could introduce the value of K in a manner of Holmstrom and Milgrom (1991), which is not pursued here.

The next section will solve the outcome for each of the regimes.

## II.2 The Four Regimes

We apply backward induction to solve the outcome for each regime.

### II.2.1 Independent Contractor (Regime 1)

In regime 1, A is an independent contractor of P and has both the payoff rights and the ownership of A. Then at date 4, A owns the product of the chosen project. If it is the general project, then he obtains the general product and sells it to the outside market at price $v_{in}$. If the project is "cd", he obtains the special product that only P demands. Then the two bargain over the price of the specific good. Given the bargaining power distribution is 05-0.5, the price is $\frac{1}{2}v_{cd}$.

At date 3, the effort level is chosen to maximize $v_{in}(i,e;s) - c_e(e)$ with the general project, or to maximize $\frac{1}{2}v_{cd}(i,e) - c_e(e)$ with the specific project. Note that $v_{in}(i,e;s) = sv_{cd}(i,e)$, so the two problems can be unified. Let $V(i,s)$ be the value of the problem $\max_e sv_{cd}(i,e) - c_e(e)$, and $e(i,s)$ be the maximizer. When we are discussing what happens after date 1, argument $i$ is neglected for simplicity. Then A chooses effort level of $e(s)$ with the general project and that of $e(0.5)$ with the specific project. And $V(s)$ is the social value of the general project in $s$ at date 2 and that of the specific project is $V(\frac{1}{2}) + \frac{1}{2}v_{cd}(e(0.5))$. Let $\widetilde{s}$ be the solution of $V(\frac{1}{2}) + \frac{1}{2}\widehat{v}_{cd} = V(s)$. Then $\widetilde{s} > \frac{1}{2}$. Then, the efficient project is the specific one if $s < \widetilde{s}$ and is the general one if $s > \widetilde{s}$.

At date 2, P and A bargain on the project to be done, as follows. Since under the regime A owns the capital and the final product, he can go straightforwardly for the general project, if he wants. Or he chooses to bargain with P on the price which she pays for him to do the specific project. If he chooses so, the nature decides who has the chance to make a take-it-or-leave-it (tioli) offer to the other. With probability 0.5, A offers tioli to P; if she takes it, the specific project is chosen; if she leaves it, A comes back to the general project. With probability 0.5, P offers tioli to A; if A takes it, the specific project is chosen; if A leaves it, he comes back to the general project. Thus game tree is the following:

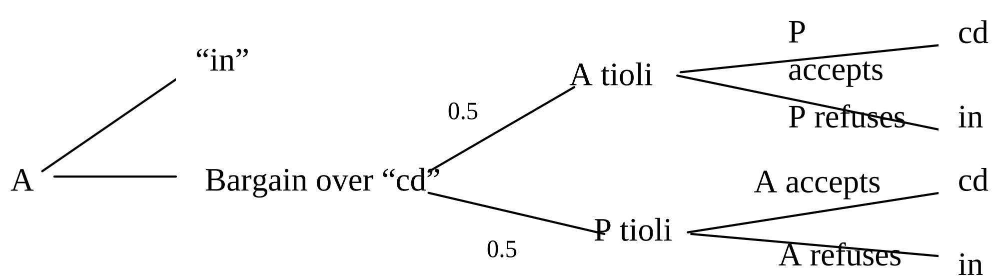


Figure 5: game tree of bargaining on the project choice under regime 1; "in" represents the general project and "cd" the specific

Notice that in the game above at stage 1 strategy "in" is weakly dominated by "Bargain over cd" for any realized state $s$, because choosing to bargain A can still pick "in", if he wishes, simply by rejecting P's offer or tendering P with an unacceptable offer. Thus, at stage 1, A always chooses to bargain and this choice signals nothing of A's private information. Denote by $T$ the price P pays to A. From the specific project, P obtains one half of the specific product's value $\widehat{v}_{cd} \equiv v_{cd}(e(0.5))$. Therefore, she will reject any asked price $T > \frac{1}{2}\widehat{v}_{cd}$, and A will ask $\frac{1}{2}\widehat{v}_{cd}$ when he offers tioli and wants it to be accepted, which then gives him $V(\frac{1}{2}) + \frac{1}{2}\widehat{v}_{cd}$. He obtains $V(s)$ if the tioli is rejected. So his payoff is $\max\{V(\frac{1}{2}) + \frac{1}{2}\widehat{v}_{cd}, V(s)\}$ when getting the chance to offer tioli. The argument also shows that when A offers tiolio, the project choice is efficient: the specific project is chosen iff $s \leq \widetilde{s}$. This is intuitive, as P has no private information.

Consider the case where P offers tioli. A gets $V(s)$ from the general project in state $s$, and $V(\frac{1}{2}) + T$ from accepting P's offer. Then he accepts her offer if and only if $V(\frac{1}{2}) + T \geq V(s)$. The following happens.

**Lemma 9** *P will offer $T = 0$ and $A$ accepts it iff $s \leq 0.5$.*

***Proof.*** *When having the chance to offer tioli, P's problem is* $\max_T \Pr(V(\frac{1}{2}) + T \geq V(s))(\frac{1}{2}\widehat{v}_{cd} - T)$. *Notice that the solution of the problem is internal, and thus satisfies the first order condition (FOC). Do the variable transformation* $T = V(t) - V(\frac{1}{2})$. *Then the problem becomes* $\max_t \Pr(t \geq s)(\frac{1}{2}\widehat{v}_{cd} + V(\frac{1}{2}) - V(t)) = \max_t t(\frac{1}{2}\widehat{v}_{cd} + V(\frac{1}{2}) - V(t))$, *given $s$ is distributed uniformly. By envelop theorem,* $V'(t) = v_{cd}(e(t))$. *The FOC of the problem is*

$$\frac{1}{2}\widehat{v}_{cd} + V(\frac{1}{2}) - V(t) - tv_{cd}(e(t)) = 0.$$

*It is easy to see that $t = \frac{1}{2}$ is a solution as $\widehat{v}_{cd} = v_{cd}(e(0.5))$. And it is the unique solution: $V(t) + tv_{cd}(e(t))$ is an increasing function of $t$ as $e(t)$ is increasing.* ■

Summarize the two case. if $s \leq 0.5$, the specific project is chosen definitely; if $0.5 < s \leq \widetilde{s}$, it is chosen with probability $\frac{1}{2}$; if $\widetilde{s} < s$, it is never chosen. Then

**Corollary 1** *Under regime 1, ex ante there is loss of control with probability* $1 - \widetilde{s} + \frac{\widetilde{s}-0.5}{2}$.

Note that when $0.5 < s \leq \widetilde{s}$, the specific project is not chosen when P offer tioli, even though it is the efficient one. Thus,

**Corollary 2** *Regime 1 induces too little coordination; whenever the specific project is efficient, it is always chosen, while with probability* $\frac{\widetilde{s}-0.5}{2}$, *the specific project is efficient but not chosen.*

Then we move on to the ex ante incentive problem at date 1. At date 2, with probability $\frac{\widetilde{s}+0.5}{2}$ "cd" is chosen and the total surplus is $V(i, 0.5) + \frac{1}{2}v_{cd}(i, e(i, 0.5))$. And if $s > \widetilde{s}$ "in" is definitely chosen and the expected surplus is $E(V(i,s)|s > \widetilde{s})$; if $0.5 < s \leq \widetilde{s}$ "in" is chosen with probability 0.5 and the expected surplus is $E(V(i,s)|0.5 < s < \widetilde{s})$. Thus, in regime 1, the total surplus expected at date 1 is $W^1(i) = \frac{\widetilde{s}+0.5}{2}[V(i, 0.5) + \frac{1}{2}v_{cd}(i, e(i, 0.5))] + (1 - \widetilde{s})E(V(i,s)|s > \widetilde{s}) + \frac{\widetilde{s}-0.5}{2}E(V(i,s)|0.5 < s < \widetilde{s})$, where $\widetilde{s}$ is also a function of $I$. When A offers tioli, his payoff is $\max\{V(\frac{1}{2}) + \frac{1}{2}\widehat{v}_{cd}, V(s)\}$ and when P offers tioli, his payoff is $\max\{V(\frac{1}{2}), V(s)\}$. Therefore, the expected payoff of A at date 1, if he chooses $i$, is $U^1(i) = \frac{1}{2}E(\max\{V(\frac{1}{2}) + \frac{1}{2}\widehat{v}_{cd}, V(s)\} + \max\{V(\frac{1}{2}), V(s)\})$. At date 1, the investment level $i^1$ solves $\max_i U^1(i) - c_i(i)$. Then, at date 0, the total surplus is $W^1(i^1)$.

### II.2.2 M-form Organization (Regime 2)

In this regime, A is a division of an M-form organization; ownership of physical capital is centralized in the hands of P, but A keeps an independent account, in the form of owning the final product.

At date 4, A owns the product of the chosen project, as under regime 1. So he gets $v_{in}$ from project "in" and $\frac{1}{2}v_{cd}$ from project "cd". And at date 3, the effort level is decided accordingly, as under regime 1.

At date 2, the difference from regime 1 presents itself. Under regime 1, the default choice is "in", when A chooses not to bargain with P or bargaining fails to reach any agreement, because he owns K there and can work away with it to supply the market. Under regime 2, in contrast, P owns K and the default choice is "cd", which is the project benefiting her before any side payment from A; if A wants to do "in", he has to make the side payment to buy P's approval. Thus, the change of ownership of K alters the default project. This difference gives P more control over the project choice.

The bargaining process under regime 2 is similar to that of regime 1, with the difference in the default project. At stage 1, A chooses to do "cd" directly or to bargain over the price which he pays P for her approval of "in". If he chooses the latter, then with probability 0.5, A offers P a tioli; if she takes it, "in" is chosen; if she leaves it, A comes back to project "cd". And with probability 0.5, P offers A a tioli price; if he takes it, "in" is chosen; if leaving it, he comes back to "cd". The game tree is the following:

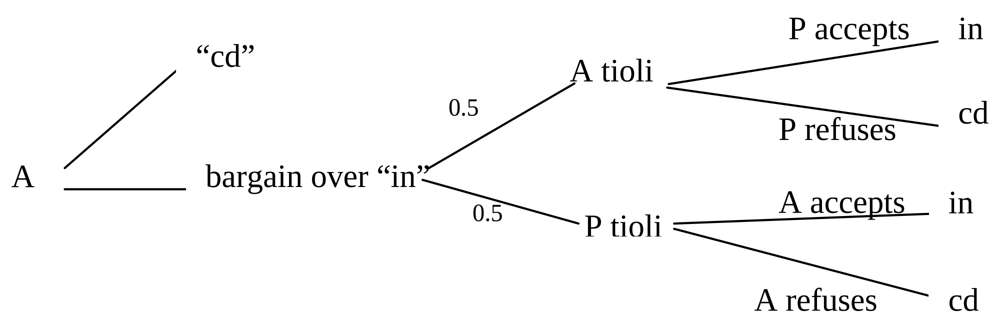


Figure 6: the game tree of bargaining on the project choice under regime 2, where "cd" represents the specific project and "in" the general

Similarly, strategy "cd" is weakly dominated to A in state $s$ at stage 1. Thus A chooses always to bargain and this signals nothing of his private information. If "cd" is chosen, he gets $V(\frac{1}{2})$ and P gets $\frac{1}{2}\widehat{v}_{cd}$. She gets nothing directly from "in". To buy her approval for it, A has to pay $T \geq \frac{1}{2}\widehat{v}_{cd}$ and will actually pay $T = \frac{1}{2}\widehat{v}_{cd}$, if he offers tioli. Thus, A gets $V(s) - \frac{1}{2}\widehat{v}_{cd}$ if he offers tioli and wants to do "in". Remember $V(s) - \frac{1}{2}\widehat{v}_{cd} > V(\frac{1}{2})$ iff $s > \widetilde{s}$. Thus, when A offers tioli, "cd" is chosen iff $s \leq \widetilde{s}$, without efficiency loss, as under regime 1. What happens when P offers tioli is summarized in the lemma below.

**Lemma 10** *P offers price $\widehat{T} > \frac{1}{2}\widehat{v}_{cd}$. There exists some $\widehat{s} < 1$ such that A accepts the offer iff $s \geq \widehat{s}$.*

**Proof.** By the argument above, A accepts price T iff $V(s) - T \geq V(\frac{1}{2})$. If he takes the offer, P gets $T$; if otherwise she gets $\frac{1}{2}\widehat{v}_{cd}$. Thus her problem is to choose $T$ to maximize $f(T) \equiv \Pr(V(s) - T \geq V(\frac{1}{2}))T + [1 - \Pr(V(s) - T \geq V(\frac{1}{2}))]\frac{1}{2}\widehat{v}_{cd}$.

$T < \frac{1}{2}\widehat{v}_{cd}$ is never optimal, since P gets $\frac{1}{2}\widehat{v}_{cd}$ from the default option. $f(\frac{1}{2}\widehat{v}_{cd}) = \frac{1}{2}\widehat{v}_{cd}$. And note that $V(1) = v_{cd}(e(1)) - c_e(e(1)) > v_{cd}(e(\frac{1}{2})) - c_e(e(\frac{1}{2})) = \frac{1}{2}\widehat{v}_{cd} + V(\frac{1}{2})$. Thus if $T$ is little bigger than $\frac{1}{2}\widehat{v}_{cd}$, $\Pr(V(s) - T \geq V(\frac{1}{2})) > 0$, and then $f(T) > \frac{1}{2}\widehat{v}_{cd}$. Therefore, the optimal price $\widehat{T} > \frac{1}{2}\widehat{v}_{cd}$.

On the other hand, P will never let $T$ be so high that $\Pr(V(s) - T \geq V(\frac{1}{2})) = 0$, and then $f(T) = \frac{1}{2}\widehat{v}_{cd}$. That is, $\widehat{T} < V(1) - V(\frac{1}{2})$. Let $\widehat{s}$ be the solution of $V(s) - \widehat{T} = V(\frac{1}{2})$. Then, $\widehat{s} < 1$, and A accepts P's offer if and only if $s \geq \widehat{s}$. ■

Lemma 2 is intuitive. When A offers tioli, the offer is $\frac{1}{2}\widehat{v}_{cd}$, which should be strictly less than the price asked by P when she offers tioli.

$V(\widehat{s}) - \widehat{T} = V(\frac{1}{2})$. Remember, $\widetilde{s}$ is the solution of $V(s) - \frac{1}{2}\widehat{v}_{cd} = V(\frac{1}{2})$. Then $\widetilde{s} < \widehat{s}$ because $\widehat{T} > \frac{1}{2}\widehat{v}_{cd}$. Overall, if $s \leq \widetilde{s}$ "cd" is chosen definitely; if $\widetilde{s} < s \leq \widehat{s}$ it is chosen when P is offering tioli and "in" is chosen when A is offering tioli; and if $\widehat{s} < s$ "in" is chosen definitely.

**Corollary 3** *In regime 2 with probability $1 - \widehat{s} + \frac{\widehat{s} - \widetilde{s}}{2}$ there is loss of control.*

Remember that if $\widetilde{s} < s$ "in" is socially efficient. But in regime 2 if $\widetilde{s} < s \leq \widehat{s}$ it is chosen only with probability $\frac{1}{2}$. Therefore,

**Corollary 4** *Regime 2 induces too much coordination; the specific project is chosen whenever it is efficient, and with probability* $\frac{\widehat{s}-\widetilde{s}}{2}$, *it is chosen even if it is not efficient.*

When $s < \widetilde{s}$ and with probability 0.5 when $\widetilde{s} \leq s < \widehat{s}$, "cd" is chosen without any side payment and hence A's payoff is $V(i, 0.5)$; with probability 0.5 when $\widetilde{s} \leq s < \widehat{s}$, A pays to P $0.5\widehat{v}_{cd}$ to do "in", by which he gains $V(i,s) - 0.5\widehat{v}_{cd}$; and when $\widehat{s} \leq s$, A pays either $0.5\widehat{v}_{cd}$ or $\widehat{T}$ to do "in" and obtains in expectation $V(i,s) - \frac{\widehat{T}+0.5\widehat{v}_{cd}}{2}$. Thus at date 1, if choosing human capital investment $i$, A expects to obtain $U^2(I) = \frac{\widehat{s}+\widetilde{s}}{2}V(I, 0.5) + \frac{\widehat{s}-\widetilde{s}}{2}E(V(I,s) - 0.5\widehat{v}_{cd}|\widehat{s} \geq s > \widetilde{s}) + (1 - \widehat{s})E(V(I,s) - \frac{\widehat{T}+0.5\widehat{v}_{cd}}{2}|s > \widehat{s})$. The investment level, $i^2$, thus solves problem $\max_i U^2(i) - c_i(i)$. The total surplus at date 1 from investment $i$ is $W^2(i) = \frac{\widehat{s}+\widetilde{s}}{2}[V(i, 0.5) + 0.5\widehat{v}_{cd}] + (1 - \widehat{s})E(V(i,s)|s > \widehat{s}) + \frac{\widehat{s}-\widetilde{s}}{2}E(V(i,s)|\widehat{s} \geq s > \widetilde{s})$. At date 0, the equilibrium surplus is $W^2(i^2)$.

### II.2.3 Regimes 3 and 4

Under both regimes, at date 4 since P has payoff rights, A just deliver whatever produced to P. Thus A gets 0, regardless of his effort level and investment level. Thus he chooses the lowest possible effort doing the project, that is, $e = 0$ at date 3.

At date 2, since A always gets 0 at date 4 doing whatever project, he is indifferent with the project to do and just follows P's requirement. She will choose "cd" certainly. Then at date 2 the total surplus is $v_{cd}(0) - c_e(0) < v_{cd}(e(0.5)) - c_e(e(0.5))$.[16] The right hand side is the minimum surplus under the above two regimes (the surplus for $s \leq 0.5$). Therefore, under both regimes 3 and 4 there is no loss of control, but a stark loss of ex post incentive, which makes regimes 3 and 4 dominated by regime 1 or 2 in ex post efficiency.

Regimes 3 and 4 also induces stark ex ante incentive loss, due to the specific assumption that the two incentive variables are complement. Remember $v_{cd}(i, e) = v(i, e) + B$. $B$ has no

[16] $e(s)$ satisfies the FOC $sv'_{cd}(e) = c'_e(e)$. Therefore $v_{cd}(e(s)) - v_{cd}(e(0)) = \int_0^s v'_{cd}e'(t) = \int_0^s \frac{c'_e}{t}e'(t) > \int_0^s c'_e e'(t) = c_e(e(s)) - c_e(e(0))$, any $s$.

incentive effect and $v(i, 0) = 0$ for any $i$. Since A is going to choose $e = 0$ at date 3, he has no incentive to make any human capital investment at date 1, that is, $i = 0$. Summarily we have

**Corollary 5** *Under both regimes 3 and 4 there is no loss of control but a huge loss of incentive, both ex ante and ex post. The two regimes are thus dominated by regime 1 or 2.*

We move on to compare the four regimes to find the equilibrium regime. By this corollary, I only need to compare between regimes 1 and 2. Following GHM, regime 2 is called "integration" and regime 1 "non-integration".

## II.3 The Comparisons

We saw regime 1 induces more control than regime 2. Too much control is not good, because it entails incentive loss: doing the specific project, half of the value is appropriated by P while doing the general project, A reaps the full value, and hence the more likely is the specific project chosen, the higher is the incentive loss. There are two dimensions of incentive, $i$ and $e$. They could interact in a complex way, to avoid which, I will separate the two dimensions: when considering the case of one variable, the level of the other will be fixed. First consider the case of the effort and thus fix the choice of ex ante investment.

### II.3.1: Control versus Ex Post Incentive ($e$)

In this subsubsection, suppose $i = \bar{i}$ (by, for example, assuming $c_i(i) = \infty$ for $i > \bar{i}$; $= 0$ for $i \leq \bar{i}$). As before, for simplicity, it will be suppressed during the discussion of the subsection below. Then, the second best project choice is "cd" if $s \leq \widetilde{s}$ and "in" if $s > \widetilde{s}$. Different from the second best, "cd" is chosen only with probability 0.5 if $0.5 < s < \widetilde{s}$ under regime 1, and is still chosen with the probability 0.5 if $\widetilde{s} < s < \widehat{s}$ under regime 2. Figure 3 below illustrates the probability the specific project is chosen as a function of $s$ in the three cases.

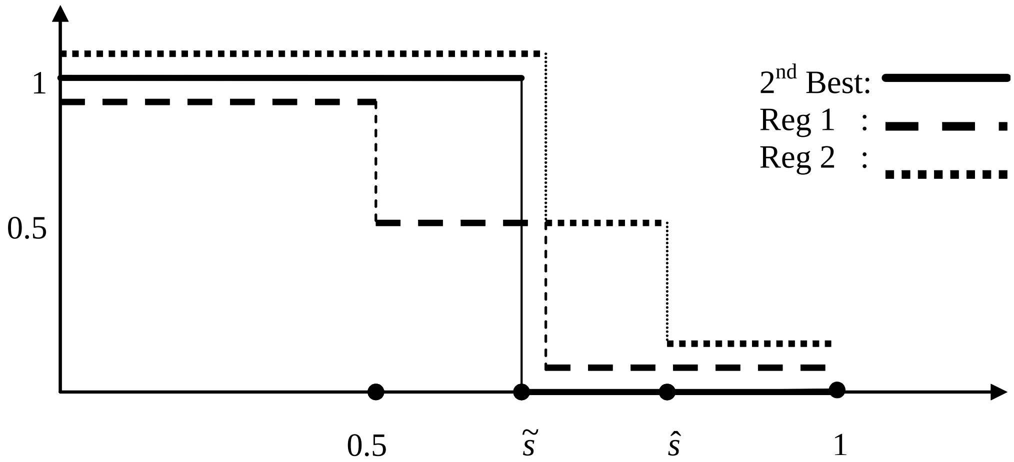


Figure 7: the comparison of the second best, regime 1 and regime 2; the vertical axis is the probability of choosing "cd" and the horizontal one is $s$.

According to the figure, the following proposition is straightforward.

**Proposition 5** *Fix $i = \bar{i}$. Compared with non-integration, integration brings about better control, in the sense that the specific project is chosen at a higher probability, but induces loss in the effort level. Overall, Integration induces too much coordination but Non-integration too little.*

The proposition hints that integration happens iff the coordination benefit is larger enough. Remember $v_{cd}(e) = v(e) + B$ and here $B$ measures the importance of coordination and has no incentive effect. Regime 2 arises if and only if it generates a higher social surplus than regime 1. Then what the proposition above hints is strictly expressed in the following.

**Proposition 6** *If $c_e''' \geq 0$ and $v''' \leq 0$, $\frac{d(W^2 - W^1)}{dB} > \chi$ for some $\chi > 0$. That is, integration arises in equilibrium iff the coordination benefit ($B$) is larger than some critical level.*

**Proof.** The proof is put into the appendix. ■

We finish examining the trade-off between control and ex post incentive. We move on to examine the case of the ex ante incentive.

### II.3.2 Control versus Ex Ante Incentive $(i)$

In this subsection, we assume $e = \widetilde{e}$ always, and for simplicity, we neglect argument $\widetilde{e}$ in all functions and without loss of generality let $c_e(\widetilde{e}) = 0$. Thus, if at date 1 A invests $i$, the social value of "cd" is $v_{cd}(i) \equiv v(i) + B$.

Since there is no ex post incentive problem, $V(i,s) = sv_{cd}(i)$ for all $s$. In this setting, it is straightforward to verify $\widetilde{s} = \widehat{s} = 1$. Because there is no problem of ex post incentive, the specific project is always efficient. As this is the common knowledge, it cannot be negotiated away if it is the default project. Therefore, regime 1 implements the second best project choice. However, under regime 1, where the default project is the general one, there is positive probability "cd" is not chosen when P offers tioli, as she trade-off the probability A accepts the offer with the value she obtains on his acceptance. Therefore, regime 2 strictly dominates regime 1 in ex post efficiency. However, regime 2 induces loss in ex ante incentive.

Under Regime 1, the social surplus is $W^1(i) = \frac{15}{16}v_{cd}(i)$, where the margin $\frac{15}{16}$ comes as follows. Under the regime, with probability $\frac{3}{4}$, "cd" is chosen and so the margin is 1 and with probability $\frac{1}{4}$ (when $s \geq 0.5$ and A offers tioli), "in" is chosen and the margin is $E(s|0.5 \leq s \leq 1) = \frac{3}{4}$; therefore, the overall margin is $\frac{3}{4} + \frac{1}{4} \cdot \frac{3}{4} = \frac{15}{16}$. And A's payoff is $U^1(i) = \frac{9}{16}v_{cd}(i)$: when "cd" is chosen the incentive margin to A is $\frac{1}{2}$, and when "in" is chosen, the margin, as calculated above, is $\frac{3}{4}$; therefore, the overall margin is $\frac{3}{4} \cdot \frac{1}{2} + \frac{1}{4} \cdot \frac{3}{4} = \frac{9}{16}$. Under Regime 2, the social surplus is $W^2(i) = v_{cd}(i)$, as "cd" is chosen with probability 1, and A's payoff is $U^2(i) = \frac{1}{2}v_{cd}(i)$. The optimal investment level under regime 1 is $i^1 = \arg\max \frac{9}{16}v_{cd}(i) - c_i(i)$ and that under regime 2 is $i^2 = \arg\max \frac{1}{2}v_{cd}(i) - c_i(i)$.

**Lemma 11** $i^1 > i^2$.

**Proof.** : $i^k$ satisfies the first order condition $\alpha^k v'(i) = c_i'(i)$, for $k = 1, 2$, where $\alpha^1 = \frac{9}{16}$ and $\alpha^2 = \frac{1}{2} < \alpha^1$. If $i(\alpha)$ is the solution of $\alpha v'(i) = c_i'(i)$, then $\frac{di}{d\alpha} = \frac{v'}{c'' - \alpha v''} > 0$. Then $i^1 = i(\alpha^1) > i(\alpha^2) = i^2$. ■

$B$ does not have any incentive effect and hence $i^1$and $i^2$ are independent of $B$. Then the difference between the ex ante total social surplus is $W^1 - W^2 = W^1(i^1) - W^2(i^2) = \frac{15}{16}(v(i^1) + B) - [v(i^2) + B]$. Thus $\frac{d(W^1 - W^2)}{dB} = -\frac{1}{16} < 0$. That is, $W^2 > W^1$ iff the importance of coordination $(B)$ is large enough.

Altogether, we have,

**Proposition 7** *Fix $e = \widetilde{e}$. Comparing with non-integration, integration generates no loss of control, but a loss in ex ante human capital investment. Integration arises iff the benefit of coordination is large enough to outweigh the cost of the incentive loss.*

## II.4 Conclusion and Empirical Evidences of Chapter Two

The literature on the theory of the firm concentrates on incentive problem. There the allocation of physical capital ownership is driven by incentive balance, either between different tasks or between different players. This chapter argues that incentive is just one side of allocating the ownership rights and the other side is control and coordination. The main message is that these two sides are in trade-off; better control is at price of worse incentive.

In this chapter, the economy consists of a principal-agent relationship and a physical capital. Control/coordination side is to ensure the agent to choose the project that coordinates with the principle's integrated need, rather than the independent project; incentive side is to motivate the agent either ex ante to make human capital investment or ex post to exert effort doing the chosen project. Depending on the allocation of ownership rights and payoff rights of the physical capital, the chapter considers four regimes. In this framework, it is shown that it is efficient to give the payoff rights always to the agent. Thus the key is who owns the capital. If the agent owns the capital, then he is an independent contractor to the principal and the regime is non-integration. If the principal owns it, then integration happens and the agent is a division of a M-form firm as he keeps the payoff rights of his production but the capitals are centralized in the principal's hands. Comparing with non-integration, integration brings better control and thus coordination but induces either ex ante or ex post incentive loss. Thus integration arises if and only if the coordination benefits are large enough to outweighs the incentive loss.

Below I present some empirical evidences to support the main conclusion that ownership structure of physical capitals is determined by the trade-off between coordination benefits and incentive loss. These evidences are hard to be explained by the other theories of the firm.

### II.4.1 GM-Fisher Re-examined

The event that General Motors acquired all Fisher Bodies interest in 1926 is extensively cited in the literature on theory of the firm since Klein et al (1978). In 2000 three papers published by Coase, Freeland, and Casadesus-Masanell & Spulber respectively in Journal of Law and Economics reexamine this classic story.[17] Their common point is that hold-up problem and the relationship-specific physical investment were not problems at all when GM acquired Fisher Body. There obviously existed no important incentive problem in this instance either. About the motivation of integration, Coase says little; Freeland's point is that "the primary factors leading to vertical integration were GM management's fears over the Fisher brothers' impending departure, coupled with problems of financing new body plants";[18] Casadesus-Masanell & Spulber hold that "vertical integration was directed at improving coordination of production and inventories, assuring GM of adequate supplies of auto bodies, and providing GM with access to the executive talents of the Fisher brothers".[19]

In a word, integration was implemented mostly for coordination, in the design of car bodies

[17] See Coase, R., The Acquisition of Fisher Body by General Motors, 15-31; Freeland, R., Creating Holdup through Vertical Integration: Fisher Body Revisited, 33-66; Casadesus-Masanell, R. & D. Spulber, The Fable of Fisher Body, 67-104.

[18] Page 33 supra.

[19] Page 67 supra.

and the supplies of closed bodies, to which it is critical to retain and control Fisher brothers' human capitals that were indispensable to the production of closed bodies.[20] In addition, the integration occurred in 1926 because about that time closed bodies were coming to have strategic importance.[21] That is, the benefits of coordination became high. Below we elaborate on these two points.

From 1924, the automobile market began to transform, "the design and the styling of closed bodies became the primary method of achieving product differentiation and defining a new line of cars".[22] Acquiring Fisher Body, GM not only "increased (its) output but also deprived competitors of closed-body capacity", thus establish its competitive advantage. This is the second point.

For the first point, the three papers point out two kinds of coordination between GM and Fisher Body. One was the technological coordination. Responding to that transformation in auto market in 1924, GM took the "policy of introducing annual model changes…".[23] Then "with annual model changes, redesigns of chassis and bodies would require ongoing consultation and coordination between Fish and the car divisions." It is hard to contract on design and innovation since they are notoriously difficult to foresee and describe, on which GM wanted a bigger say than it had before.

The other kind of coordination is for the sake of competition. GM wanted to cut its competitors' access to Fisher Body, and more importantly, to Fisher brothers' human capital. The former end could be accomplished by an exclusive dealing contract that bound Fisher Body to supply only GM. The draw back of the contract is that it cannot prevent Fisher brothers from

[20] "GM's management believed that Fisher's physical assets would remain relatively useless without the continued involvement of the Fishers", Page 53 supra.

[21] Freeland, Page 52, "A second factor contributing to vertical integration was Fisher's increasing strategic importance".

[22] Freeland, pp 52.

[23] Freeland, pp 50.

serving GM's competitors in a way that was dispensed with Fisher Body. To effectively control the human capital of Fisher brothers, it was necessary for GM to acquire all the capital they operated with and make them its employees (and also shareholders).

Thus GM-Fisher story is thus an evidence to the theory's assertion that integration is to accomplish the benefit of coordination, and through enhanced control over the agent's human capital.

### II.4.2 Retail Contracting

Manufacturers sell their product to consumers through the retail outlets owned by themselves (vertical integration) or through independent retailers (non-integration). Extensive empirical work has been done on this choice. Lafontaine and Slade (1997) provide a good survey. In retail contracting, as they point out, generally there are no important relationship-specific assets or investment. Lafontaine & Shaw (2001) , using an extensive longitudinal data set on franchising firms, show that after eight or more years stable development franchisors maintain a stable rate of company-owned outlets to the franchised ones. The stable rates vary considerably across sectors, and they find that brand-name value is a primary determinant, high brandname value franchisors targeting high rates of company ownership. They argue that that is because high-value franchisors need to exert more direct managerial control over outlets to avoid or reduce the free riding of franchisees on brandname value. The brand-name value is therefore a good proxy for the benefit of coordination and their argument fits into the theory of the chapter very well. In some cases, the effects on brandname value are measured by "outlet size" or "previous experience required". Lafontaine & Shaw (2001) pointed out the effects of these two variables on company ownership is inconsistent with agency theory, which predicts that high monitoring costs implied by big size or high managerial experience tend to lessen company ownership.

# Chapter Three: The Allocation of the Liability to Investors: Why Financial Intermediation?

## III.0 Introduction of the Chapter

This chapter offers a new approach to compare various modes of finance; this approach potentially has a broad range of future applications. The many financial modes of our world can roughly be divided into two classes, direct finance and financial intermediation (FI). Under the former, the capital goes from investors directly to entrepreneurs; under the latter, it goes first to an intermediary, such as a bank, and then is invested by the intermediary in entrepreneurs. How to differentiate these two classes of financial modes poses a challenge at first place. Apparently, the capital changes hands once more under FI, but this simple fact can hardly make any difference. Some, such as Diamond (1984), maintains that the difference is that the intermediary provides not only intermediation of the flow of the capital, but also some extra service, such as monitoring, which helps with some friction of finance; this extra service enables FI to dominate direct finance despite of it adding one more level of agency problem. This extra-service point of view, however, preassumes the service and the intermediation of capital flow are bound together. In principle, this is not always the case; alternatively, the service can be provided by a specialist, while the capital goes from investors directly to entrepreneurs. These two alternatives have been recognized by Holmstrom and Tirole (1997)[24], but cannot be distinguished in their paper.

This chapter is the first to propose that financial modes are decided by who takes the liability to repay investors. Where the service provider alone takes it, the mode is of FI; where entrepreneurs, by some means, take it, the mode is of direct finance. Comparing various financial modes consists in comparing the according allocations of the liability, whereas the way of capital flow

[24] In their paper, the service is also called "monitoring", but means something different from that in Diamond (1984); the effect of monitoring in the former is rather to address a moral hazard problem of entrepreneurs.

makes no difference but hallmarks the liability allocation. It is the economics of this liability allocation that decides one mode or another rises up in equilibrium.

This chapter illustrates one case of such economics in an economy where it is costly for the investors to verify the entrepreneurs' outputs (through auditing)[25], but it is cheap for some expert to observe the outputs through monitoring. By the argument of Diamond (1984)[26], FI, owing to providing a service of monitoring, improves efficiency over the mode under which each entrepreneur is financed independently. From the perspective of this chapter, he has compared two allocations of the liability: one, the expert, as the service provider, takes the liability to repay the investors of all entrepreneurs; the other, each entrepreneur independently takes the liability to repay his investors. Under sufficient diversification, the former dominates the latter, as he has shown[27]. This, however, does not ensure the viability of FI, as he claims; to be viable, FI has to dominate other possible financial modes. This chapter exhausts all the possibilities of financial modes for the two-entrepreneur case[28] and finds that the race is between FI and that

[25] Townsend (1979) for the first time studies the contracting problem with this type of fricitions. Mookerjee and Png (1989) extend it considering the case with stochastic auditing strategies.

[26] There are differences between the set-up of Diamond (1984) and that of this paper, such as he uses non-peculiar punishments rather than auditing costs. But in spirit the two set-ups are equivalent, and this paper will replicate his key results.

[27] Diamond (1984) is the first to show that FI dominates independent finance under sufficient diversification. Following him, Williamson (1986), Krasa and Villamil (1992) and Hellwig (2000) among others discuss the optimal contracts of the bank under various circumstances. But they, including Diamond, do not consider alternative modes of finance that also accomodate monitoring and implement the benefit of diversification.

[28] Bond (2004) also endogenizes various modes of finance for a two-entrepreneur case with a Townsendian friction. The fundamental difference is in the information technology. In his paper, the disclosure costs are proportional to the number of agents to whom the truth is disclosed, while they are constant here. As a result, the driving force of that paper is to raise more investors into senior classes, where disclosure happens under fewer contingencies, while seniority plays no role here. Moreover, it is hard to apply his approach for the case of other extra services, such as screening, while the applications of this paper's approach is straightforward and is

under which the entrepreneurs jointly take the liability under monitoring of the expert. This mode is called "Conglomeration", since it features a conglomerate where each project becomes a division, the entrepreneur the division manager, and the monitor sits in the headquarters in charge of finance.

This chapter then passes on to compare FI to Conglomeration for the case of a large number of entrepreneurs, and finds the following. Conglomeration also implements the benefit of diversification so that diversification in itself does not drive FI, as Diamond (1984) claims; actually, the two modes are equally good under perfect diversification. Rather, FI is driven by a Number Advantage: under FI the investors, when default is declared, audit only *one* intermediary (the bank), rather than *many* entrepreneur projects; the economy this brings about is measured by the ratio of the total costs of auditing all the projects financed by the bank over the costs of auditing this bank. This advantage owes to the banking technology that sets up a unified and rationalized book of all its assets, which can be easily checked. FI dominates Conglomeration only if its Number Advantage is beyond some critical level. Surprisingly, the critical level depends on the monitoring costs per project (bear in mind that monitoring is provided under both modes). Under sufficient but still imperfect diversification, the bigger the costs, the lower the critical level and thus the more likely FI wins the race. When the costs are negligible, FI dominates Conglomeration only if the elasticity of auditing costs to asset scale for the bank is not more than one half. That is, for instance, if the bank assets expand 100 times, then the costs of auditing the bank should not increase by more than $10(=\sqrt{100})$ times. It is hardly credible that organizing the bank can technically brings down auditing costs to such a scale. However, when the monitoring costs are positive, however small, FI can dominate Conglomeration even if the elasticity equals one and thus the costs of auditing the bank increase linearly with its asset scale. The monitoring costs here are the expenses with which a financial expert acquires soft information of the outcome of a project, and hence measure the complexity of the project in

suggested as below.

particular and of the economic system in general. The comparative statics, therefore, explains the growing dominance of FI in this increasingly complex world.

Conglomeration is a prevalent mode in real life intended to implement the benefit of diversification, but is overlooked by the literature that focuses this benefit to justify FI because of its failure to consider the general problem of allocating the liability. A case of this mode, which is not even attached with the name "conglomerate", is the American clearinghouse system. Before the Federal Reserve System was founded, during banking panics banks submitted collateral to the clearinghouse of which they were members. The clearinghouse issued papers upon the portfolio of these assets. These papers were initially used only in interbank clearing, to save cash for depositors, and after the panic of 1893, they were also issued as "cash" directly to depositors. The clearinghouse was responsible for scrutinizing the quality of all the collateral assets, but did not set up the book of their revenues that is necessary to form a unified asset. This was thus an mode of Conglomeration, where the member banks took the joint liability and the monitor was the clearinghouse[29].

To consider the allocation of the liability, much more than only generating new insights in that Townsendian economy, leads to a broad range of new research on the organization of financial markets. For each service that addresses some friction of finance, an examination similar to that of this chapter could be carried out. For example, let the friction be that the quality of a project is unobservable to the investors, but is observed (possibly with noise) to some expert. There are two modes of organizing the screening service. Either he sells directly his knowledge of good projects to investors, and thus becomes a rating agency; or he uses his knowledge to invest, with the capital of investors, and thus becomes a banker or fund manager. The difference between the two modes consists, again, in the allocation of the liability to repay the investors; under the former, it is taken by the entrepreneurs, while under the latter, it is taken by the expert.

[29] See Gorton (1985), and Gorton and Winton (2002) (page 70-72) for details.

This line of consideration leads us to compare, for the first time, the rating agency to the bank. Moreover, to consider the allocation of liability provides a new perspective on the theory of the firm (see Gibbons (2004) for a survey). One branch of the literature (Grossman and Hart (1986) and Hart and Moore (1990)) delineates the boundary of the firm according to the allocation of ownership of physical capital. This chapter suggests that it be delineated according to the allocation of liabilities to a third party (for example, the customers). Basically, if party A takes (uncontractible) liabilities of party B's work, then B is an employee of A, while if B himself takes them, he is an independent contractor. Where the incentive to avoid these liabilities matters, we will reach a theory in the manner of Grossman and Hart (1986).

The rest of the chapter is organized as follows. Subsection III.1 sets up the model. Subsection III.2 examines Independent Finance as the benchmark. Subsection III.3 examines FI. Subsection III.4 examines Conglomeration and then compares it with FI. Concluding remarks and further discussions are given in subsection III.5. Technical proofs are relegated to the appendix.

## III.1 The Model

The model bears a little variation from Diamond (1984) and is isomorphic to it.

## Agents and Production

There are two dates, $\mathrm{T}_0$ for investment and $\mathrm{T}_1$ for return, and three classes of risk neutral agents: entrepreneurs, investors and an expert.

There are $N \geq 2$ entrepreneurs, $E_1, E_2 ... E_N$. Each has an independent and identical project. Each project needs a unit capital to invest, and returns $R$ with probability $q$, and nothing with probability $1 - q$. All entrepreneurs are penniless at $\mathrm{T}_0$. There are infinite potential investors, each of whom has a small amount of capital, but the aggregate capital is well sufficient to finance all the projects. The expert, called Ms X, has neither physical capital nor projects, but has the human capital of monitoring, which has the same meaning as it has in Diamond (1984). So here delegated monitoring is directly assumed, while it is justified with an argument of cost replication in Diamond (1984)[30].

All agents are protected by limited liability, and no one discounts across the two dates. Entrepreneurs are assumed to have all the bargaining power, that is, equilibrium will be driven by maximizing their expected profits.

## Information Structure and Technologies

The only friction to finance a project is that only the entrepreneur costlessly observes its outcome, success or failure. For others to find out the outcome, two information technologies are available, with different costs and information strength. The weak technology is monitoring. If the expert has been monitoring a project from $T_0$ on, then she knows its outcome at $T_1$ through her personal experience, but she is not able to convince others of what she knows. The strong technology is auditing, which discloses the outcome to the public after it is realized. Accordingly, the monitoring costs per project, denoted by $m$, are close to 0 and much less than the auditing costs, denoted by $C$. Only Ms X knows how to monitor, but the investors can access auditing, provided they can afford $C$ collectively[31]. The auditing costs here correspond to the non-pecuniary penalties of Diamond (1984) upon an entrepreneur. Both are the deadweight costs incurred when (actual) default happens; in both set-ups, to avoid these costs through the cheap monitoring is the driving force for various financial modes. There is only one difference:

[30] This argument is not as convincing as it looks. If one million investors need to know something, it is not necessary for each of them to pay the information cost; it would be enough that a hundred randomly chosen investors are paid to investigate the thing, and then to independently report their findings to the rest of the investors. Cross check will stop anyone from lying, and there is no other incentive problem if the action of investigation is contractible.

[31] Notice that here the auditing costs do not vary with the number of the agents to whom the truth is disclosed, as is assumed by Bond (2004).

the auditing costs are borne by the investors, while the penalties are borne by the entrepreneur. This difference does not affect equilibrium since the auditing costs are ultimately borne by the entrepreneur through the individual rationality constraint of the investors.

Ms X observes the outcome of a project at minor costs. If she never colludes with its entrepreneur, the investors can simply rely on her word of mouth to know the outcome and the unobservability is not a friction any more. To exclude such a trivial solution, this chapter is going to allow all possible collusion between Ms X and the entrepreneurs. For that purpose, it is assumed that any side transfers between some or all of these non-investors are costlessly observable to no one but the parties involved. Denote by $C_K$ the costs of auditing Ms X's account if she has monitored $K$ projects and been repaid by the $K$ entrepreneurs. Notice that the problem of collusion plainly precludes Maskin-Moore-Repullo mechanisms from functioning here[32]. For simplicity, it is assumed that the action of monitoring is contractible, since the according moral hazard problem of Ms X is not a necessary part of this chapter. Assume $C_K \geq C$ for $K \geq 2$ and $C_1 = C$ (that is, there is no technological benefit for a single entrepreneur to use monitoring service[33]). This $C_K$ corresponds to the non-pecuniary penalties of Diamond (1984) upon the intermediary, as $C$ corresponds to the penalties upon the entrepreneur. However, there is another difference here: the penalties decrease with the payment from the intermediary to the investors in his paper[34], but $C_K$ is invariant with it; introducing the auditing costs in that way would only add technical complication.

[32] Any two non-investors who have the same information would act as one party. Therefore, there is no way to design a mechanism to elicit the information.

[33] Diamond assumes the same, as he says "(T)he intermediary is not viable [for one entrepreneur case] because it incurs at least as high a deadweight cost [as an entrepreneur]..." (page 400).

[34] In Diamond's paper, the penalties are set by the investors, in such a way that the intermediary always loses overall as much as the total face value of its debts, whenever it defaults any part of them. In this way, the intermediary has no incentive to lie.

A metaphor may be helpful. Assume each of the entrepreneurs and Ms X has a box (or pocket). An entrepreneur knows what is in his box, but cannot see into the other boxes. Ms X knows what is in her box and can also see into an entrepreneur-box at minor costs $m$. The investors have to spend $C$ to open an entrepreneur-box, and $C_K$ to open the X-box if it has been filled by $K$ entrepreneurs.

Additional assumptions are laid out below.

Assumption 1: At $T_0$ the investors commit to playing a *stochastic* auditing strategy.

The investors are able to commit, since the action and the costs of auditing are verifiable and thus the investors have no difficulty in overcoming the collective action problem in connection with the commitment. Hereinafter, they will be dealt with as one party. This assumption of commitment facilitates the approach of mechanism design. Without it, I have to analyze a two-stage game, which is technically more complicated, but the main insights of this chapter will be passed on. Diamond (1984) dispenses with this assumption, since the deadweight costs are imposed by the investors at no expense, so they will impose the penalties whenever there is default. In this chapter, in contrast, investors, having decided to impose auditing, must bear the costs. The other point of Assumption 1 is that the investors enforce auditing stochastically, whereas in Diamond (1984), they impose the penalties deterministically. This assumption of stochastic auditing not only facilitates the mechanism design approach, but also uncovers the indispensable role of monitoring (see the remark following Lemma 2).

Assumption 2: $S \equiv qR - (1-q)C \geq 1$ and $(1-q)C \geq \frac{qR}{2}$.

Basically $S$ is the "surplus" of a project and $S \geq 1$ ensures it is worthy of being financed. The other part of the assumption ensures that the auditing costs are significant enough to leverage up various modes of finance.

Assumption 3: Securities issued to investors must bear repayments that weakly increase with the economic fundamental.

That is, the investors are repaid more when more projects succeed, which is a feature of realistic securities. This assumption restricts the feasible sets of contracts to investors in the problems of designing optimal contracts. Wang (2007) examines what happens without this assumption for the two-entrepreneur case: then, under FI, besides the contractual arrangement considered by Diamond (1996), another one (called "Fund" in Wang (2007)) could be optimal, according to which the investors are repaid with the most when only one project succeeds. As is pointed out by Wang (2007) as well as Innes (1990), when Ms X could collude with the investors, what is assumed in Assumption 3 follows.

Bear in mind that this assumption is needed only because this chapter does not presume the contract of one entrepreneur with the intermediary to be independent of another entrepreneur. If that is presumed, as Diamond (1984, 1996) and those who followed did, Assumption 3 automatically holds: then, the more entrepreneurs succeed, the more the intermediary obtains and accordingly the more it pays out to the investors.

## Timing

**$\mathbf{T}_0$ Morning**: The entrepreneurs *cooperatively* decide how to get financed[35]. They decide who takes the liability to repay investors and thus what the collateral is. Then they design accordingly the contractual arrangement (the mechanism) between them and the investors, and Ms X if the monitoring service is used.

**$\mathbf{T}_0$ Afternoon:** The securities are issued to investors. After buying the securities, they commit to a stochastic auditing strategy.

**$\mathbf{T}_1$ Morning**: The outcomes of all the projects are realized. Non-investors could arrange various sorts of collusion.

[35] At this time, they act as one and the same designer. I abstract away the game probably played between them at this time, as it would be very complicated to take it into account. For example, one entrepreneur's contracts could be contingent on the others', and vice versa, resulting in a problem of infinite recursiveness.

**$T_1$ Afternoon:** The liable entity reports the performance of the collateral to the investors and is ready to repay them accordingly. Contingent on the report, they audit the collateral according to the committed strategy, and if the auditing uncovers any fraud in the report, they appropriate the *whole* collateral[36].

## The Liability Allocations and the Organizations

The mode of organizing finance and the monitoring service is decided by who takes the liability to repay investors. The investors hold claims upon the revenues of the assets of the liability taker(s); these assets are thus the collateral to secure the claims. That an asset is (a part of) the collateral has three implications in this economy:

I, the investors audit it contingent on the report of the performance of the collateral;

II, they appropriate the whole asset whenever auditing uncovers any fraud in the report;

III, any bit of the asset cannot be disposed of before the investors are satisfied, that is, either their claims are fully repaid, or they finish auditing and appropriating the whole collateral.

These three implications decide the economics of the allocation of the liability to repay investors. Based on the allocation of the liability, we can envisage the following modes.

(1) Independent Finance (IF): each project is financed independently and is the collateral upon which the entrepreneur takes the liability to repay its investors.

[36] Any securities to the investors must entail the rights to audit the collateral and furthermore to appropriate a part or the whole of it whenever auditing uncovers a fraud, because this is the only way to incentivize truth-reporting. Ex ante, the harsher the punishment, the lower the incentive to lie, and hence the less the auditing needed. That benefits the entrepreneurs, for the price of the securities has to compensate the auditing costs the investors expect to incur. Therefore, optimally the securities will entail the maximum punishment, that is, a fraud triggers the appropriation of the *whole* collateral.

A caveat about appropriation is needed. The investors do not know the value of the whole collateral if they only audit a part of it when uncovering a fraud. In this situation, the implicit assumption is that they commit to auditing the rest of the collateral whenever uncovering a fraud, to appropriate the whole collateral. This implicit assumption is off-equilibrium path.

(2) Joint Liability without Monitoring: the entrepreneurs, without using the monitoring service, take the joint liability to repay all the investors upon the portfolio of all the projects. This mode, as will be shown, is equivalent to (1).

(3) FI: as discussed by Diamond (1984), Ms X alone takes the liability to repay all the investors, upon the bank asset that is formed by her investment in the projects.

(4) Conglomeration: the liability is taken upon the portfolio of all the projects by a conglomerate, where each project is a division managed by the entrepreneur and Ms X becomes the headquarters monitoring the divisions; Conglomeration differs from mode (2) in monitoring being provided.

(5) Mode M$(N, K)$, for $K = 1, 2, ...N - 1$ : these are the modes mixed between (3) and (4). Under M$(N, K)$, the liable entity consists of Ms X and $K$ entrepreneur, and both runs these entrepreneurs' $K$ projects and finances the other $N - K$ projects as the intermediary. The collateral then consists of the $K$ directly financed projects and the intermediary asset.

Diamond (1984), not explicitly considering the allocation of the liability, examined only FI and IF[37]. As regards these two modes, all his results will be transplanted here. In fact even the proof of his Proposition 2 can be transplanted to prove its counterpart in this chapter, as is shown in the Appendix.

[37] Diamond (1984) possibly realized the existence of other financial modes, as he said "(T)he costs of delegation are analysed *when* the monitor is a financial intermediary..." (page 398, italiced by this paper). Even so, he did not consider them, possibly because he thought "...asymptotically no *other* delegated monitoring structure will have lower costs" (page 395, italiced by this paper). However, this assertion is incorrect; as will be shown, under large but still finite diversification, FI can be dominated by Conglomeration, though the two modes are equally good under infinite diversification.

### III.2 Benchmark: Independent Finance (IF)

In IF, each project is financed independently and is the collateral. In this mode, Ms X adds no value to any entrepreneur and hence is in idle, because she and the entrepreneur will act as one party due to perfect collusion between the two.

A project has two states, success and failure. Only in the state of success are the investors repaid, and the amount of the repayment, denoted by $d$, defines the security. If the entrepreneur reports a failure, the investors audit the project with probability $l$. A mechanism is then represented by $(d, l)$.

In the state of success, if the entrepreneur reports the truth, he lays out $d$ to clear his liability. If he lies and claims to have failed, with probability $l$, the project is audited, which uncovers the lie and causes the whole revenue, $R$, to be appropriated, whereas with probability $1 - l$, the project is not audited and he escapes the liability. Thus, if lying he expects to outlay $lR$. The incentive compatibility constraint (IC) for truth telling is, therefore, $d \leq lR$.

With probability $q$ the project succeeds and the investors are then repaid with $d$. With probability $1 - q$ it actually fails and they not only get repaid with nothing but also incur the auditing costs, $C$, with probability $l$. Thus, the expected benefit of financing the project is $qd - (1 - q)lC$, and the individual rationality constraint (IR) is $1 \leq qd - (1 - q)lC$.

Each entrepreneur chooses $(d, l)$ to minimize $d$ subject to the IC and IR. Substitute the IC into the IR, $1 \leq qd - (1-q)lC \leq qlR - (1-q)lC$, which implies $\frac{1}{S} \leq l$, where $S \equiv qR - (1-q)C$ is assumed to be no less than 1. Then, by the IR, $\frac{R}{S} \leq d$. Thus, the optimal mechanism of IF is $(d^I = \frac{R}{S}, l^I = \frac{1}{S})$, and under IF a successful entrepreneur outlays $d^I = \frac{R}{S}$.

Under IF, whenever a project fails, it is audited with probability $l^I$. If some other successful entrepreneurs repay its liability and save the project from being audited, then the auditing costs of $l^I C$ are saved. This is exactly the benefit of cross subsidization discussed by Diamond (1984, 1996). Can the entrepreneurs, without involving Ms X, materialize this benefit with a mode of joint liability? The answer is No, as shown below.

The collateral is the portfolio of their projects and has $N+1$ states, $s = 0, 1, 2, ..N$, defined by the number of successful projects, and occurring with probability $p_N^s = C_N^s q^s (1-q)^{N-s}$, where $C_N^s$ is the number of combinations of picking $s$ items out of $N$. A symmetric mechanism of the joint liability mode is represented by $\{D_s, l_s\}_{s=0,1...,N}$ : in state $s$, the investors are repaid with $D_s$, to which each successful entrepreneur equally contributes $\frac{D_s}{s}$, and they audit a (reportedly) failed project with probability $l_s$. $D_0 = 0$ due to limited liability. A successful entrepreneur, if telling truthfully his success, expects to outlay $d^J = \sum_{s=0}^{N-1} p_{N-1}^s \frac{D_{s+1}}{s+1}$. If lying, his project is audited with probability $l^J = \sum_{s=0}^{N-1} p_{N-1}^s l_s$, which is also the probability of an actually failed project being audited in equilibrium. Then, the IC is $l^J R \geq d^J$. The IR is $\frac{1}{N}\sum_{s=0}^{N} p_N^s D_s \geq 1 + \frac{C}{N}\sum_{s=0}^{N} p_N^s l_s (N-s)$. Notice that $p_{N-1}^s \frac{1}{s+1} = \frac{1}{Nq} p_N^{s+1}$. Substitute this into the formula of $d^J$ in the IC, $l^J R \geq \frac{1}{Nq}\sum_{s=0}^{N-1} p_N^{s+1} D_{s+1}|_{D_0=0} = \frac{1}{Nq}\sum_{s=0}^{N} p_N^s D_s|_{\text{the IR}} \geq \frac{1}{q}[1 + \frac{C}{N}\sum_{s=0}^{N} p_N^s l_s (N-s)] = \frac{1}{q}[1 + C(1-q)\sum_{s=0}^{N-1} p_{N-1}^s l_s] = \frac{1}{q}[1 + (1-q)Cl^J]$, where for the second to last equality I apply $\frac{p_N^s (N-s)}{N} = (1-q)p_{N-1}^s$ and $p_N^s l_s (N-s)|_{s=N} = 0$. Check the two ends of the chain above, $l^J \geq \frac{1}{qR-(1-q)C} = l^I$. Thus, a failed project is audited with probability of at least $l^I$. Hence, the joint liability mode cannot save the auditing costs. On the other hand, $\{D_s = sd, l_s = l_I\}_{s=0,1...,N}$, the aggregate of the optimal mechanism of IF, is always feasible, and thus under the joint liability mode the entrepreneurs are never worse off then under IF. Therefore, the two are equivalent. To summarize,

**Lemma 12** *Under IF a successful entrepreneur outlays $d^I = \frac{R}{S}$. Without Ms X, the mode of joint liability is equivalent to IF.*

The assumption of stochastic auditing is indispensable to the second part of the lemma. It does not hold sometimes, if only deterministic auditing is allowed. Here is a counter example for the case of $N = 2$. Let $D_2 = 2D_1 = 2D; l_0 = 1, l_1 = l_2 = 0$. The IR is $2qD \geq 2 + 2(1-q)^2 C$ and the IC is $(1-q)R \geq D$. If $2q(1-q)R \geq 2 + 2(1-q)^2 C \Leftrightarrow (1-q)S \geq 1$, then with a range of $(q, R, C)$, the IR and the IC are satisfied, and hence auditing occurs only in state 0 under the joint liability mode, where it occurs in both states 0 and 1 under IF.

How does monitoring help materialize the benefit of cross subsidization? Because joint liability saves a failed project from being audited, it gives a successful entrepreneur higher incentive to hide his success. This incentive compatibility problem dissipates all the benefit of cross subsidization. Where Ms X knows the outcome of each project through monitoring, an entrepreneur has to buy his silence to hide success, which lessens the incentive to do that.

There are various modes of accommodating the monitoring service. First, as presupposed by Diamond (1984), there is the mode of FI, where Ms X is a banker. This mode is examined in the next section, which, besides re-deriving the results of Diamond (1984), solves for the speed of convergence and the order of the rent to the banker.

## III.3 Diamond World: B-Model

Under this mode, Ms X alone takes the liability to repay investors. At $T_0$, she issues securities in exchange for the investors' capital and then invests it to finance the $N$ projects; this investment forms the bank asset, the collateral under this mode. At $T_1$, she collects repayment funds from the entrepreneurs, reports the funds and uses them to settle her liability to the investors. The investors audit the bank asset with a probability contingent on the report, at costs $C_N \geq C$.

**Definition 3** *FI has "Number Advantage", if $C_N < NC$.*

$NC$ are the total costs of auditing $N$ projects, each at costs $C$. Compared to IF, FI reduces the number of to-be-audited boxes from $N$ (entrepreneur-boxes) to 1 (X-box). If $C_N < NC$, then organizing FI *technically* saves auditing costs. In Diamond (1996), $C_2 = 2C$[38], and hence no number advantage is assumed.

[38] There the costs are the destruction of the low output, $L$. Thus $C = L$ and $C_2 = 2L$.

Below I conduct a fully fledged analysis for the case of $N = 2$ and then describe what happens for the case of $N$ being large. In this section, the marks of some equations are suffixed with "b" to indicate that they are peculiar to the mode of FI, where Ms X is the "b"ank.

### III.3.1 $N = 2$ Case

The economy has three states $s = 0, 1$, and 2, defined by the number of successful projects. A general asset contract of the bank is $\{d_s\}_{s=0,1,2}$, and a liability contract is $\{D_s\}_{s=0,1,2}$, where in state $s$, each successful entrepreneur repays the bank with $d_s$ and it then passes $D_s$ to the investors. At $\mathrm{T}_0$, the investors, facing $\{d_s, D_s\}_{s=0,1,2}$, commit to auditing the bank with probability $l_s$ in reported state state $s$. A mechanism is thus $\{d_s, D_s; l_s\}_{s=0,1,2}$.

By limited liability for X and the entrepreneurs,

$$D_0 = d_0 = 0,\ D_1 \leq d_1 \leq R \text{ and } D_2 \leq 2d_2 \leq 2R \tag{LL}$$

According to implication III of being the collateral, the bank always extracts as many funds from the entrepreneurs as possible. There is no such collusion in which the entrepreneurs fill nothing into the X-box (i.e. the bank asset) first, making Ms X declare default and her box audited always, and then compensate her after the auditing; because under FI, by implication III, they are free to and will dispose of whatever remains in their boxes while the bank is being audited, leaving nothing for the compensation afterwards. So, in state $s$, the bank has in its box (at least) $sd_s$.

Consider the IC for the bank to truthfully report the return of its asset (the collateral). Suppose the state is 2. If honoring the contracts, the bank repays the investors with $D_2$. If instead it lies that the state is 0, then with probability $l_0$, the investors audit the bank asset and appropriate all its revenues, $2d_2$ in this state, while otherwise the bank escapes the liability payments. Thus, if telling the lie, the bank expects to outlay $l_0 \cdot 2d_2$. The IC for the bank not to misreport state 2 to be 0 is, thus,

$$l_0 \cdot 2d_2 \geq D_2 \tag{G20b}$$

Alternatively the bank could lie that the state is 1. Then, with probability $l_1$, it loses $2d_2$, as the lie is uncovered, and with probability $1 - l_1$, the report is accepted as the truth and accordingly the bank outlays $D_1$ to the investors. The IC for not to misreport state 2 to 1 is, thus,

$$l_1 \cdot 2d_2 + (1 - l_1)D_1 \geq D_2 \tag{G21b}$$

When the true state is 1, misreporting it to state 2 is never profitable, since by Assumption 3, $D_2 \geq D_1$. Thus, if state 2 is reported, it must be the truth and $l_2 = 0$. If the bank lies that the state is 0, it expects to lose the whole asset, now worth $d_1$, with probability $l_0$, and nothing otherwise. Thus, the IC for not to misreport state 1 to 0 is

$$l_0 d_1 \geq D_1 \tag{G10b}$$

The investors obtain $D_1$ and $D_2$ in state 1 (with probability $2q(1-q)$) and state 2 (with probability $q^2$) respectively. They expect to incur auditing costs of $[(1-q)^2 \cdot l_0 + 2q(1-q)l_1]C_2$. The IR for them to invest in the bank is

$$q^2 D_2 + 2q(1-q)D_1 \geq 2 + [(1-q)^2 \cdot l_0 + 2q(1-q)l_1]C_2 \tag{IR-Ib}$$

Facing the bank contract $C \equiv \{d_s, D_s\}_{s=0,1,2}$, the investors commit to the strategy $\{l_s\}_{s=0,1,2}$ that minimizes the auditing costs subject to the ICs and IR above. Denote the optimal strategy by $\{l_s(C)\}_{s=0,1,2}$. As mentioned above, $l_2(C) = 0$.

The three ICs above guarantee that any misreport of the state does not make the investors lose. Even when these ICs hold, a successful entrepreneur, however, could still be exploited by Ms X through partial collusion[39]. Suppose the true state is 2. She may arrange collusion in

[39] A failed entrepreneur has nothing to be exploited. It is called "partial" as the collusion does not involve all the non-investors.

which she collects $t < d_2$ from $E_1$ to buy his silence when she declares state 1 to the investors *and* $E_2$. Then she collects $d_1$ from $E_2$, besides $t$ from $E_1$, and repays the investors with $D_1$, if this fraud is not uncovered, which happens only with probability $1 - l_1$; otherwise, she loses all. The IC for no partial collusion of misreporting state 2 to 1 is, thus, $2d_2 - D_2 \geq (1 - l_1)(d_1 + t - D_1)$ for any $t < d_2$. Equivalently,

$$2d_2 - D_2 \geq (1 - l_1)(d_1 + d_2 - D_1) \tag{P21b}$$

When the true state is 1, Ms X could arrange partial collusion in which she gives $\epsilon$ to the failed entrepreneur and buy his silence when declaring state 2[40]. She then collects $d_2$ from him and outlays $D_2$ to the investors, who, happy to obtain $D_2$, will not audit the bank. The IC for no partial collusion of misreporting state 1 to be 2 is, thus, $d_1 - D_1 \geq d_2 - D_2 - \epsilon$ for any $\epsilon > 0$. Equivalently,

$$d_1 - D_1 \geq d_2 - D_2 \tag{P12b}$$

The investors do not care how the bank deals with the entrepreneurs and thus do not take into account the two partial collusion proof constraints when choosing the auditing strategy. They actually get more from partial collusion if the constraints are not satisfied: in state 1, they obtain $D_2$ instead of $D_1$; in state 2, if $(1 - l_1)(d_1 + t - D_1) > 2d_2 - D_2$ (the partial collusion occurs), they obtain $l_1(d_1 + t) + (1 - l_1)D_1$, which is more than $D_2$[41]. However, auditing by the investors help the entrepreneurs prevent the partial collusion; the higher is $l_1$, the looser is (P21b).

The last constraint to consider is the IR for Ms X. She keeps the difference between the asset paid-in and liability paid-out. She incurs monitoring costs for the two projects. The IR for her

[40] Notice that misreporting state 1 to 2 is never profitable to the coalition of all the non-investors, but it could be profitable to a subcoalition of Ms X and the failed entrepreneur at the loss of the successful one.

[41] The inequality, by (G21b), is implied by $d_1 + t > 2d_2$, which is derived as follows. $(1 - l_1)(d_1 + t - D_1) > 2d_2 - D_2 \Leftrightarrow (1 - l_1)(d_1 + t) > 2d_2 + (1 - l_1)D_1 - D_2 \geq (1 - l_1) \cdot 2d_2$, where $(1 - l_1)D_1 - D_2 \geq -l_1 \cdot 2d_2$ is implied by (G21b).

is

$$q^2(2d_2 - D_2) + 2q(1-q)(d_1 - D_1) \geq 2m \tag{IR-X}$$

The entrepreneurs' problem is then given by

**Problem 4** $\min_{\{d_s;D_s,l_s\}_{s=0,1,2}} 2q(1-q)d_1 + q^2 \cdot 2d_2$, *subject to*

*(1): $l_s = l_s(C)$ for $s = 0, 1, 2$, where $C \equiv \{d_s, D_s\}_{s=0,1,2}$.*

*(2): (LL), (P21b), (P12b), (IR-X), and $D_1 \leq D_2$.*

**Lemma 13** *The optimal mechanism of FI is $D_1 = D_2 = d_1 = d_2 = d^B = \frac{2+(1-q)^2 C_2}{2q-q^2}$; $l_0 = 1, l_1 = l_2 = 0$.*

**Proof.** See the appendix. The binding constraints are (G21b), (G10b), (P21b), and (IR-Ib). ■

This mechanism is the same as that of Diamond (1996), though he assumes $d_1 = d_2$. The mechanism is driven by the trade-off between the auditing costs and the rent to Ms X; the rent is due to her advantage of being the only one who knows of the overall state before auditing. However, auditing discloses to the public what she knows, and thereby reduces her information advantage. Thus, the more auditing is exercised, the less rent she gains. If concern about the auditing costs dominates, then the mechanism is as above, which gives Ms X a net rent of $q^2 d^B - 2m$, but triggers auditing only in state 0. If instead concern about the rent dominates, the optimal mechanism would have $l_0 = l_1 = 1$, which gives her no rent but triggers auditing in both states 0 and 1. Assumption 2 ensures the former concern dominates.

A successful entrepreneur outlays $d^B = \frac{2+(1-q)^2 C_2}{2q-q^2}$ under FI. FI outperforms IF iff $d^B \leq d^I \Leftrightarrow \frac{C_2}{C} \leq \frac{2(1-q)-q^2\frac{R}{C}}{(1-q)^2 S}$. By Assumption 2, $\frac{2(1-q)}{q} \geq \frac{R}{C} > \frac{1-q}{q}$ and $S \geq 1$, which implies $0 < \frac{2(1-q)-q^2\frac{R}{C}}{(1-q)^2 S} < \frac{2-q}{1-q}$. As $\frac{C_2}{C} \geq 1$, FI never always outperforms IF. On the other hand, if $2 < \frac{C_2}{C} < \frac{2-q}{1-q}$, that is, if FI has no number advantage, it still has a positive chance of defeating IF, because it materializes the benefit of cross subsidization, i.e. auditing only happens in state 0. To sum up,

**Corollary 6** *: For $N = 2$, a successful entrepreneur outlays $d^B = \frac{2+(1-q)^2 C_2}{2q-q^2}$. FI never always outperforms IF, but it has a positive chance of winning even without number advantage.*

Notice that the auditing probability under IF, $l^I = \frac{1}{S}$, decreases with $S$, whereas that under FI, $l_0^B = 1$, is independent of $S$. That is due to the difference in the collateral. Under IF, the collateral of each entrepreneur is his pocket, namely his project; the plumper the pocket (the higher is $S$), the less is auditing needed. In contrast, under FI, the collateral is the pocket of Ms X, which is filled by the entrepreneurs. To have her pocket audited with a lower probability, they have to fill more funds into it, which they definitely dislike, so they pick $l_0^B = 1$.

### III.3.2 Large $N$ Case

The economy has $N+1$ states, state $s = 0, 1...N$, occurring with probability $p_N^s$ (hereinafter it is simplified as $p_s$). A general mechanism is $\{d_s, D_s; l_s\}_{s=0,1,...,N}$. Similar to (G21b), the IC for not to misreport the true state $s$ to state $t$ is

$$l_t \cdot sd_s + (1 - l_t) \cdot D_t \geq D_s \tag{Gstb}$$

As suggested by $N = 2$ case, the optimal mechanism in this case is of Diamond (1984): $d_s = d; D_s = \begin{Bmatrix} sd, \text{for } s < k \\ kd, \text{for } s \geq k \end{Bmatrix}, l_s = \begin{Bmatrix} 1, \text{for } s < k \\ 0, \text{for } s \geq k \end{Bmatrix}$ for some $k$. That is, a successful entrepreneur always repays $d$ to the bank, and the bank's liability contract is debt, with the total face value $F = kd$, and whenever the debt is not fully repaid, the bank is audited with probability 1. The Diamondian mechanism satisfies all (G$st$b). It is partial collusion proof as well, because the outlay of a successful entrepreneur is independent of her report of the economy's state. Only the two IRs, for the investors and for Ms X respectively, are left to check.

The IR for the investors is binding, as follows:

$$d[k \sum_{s \geq k} p_s + \sum_{s \leq k-1} sp_s] = N + C_N \sum_{s \leq k-1} p_s \tag{IR-I-Nb}$$

Whether the IR for X is binding depends whether $m = 0$ or $m > 0$. Consider the former first.

**III.3.2.1 $m = 0$ Subcase** The IR for X is never binding. (IR-I-Nb) determines a function $d(k)$. Auditing happens in states $s \leq k-1$ and Ms X receives a rent in states $s > k$. Thus the optimal $k$ that minimizes $d(k)$ is decided by the trade-off between the auditing costs and the rent to Ms X, as in the case of $N = 2$.

Divide by $N$ both sides of (IR-I-Nb) and apply the Central Limit Theorem (CLT), $\frac{s-Nq}{\sqrt{Nq(1-q)}} \sim N(0,1)$, of which the dense and cumulative distribution functions are $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ and $\Phi$ respectively, and let $k = Nq + h\sqrt{Nq(1-q)}$. Then, the outlay $d(k)$ becomes

$$d(h) = \frac{1 + \frac{C_N}{N}\Phi(h)}{q + \sqrt{\frac{q(1-q)}{N}}(h(1-\Phi(h)) - \phi(h))} \tag{ob}$$

Given $N$, the optimal $h$ satisfies the first order condition (FOC):

$$\frac{C_N}{N}\phi(h)[q + \sqrt{\frac{q(1-q)}{N}}(h(1-\Phi(h)) - \phi(h))] = \sqrt{\frac{q(1-q)}{N}}[1 + \frac{C_N}{N}\Phi(h)](1-\Phi(h))$$

To solve $h$ from the FOC explicitly, I assume $\lim_{N\to\infty} \frac{C_N}{N^\alpha} = z > 0$ for some $\alpha \in (0,1]$. Any $\alpha < 1$ means a big number advantage. Suppose $\frac{C_N}{N}\Phi(h) = o(1)$, which is obvious for $\alpha < 1$ and to be verified for $\alpha = 1$. Then, the right hand side (RHS) of the FOC $\approx \sqrt{\frac{q(1-q)}{N}}(1-\Phi(h))$[42]. Suppose $\frac{h}{\sqrt{N}} = o(1)$ (to be verified later), which, in combination with $\frac{\phi(h)}{\sqrt{N}} = o(1)$, implies that the left hand side (LHS) of the FOC $\approx \frac{C_N}{N}\phi(h)q$. Substituting $C_N = zN^a$, the FOC is simplified to

$$\frac{qz}{\sqrt{q(1-q)}} N^{\alpha-0.5} = \frac{1-\Phi(h)}{\phi(h)} \tag{hb}$$

**Lemma 14** *The solution of (hb) is* $h_N^B = \left\{ \begin{array}{ccc} \frac{\sqrt{q(1-q)}}{qz} N^{0.5-\alpha} + o & & \alpha < 0.5 \\ \widehat{h} & \text{if} & \alpha = 0.5 \\ -\sqrt{(2\alpha-1)\log N} + o & & \alpha > 0.5 \end{array} \right\}$, *where* $\widehat{h}$ *is the unique solution of* $\frac{qz}{\sqrt{q(1-q)}} = \frac{1-\Phi(h)}{\phi(h)}$.

[42] Hereinafter, the notation $y \approx x$, sometimes denoted as $y = x + o$, means $\frac{y-x}{x} \to 0$ if $x \neq 0$, or $y \to 0$ if $x = 0$. The notation $y = O(x)$ means $y \approx \lambda x$ for some $\lambda > 0$.

**Proof.** See appendix. ■

So indeed $\frac{h}{\sqrt{N}} = o(1)$ and $\frac{C_N}{N}\Phi(h) = z\Phi(h) = o(1)$ for $\alpha = 1$. A successful entrepreneur outlays $d_N^B = d(h_N^B)$. By (ob) and Lemma 3[43],

$$d_N^B = \frac{1}{q} + \left\{ \begin{matrix} O(\sqrt{\frac{1}{N^{1-\alpha}}}), \text{ if } \alpha \leq 0.5 \\ O(\sqrt{\frac{(2\alpha-1)\log N}{N}}) \text{ if } \alpha > 0.5 \end{matrix} \right\} \qquad \text{(db)}$$

(db) gives the speed of $d_N^B$ converging to $\frac{1}{q}$.

The rent to Ms X is $V_N = \sum_{s \geq k} d_N^B (s-k) p_s$. Apply the CLT and the fact that $d_N^B \approx \frac{1}{q}$, and let $s = Nq + t\sqrt{Nq(1-q)}$. We have $V_N \approx \frac{\sqrt{Nq(1-q)}}{q} \int_{h_N^B}^{\infty} (t - h_N^B)\phi(t)dt$. The integration equals $\phi(h_N^B) - h_N^B(1 - \Phi(h_N^B))$. It converges to $\phi(\widehat{h}) - \widehat{h}(1-\Phi(\widehat{h}))$, for $\alpha = 0.5$, and to $-h_N^B$ for $\alpha > 0.5$ ($h_N^B \to -\infty$); if $a < 0.5$ and thus $h_N^B = O(N^{0.5-\alpha}) > 0$, it is smaller than $\phi(h_N^B)$, which multiplied by $\sqrt{N}$ still goes to 0. Therefore, $V_N = \left\{ \begin{matrix} o & & \alpha < 0.5 \\ O(\sqrt{N}) & \text{if} & \alpha = 0.5 \\ O(\sqrt{(2\alpha-1)N\log N}) & & \alpha > 0.5 \end{matrix} \right\}$. This gives the order of the rent to Ms X.

**III.3.2.2** $m > 0$ **Subcase** Without considering the IR for X, the rent to her is at most in the order of $\sqrt{N \log N}$. It is dominated by $Nm$ for large $N$ if $m > 0$. Therefore, if $m > 0$, the IR for X is binding,

$$d\sum\nolimits_{s \geq k} (s-k)p_s = Nm \qquad \text{(IR-X-Nb)}$$

We have two equations, (IR-I-Nb) and (IR-X-Nb), for the two unknowns, $d$ and $k$. The unique solution is denoted as $\widehat{d}$ and $\widehat{k}$[44]. Given no rent to Ms X, the costs of FI consist of the monitoring costs ($Nm$ in total) and the auditing costs. Now I go on to figure out the latter. When $N$ goes to infinity, the average expected auditing costs go to zero. Thus, the total face value of the debt is

[43] $\frac{1+x}{q+y} \approx \frac{1}{q}(1+x)(1-\frac{y}{q}) \approx \frac{1}{q} + \frac{x}{q} - \frac{y}{q^2}$ if $x, y \approx 0$. For the fraction of $d_N^B$, $x = o(y)$ when $\alpha < 0.5$, and $y = o(x)$ when $\alpha > 0.5$.

[44] By (IR-I-Nb), $d$ is decreasing with $k$, but by (IR-X-Nb), it is increasing. Thus, there is a unique pair of $d$ and $k$ that satisfies both equations.

$F = N + o$, and the principal part of the repayment from each successful entrepreneur offsets the sum of the investment costs plus the monitoring costs, that is, $\widehat{d} = \frac{1+m}{q} + o$. In this mechanism, $\widehat{k}\widehat{d} = F$, so $\widehat{k} = \frac{qN}{1+m} + o$. By the CLT, the probability of auditing the bank approximates to $\Phi(\frac{\widehat{k}-Nq}{\sqrt{Nq(1-q)}}) = \Phi(\frac{-m}{1+m}\sqrt{\frac{qN}{1-q}})$, which multiplied by $C_N$ gives the total expected auditing costs for this subcase. Therefore, an successful entrepreneur outlays

$$\widehat{d} \approx \frac{1+m}{q} + \Phi(\frac{-m}{1+m}\sqrt{\frac{qN}{1-q}})\frac{C_N}{qN} \tag{db'}$$

**III.3.2.3 Summary for large $N$ Case** In either subcase, when $N$ goes to infinity, $d$ goes to $\frac{1+m}{q}$, even if $\frac{C_N}{NC} > 1$ ($\alpha = 1$ and $z > 1$), that is, even if FI has no number advantage. Thus FI always outperforms IF, as $\frac{1+m}{q} < \frac{R}{S}$[45], which is the point of Diamond (1984). The following corollary summarizes the two subcases:

**Corollary 7** *For large $N$, no matter if FI has number advantage, $d^B_N$ converges to $\frac{1+m}{q}$, as shown in (db) for $m = 0$ and in (db') for $m > 0$. The rent to Ms X is in the order of at most $\sqrt{N \log N}$ if $m = 0$ and is 0 if $m > 0$. In both subcases FI always outperforms IF.*

Diamond claims that, as FI outperforms IF, it is viable; not because of the number advantage, but because of the benefit of cross subsidization amplified by the LLN. However, FI is just one mode of materializing the benefit. The next section examines an alternative mode, which is as much blessed by the LLN as FI. To outperform this alternative mode, the only pillar of FI is number advantage.

## III.4 Conglomeration: H-Model

In this mode, the entrepreneurs, plus Ms X, form a conglomerate, where each project becomes a division, the entrepreneur the division manager, and Ms X sits in the headquarter monitoring the

[45] Because $m \approx 0$ and $\frac{1}{q} < \frac{R}{S}$; the latter $\Leftrightarrow S < qR \Leftrightarrow qR - (1-q)C < qR$.

divisions. Here funds could flow, differently from under FI, directly between the investors and the entrepreneurs. But this difference does not matter at all; what matters is the difference in the collateral. To highlight this point, I suppose Ms X also intermediates under Conglomeration, collecting and distributing the investment capital at $T_0$, and collecting and distributing the repayment funds at $T_1$; in some sense, she is the chief financial officer of the conglomerate. The only difference from FI, then, is that under Conglomeration, the collateral is not the pocket of Ms X, but the portfolio of all the $N$ projects.

Again, I do a fully fledged analysis with the case of $N = 2$ and then move on to the case of large $N$. In this section, the marks of some equations are suffixed with "h" to indicate that they are peculiar to the mode of Conglomeration, where Ms X is sitting in the "h"eadquarters.

### III.4.1 $N = 2$ Case

Parallel to subsection 4.1, the economy has three states, $s = 0$, 1 and 2, and a mechanism is $\{d_s, D_s; l_s\}_{s=0,1,2}$, where in state $s$, a successful division gives the headquarter $d_s$ and the headquarter repays the investors with $D_s$. $l_s$ has a different meaning, due to the difference in the collateral: $l_0/l_2$ is the probability of auditing *each* project if state 0/2 is reported and $l_1$ the probability of auditing the reportedly failed project if state 1 is reported[46]. Limited liability for X and the entrepreneurs is (LL), the same as under FI.

Consider the ICs for the liable entity, the conglomerate now, to truthfully report the state to the investors. Suppose the true state is 2. The conglomerate outlays $D_2$ if honoring the contracts. If instead it (mis)reports state 0, each of the two projects is audited with probability $l_0$, and only when both are not audited, the fraud is not uncovered and the conglomerate outlays nothing; otherwise, it loses all the revenues $2R$[47]. Therefore, the IC for not to misreport state 2

[46] If the report is $\{s, f\}$ (project 1 succeeds and 2 fails), the truth is either $\{s, f\}$ or $\{s, s\}$. In other words, if the truth is $\{f, f\}$ or $\{f, s\}$, there is no incentive to misreport it to be $\{s, f\}$. Thus, only the reportedly failed project is audited in state 1.

[47] Here a caveat helps. If the investors actually audit one project only, they find the report of state 0 is a lie, but

to be 0 is

$$(2l_0 - l_0^2) \cdot 2R \geq D_2 \tag{G20h}$$

If the conglomerate lies that the state is 1, then with probability $l_1$, the reportedly failed project is audited, the fraud uncovered, and the conglomerate loses $2R$; otherwise it outlays $D_1$, based on the reported state. The IC for not to misreport state 2 to 1 is, thus,

$$l_1 \cdot 2R + (1 - l_1) \cdot D_1 \geq D_2 \tag{G21h}$$

Similarly, when the state is 1, the IC for not to misreport state 1 to state 0 is

$$l_0 R \geq D_1 \tag{G10h}$$

Again, misreporting state 1 as 2 is never profitable, and thus $l_2 = 0$.

The investors, facing the security $C \equiv \{D_s\}_{s=0,1,2}$, commit to the strategy that minimizes the expected auditing costs $[(1-q)^2 \cdot 2l_0 + 2q(1-q)l_1]C$ subject to the three ICs above and the following IR:

$$q^2 D_2 + 2q(1-q)D_1 \geq 2 + 2[(1-q)^2 l_0 + q(1-q)l_1]C. \tag{IR-Ih}$$

Let the optimal strategy be $\{l_s(C)\}_{s=0,1,2}$. We have known $l_2(C) = 0$.

Move on to consider partial collusion. The investors want the conglomerate to truthfully report the state, but do not care how the internal contracts, $\{d_s\}_{s=0,1,2}$, are arranged; in fact, so long as (G20h), (G21h), and (G10h) are satisfied, they obtain no less than $D_s$ from any partial collusion. However, a successful entrepreneur could be exploited by Ms X through partial collusion. Suppose the true state is 2. Honoring the contracts, Ms X obtains $2d_2 - D_2$. From partial collusion in which she collects $t < d_2$ from $E_2$ to buy his silence when she declares state 1, she obtains $(t + d_1 - D_1)(1 - l_1)$. Previously, under FI, $E_2$ always gains net $d_2 - t$ from the

they do not know the revenues of the other project. Then, how can they appropriate these revenues? Here the implicit assumption is that whenever they have uncovered a fraud, they commit to auditing the whole collateral (here the two projects), for the purpose of appropriation. This commitment is off the equilibrium path.

collusion, even when the fraud is uncovered, since his pocket is not subject to appropriation there. On the contrary, it is here; thus he obtains $(1-l_1)(R-t)$, rather than $R-t$, with the collusion, while $R-d_2$ without. The IC for no partial collusion of misreporting state 2 to 1 is, therefore, $2d_2 - D_2 \geq (1-l_1)(t+d_1-D_1)$ for any $t$ such that $(1-l_1)(R-t) \geq R-d_2$ . Equivalently,

$$d_2 + R - D_2 \geq (1-l_1)(d_1 + R - D_1) \tag{P21h}$$

Similarly, the IC for no partial collusion of misreporting state 1 to 2 is

$$d_1 - D_1 \geq d_2 - D_2 \tag{P12h}$$

Lastly, the IR for Ms X is (IR-X), the same as under FI. The entrepreneurs' problem under Conglomeration, is then

**Problem 5** $\min_{\{d_s;D_s,l_s\}_{s=0,1,2}} 2q(1-q)d_1 + q^2 \cdot 2d_2$, *subject to*

*(1): $l_s = l_s(C)$ for $s = 0,1,2$, where $C = \{D_s\}_{s=0,1,2}$.*

*(2): (LL), (P21h), (P12h), (IR-X), and $D_1 \leq D_2$ (Assumption 3).*

**Lemma 15** *The optimal mechanism of Conglomeration is: $D_1 = D_2 = d_1 = d_2 = d^H = \frac{2R}{q^2R+2(1-q)S}$; $l_0 = \frac{2}{q^2R+2(1-q)S}$, $l_1 = l_2 = 0$.*

The proof is relegated in the appendix. Similarly, (G21h), (G10h), (P21h), and (IR-Ih) are binding. The mechanism is similar to that under FI. Again Ms X gains in net $q^2d^h - 2m$.

Under Conglomeration, a successful entrepreneur outlays $d^H = \frac{2R}{q^2R+2(1-q)S}$. Since $d^H \leq d^I$[48], Conglomeration always outperforms IF. Intuitively, it cannot be worse than joint liability without monitoring (equivalent to IF), as the latter can be regarded as a special case of the former where Ms X's advice is ignored. FI outperforms Conglomeration, if and only if $d^B \leq d^H$, equivalent to

[48] $\frac{2R}{q^2R+2(1-q)S} \leq \frac{R}{S} \Leftrightarrow 2S \leq q^2R + 2(1-q)S \Leftrightarrow 2S \leq qR \Leftrightarrow 2qR - 2(1-q)C \leq qR \Leftrightarrow qR \leq 2(1-q)C$, where the last inequality is assumed in Assumption 2.

$\frac{C_2}{2C} \leq \frac{2}{q^2R+2(1-q)S}$. Notice that $\frac{2}{q^2R+2(1-q)S} \leq 1$[49]. Therefore, only if FI has "Number Advantage", it has a chance of defeating Conglomeration. On the other hand, if $\frac{C_2}{C} < 2$, there exists a range of $(q, R)$ under which $d^B \leq d^H$.

**Proposition 8** *For $N = 2$, $d^H = \frac{2R}{q^2R+2(1-q)S}$. Conglomeration always outperforms IF. FI has a chance of defeating Conglomeration if and only if it has Number Advantage.*

The comparison of the $(\mathrm{G}ij\mathrm{h})_{0\leq j<i\leq 2}$ and (IR-Ih) to their counterparts under FI gives intuition on how the collateral makes difference. These $(\mathrm{G}ij)$ are tighter under FI than under Conglomeration, since, compared to the $(\mathrm{G}ij\mathrm{h})$, each $(\mathrm{G}ij\mathrm{h})$ has the same RHS, but a smaller LHS, for two reasons. (1) For the pairs of (G10) and (G21), it is because $d_1 \leq R$, or $2d_2 \leq 2R$, that is, the collateral under FI (the bank asset) is always worth less than that under Conglomeration (the portfolio of the projects); because the bank asset is filled with the revenues out of the projects. This advantage of Conglomeration is called "Collateral Advantage". (2) For the pair of (G20), $l_0 \cdot 2d_2 < (2l_0 - l_0^2) \cdot 2R$ because not only $2d_2 < 2R$, but also $l_0 \leq 2l_0 - l_0^2$. The latter inequality is due to the fact that the collateral of Conglomeration is spread across the two projects, and auditing any one uncovers the fraud. Thus, given $l_0$, the probability of uncovering the fraud is higher under Conglomeration than under FI. This advantage of Conglomeration is called "Spread Advantage". On the other hand, the LHSs of the two (IR-I) are the same, but in the RHS of (IR-Ib), the total amount of auditing, $(1-q)^2 \cdot l_0 + 2q(1-q)l_1$, is smaller than $2(1-q)^2 \cdot l_0 + 2q(1-q)l_1$, the amount in the RHS of (IR-Ih). That is because in state 0, the *two* projects are audited under Conglomeration, while under FI, it is always *one* bank asset. This advantage of FI is exactly "Number Advantage". For the two-entrepreneur case, (G20) is not binding in either mode, and hence the spread advantage makes no difference. Then, the race of FI against Conglomeration is decided by the strength of number advantage relative to collateral

[49] $q \cdot qR + (1-q) \cdot 2S \geq 2$ since $S \geq 1$ and $qR \geq 2$.The latter is because $1 \leq qR - (1-q)C \leq qR - 0.5qR = 0.5qR$, where the second inequality applies $2(1-q)C \geq qR$.

advantage, as shown in the inequality $\frac{C_2}{2C} \leq \frac{2}{q^2 R + 2(1-q)S}$.

### III.4.2 Large $N$ Case

A general mechanism is $\{d_s; D_s, l_s\}_{s=0,1,\dots,N}$. $l_s$ now is the probability of auditing a reportedly failed project[50]. Consider the IC for the conglomerate to truthfully report to the investors. Suppose the true state is $s$. If honoring the contracts, the conglomerate outlays $D_s$. If it lies that the state is $t < s$, then each of the $s - t$ actually successful but reportedly failed projects is audited with probability $l_t$, and auditing any one leads to the appropriation of all the projects, worth $sR$; otherwise it outlays $D_t$. Therefore, the IC for not to misreport state $s$ to be $t$ is

$$(1 - (1 - l_t)^{s-t}) \cdot sR + (1 - l_t)^{s-t} D_t \geq D_s \qquad \text{(Gsth)}$$

Compare the LHS of (G$st$h) to that of (G$st$b), $l_t \cdot sd_s + (1 - l_t) \cdot D_t$. Collateral advantage of Conglomeration is shown in $sR \geq sd_s$. The spread advantage is shown in $1 - (1 - l_t)^{s-t} \geq l_t$; the bigger is $s - t$, the more significant is this advantage.

Consider the following mechanism. $d_s = \begin{Bmatrix} 0, \text{for } s < k \\ d, \text{for } s \geq k \end{Bmatrix}$, $D_s = \begin{Bmatrix} 0, \text{for } s < k \\ kd, \text{for } s \geq k \end{Bmatrix}$ and $l_s = \begin{Bmatrix} l_s, \text{for } s < k \\ 0, \text{for } s \geq k \end{Bmatrix}$ for some critical value $k$. This mechanism exploits spread advantage to the most[51], and I guess is asymptotically optimal. For FI to outperform Conglomeration with its optimal mechanism, it has to outperform Conglomeration with this mechanism. This mechanism is immune to partial collusion, since for the true state $s < k$, Ms X never lies that it is a state $t \geq k$, because she has no means to afford $kd$, and for $s \geq k$, the outlay of a successful entrepreneur is fixed at $d$, independent of Ms X's report of the overall state.

All (G$kt$h) for $t < k$ are binding: $kd = (1-(1-l_t)^{k-t}) \cdot kR$. It follows that $l_t = 1-(1-\frac{d}{R})^{\frac{1}{k-t}}$. The binding IR for the investors is then $dk \sum_{s \geq k} p_s = N + C \sum_{s \leq k-1} p_s (N - s)(1 - (1 - \frac{d}{R})^{\frac{1}{k-s}})$. It would be hard to solve $d$ out of this equation. But notice that $\frac{d}{R}\frac{1}{k-s} \leq 1 - (1 - \frac{d}{R})^{\frac{1}{k-s}} \leq$

[50] Because $D_s$ increases with $s$, no actually failed project will be misreported as being successful.

[51] By the binding (G$st$h), $l_t = 1 - \sqrt[\frac{1}{k-t}]{\frac{kR - D_k}{kR - D_t}}$ increases with $D_t$. Therefore, $D_t = 0$ minmizes $l_t$.

$\log(1-\frac{d}{R})^{-1}\frac{1}{k-s}$[52]. Since this chapter is interested finding the necessary condition for FI to outperform Conglomeration, hereinafter I let $l_s = \log(1-\frac{d}{R})^{-1}\frac{1}{k-s}$. The bigger are these $l_s$, the worse is the performance; if FI loses to Conglomeration for these $l_s$, it loses for the true $l_s$. As will be shown, $d \to \frac{1+m}{q}$. Thus $\log(1-\frac{d}{R})^{-1} \to \log(1-\frac{1+m}{qR})^{-1} \equiv \gamma$. Then the IR for the investors becomes

$$dk \sum_{s \geq k} p_s = N + \gamma C \sum_{s \leq k-1} p_s \frac{N-s}{k-s} \tag{IR-I-Nh}$$

The IR for X being binding or not depends on whether $m = 0$. Again, there are two subcases.

**III.4.2.1** $m = 0$ **Subcase** Here the IR-X is never binding. A successful entrepreneur outlays $d$ with probability $\sum_{s \geq k} p_s$. The expected outlay is, by (IR-I-Nh), $d(k) = \frac{N+\gamma C \sum_{s \leq k-1} p_s \frac{N-s}{k-s}}{k}$. Apply the CLT and let $k = Nq + h\sqrt{Nq(1-q)}$ and $s = Nq + t\sqrt{Nq(1-q)}$. Then, $\sum_{s \leq k-1} p_s \frac{N-s}{k-s} \approx \int_{-\infty}^{h-\frac{1}{\sqrt{Nq(1-q)}}} \frac{\sqrt{N}(1-q)-t\sqrt{q(1-q)}}{(h-t)\sqrt{q(1-q)}}\phi(t)dt \approx \int_{-\infty}^{h-\frac{1}{\sqrt{Nq(1-q)}}} \frac{\sqrt{N}(1-q)}{(h-t)\sqrt{q(1-q)}} \equiv \sqrt{\frac{N(1-q)}{q}}G(h)$, and the expected outlay becomes

$$d(h) = \frac{1+\gamma C\sqrt{\frac{1-q}{Nq}}G(h)}{q+h\sqrt{\frac{(1-q)q}{N}}} \tag{oh}$$

Given $N$, the optimal $h$ that minimizes $d(h)$ satisfies the FOC: $\gamma CG'(h)(1+h\sqrt{\frac{1-q}{Nq}}) = 1 + \gamma C\sqrt{\frac{1-q}{Nq}}G(h)$. Solving it for $h$ gives:

**Lemma 16** *The optimal* $\widetilde{h} = -\sqrt{2\log\log N} + o$. *Moreover,* $\frac{1}{\sqrt{N}}G(\widetilde{h}) = o(\frac{-\widetilde{h}}{\sqrt{N}})$.

**Proof.** See Appendix. ■

By (oh) and the latter half of Lemma 5, $d_N^H = d(\widetilde{h}) \approx \frac{1}{q+\widetilde{h}\sqrt{\frac{(1-q)q}{N}}} \approx \frac{1}{q}(1-\widetilde{h}\sqrt{\frac{(1-q)}{Nq}})$. By the former half,

$$d_N^H = \frac{1}{q} + \sqrt{\frac{\log\log N}{N}}\sqrt{\frac{2(1-q)}{q^3}} + o \tag{dh}$$

[52] That is because $(1-x)\mu \leq 1-x^\mu \leq -\mu\log x$ for $0 < x, \mu \leq 1$. For the former inequality, let $f(x) = 1-x^\mu-(1-x)\mu$. $f(1) = 0$, and $f'(x) = -\mu(x^{\mu-1}-1) < 0$ for $x < 1$. Therefore, $f(x) > 0$ if $x < 1$. For the latter, let $\beta = x^\mu$ and $f(\beta) = -\log\beta-(1-\beta)$. $f(1) = 0$, and $f' = 1-\frac{1}{\beta} < 0$ for $\beta < 1$. Therefore, $f(\beta) > 0$ if $\beta < 1$.

Compare (dh) to (db). $d_N^B < d_N^H$ if and only if $\alpha \le \frac{1}{2}$, where $\alpha = \lim_{N\to\infty} \frac{\log C_N}{\log N}$ is the elasticity of the auditing costs to the asset scale for the bank. That is,

**Proposition 9** *For large $N$, if $m = 0$, FI outperforms Conglomeration only if the elasticity of the auditing costs to the asset scale for the bank is no bigger than* 0.5.

The elasticity being no bigger than 0.5 means that with a 100 times expansion of the bank asset, the costs of auditing the bank increase by no more than $\sqrt{100} = 10$ times. It seems hopeless to satisfy this condition. However, I am going to show that with a positive $m$, the condition for FI to outperform Conglomeration is much less demanding.

As calculated in subsection 4.2.1, the rent to Ms X is in the order of $\sqrt{N(-h)} = O(\sqrt{N \log\log N})$.

**III.4.2.2 $m > 0$ Subcase** Again, the rent is dominated by $Nm$, the total monitoring costs. Thus, the IR-X is binding in this case: $d\sum_{s\ge k}(s-k)p_s = Nm$. The two IRs decide a unique pair of $d, k$. As in subsection 4.2.2, $d \approx \frac{1+m}{q}$ and $k \approx \frac{N}{d} \approx \frac{qN}{1+m}$, and the probability of default, $\sum_{s\le k-1} p_s$, is approximately $\Phi(\frac{-m}{1+m}\sqrt{\frac{qN}{1-q}})$, the same as that under FI.

The expected auditing costs in this subcase are $\gamma C \sum_{s\le k-1} p_s \frac{N-s}{k-s} < \gamma NC \sum_{s\le k-1} p_s \frac{1}{k-s} = \gamma NC \sum_{s\le k-1} p_s \cdot E(\frac{1}{k-s}|s \le k-1) \equiv \overline{C_v}$.

**Lemma 17** $E(\frac{1}{k-s}|s \le k-1) \to \frac{m}{1-q}\log\frac{1-q+m}{m}$, *when* $N \to \infty$.

**Proof.** See the appendix. ■

FI has a chance of defeating Conglomeration only if the expected auditing costs under FI, $\Phi(\frac{-m}{1+m}\sqrt{\frac{qN}{1-q}})C_N$, are no bigger than the upper bound $\overline{C_v}$, which is equivalent to $C_N \le \gamma NC\frac{m}{1-q}\log\frac{1-q+m}{m}$ by this lemma. Thus,

**Proposition 10** *For large $N$, if $m > 0$, FI outperforms Conglomeration only if* $\frac{C_N}{NC} < \frac{\gamma}{1-q} m \log\frac{1-q+m}{m}$.

This proposition says that FI win the race against Conglomeration only if the number advantage is beyond some critical level. This level becomes more demanding and the chance for

FI to win becomes slimmer, when $m$ is smaller, as $\frac{\gamma}{1-q} m \log \frac{1-q+m}{m}$ becomes smaller[53], and when $m \to 0$, $\frac{\gamma}{1-q} m \log \frac{1-q+m}{m} \to 0$ and hence the chance approaches 0, if $C_N$ is in the order of $N$; indeed, if $m = 0$, by the discussion the last subcase, to give FI any chance of winning the race, $C_N$ has to be in the order of at most $\sqrt{N}$.

## III.5 Conclusion and Discussions of Chapter Three

This chapter puts at the core the allocations of the liability to the investors, which decides the organization of financial markets. It has discussed the four allocations: (1) Independent Finance (IF), where each entrepreneur takes the liability independently to his investors based upon his project alone; (2) joint liability without the expert of monitoring, where the entrepreneurs take the liability jointly upon the pool of their projects; (3) B-model, where the expert takes the liability upon the intermediary asset; (4) H-model, where the entrepreneurs and the expert forms a conglomerate to take the liability upon the pool of the projects. Diamond (1984) considers (1) and (3). Showing that (3) dominates (1) under sufficient diversification, it claims that that ensures the viability of B-model. This chapter shows that (1) is equivalent to (2) and is dominated by (4). Thus the real race is between (3) and (4). To rise up in equilibrium, B-model has to win over this conglomerate H-model.

the chapter shows that H-model also materializes the benefit of diversification. Under perfect diversification it is as good as B-model. Therefore, diversification does not push B-model up to equilibrium. Under imperfect diversification, the chapter finds that B-model is actually supported by "Number Advantage", that organizing B-model reduces auditing costs because then only one bank asset, rather than many entrepreneur projects, needs to be audited when default is declared. This advantage has to be large enough for B-model to win over H-model to rise in equilibrium. And the bigger are the monitoring costs, the more chance does B-model have to

[53] $\gamma = \log(1 - \frac{1+m}{q})^{-1}$ increases with $m$. $\{m \log \frac{1-q+m}{m}\}' = \log \frac{1-q+m}{m} - \frac{1-q}{1-q+m} \equiv \log(1+x) - \frac{x}{1+x}$, where $x = \frac{1-q}{m}$. $f(x) = \log(1+x) - \frac{x}{1+x} > 0$ for $x > 0$, since $f(0) = 0$ and $f' = \frac{1}{1+x} - \frac{1}{(1+x)^2} > 0$.

win; if the costs are negligible, it wins only if the elasticity of auditing costs to the asset scale in a bank is no bigger than one half. That is, if the bank's asset expand 100 times, auditing costs increase not more than $\sqrt{100} = 10$ times. It hard to believe that organizing bank could reduce auditing costs to such a scale.

Some discussions of the chapter are put below.

### III.5.1 Other Allocations of the Liability

Given the main text has discussed the four above allocations of the liability, a natural question is: do they exhaust all the possibilities? The answer is no. I illustrate it with an example of $N = 2$. Here is allocation (5), where an entrepreneur, say $E_1$, and Ms X forms an entity to take the liability. Hence, $E_2$'s project is invested by this entity, not directly by the investors. The collateral asset is project 1 plus the intermediary asset in project 2. Then I claim that (1)-(5) exhaust all the possibilities of allocations of the liability.

Consider what could be the collateral asset. In the end, the revenue comes from successful projects. That is, the repayment fund to investors is either directly from the projects, or indirectly from the asset invested in them, or from the mix between the former two cases. For the first case where the projects are the collateral assets, the two projects are either separated into two independent collaterals, which gives rise to allocation (1), or tied together into one collateral asset, which gives rise to (2) or (4), depending whether Ms X is involved. In the second case the investment in the two projects forms the collateral asset. That leads to FI. Moreover, only Ms X has the information advantage that is necessary to do FI. Therefore, the second case corresponds exactly to allocation (3). In the mixed case, the collateral asset is one project plus the intermediary asset within the other project. Again, the intermediary has to be Ms X. Thus, the liability is allocated to the combination of her and the entrepreneur of the directly financed project. This case is exactly allocation (5).

The mixed model, though not considered in the main text, does not affect the conclusions, because it is dominated by H-model (allocation (4)). Its collateral consists of one project plus the intermediary asset in the other project. Thus in case of default, the investors have to audit two assets. That means that comparing to H-model, the mixed model has no "number advantage". However, as B-model, its collateral is always worth less than the pool of the projects, the collateral of H-model. Thus, it is dominated by H-model.

### III.5.2 The Implications on the Theory of the Firm

The allocation of the liability could provide a new perspective for the theory of the firm. The core of the theory is to differentiate the market (or contractual) relationship from employment relationship and compare their efficiencies. The literature focuses on the allocation of (residual) control/decision rights, particularly ownership rights (Grossman and Hart (1986), Hart and Moore (1990) and see more literature in Gibbons (2004)). Roughly, if party A owns some capital that is indispensable for party B to work, then B is an employee of A; otherwise, B is an independent contractual party. The allocation of the liability to some third party (like customers) could be another perspective to differentiate the two relationships. If A takes the liabilities of B's work, then B is an employee of A; otherwise, B is an independent contractual party.

# Conclusion of the Thesis

The thesis examines the nature of the organization, both as a whole and as a stage set up for the members to interact. Chapter One considers why and how the organization as a whole, represented by its name, holds reputation, like a natural person, even thought it, unlike the person, has no fixed self, or "type" as is called in economics. The chapter finds that having names hold reputations improves the economic efficiency, because reputations, if only held by natural persons, will die with the persons, but if held by names, they live longer than the persons, since names are inanimate artefacts and can technically live forever. The chapter finds that organizational reputation is driven by two mechanisms. One is the value-adding mechanism: good types are more capable of adding value to the names than the bad types, because they are more likely to succeed in producing high quality products. The other is the commitment mechanism: buying highly reputable names is equivalent to committing to price the products honestly and only good sellers are willing to make the commitment, thus sorted out by these names. The chapter also derives the dynamics of organizational reputation in the socially best equilibria and find these dynamics similar to the dynamics in which personal reputation evolves.

Chapter Two and Three examine how different parties are glued together into an organization. Chapter Two considers the optimal allocation of ownership of physical capital. The allocation decides the boundary of the firm. Its effect on control receives little attention in the literature and is the focus of the chapter. Control means here to affect the project choice of the agent, between the general one with marketable product and the specific one with the product valuable only to the principal. The value of either depends on the agent's private choice of ex ante human capital investment and/or of ex post effort, which give rise to the incentive side of the set-up. Given the levels of the investment and the effort, the specific product is worth more than the general product. A physical capital is indispensable for the agent to do either project. The chapter examines when the principal owns the capital in equilibrium and when the agent owns. It concludes that the principal ownership improves control, in the sense that the specific project is chosen with a higher probability, and yet reduces incentive of the agent, compared to the agent ownership; thus the former, called "integration", happens iff the benefit of coordination outweighs the loss in incentive. The chapter also provides a rational for M-form organizations, with centralized ownership of physical capital to facilitate coordination, and payoff rights remained to divisions to give them incentive.

While in Chapter Two the glue is ownership of physical capital, Chapter Three provides a new perspective, where the glue is the liability to investors, and challenges the view that the benefit of diversification drives Financial Intermediation (FI), a view first established Diamond (1984) and well accepted by the literature. This chapter argues that there is indeed the benefit of diversification, but it is also implemented by an arrangement of direct finance, called "Conglomeration, which differs from FI in who takes the liability to investors. Under FI, it is taken by the monitor alone, who becomes the bank under FI, while under conglomerate it is taken by the entrepreneurs and the monitor altogether. The assets of the liability taker are the collateral that secures the investors' claims. Therefore, under FI, the collateral is the bank asset, while under conglomerate it is the pool of all the entrepreneur projects. The trade-off between Conglomeration is as follows. First, if default is declared, the investors audit one bank asset under FI but many entrepreneur projects under conglomerate. That is, FI has "Number Advantage". Second, threat of losing the collateral gives the liable agents incentive to report the true revenue. The higher the value of the collateral, the less the incentive to lie, thus the lower the probability auditing is exercised. The bank asset is a part of the pool of the projects and hence is always worth less than the latter. Therefore, Conglomeration has "Collateral Advantage". Under the perfect diversification, Conglomeration is as good as FI; when the number of the entrepreneurs is large but still finite, FI dominates Conglomeration only if its number advantage is larger than a critical value, which depends on monitoring costs. The larger the costs, the higher the chance FI dominates Conglomeration, which explains the increasing prevalence of FI over time.

# Appendices

## A: an Example of Long Dynamics in the Basic Model of Chapter One

The analysis of this example applies (5), which is expounded in subsection 3.2. The dynamics involves four states, as follows. In the equilibrium good sellers buy names of all the states.

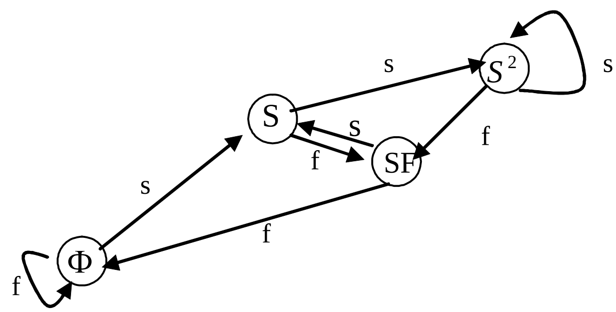


Figure A: the Four States Dynamics

**Claim**: This dynamics implements the second best surplus, $rq\pi$, if and only if it degenerates to the two-state dynamics illustrated in Figure 1, with $SF$ equivalent to $\Phi$ and $S$ to $S^2$.

**Proof**: Consider which levels of efficiency this dynamics can implement. By (5), good sellers buying $h$-names obtain $R_G(h) = rq(p_{hs} - p_{hf}) - \Delta_h$, where $\Delta_h \equiv \max\{p_h - \pi - rp_{hf}, 0\}$. As names of all the four states are bought by good sellers on the equilibrium path, $R_G(h) = R_G(h')$ for any $h, h'$, that is,

($h$-$h\prime$): $rq(p_{hs} - p_{hf}) - \Delta_h = rq(p_{h's} - p_{h'f}) - \Delta_{h'}$.

Notice that $\Delta_\phi = 0$, $p_f = 0$, $p_{s^2f} = p_{sf}$, and $p_{s^2s} = p_{s^2}$. So we have the following two equations.

($\phi$-$s$): $rqp_s = rq(p_{s^2} - p_{sf}) - \Delta_s$.

($s$-$s^2$): $rq(p_{s^2} - p_{sf}) - \max\{p_s - \pi - rp_{sf}, 0\} = rq(p_{s^2} - p_{sf}) - \max\{p_{s^2} - \pi - rp_{sf}, 0\}$.

($s$-$s^2$) implies that its two "max" terms are equal. They equal 0, otherwise $p_{s^2} = p_s$, which means $s^2$-names are equivalent to $s$-names: both have the same value and evolve into the same state after either a success or a failure. $\max\{p_{s^2} - \pi - rp_{sf}, 0\} = 0 \Rightarrow$

(A1): $p_{s^2} - \pi - rp_{sf} \leq 0$.

Substitute $\Delta_s = 0$ into ($\Phi$-$S$),

(A2): $p_{s^2} = p_s + p_{sf}$.

Substitute (A2) into (A1), then

(A3): $p_s \leq \pi - (1-r)p_{sf}$.

The surplus implemented by this dynamics is $R_G(\phi) = rqp_s$. By (A3), $rqp_s = rq\pi \Rightarrow p_{sf} = 0$. This in combination with (A2) implies $p_{s^2} = p_s$. But then $SF$-names are equivalent to $\Phi$-names and $S^2$-names to $S$-names. Q.E.D.

## B: the Sufficiency of the Condition in Proposition 3

To prove the sufficiency, we construct a series of Norm Equilibria whenever $\tau > \frac{(qv-r\pi)(1-rq)}{r(1-q)[qv-(1-rq)\pi]}$. They are supported by the following dynamics, indexed by $N$.

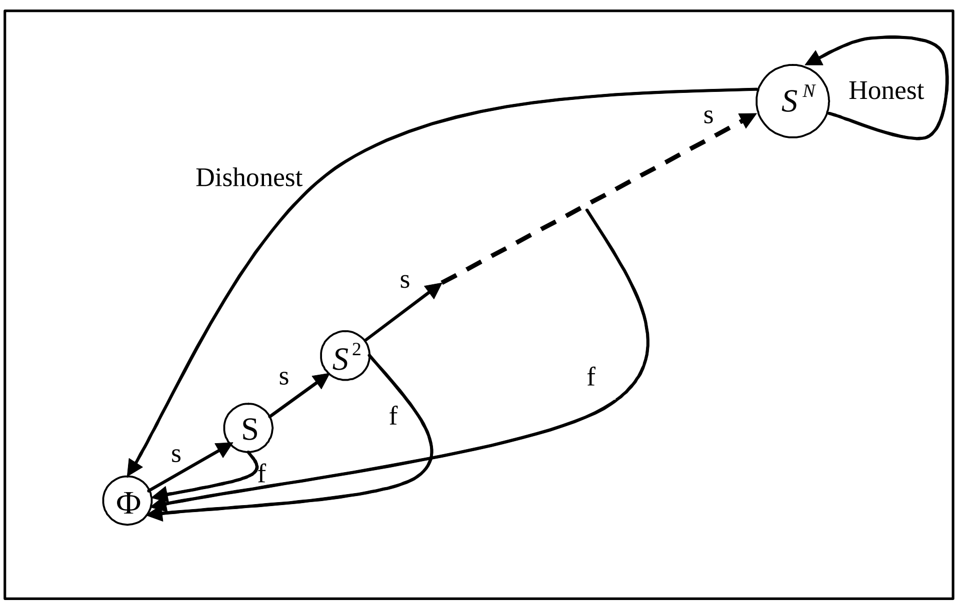


Figure B: $N-$Dynamics

In N-dynamics, there are $N+1$ states, $S^0 \equiv \Phi, S^1, S^2 ... S^N$, all bought by good sellers. For $n = 0, 1 ... N-1$, $S^n$ become $S^{n+1}$ after a success and $\Phi$ after a failure, and are thus non-commitment names. $S^N$ are supposed to be commitment names. The value of them does not sway with performance ($p_{s^N h} = p_{s^N}$ for any $h$), provided they set price 0 for useless widgets. They are destroyed into $\Phi$-names if and only if they unintentionally price the useless widgets at $\overline{w}$.

**Claim**: Whenever $\tau > \frac{(qv-r\pi)(1-rq)}{r(1-q)[qv-(1-rq)\pi]}$, there exists some $N_0 \geq 2$ such that for any $N \geq N_0$, $N$-dynamics above is in equilibrium and implements $R_G > rq\pi$.

**Proof:** We first figure out the $p_s$ and $p_{s^N}$, according to the equilibrium constraints, and then check following two points. (1) $p_{s^N f} \geq \frac{\overline{w}}{r}$, so that $S^N$-names are indeed commitment names; and (2) the dynamics' $R_G$ is strictly bigger than $rq\pi$. (2) is equivalent to $p_s > \pi$, since $R_G = rqp_s$, which is derived by applying (5) to $h = \phi$. (1) is equivalent to $p(s^N) \geq \frac{\overline{w}}{r}$, since $p(s^N f) = p(s^N)$.

$h = s^1, s^2 ... s^{N-1}$ are non-commitment names. Apply (5) to them and notice $R_G = rqp_s$ and $p_{hf} = 0$, $p(s^{n+1}) = \frac{p(s^n)}{rq} + \frac{rqp_s - \pi}{rq}$, for $n = 1, 2, ..., N-1$, where $p(s^n) > p_s > \pi$ is applied. It follows that

(B1): $p(s^N) = \frac{p_s - \pi}{1-rq}(\frac{1}{rq})^{N-1} + \frac{\pi - rqp_s}{1-rq}$.

On the other hand, as $s^N$ are supposed to be commitment names, they satisfy (6). Substituting $p(s^N f) = p(s^N)$ and $R_G = rqp_s$ into it, we have

(B2): $\pi - (1 - rq - r(1-q)\tau)p(s^N) = rqp_s$.

From (B1) and (B2), $p_s = \frac{1-rq-r(1-q)\tau+r(1-q)\tau(rq)^{N-1}}{1-rq-r(1-q)\tau+r(1-q)\tau(rq)^N}\pi > \pi$; thus, point (2) above is checked. Moreover, when $N \to \infty$, $p_s \downarrow \pi$, and hence by (B2), $p(s^N) \uparrow \frac{1-rq}{1-rq-r(1-q)\tau}\pi$. $\frac{1-rq}{1-rq-r(1-q)\tau}\pi > \frac{\overline{w}}{r} \Leftrightarrow \tau > \frac{(qv-r\pi)(1-rq)}{r(1-q)[qv-(1-rq)\pi]}$. Therefore, whenever this condition holds, for large enough $N$, $p(s^N) \geq \frac{\overline{w}}{r}$ and point (1) is checked. Q.E.D.

## The Proof of Lemma 3 of Chapter One

Given any h-names, define $W^t(h)$ the total values of the names that evolves from one unit of the $h$-names on the $t$th period after. Formally, $W^t(h) = \sum_{h' \in H^t} \rho(h')p(hh')$. For example, $W^0(h) = p_h$ and $W^1(h) = q\lambda_h p(hs) + (1 - q\lambda_h)p(hf)$. No Ponzi requires that $\lim_{t\to\infty} r^t W^t(h) = 0$ for any

$h-$names.

**Claim 1**: $W^1(h) \geq \frac{p_h-\pi}{r}$ for any h-names.

**Proof**: Consider all the sellers who own the unit of h-names. The sum of their return is $-p_h + \Pi_h + rW^1(h)$, where $\Pi_h$ is the sum of their profit from selling the widgets. In equilibrium any seller's return is non-negative and his profit is no bigger than $\pi$. Thus, $-p_h + \Pi_h + rW^1(h) \geq 0$ and $\Pi_h \leq \pi$, which implies claim 1. q.e.d.

**Claim 2**: $W^{t+1}(h) \geq \frac{W^t(h)-\pi}{r}$ for any $t$ and any h-name.

**Proof**: By mathematical induction. For $t = 0$, the claim is exactly claim 1. Assume the claim is true for $t = k-1$. Then consider the case $t = k$. Remember $W^{k+1}(h) = q\lambda_h W^k(hs) + (1 - q\lambda_h)W^k(hf)$. Then by induction assumption, $W^{k+1}(h) \geq q\lambda_h \frac{W^{k-1}(hs)-\pi}{r} + (1-q\lambda_h)\frac{W^{k-1}(hf)-\pi}{r} = \frac{q\lambda_h W^{k-1}(hs)+(1-q\lambda_h)W^{k-1}(hf)-\pi}{r} = \frac{W^k(h)-\pi}{r}$. Then the claim holds true for $t = k$. q.e.d.

The following claim is used as a technical tool.

**Claim 3** (Comparison Lemma): Suppose sequence $\{x_t\}$ is defined as follows. $x_0 = p_h = W^0(h)$ and $x_{t+1} = \frac{x_t-\pi}{r}$ for $t \geq 0$. Then $W^t(h) \geq x_t$ for any $t \geq 0$.

**Proof**: By mathematical induction. $t = 0$, that is true by assumption. Assume the claim is true for $t = k$. Then consider the case $t = k+1$. By claim 2, $W^{k+1}(h) \geq \frac{W^k(h)-\pi}{r} \geq \frac{x_k-\pi}{r} = x_{k+1}$, where the second inequality applies the induction assumption. q.e.d.

Then I can prove the lemma, which is about the upper bound of the prices of names.

**Claim 4**: for any h-name, $p_h \leq \frac{\pi}{1-r}$.

**Proof**: Suppose on the contrary for some h-names, $p_h = W^0(h) > \frac{\pi}{1-r}$. Then compute the sequence defined in claim 3. It is easy to get that $x_t = (\frac{1}{r})^t \frac{p_h(1-r)-\pi}{1-r} + b$ for some $b$. Then $r^t W^t(h) \geq r^t x_t \longrightarrow \frac{p_h(1-r)-\pi}{1-r} > 0$. That is, No Ponzi condition is violated. Q.E.D.

## The Proof of Lemma 5 of Chapter One

The proof is similar to the proof of Proposition 2. $P \geq \frac{\overline{w}}{r} > qv$. Thus $P - c > qv - c = \pi$. In Norm Equilibria $R_G > rq\pi$. For any $\varepsilon$ such that $0 < \varepsilon < \min\{c, \frac{R_G}{rq} - \pi\}$, find $h^*$-names such that $p(h^*) > P - \varepsilon$. Then, firstly, $p(h^*) > \pi$ as $P - \varepsilon > P - c > \pi$. Secondly, the names cannot be bought by bad sellers only, since otherwise $p(h^*f) = \frac{p(h^*)+c}{r} > \frac{P}{r} > P$, contradictory to the definition of $P$. Thirdly, $h^*$-names are commitment names. Otherwise, as good sellers buy the names, by (5), $p(h^*s) = p(h^*f) + \frac{R_G}{rq} + \frac{1}{rq}\Delta_h$, which takes the minimum value at $p(h^*f) = \frac{p_h - \pi}{r}$. Therefore $p(h^*s) \geq \frac{p(h^*)-\pi}{r} + \frac{R_G}{rq} > p(h^*) - \pi + \frac{R_G}{rq} > p(h^*) + \varepsilon > P$, a contradiction again.

Then, the $h^*$-names satisfy (6). $R_G = -p(h^*) + \pi + rqp(h^*s) + r(1-q)\tau p(h^*f) < -P + \varepsilon + \pi + [rq + r(1-q)\tau]P \Rightarrow P < \frac{\pi - R_G + \varepsilon}{1 - rq - r(1-q)\tau}$. Let $\varepsilon \to 0$, $P \leq \frac{\pi - R_G}{1 - rq - r(1-q)\tau}$. Q.E.D.

## The Proof of Lemma 7 of Chapter One

Consider Problem 4-n: $\min_{\{p_h\}_{h \in H}} P$, s.t. $p_\phi = 0$; (9) if $p_h < P$; (10); $p_{hf} \leq p_h$; and $l = n$. The problem is derived from Problem 3 by dropping the constraint $P \geq \overline{w}$ and imposing the constraint $l = n$. Denote the value of the problem by $V_n$.

**Claim 1**: If $V_2 \geq \overline{w}$, then the solutions of Problem 4-2 are solutions of Problem 3. Hence the constraint $P \geq \overline{w}$ is not binding in Problem 3.

**Proof**: By (7), $p_{hs} > p_h$ for names not of the top value. Thus $V_n > V_{n-1}$. It follows that if we solve the problem that is derived from Problem 3 by dropping the constraint $P \geq \overline{w}$, the value of this problem is $V_2$. If $V_2 \geq \overline{w}$, this constraint is not binding and the solutions of this problem are solutions of Problem 3. q.e.d.

**Claim 2**: If $\pi \geq \frac{q+2\beta}{2-q}\overline{w}$, $V_2 \geq \overline{w}$.

**Proof**: Solve Problem 4-2. Because $p_{hs} > p_h$ for non-top names and $l = 2$, the first top names are $s^2$-names. We saw in the proof of Lemma 5 $\min p_{s^2} = 2p_s - \pi$, when $p_{sf} = p_s - \pi$. On the other hand, as $p_{s^2} = P$, by (10), $qp_s = -\beta p_{s^2} + \pi$. From these two equations, $p_s = \frac{1+\beta}{q+2\beta}\pi$

and $p_{s^2} = \frac{2-q}{q+2\beta}\pi$. If $\pi \geq \frac{q+2\beta}{2-q}\overline{w}$, $V_2 = p_{s^2} \geq \overline{w}$. q.e.d.

Combining the two claims, we know that the solutions of Problem 4.2 are solutions of Problem 3, if $\pi \geq \frac{q+2\beta}{2-q}\overline{w}$. In the proof of Claim 2, we have found $p_s$, $p_{s^2}$, and $p_{sf} = p_s - \pi = \frac{(1-q)\tau}{q+2\beta}\pi$. To construct fully the solution dynamics of Problem 4-2, we move on to specify the values of other names. Since $p_{s^2} = P$, $p_{s^2h} = P$ for any $h$, provided the widgets are priced honestly. The specification of $p_{sfh}$ for $h \neq \phi$ has some degrees of freedom. If $p_{sf} \leq \pi \Leftrightarrow \tau \leq \frac{2-q}{3-q}$, we set $p_{sff} = 0$ and $p_{sfs} = p_s$. That is, $sf$-names become new names after a failure and $s$-names after a success. The constraint applicable to $h = sf$ is (9), which is satisfied with these prices so specified. The full dynamics are illustrated by figure 2 in the main context. If $p_{sf} > \pi$, set $p_{sff} = p_{sf}$ and $p_{sfs} = P = p_{s^2}$. That is, $sf$-names remain $sf$-names after a failure and become top names after a success. (9) for $h = sf$ is satisfied again with these prices. The dynamics is illustrated by figure 3.

$S^2$-names, the top names and the only commitment names, sort out good sellers through the commitment mechanism by Lemma 4. For non-commitment names, $\lambda_h = \frac{w_h}{qv} = \frac{p_h - p_{hf} + c}{qv}$, if bad sellers ever buy the $h$-names. Then, $\lambda_S = \frac{p_s - p_{sf} + c}{qv} = 1$, that is, $S$-names also sort out good sellers. $\lambda_\Phi = \frac{c}{qv} < 1$ and $\lambda_{SF} = \frac{p_{sf}+c}{qv} \leq 1$ when $\tau \leq \frac{2-q}{3-q}$ and $\lambda_{SF} = \frac{c}{qv} < 1$ otherwise. Therefore, bad sellers buy new names and $SF$-names. Q.E.D.

## The Proof of Lemma 8 of Chapter One

To ease notations, let $p_n$ denote $p_{s^n}$ and $f_x$ denote $\frac{\partial f}{\partial x}$. (11) becomes $p_{n+1} = \max(p_n, \frac{p_n - \pi}{q}) + p_1$, for $n = 1, 2, ..., N-1$. These equations define $p_N$ as a function of $p_1$, $\pi$, and $N$. That is, $p_N = f(p_1, \pi, N)$. Obviously, $f_{p_1} > 0$, $f_\pi < 0$, and $f_N > 0$. Let the inverse function be $p_1 = g(p_N, \pi, N)$. That is, $P \equiv f(g(P, \pi, N), \pi, N)$, where we substitute $P = p_N$. Then $g_P = \frac{1}{f_{p_1}} > 0$, $g_\pi = \frac{-f_\pi}{f_{p_1}} > 0$, and $g_N = \frac{-f_N}{f_{p_1}} < 0$. Moreover, $P(\pi, N)$ is implicitly defined by (12): $-\beta P + \pi = qg(P, \pi, N) \Leftrightarrow F(P, \pi, N) \equiv \beta P + qg(P, \pi, N) - \pi = 0$. Then, $P_N = \frac{-F_N}{F_P} = \frac{-qg_N}{\beta + qg_P} > 0$, as $g_N < 0$ and $g_P > 0$, which proves the first half of the lemma.

For the second half, remember that $N(\pi)$ is implicitly defined by $P(\pi, N) = \overline{w}$. $N'(\pi) = \frac{-P_\pi}{P_N}$. We saw $P_N > 0$. To prove $N'(\pi) < 0$, it suffices to prove $P_\pi > 0$. By implicit function theorem, $P_\pi = \frac{-F_\pi}{F_P} = \frac{1-qg_\pi}{\beta+qg_P}$. The dominator was knew to be positive. The nominator is also positive, by the claim below.

**Claim:** $g_\pi < \frac{1}{q}$.

**Proof:** Since $g_\pi = \frac{-f_\pi}{f_{p_1}}$ and $f_{p_1} > 0$, it suffices to prove that $-f_\pi < \frac{1}{q} f_{p_1}$. To do that, we apply mathematical induction as to $N$. For $N = 2$, $f(p_1, \pi, N) = \max(p_1, \frac{p_1-\pi}{q}) + p_1 = \left\{ \begin{array}{ccc} 2p_1 & & p_1 \leq \frac{\pi}{1-q} \\ & \text{, if} & \\ \frac{p_1-\pi}{q} + p_1 & & p_1 > \frac{\pi}{1-q} \end{array} \right\}$. So $f_\pi = \left\{ \begin{array}{ccc} 0 & & p_1 \leq \frac{\pi}{1-q} \\ & \text{, if} & \\ \frac{-1}{q} & & p_1 > \frac{\pi}{1-q} \end{array} \right\}$ and $f_{p_1} = \left\{ \begin{array}{ccc} 2 & & p_1 \leq \frac{\pi}{1-q} \\ & \text{, if} & \\ \frac{1}{q}+1 & & p_1 > \frac{\pi}{1-q} \end{array} \right\}$. It is obvious that $-f_\pi < f_{p_1} < \frac{1}{q} f_{p_1}$. Assume it holds true for $N = k-1$. Consider the case of $N = k$. To ease notations, we keep $N$ but suppress other arguments $p_1$ and $\pi$; for example, $f_\pi(N) \equiv f_\pi(p_1, \pi, N)$. Then, $f(k) = \max(f(k-1), \frac{f(k-1)-\pi}{q}) + p_1 = \left\{ \begin{array}{ccc} f(k-1)+p_1 & & f(k-1) \leq \frac{\pi}{1-q} \\ & \text{, if} & \\ \frac{f(k-1)-\pi}{q} + p_1 & & f(k-1) > \frac{\pi}{1-q} \end{array} \right\}$.

So, $f_\pi(k) = \left\{ \begin{array}{ccc} f_\pi(k-1) & & f(k-1) \leq \frac{\pi}{1-q} \\ & \text{, if} & \\ \frac{f_\pi(k-1)-1}{q} & & f(k-1) > \frac{\pi}{1-q} \end{array} \right\}$ and $f_{p_1}(k) = \left\{ \begin{array}{ccc} f_{p_1}(k-1)+1 & & f(k-1) \leq \frac{\pi}{1-q} \\ & \text{, if} & \\ \frac{f_{p_1}(k-1)}{q} + 1 & & f(k-1) > \frac{\pi}{1-q} \end{array} \right\}$. When $f(k-1) \leq \frac{\pi}{1-q}$, $-f_\pi(k) = -f_\pi(k-1) < \frac{1}{q} f_{p_1}(k-1) < \frac{1}{q}(f_{p_1}(k-1)+1) = \frac{1}{q} f_{p_1}(k)$, where the first inequality applies the induction assumption. When $f(k-1) > \frac{\pi}{1-q}$, $-f_\pi(k) = \frac{1}{q}(-f_\pi(k-1)+1) < \frac{1}{q}(\frac{f_{p_1}(k-1)}{q}+1) = \frac{1}{q} f_{p_1}(k)$. Therefore, $-f_\pi < \frac{1}{q} f_{p_1}$ holds true for $N = k$. q.e.d.

By the claim, $P_\pi = \frac{1-qg_\pi}{\beta+qg_P} > 0$. Therefore, $N'(\pi) = \frac{-P_\pi}{P_N} < 0$. Q.E.D.

## The Proof of Proposition 6 of Chapter Two

To simplify notations, let $e_s = e(s), v_s = v(e(s))$ and $W(s) = \max_e sv(e) - c_e(e)$. Then $V(s) = W(s) + sB$, $e_0 = v_0 = 0$. Remember $\widehat{v}_{cd} = v_{cd}(e_{0.5}) = v_{0.5} + B$.

By Figure 4, the difference between the two regimes is that for $0.5 < s < \widehat{s}$, "cd" is chosen in regime 2 with the probability 0.5 higher than in regime 1. If "cd" is chosen, the total surplus of the P-A relationship is $V(0.5) + 0.5\widehat{v}_{cd}$; if "in" is chosen, the total surplus is $V(s)$. Therefore, the difference of the surplus between the two regimes satisfies $2(W^1 - W^2)(B) = \int_{0.5}^{\widehat{s}} V(s) - (V(0.5) + 0.5\widehat{v}_{cd})ds = \int_{0.5}^{\widehat{s}} W(s) - W(0.5) - v_{0.5} + (s-1)Bds$. Then,

(A0): $2\frac{d(W^1-W^2)(B)}{dB} = \int_{0.5}^{\widehat{s}}(s-1)ds + (V(\widehat{s}) - V(0.5) - 0.5\widehat{v}_{cd})\frac{d\widehat{s}}{dB}$.

The first part of the right hand side is negative. But $V(\widehat{s}) - V(0.5) - 0.5\widehat{v}_{cd} = V(\widehat{s}) - V(\widetilde{s}) > 0$ as $V(\widetilde{s}) = V(0.5) + 0.5\widehat{v}_{cd}$. Thus we have to prove the first part outweighs the second one, to prove that there is a negative upper bound for $\frac{d(W^1-W^2)(B)}{dB}$. For this purpose, we have to estimate $\frac{d\widehat{s}}{dB}$ and $V(\widehat{s}) - V(0.5) - 0.5\widehat{v}_{cd}$.

By lemma 2, $\widehat{s}$ is defined by $V(\widehat{s}) = \widehat{T} + V(\frac{1}{2})$ where $\widehat{T} > 0.5\widehat{v}_{cd}$.is the solution of P's problem when she is offering the tioli $\max_T \Pr(V(s) - T) \geq V(0.5))(T - 0.5\widehat{v}_{cd})$. By variable transformation $V(s) = T + V(0.5)$, the problem becomes $\max_s(1-s)(V(s) - V(0.5) - 0.5\widehat{v}_{cd})$ and $\widehat{s}$ is the solution. By the argument of lemma 2, it satisfies the first order condition $-V(\widehat{s}) + V(0.5) + 0.5\widehat{v}_{cd} + (1-\widehat{s})(v_{\widehat{s}} + B) = 0 \Leftrightarrow -W(\widehat{s}) + W(0.5) + 0.5v_{0.5} + (1-\widehat{s})(v_{\widehat{s}} + 2B) = 0$, given that $V'(s) = v_{cd}(e_s) = v_s + B$. Applying $W' = v_s$, by implicit function theorem, we have

(A1): $\frac{d\widehat{s}}{dB} = \frac{2(1-\widehat{s})}{2v_{\widehat{s}} + 2B - (1-\widehat{s})v'_{\widehat{s}}e'_{\widehat{s}}}$.

*Lemma A1*: when $c_e'''(.) \geq 0$ and $v'''(.) \leq 0$, $e_s$ is concave.

*Proof:* $e_s$ is decided by the first order condition $sv' = c_e'$. Then $e''(s) = \frac{v''(c''-sv'')-v'(c'''-v''-sv''')}{(c''-sv'')^2} < 0$ given the conditions in the lemma and $v' > 0, c'' > 0, v'' < 0$. QED.

As $v(s)$ is the compound of a concave function $v(e)$ with another concave function $e(s)$, by the lemma $v_s$ is also concave.

*Lemma A2*: $\frac{d\widehat{s}}{dB} < \frac{2(1-\widehat{s})}{v_{\widehat{s}}}$.

*Proof*: As $v_s$ is also concave and $v_0 = 0$, $v_s \geq v'_s e'_s s$. Obviously $\widehat{s} > 0.5$. Therefore $v_{\widehat{s}} \geq v'_{\widehat{s}}e'_{\widehat{s}}\widehat{s} > v'_{\widehat{s}}e'_{\widehat{s}}(1-\widehat{s})$. Then $2v_{\widehat{s}} + 2B - (1-\widehat{s})v'_{\widehat{s}}e'_{\widehat{s}} > v_{\widehat{s}}$. This lemma then follows (A1). QED.

*Lemma A3:* If $f(t)$ is concave and $f(0) = 0$, $\frac{f(s)}{f(t)} \geq \frac{s}{t}$ for $0 < s < t$.

*Proof*: $f(s) = f(\frac{s}{t}t + \frac{t-s}{t}0) \geq \frac{s}{t}f(t) + \frac{t-s}{t}f(0)$. QED.

*Lemma A4*: if $f(t)$ is concave, then $\int_a^b f(t)dt \leq f(\frac{b+a}{2})(b-a)$.

*Proof*: for any $x \in [0, \frac{b-a}{2}]$, $f(\frac{b+a}{2}) - f(\frac{b+a}{2} - x) \geq f'(\frac{b+a}{2})x \geq f(\frac{b+a}{2} + x) - f(\frac{b+a}{2})$. QED.

*Lemma A5*: $\widehat{s} \geq \frac{\sqrt{3}}{2}$.

Proof: Since $v_s$ is also concave, $v_{cd}(e_s) = v_s + B$ is also concave. Then $V(\widehat{s}) - V(0.5) > 0.5v_{cd}(e_{0.5}) \Leftrightarrow \int_{0.5}^{\widehat{s}} v_{cd}(e_s) > 0.5v_{cd}(e_{0.5})|_{\text{by lemma A4}} \Rightarrow v_{cd}(e(\frac{\widehat{s}+0.5}{2}))(\widehat{s} - 0.5) > 0.5v_{cd}(e_{0.5}) \Rightarrow 2(\widehat{s} - 0.5) > \frac{v_{cd}(e_{0.5})}{v_{cd}(e(\frac{\widehat{s}+0.5}{2}))}|_{\text{by lemma A3}} > \frac{0.5}{\frac{\widehat{s}+0.5}{2}}$. From the last inequality, we have $\widehat{s}^2 - 0.25 > 0.5$. QED.

We have done enough to estimate $\frac{d\widehat{s}}{dB}$. Now we come to $V(\widehat{s}) - V(0.5) - 0.5\widehat{v}_{cd}$.

*Lemma A6*: $V(\widehat{s}) - V(0.5) - 0.5\widehat{v}_{cd} < \widehat{s}v_{\widehat{s}} - v_{0.5}$.

*Proof*: $V(\widehat{s}) - (V(0.5) + 0.5\widehat{v}_{cd}) = \widehat{s}(v_{\widehat{s}} + B) - c_e(e_{\widehat{s}}) - (v_{0.5} + B - c_e(e_{0.5})) = \widehat{s}v_{\widehat{s}} - v_{0.5} - (1 - \widehat{s})B - (c_e(e_{\widehat{s}}) - c_e(e_{0.5})) < \widehat{s}v_{\widehat{s}} - v_{0.5}$, since $0.5 < \widehat{s} < 1$. Q.E.D.

By (A0), Lemmas A3 and A6, $2\frac{d(W^1-W^2)(B)}{dB} < -(\widehat{s} - 0.5)(1 - \frac{\widehat{s}+0.5}{2}) + (\widehat{s}v_{\widehat{s}} - v_{0.5})\frac{2(1-\widehat{s})}{v_{\widehat{s}}} \leq -(\widehat{s}-0.5)(1-\frac{\widehat{s}+0.5}{2}) + 2(1-\widehat{s})(\widehat{s} - \frac{0.5}{\widehat{s}})$, where the last inequality applies $\frac{\widehat{s}v_{\widehat{s}}-v_{0.5}}{v_{\widehat{s}}} = \widehat{s} - \frac{v_{0.5}}{v_{\widehat{s}}} \leq \widehat{s} - \frac{0.5}{\widehat{s}}$ by lemma A3. To show that $\frac{d(W^1-W^2)(B)}{dB}$ is upper bounded by a strictly negative constant, it suffices to show that $\min_{1\geq\widehat{s}\geq\frac{\sqrt{3}}{2}}(\widehat{s} - 0.5)(1 - \frac{\widehat{s}+0.5}{2}) - 2(1 - \widehat{s})(\widehat{s} - \frac{0.5}{\widehat{s}}) > 0$, where $\widehat{s} \geq \frac{\sqrt{3}}{2}$ is from lemma A5. Rearrange the terms, $(\widehat{s} - 0.5)(1 - \frac{\widehat{s}+0.5}{2}) - 2(1 - \widehat{s})(\widehat{s} - \frac{0.5}{\widehat{s}}) = \frac{1}{4\widehat{s}}g(\widehat{s})$, where $g(s) = 6s^3 - 4s^2 - 5.5s + 4$. To show $\min_{1\geq s\geq\frac{\sqrt{3}}{2}} \frac{g(s)}{4s} > 0$, it suffices to show $\min_{1\geq s\geq\frac{\sqrt{3}}{2}} g(s) > 0$. It is easy to check that in $[0.5, 1]$, $g'(s) = 0$ has unique solution $s = \frac{4+\sqrt{115}}{18} < \frac{\sqrt{3}}{2}$, and that $g'(1) > 0$. Therefore $g'(s) > 0$ for $s \in [\frac{\sqrt{3}}{2}, 1]$. Therefore $\min_{1\geq s\geq\frac{\sqrt{3}}{2}} g(s) = g(\frac{\sqrt{3}}{2}) > 0$. Let $\chi = \frac{1}{2}\min_{1\geq s\geq\frac{\sqrt{3}}{2}} \frac{g(s)}{4s} > 0$. Then we have proven that $\frac{d(W^1-W^2)(B)}{dB} < -\chi$. **Q.E.D.**

## The Proof of Lemma 13 of Chapter Three (the optimal contracts under B-model)

$l_0(C)$ and $l_1(C)$ are decided as follows. For the investors' problem, (G21b) is binding; otherwise, $l_1$ could be lowered, which only loosens (IR-Ib). Thus, we have

(A1) $D_2 = 2l_1 d_2 + (1 - l_1)l_0 d_1$.

Similarly (G10b) is binding; otherwise, $l_0$ can be lowered. That is,

(A2): $D_1 = l_0 d_1$.

(P21b) is binding. Otherwise, consider the mechanism $D_1 = D_2 = d_1 = 2d_2; l_0 = 1, l_1 = l_2 = 0$. It implements the benefit of cross subsidization, but does not give the rent to Ms X, which is impossible to achieve. This mechanism satisfies all the constraints but (P21b). Therefore, (P21b) must be binding:

(A3): $2d_2 - D_2 = (1 - l_1)(d_1 + d_2 - D_1)$.

Lastly, (IR-Ib) is binding:

(A4): $q^2 D_2 + 2q(1 - q)D_1 = 2 + C_2[1 - q)^2 l_0 + 2q(1 - q)l_1]$.

All other constraints, (G20b), (P12b) and (IR-X), will be verified not binding. The entrepreneurs' problem is to minimize $d^B = qd_2 + (1 - q)d_1$, s.t. (A1)-(A4).

Substituting (A1) and (A2) into (A3), we get a link between $d_1$ and $d_2$ : $2d_2 - [2l_1 d_2 + (1 - l_1)l_0 d_1] = (1-l_1)[d_2+(1-l_0)d_1] \Leftrightarrow (1-l_1)[2d_2-l_0 d_1] = (1-l_1)[d_2+(1-l_0)d_1] \Leftrightarrow (1-l_1)(d_2-d_1) = 0$. Then, two subcases could arises, either $l_1 = 1$, or $d_2 = d_1$. They are examined one by one.

If $d_2 = d_1 = d^B$, by (A1) $D_1 = l_0 d^B$ and by (A2) $D_2 = (2l_1 + l_0 - l_0 l_1)d^B$. Substitute these into (A4), we have $d^B = \frac{1}{q}\frac{2+C_2[(1-q)^2 l_0+2q(1-q)l_1]}{2ql_1+(2-q)l_0-ql_0 l_1}$. Applying $(\frac{a+bx}{c+dx})' = \frac{bc-ad}{(c+dx)^2}$, we have $\frac{\partial d^B}{\partial l_0} = \frac{1}{q}\frac{-2(2+q+ql_1)+2q(1-q)C_2 l_1(ql_1-1)}{[2ql_1+(2-q)l_0-ql_0 l_1]^2}$. $-2(2+q+ql_1) \leq 0$ and $ql_1 - 1 < 0$, so that $\frac{\partial d^B}{\partial l_0} < 0$ and the optimal $l_0 = 1$. Similarly, $\frac{\partial d^B}{\partial l_1}|_{l_0=1} = \frac{-2+(1-q)C_2(3-q)}{[2ql_1+(2-q)l_0-ql_0 l_1]^2} > 0$, where $C_2(1-q)(3-q) > 2$ because $C_2 \geq C \geq \frac{1}{1-q}$ (by Assumption 2 $(2(1-q)C-(1-q)C \geq qR-(1-q)C \geq 1)$. Therefore, the optimal $l_1 = 0$.

Substitute $l_0 = 1$ and $l_1 = 0$ into the formula of $d^B$, and we have $d^B = \frac{2+C_2(1-q)^2}{2q-q^2} = d_2 = d_1$, and into (A1) and (A2) we have $D_1 = D_2 = d_1 = d^B$.

If $l_1 = 1$, by (A2) $D_2 = 2d_2$. As $D_2 \le 2l_0 d_2$ ((G20b)), we get $l_0 = 1$. By (A1) $D_1 = d_1$. Then the objective $d^B = q\frac{D_2}{2} + (1-q)D_1$. The RHS of (IR-Ib) equals $2qd^B$. As $l_0 = l_1 = 1$, the LHS of (A4) equals $2 + C_2[(1-q)^2 + 2q(1-q)]$. Therefore $d^B = \frac{2+C_2(1-q^2)}{2q}$.

Compare the two subcases. $\frac{2+C_2(1-q^2)}{2q} > \frac{2+C_2(1-q)^2}{2q-q^2} \Leftrightarrow C_2(1-q)(3-q) > 2$, which is proven to hold as above. Thus, the first subcase is the solution of the minimization problem. Therefore the optimal mechanism is $d_2 = d_1 = D_1 = D_2 = d^B = \frac{2+C_2(1-q)^2}{2q-q^2}$, $l_0 = 1$ and $l_1 = 0$. It is easy to check that the mechanism satisfies (G20b) and (P12b), and Ms X obtains net rent so that her IR is not binding.

Q.E.D.

## The Proof of Lemma 14 of Chapter Three

If $\alpha = 0.5$, (hb) becomes $\frac{qz}{\sqrt{q(1-q)}} = \frac{1-\Phi(h)}{\phi(h)}$. $\lim_{h\to-\infty} \frac{1-\Phi(h)}{\phi(h)} = \infty$ and by L'Hospital's rule, $\lim_{h\to+\infty} \frac{1-\Phi(h)}{\phi(h)} = \lim_{h\to+\infty} \frac{-\phi(h)}{-\phi(h)h} = 0$. And I am going to prove that $\frac{1-\Phi(h)}{\phi(h)}$ is decreasing, which implies the equation has a unique solution. $\{\frac{1-\Phi(h)}{\phi(h)}\}' = \frac{(1-\Phi(h))h-\phi(h)}{\phi(h)}$. $\{(1-\Phi(h))h - \phi(h)\}' = 1-\Phi(h) > 0$, and $\lim_{h\to+\infty}(1-\Phi(h))h-\phi(h) = \lim_{h\to+\infty}(1-\Phi(h))h = \lim_{h\to+\infty} \frac{1-\Phi(h)}{h^{-1}}|_{\text{L'Hospital}} = \lim_{h\to+\infty} h^2\phi(h) = 0$. Therefore, $(1-\Phi(h))h - \phi(h) < 0$ for finite $h$. Then $\{\frac{1-\Phi(h)}{\phi(h)}\}' < 0$.

If $\alpha > 0.5$, By (hb) when $N \to \infty$, $\frac{1-\Phi(h)}{\phi(h)} \to \infty$. Thus $h_N^B \to -\infty$. Then $1 - \Phi(h) \to 1$ and (hb) becomes $\phi(h)^{-1} = \frac{qz}{\sqrt{q(1-q)}} N^{\alpha-0.5}$. As $\phi(h) = \frac{1}{\sqrt{2\pi}} e^{-\frac{h^2}{2}}$, $\sqrt{2\pi} e^{\frac{h^2}{2}} = \frac{qz}{\sqrt{q(1-q)}} N^{\alpha-0.5} \Rightarrow \frac{h^2}{2} \approx (\alpha - 0.5)\log N \Rightarrow h \approx -\sqrt{(2\alpha-1)\log N}$.

Q.E.D.

## The Proof of Lemma 15 of Chapter Three (the optimal contracts under H-model)

Many steps are parallel to the proof of Lemma 2. We figure out $l_0(D_1, D_2)$ and $l_1(D_1, D_2)$ first. (G21h) is binding, to minimize $l_1$, which pins down $l_1 = \frac{D_2 - D_1}{2R - D_1}$. $l_0$ is present in both (G10h) and (G20h), which implies $l_0 \geq \frac{D_1}{R}$ and $l_0 \geq 1 - \sqrt{1 - \frac{D_2}{2R}}$ respectively. Thus the minimum $l_0 = \max(\frac{D_1}{R}, 1 - \sqrt{1 - \frac{D_2}{2R}})$. Then, given $(D_1, D_2)$, the optimal $l_0$ and $l_1$ are

(B1): $(l_0, l_1) = (\max(\frac{D_1}{R}, 1 - \sqrt{1 - \frac{D_2}{2R}}), \frac{D_2 - D_1}{2R - D_1})$.

Let $C_v = 2C[(1-q)^2 l_0 + q(1-q)l_1]$ (the total auditing costs), $m_1 = d_1 - D_1$ (the rent to Ms X in state 1), $m_2 = 2d_2 - D_2$ (the rent in state 2), and $V = 2q(1-q)m_1 + q^2 m_2$ (the total rent). Then the total financial costs are $2 + C_v + V$, which the entrepreneurs want to minimize. Using these notations, (P21h) becomes $m_2 \geq 2(1 - l_1)m_1 + D_2 - 2l_1 R$, and (P12h) $m_1 \geq \frac{m_2 - D_2}{2}$. The entrepreneurs want to minimize $(m_1, m_2)$ subject to nonnegative constraint. Let $m_1 = 0$ and $m_2 = D_2 - 2l_1 R$, which makes (P21h) binding. Then by (B1), $m_2 = \frac{(2R - D_2)D_1}{2R - D_1} \geq 0$, and (P12h) is equivalent to $0 \geq -l_1 R$, unbinding. Thus, in the optimization,

(B2): $(m_1, m_2) = (0, D_2 - 2l_1 R)$.

Lastly, the (IR-Ih) is binding. Thus,

(B3): $q^2 D_2 + 2q(1-q)D_1 = 2 + 2[(1-q)^2 l_0(D_1, D_2) + q(1-q)l_1(D_1, D_2)]C$.

(B3) implicitly defines a function $D_2(D_1)$. $\{(D_1, D_2) | D_2 = D_2(D_1)\}$ is the set of all feasible securities. If the repayment in state 1 ($D_1$) decreases, as compensation, the repayment in state 2 ($D_2$) has to increase. Use "$'$" represents the derivative with respect to $D_1$. Then, $D_2' < 0$.

The entrepreneurs' problem becomes: $\min_{D_1, D_2} V + C_v$, s.t. (B1)-(B3).

Lemma 4 asserts that the minimization happens at $D_1 = D_2$. As $D_1 \leq D_2$ is assumed, to prove that, it suffices to show that $(V + C_v)' < 0$ everywhere. As $l_0 = \max(1 - \sqrt{1 - \frac{D_2}{2R}}, \frac{D_1}{R})$, I consider two subcases depending whether $1 - \sqrt{1 - \frac{D_2}{2R}} \leq \frac{D_1}{R}$ or not.

Consider first the subcase where $1-\sqrt{1-\frac{D_2}{2R}} \leq \frac{D_1}{R}$ and thus $l_0 = \frac{D_1}{R}$. To get an explicit trade-off between the rent ($V$) and the auditing costs ($C_v$), notice that $V = q^2 m_2 = q^2 D_2 - 2q^2 R l_1 \Rightarrow q^2 D_2 = V + 2q^2 R l_1$. And $l_0 = \frac{D_1}{R} \Rightarrow D_1 = l_0 R$. Let $l \equiv q l_1 + (1-q) l_0$ and substitute these into (B3), we get $V + 2q^2 l_1 R + 2q(1-q) l_0 R = 2 + 2(1-q)Cl \Leftrightarrow V + 2qRl = 2 + 2(1-q)Cl \Leftrightarrow V + 2(qR - (1-q)C)l = 2$. As $l = \frac{C_v}{2(1-q)C}$, it follows that $V + \frac{qR-(1-q)C}{(1-q)C} C_v = 2$. Then $(V + C_v)' = \frac{qR-2(1-q)C}{qR-(1-q)C} V'$. By Assumption 2, $qR - 2(1-q)C \leq 0$. $V = q^2 m_2$. Notice that $m_2 = D_2 - 2l_1 R = \frac{(2R-D_2)D_1}{2R-D_1}$. $\frac{\partial m_2}{\partial D_1} > 0$, $\frac{\partial m_2}{\partial D_2} < 0$, and $D_2' < 0$. Hence $V' = q^2 m_2' > 0$ and $(V + C_v)' < 0$.

Consider the subcase where $1 - \sqrt{1-\frac{D_2}{2R}} > \frac{D_1}{R}$ and thus $l_0 = 1 - \sqrt{1-\frac{D_2}{2R}}$. Then, $l_0' = \frac{dl_0}{dD_2} D_2' < 0$. As $V + C_v = q^2 D_2 - 2q^2 R l_1 + 2C[(1-q)^2 l_0 + q(1-q) l_1] = q^2 D_2 - 2qS l_1 + 2C(1-q)^2 l_0$ (remember $S = qR - (1-q)C$), $(V + C_v)' = q^2 D_2' - 2qS l_1' + 2C(1-q)^2 l_0' < q^2 D_2' - 2qS l_1'$. It suffices to prove that $q^2 D_2' - 2qS l_1' < 0$. By (B3), $q^2 D_2' + 2q(1-q) = 2(1-q)^2 C l_0' + 2q(1-q) C l_1'$. $q^2 D_2'$ is smaller than the LHS of this equation, and the RHS is smaller than $2q(1-q)Cl_1'$, as $l_0' < 0$. Therefore, $q^2 D_2' < 2q(1-q)Cl_1'$. Then, $q^2 D_2' - 2qS l_1' < 2q(1-q)Cl_1' - 2qS l_1' = 2q[(1-q)C - S] l_1' < 0$, where for the last inequality we applies $(1-q)C - S = 2(1-q)C - qR \geq 0$ and $l_1' < 0$ ($l_1 = \frac{D_2 - D_1}{2R - D_1}$ so that $\frac{\partial l_1}{\partial D_1} < 0$ and $\frac{\partial l_1}{\partial D_2} > 0$). Therefore, $(V + C_v)' < 0$ in this subcase.

Summing up, the solution to the entrepreneurs' problem is $D_1 = D_2 = D$. Accordingly, by (B1), $l_1 = 0; l_0 = \frac{D}{R}$. Substituting all these into (B3), we have $[q^2 + 2q(1-q)]D = 2 + 2C(1-q)^2 \frac{D}{R}$, which implies $D = \frac{2R}{q^2 R + 2(1-q)S}$. By (B2), $m_1 = 0; m_2 = D_2 = D$. Then $d_1 = D_1 + m_1 = D; d_2 = \frac{m_2 + D_2}{2} = D$. $D \leq R$ (limited liability for the entrepreneurs) is satisfied, since $q^2 R + 2(1-q)S \geq 2$, which holds true because $2S \geq 2$ and $qR \geq 2$ by Assumption 2.

Q.E.D.

## The Proof of Lemma 16 of Chapter Three

The FOC is $1 = \frac{C}{\sqrt{q(1-q)}R}\frac{1}{(1-\Phi(h))^2}[\sqrt{\frac{1}{N}}G(h)\phi(h) + (\sqrt{\frac{1}{N}}G(h))'(1-\Phi(h))]$. To simplify the FOC, we have the following lemma.

**Lemma A1:** When $N \to \infty$, if $h \to -\infty$ and $h\log N = o(\sqrt{N})$, then $\sqrt{\frac{1}{N}}G(h) \approx \sqrt{\frac{1-q}{q}}\phi(h)\log\sqrt{Nq(1-q)}$ and $(\sqrt{\frac{1}{N}}G(h))' \approx -\sqrt{\frac{1-q}{q}}\phi(h)h\log\sqrt{Nq(1-q)}$.

**Proof:** Let $\varepsilon = \frac{1}{\sqrt{Nq(1-q)}}$. Then $\sqrt{\frac{1}{N}}G(h) = \int_{-\infty}^{h-\varepsilon}\frac{(1-q)-\frac{t}{\sqrt{N}}\sqrt{q(1-q)}}{(h-t)\sqrt{q(1-q)}}\phi(t)dt \approx \int_{-\infty}^{h-\varepsilon}\frac{(1-q)}{\sqrt{q(1-q)}(h-t)}\phi(t)dt = \sqrt{\frac{1-q}{q}}\int_{-\infty}^{h-\varepsilon}\frac{1}{h-t}\phi(t)dt$, where the first "$\approx$ " applies $\frac{t}{\sqrt{N}} = o(1-q)$. And similarly, $(\sqrt{\frac{1}{N}}G(h))' \approx \sqrt{\frac{1-q}{q}}(\int_{-\infty}^{h-\varepsilon}\frac{1}{h-t}\phi(t)dt)' = \sqrt{\frac{1-q}{q}}[-\log\varepsilon\phi(h-\varepsilon) - \int_{-\infty}^{h-\varepsilon}\log(h-t)\phi(t)tdt]' = -\sqrt{\frac{1-q}{q}}\int_{-\infty}^{h-\varepsilon}\frac{\phi(t)t}{h-t}dt$. I show that under the conditions of the lemma, $\int_{-\infty}^{h-\varepsilon}\frac{1}{h-t}\phi(t)dt \approx -\log\varepsilon\phi(h)$. Using integration by parts, $\int_{-\infty}^{h-\varepsilon}\frac{1}{h-t}\phi(t)dt = -\log\varepsilon\phi(h-\varepsilon) - \int_{-\infty}^{h-\varepsilon}\log(h-t)\phi(t)tdt$. Notice that when $\varepsilon \to 0$ and $h \to -\infty$, both integrations of the equation are dominated by the value of the integrate functions at $t = h-\varepsilon$. For the former, the value is $\frac{1}{\varepsilon}\phi(h-\varepsilon)$ and for the latter it is $\log\varepsilon(h-\varepsilon)\phi(h-\varepsilon)$. By the conditions of the lemma, $\log\varepsilon(h-\varepsilon) = o(\frac{1}{\varepsilon})$, Thus the latter integration is in the lower rank of the former one. Therefore, $-\log\varepsilon\phi(h-\varepsilon)$ is the principal part of $\int_{-\infty}^{h-\varepsilon}\frac{1}{h-t}\phi(t)dt$.

$-\int_{-\infty}^{h-\varepsilon}\frac{\phi(t)t}{h-t}dt = \int_{-\infty}^{h-\varepsilon}\phi(t)td\log(h-t) = \phi(h-\varepsilon)(h-\varepsilon)\log\varepsilon + \int_{-\infty}^{h-\varepsilon}\log(h-t)(t^2-1)\phi(t)dt$. A similar argument establishes that $-\int_{-\infty}^{h-\varepsilon}\frac{\phi(t)t}{h-t}dt \approx \phi(h)h\log\varepsilon$, which proves the second part of the lemma. q.e.d.

We verify it later that the conditions of Lemma A1 are satisfied. By the lemma, the FOC becomes $1 = \frac{C}{qR}\log\sqrt{Nq(1-q)}\frac{-\phi(h)h(1-\Phi(h))+\phi(h)^2}{(1-\Phi(h))^2}$. Thus, when $N \to \infty$, $h \to -\infty$ or $+\infty$. As the default probability is $\Phi(h)$, $h \to +\infty$ is never possible. Another way to see this is to figure out $\frac{\phi(h)^2-\phi(h)h(1-\Phi(h))}{(1-\Phi(h))^2} \to 1$ when $h \to +\infty$. Therefore, $h \to -\infty$. Thus, $1-\Phi(h) \to 1$ and $\phi(h)^2 = o(\phi(h)h)$ .The FOC becomes $1 = \frac{C}{qR}\log\sqrt{Nq(1-q)}\phi(h)(-h) \Rightarrow$

(C1): $\phi(h)^{-1} \approx \frac{C}{2qR}(-h)\log N$.

That implies that $\frac{h^2}{2} \approx \log(-h)+\log\log N$. Therefore, $-h = o(\log N)$, and $h = -\sqrt{2\log\log N}+ o$. It is verified that the conditions of Lemma A1 are satisfied for this $h$.

By Lemma A1, $\sqrt{\frac{1}{N}}G(h) = O(\phi(h)\log N) = O(\frac{1}{-h})$, where the second equation applies $\phi(h)\log N = O(-h^{-1})$ by (C1). Then $\frac{G(h)}{N} = O(\frac{1}{-h\sqrt{N}}) = o(\frac{-h}{\sqrt{N}})$.

Q.E.D.

## The Proof of Lemma 17 of Chapter Three

As $\frac{k}{N} \to \frac{q}{1+m}$, it suffices to prove the lemma for $N = k\frac{1+m}{q}$. let $\theta = \frac{q}{1+m} = \frac{k}{N}$. An intuitive proof of the lemma is as follows. $E(\frac{1}{k-s}|s \le k-1) = \frac{p_{k-1}\cdot 1+p_{k-2}\cdot\frac{1}{2}+...+p_0\frac{1}{k}}{p_{k-1}+p_{k-2}+...+p_0} = \frac{1+\frac{p_{k-2}}{p_{k-1}}\cdot\frac{1}{2}+...\frac{p_0}{p_{k-1}}\cdot\frac{1}{k}}{1+\frac{p_{k-2}}{p_{k-1}}+...\frac{p_0}{p_{k-1}}}$. For given $N$, $\frac{p_{k-i}}{p_{k-1}} = \frac{C_N^{k-i}q^{k-i}(1-q)^{N-k+i}}{C_N^{k-1}q^{k-1}(1-q)^{N-k+1}} = (\frac{1-q}{q})^{i-1}\frac{(k-1)\cdot(k-2)\cdot...(k-i+1)}{(N-k+i)\cdot(N-k+i-1)\cdot...(N-k+2)}$, where $C_N^i = \frac{N!}{i!(N-i)!}$ is the combination number. Given $N$ is big, $\frac{k-1}{N-k+i} \approx \frac{k-2}{N-k+i-1} \approx ...\frac{k-i+1}{N-k+2} \approx \frac{k}{N-k} = \frac{\theta}{1-\theta}$. Then $\frac{p_{k-i}}{p_{k-1}} \approx (\frac{(1-q)\theta}{q(1-\theta)})^{i-1} =: \lambda^{i-1}$, where $\lambda = \frac{(1-q)\theta}{q(1-\theta)} < 1$ as $\theta < q$. Then, $E(\frac{1}{k-s}|s \le k-1) \approx \frac{1+\lambda\cdot\frac{1}{2}+...\lambda^{k-1}\cdot\frac{1}{k}}{1+\lambda+...\lambda^{k-1}} = \frac{1-\lambda}{\lambda}\frac{\int_0^\lambda \frac{1-t^k}{1-t}dt}{1-\lambda^k} \to \frac{1-\lambda}{\lambda}\log\frac{1}{1-\lambda}$ when $k \to \infty$.

For a strict proof, first I establish that $E(\frac{1}{k-s}|s \le k-1) \le \frac{1+\lambda\cdot\frac{1}{2}+...\lambda^{k-1}\cdot\frac{1}{k}}{1+\lambda+...\lambda^{k-1}}$. Notice that $\frac{(k-1)\cdot(k-2)\cdot...(k-i+1)}{(N-k+i)\cdot(N-k+i-1)\cdot...(N-k+2)} < (\frac{k}{N-k})^{i-1}$, and hence $\frac{p_{k-i}}{p_{k-1}} < \lambda^{i-1}$. Similarly, $\frac{p_{k-i}}{p_{k-j}} < \lambda^{i-j}$ for any $i > j$. The following lemma is useful to establish the inequality.

**Lemma A2**: If $\frac{a_{i+1}}{a_i} < \lambda$, then $\frac{a_1+a_2\frac{1}{2}+...a_k\frac{1}{k}}{a_1+a_2+...a_k} \ge \frac{1+\lambda\cdot\frac{1}{2}+...\lambda^{k-1}\cdot\frac{1}{k}}{1+\lambda+...\lambda^{k-1}}$.

**Proof:** By mathematical induction. For $k = 1$, that is surely true. Assume for $k$, the inequality holds true. Consider the case for $k+1$. Let $V_k = \frac{a_1+a_2\frac{1}{2}+...a_k\frac{1}{k}}{a_1+a_2+...a_k}$ and $W_k = \frac{1+\lambda\cdot\frac{1}{2}+...\lambda^{k-1}\cdot\frac{1}{k}}{1+\lambda+...\lambda^{k-1}}$. By the induction assumption $V_k \ge W_k$. Both $V_k$ and $W_k$ are the convex combination of 1 through $\frac{1}{k}$. Thus both are bigger than $\frac{1}{k+1}$. Notice that $\frac{a_{k+1}}{a_1+a_2+...a_{k+1}} < \frac{\lambda^k}{1+\lambda+...\lambda^k}$, as it $\Leftrightarrow \frac{a_1}{a_{k+1}} + \frac{a_2}{a_{k+1}} + ...1 > \lambda^{-k} + \lambda^{-(k-1)} + ...1$, which is true because $\frac{a_i}{a_{k+1}} = \frac{a_i}{a_{i+1}} \cdot \frac{a_{I+1}}{a_{i+2}} ... \frac{a_k}{a_{k+1}} > (\frac{1}{\lambda})^{k+1-i}$ for any $i = 1, 2..k$. Then $V_{k+1} = \frac{a_1+a_2+...a_k}{a_1+a_2+...a_{k+1}}V_k + \frac{a_{k+1}}{a_1+a_2+...a_{k+1}}\frac{1}{k+1} > \frac{1+\lambda+...\lambda^{k-1}}{1+\lambda+...\lambda^k}V_k + \frac{\lambda^k}{1+\lambda+...\lambda^k}\frac{1}{k+1} \ge \frac{1+\lambda+...\lambda^{k-1}}{1+\lambda+...\lambda^k}W_k + \frac{\lambda^k}{1+\lambda+...\lambda^k}\frac{1}{k+1} = W_{k+1}$, where the first inequality applies $V_k > \frac{1}{k+1}$ and the second one applies $V_k \ge W_k$. q.e.d.

By the lemma, $\frac{p_{k-1}\cdot 1+p_{k-2}\cdot\frac{1}{2}+...+p_0\frac{1}{k}}{p_{k-1}+p_{k-2}+...+p_0} \ge \frac{1+\lambda\cdot\frac{1}{2}+...\lambda^{k-1}\cdot\frac{1}{k}}{1+\lambda+...\lambda^{k-1}}$, that is $E(\frac{1}{k-s}|s \le k-1) \ge \frac{1+\lambda\cdot\frac{1}{2}+...\lambda^{k-1}\cdot\frac{1}{k}}{1+\lambda+...\lambda^{k-1}}$. Then $\lim_{k\to\infty} E(\frac{1}{k-s}|s \le k-1) \ge \lim_{k\to\infty}\frac{1+\lambda\cdot\frac{1}{2}+...\lambda^{k-1}\cdot\frac{1}{k}}{1+\lambda+...\lambda^{k-1}} = \frac{1-\lambda}{\lambda}\log\frac{1}{1-\lambda}$.

For the other direction of the inequality, I need to recover notation $p_N^s$, rather than use its simplification $p_s$. For any $L < k$, $\frac{p_N^{k-1}\cdot 1+p_N^{k-2}\cdot\frac{1}{2}+...+p_N^0\frac{1}{k}}{p_N^{k-1}+p_N^{k-2}+...+p_N^0} < \frac{p_N^{k-1}\cdot 1+p_N^{k-2}\cdot\frac{1}{2}+...+p_N^{k-L}\frac{1}{L}}{p_N^{k-1}+p_N^{k-2}+...+p_N^{k-L}}$, because the former is the convex combination of the latter and $\frac{1}{L+1}, \frac{1}{L+2}...\frac{1}{K}$, which are all smaller than the it. For this inequality, keep $L$ fixed and let $N$ (thus $k = \frac{q}{1+m}N$) goes to infinity. Then the left hand side goes to $\lim_{k\to\infty} E(\frac{1}{k-s}|s \le k-1)$. The right hand side goes to $\frac{1+\lambda\cdot\frac{1}{2}+...\lambda^{L-1}\cdot\frac{1}{L}}{1+\lambda+...\lambda^{L-1}}$, because $\frac{p_{k-i}}{p_{k-1}} = (\frac{1-q}{q})^{i-1}\frac{(k-1)\cdot(k-2)\cdot...(k-i+1)}{(N-k+i)\cdot(N-k+i-1)\cdot...(N-k+2)} \to \lambda^{i-1}$ for any given $i \le L$. Therefore, for any given $L$, $\lim_{k\to\infty} E(\frac{1}{k-s}|s \le k-1) \le \frac{1+\lambda\cdot\frac{1}{2}+...\lambda^{L-1}\cdot\frac{1}{L}}{1+\lambda+...\lambda^{L-1}}$. Let $L$ then goes to infinity. We have $\lim_{k\to\infty} E(\frac{1}{k-s}|s \le k-1) \le \frac{1-\lambda}{\lambda}\log\frac{1}{1-\lambda}$.

Therefore, $\lim_{k\to\infty} E(\frac{1}{k-s}|s \le k-1) = \frac{1-\lambda}{\lambda}\log\frac{1}{1-\lambda}$. Substitute $\lambda = \frac{(1-q)\theta}{q(1-\theta)} = \frac{(1-q)\frac{q}{1+m}}{q\frac{1+m-q}{1+m}} = \frac{1-q}{1-q+m}$. Then $\frac{1-\lambda}{\lambda}\log\frac{1}{1-\lambda} = \frac{m}{1-q}\log\frac{1-q+m}{m}$.

Q.E.D.

# References


[1] Aghion, P. and J. Tirole (1997), "Formal and Real Authority in Organizations," *Journal of Political Economy*, 105, 1-29.

[2] Anderson, E. and D. Schmittlein, (1984), "Integration of the Sales Force: An Empirical Examination", *Rand Journal of Economics*, 15, 385-395.

[3] Baker, G., and T. Hubbard (2000), "Contractibility and Asset Ownership: On-board Cumputers and Governance in U.S. Trucking," Mimeo, Harvard Business School.

[4] Baker, G., R. Gibbons and K. Murphy (2002), "Relational Contract and the Theory of the Firm", *Quarterly Journal of Economics*, 117, 39-84.

[5] Berle, A. and G. Means (1968), *The Modern Corporation and Private Property*, New York: Harcourt, Brace and World, revised edition.

[6] Bolton, P. and D. S. Scharfstein (1998), "Corporate Finance, the Theory of the Firm, and Organizations," *The Journal of Economic Perspectives,* 12, 95-114.

[7] Bond, Philip (2004), "Bank and Nonbank Financial Intermediation", *The Journal of Finance*, LIX, 2489-2529.

[8] Casadesus-Masanell, R. and D. F. Spulber (2000), "The Fable of Fisher Body", *Journal of Law and Economics*, 43, 67-104.

[9] Chandler, A. (1977), *The Visible Hand: The Managerial Revolution in American Business*, Cambridge: Harvard Uni Press.

[10] Coase, R. H. (1937), "The Nature of the Firm", *Economica*, New Series, IV, 386-405.

[11] Coase, R. H. (2000), "The Acquisition of Fisher Body by General Motors", *Journal of Law and Economics*, 43, 15-31.

[12] Deb, Joyee (2008), "Observability and Sorting in a Market for Names", mimeo, Northwestern University.

[13] Diamond, Douglas (1984), "Financial Intermediation and Delegated Monitoring", *The Review of Economic Studies*, 51, 393-414.

[14] Diamond, Douglas (1996), "Financial Intermediation and Delegated Monitoring: A Simple Case", *Economic Quarterly*, Federal Reserve Bank of Richmond, 82.

[15] Diamond, Douglas and P. H. Dybvig (1983), "Bank Runs, Deposit Insurance, and Liquidity", *The Journal of Political Economy*, 91, 401-419.

[16] Gibbons, Robert (2004), "Four Formal(izable) Theories of the Firm", mimeo, MIT.

[17] Freeland, R. F. (2000), "Creating Holdup through Vertical Integration: Fisher Body Revisited", *Journal of Law and Economics*, 43, 33-66.

[18] Gibbons, R. (2004), "Four Formal(izable) Theories of the Firm", manuscript.

[19] Grossman, S. and O. Hart (1986), "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, 94, 691-719.

[20] Hakenes, H. and M. Peitz (2007), "Observable Reputation Trading", *International Economic Review*, 48, 693-730.

[21] Hart, O. (1989), "An Economist's Perspective on the Theory of the Firm", in P. J. Buckley and J. Michie ed. *Firms, Organizations and Contracts, A Reader in Industrial Organization*, New York and London, Oxford Univ. Press, 199-218.

[22] Hart, O. (1995), Firms, *Contracts, and Financial Structure*, Oxford: Oxford University Express.

[23] Hart, Oliver (2001), "Norms and the Theory of the Firm", working paper, Harvard University.

[24] Hart, O. and B. Holmstrom (2002), "A Theory of Firm Scope," working paper.

[25] Hart, O. and J. Moore (1988), "Incomplete Contracts and Renegotiation," *Econometrica*, 56, 755-785.

[26] Hart, O. and J. Moore (1990), "Property Rights and the Nature of the Firms," *Journal of Political Economy*, 98, 1119-1158.

[27] Hayek, F. A. (1945), "The Use of Knowledge in Society ", *American Economic Review*, 35, 519-530.

[28] Hellwig, Martin (2000), "Financial Intermediation with Risk Aversion", *The Review of Economic Studies*, 67, 719-742.

[29] Hermalin, B. E. (1999), "The Firm as a Noneconomy: Some Comments on Holmstrom," *Journal of Law, Economics, and Organization*, 15, 103-106.

[30] Holmstrom, B. (1982), "Moral Hazard in Teams," *Bell Journal of Economics*, 13, 324-340.

[31] Holmstrom, B. (1999), "The Firm as a Subeconomy," *Journal of Law, Economics, and Organization*, 15, 74-102.

[32] Holmstrom, Bengt (1999), "Managerial Incentive Problems: a Dynamic Perspective", *Review of Economic Studies*, 66, 169-182.

[33] Holmstrom, B. and P. Milgrom (1991), "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, and Organization*, 7, 24-52.

[34] Holmstrom, B. and P. Milgrom (1994), “The Firm as an Incentive System”, *American Economic Review*, 84, 972-91.

[35] Holmstrom B and J. Roberts (1998), “The Boundaries of the Firm Revisited,” *Journal of Economic Perspectives*, 12, 73-94.

[36] Klein, B., R. Crawford and A. Alchian (1978), “Vertical Integration, Appropriable Rents, and the Competitive Contracting Process,” *Journal of Law and Economics*, 21, 297-326.

[37] Kreps, David (1990), “Corporate Culture and Economic Theory”, in James E. Alt and Kenneth A. Shepsle (eds.), *Perspectives on Positive Political Economy,* 90-143, (Cambridge, UK: Cambridge University Press.

[38] Krasa, Stefan and Anne P. Villamil (1992), "Monitoring the Monitor: an Incentive Structure for a financial intermediary", *Journal of Economic Theory*, 57, 197-221.

[39] Lafontaine, F. (1995), “Pricing Decisions in Franchised Chains: a Look at the Restraunt and Fast-Food Industry”, NBER working paper 5427.

[40] Lafontaine, F. and K. L. Shaw (2001), “Targeting Managerial Control: Evidence from Franchising”, NBER working paper 8416.

[41] Lafontaine, F. and M. E. Slade (1997), “Retail Contracting: Theory and Practice”, *Journal of Industrial Economics*, XLV, 1-25.

[42] Martin, R. E. (1988), “Franchising and Risk Management”, *American Economic Review*, 78, 954-968.

[43] Marvel, H. P. and Lixin Ye (2008), "Trademark Sales, Entry, and the Value of Reputation," *International Economic Review*, 547-576.

[44] Mirrlees, J. (1999), "The Theory of Moral Hazard and Unobservable Behavior, Part I," *Review of Economic Studies*, 66, 3-22.

[45] Mookherjee, Dilip and Ivan Png (1989), "Optimal Auditing, Insurance, and Redistribution", *Quarterly Journal of Economics*, 104, 399-415.

[46] Norton, S. W. (1988), "An Empirical Look at Franchising as an Organizational Form", *Journal of Business*, 61, 197-217.

[47] Rajan, R. G. and L. Zingales (2001), "The Firm as a Dedicated Hierarchy", *Quarterly Journal of Economics,* 116, 805-851.

[48] Simon, H. (1951), "A Formal Theory of the Employment Relationship", *Econometrica*, 19, 293-305.

[49] Tadelis, Steven (1999), "What is in a Name? Reputation as a Tradable Assets", *American Economic Review*, 89, 548-63.

[50] Tadelis, Steven (2002), "The Market for Reputations as an Incentive Mechanism", *Journal of Political Economy*, 92, 854-882.

[51] Tadelis, Steven (2003), "Firm Reputations with Hidden information", *Economic Theory*, 21, 635-651.

[52] Townsend, Robert M. (1979), "Optimal Contracts and Competitive Markets with Costly State auditing", *Journal of Economic Theory*, 21, 265-293.

[53] Wang, Tianxi (2007), "the allocation of the liabilities: a two entrepreneurs case", mimeo, London School of Economics.

[54] Wang, Tianxi (2007), "The Reputation of an Organization and Its Dynamics", mimeo, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1000973

[55] Williamson, O. (1971), “The Vertical Integration of Production: Market Failure Considerations”, *American Economic Review*, 61, 112-23.

[56] Williamson, O. (1991), “Comparative Economic Organization: The Analysis of Discrete Structural Alternatives”, *Administrative Science Quarterly*, 36, 269-296.

[57] Williamson, S. D. (1986), "Costly Monitoring, Financial Intermediation, and Equilibrium Credit Rationing, *Journal of Monetary Economics*, 18, 159-179.

[58] Winton, Andrew (1995), "Costly State auditing and Multiple Investors: the Role of Seniority", *Review of Financial Studies*, 8, 91-123.