Essays on Formal Transaction Cost Theory

CD101008 Yusuke Mori

Acknowledgement

This dissertation could not have been completed without the support and help from many people.

Foremost, I am deeply grateful to my advisor Hideshi Itoh. It has been a privilege to receive his invaluable guidance and genuine caring for 6 years since I enrolled in his undergraduate seminar in 2007. He has given me great freedom to pursue what I wanted to explore, always welcomed my rough research ideas, and guided me to improve and organize these ideas into research papers with insightful comments. Furthermore, I am also thankful him for having provided me precious lessons in what it is like to be in academia.

Special thanks to my second advisor Sadao Nagaoka for his beneficial suggestions and encouragement. He has been actively interested in my works and provided me with a lot of opportunities to improve them.

I owe an important debt to wonderful senior colleagues: Akifumi Ishihara, Shintaro Miura, Nozomu Muto, Takeshi Nishimura, Yasuhiro Shirata, and especially Fumitoshi Moriya. Without their kind advice, I would have had a much tougher time in the Ph.D. pursuit.

I am also indebt to Reiko Aoki, Eric Chou, Kohei Daido, Lewis Davis, Makoto Hanazono, Junichiro Ishida, Shinsuke Kambe, Simon Lapointe, Tomoharu Mori, Hodaka Morita, Takeshi Murooka, Masaki Nakabayashi, Naofumi Nakamura, Dan Sasaki, and Takashi Shimizu for their beneficial comments. I also thank the participants at Contract Theory Workshop, Contract Theory Workshop East, the Osaka Workshop on Economics of Institutions and Organizations, ESNIE 2012, 2012 Japanese Economic Association Spring and Autumn Meetings, the 6th Annual Meeting of the Association of Behavioral Economics and Finance, and 6th Annual Organizational Economics Workshop for their helpful comments.

Lastly, I would like to thank my family for all their supports.

Contents

Chapter 1
Introduction

Whether intermediate goods or services should be bought from outside suppliers or made internally is one of the central topics in organizational economics and is called ``make-or-buy decision" or ``boundary of the firm." Since the seminal work of Coase (1937), a number of approaches to the problem have been developed and these are categorized based on their focuses: ex ante incentive or ex post governance. Examples of the former approaches are the property-rights theory (e.g., Hart, 1995) and incentive-system theory (e.g., Holmstrom and Milgrom, 1991). The latter approaches, on the other hand, include transaction cost economics (TCE, e.g., Williamson, 1975, 1985, 1996). While the ex ante approaches have received theoretical attention, the ex post approaches, especially TCE, suffer from the lack of satisfactory formalization.

The aim of this dissertation is to formalize the arguments of TCE. We thus begin by reviewing them.

Transaction Cost Economics

Neoclassical economics traditionally views a firm as ``production function" or ``a technological black box in which inputs are transformed into outputs without reference to organization" (Williamson, 1996, p. 7). Coase (1937), which is recognized as the seminal work on make-or-buy decisions, points out that markets and firms are ``alternative methods of co-ordinating production" (p.388) and employ different coordination mechanisms: while markets use price mechanism, firms use authority or fiat. Coase (1937) also points out that ``there is a cost of using the price mechanism" (p. 390), and hence firms are established to economize such a cost. Coase's (1937) approach is deepened by a series of Oliver Williamson's works, such as Williamson (1975, 1985, 1996), and is named transaction cost economics.

Transaction costs include various kinds of inefficiencies, but TCE mainly focuses on inefficiencies due to ex post inefficient adaptations to unanticipated changes in trade circumstances: transaction costs include ``The ex ante costs of drafting, negotiating, and safeguarding an agreement and, more especially, the ex post costs of maladaptation and adjustment that arise when contract execution is misaligned as a result of gaps, errors, omission, and unanticipated disturbances; the costs of running the economic system" (Williamson, 1996, p. 379). TCE explains why such ex post inefficiencies occur and internal organizations are required from market failure caused by the combination of two behavioral assumptions (i.e., bounded rationality and opportunism) and ``fundamental transformation."

Bounded rationality is defined as ``limited cognitive competence to receive, store, retrieve, and process information" (Williamson, 1996, p. 377). Such a limitation makes writing ex ante complete contract impossible or prohibitively costly. In other words, the terms of the contract contain gaps and ambiguousness due to bounded rationality, and hence ex post modifications or adaptations to unanticipated changes in trade circumstances are required.

Incomplete contracts and resulting adaptations themselves do not cause any problem if each trading party aims to maximize trade surplus. However, each party has a tendency to seek self interest with guile, and hence ex post adaptations become problematic. For example, under non-integration, such ex post adaptation invites dispute over trade value (i.e., haggling). Such ``self-interest seeking with guile" (Williamson, 1996, p. 378) is called opportunism, which includes ``calculated efforts to mislead, deceive, obfuscate, and otherwise confuse" (Williamson, 1996, p. 378).

The size of ex post inefficiencies also depends on how competitive the markets are. If there are a number of possible trading partners in the markets, those who behave opportunistically will be wiped out through competition, and thus haggling becomes less costly. TCE emphasizes that what is crucial to avoid opportunistic behavior is not ex ante but ex post competitiveness. That is, ``what begins as a large numbers supply condition frequently is transformed into a small numbers exchange relation during contract execution and at contract renewal intervals" (Williamson, 1996, p. 26). Such change is called ``fundamental transformation."

Fundamental transformation is typically triggered by relationship-specific assets, which create large value only under specific trade relationship. Once such assets are acquired, it becomes costly for both a buyer and a supplier to terminate the current trade relationship (i.e., bilateral dependency arises) even if there are a large number of possible supplier before the acquisition.

TCE points out that the combination of bounded rationality, opportunism, and fundamental transformation significantly spoils market efficiency. Bounded rationality leads to incomplete contract, which requires ex post adaptation. Such ex post adaptation invites haggling over trade value because trading parties are opportunistic. Such haggling makes each party willing to waste his resources, such as time or money, to improve his bargaining power (i.e., rent seeking) and causes bargaining costs (e.g., delay in reaching agreement). When fundamental transformation and resulting bilateral dependency occur due to relationship-specific assets or other reasons, each party can exercise strong bargaining power over his partner, and hence haggling becomes more costly. TCE asserts that such ex post inefficiencies due to inefficient bargaining (haggling), from which non-integrated trading parties suffer, can be reduced or avoided by vertical integration. It is because integrated firms can implement adaptations not through haggling but by fiat, which is not available in the markets. Nevertheless, it is known that integrated firms suffer from bureaucratic costs, such as incentive degradation and logrolling, and hence all transactions cannot be carried on in a large integrated firm. Since such bureaucratic costs are said to be relatively independent of relationship-specificity (e.g., Riordan and Williamson, 1985), TCE proposes the hypotheses that the higher relationship-specificity, uncertainty, or complexity of the trade in question becomes, the more likely firms are to choose vertical integration. These hypotheses are explained as follows. Higher relationship-specificity makes fundamental transformation more likely to occur and bargaining costs larger. The higher uncertainty or complexity becomes, the more incomplete ex ante contract becomes, and thus the more likely opportunistic behavior is to occur. These hypotheses are supported by a number of empirical studies (see Lafontaine and Slade, 2007 for a survey of these studies). Despite the empirical success, a satisfactory formalization of TCE is yet to be achieved.

Topics We Address

This dissertation aims to formalize the arguments of TCE by developing formal models of ex post dispute over trade value which is invited by the unprogrammed adaptations and is one rationale for vertical integration. Throughout this dissertation, to focus on ex post inefficiencies, we do not examine ex ante inefficiency, including under-investment problems that have been extensively analyzed in the literature on a property-rights theory. See Grossman and Hart (1986), Hart and Moore (1990), and Hart (1995) for the formal models of the property-rights theory.

We will address the following three questions: how rent seeking and bargaining costs interact, why authority mitigates ex post disputes over trade value, and why some firms choose non-integration and integration alternately.

How Do Rent Seeking and Bargaining Costs Interact?

As mentioned above, various kinds of inefficiencies have been identified as transaction costs. However, existing literature on TCE examines these costs separately and is silent about how they interact. In Chapter 2, we focus on two kinds of ex post inefficiencies, namely rent-seeking costs and bargaining costs, and provide a formal model to study them in a unified way.

As mentioned above, rent seeking is resource-wasting activity to improve rent-seeker's bargaining power or payoff. It is known that rent seeking is observed under both non-integration and integration. An example of rent seeking under non-integration is the investments in the outside option which is unlikely to be exercised in equilibrium. Rent seeking under integration includes flattering those who have decision rights in an attempt to influence their decisions in rent seeker's favor. Bargaining costs, such as bargaining delay or breakdown, on the other hand, occur only under non-integration because while non-integrated firms settle ex post dispute over trade value through bargaining (haggling), integrated firms employ fiat to settle the dispute.

We show that ex post inefficient bargaining under non-integration creates a trade-off between rent seeking and bargaining costs: while non-integration suffers from bargaining delay and breakdown, which never occur under integration, it incurs lower rent-seeking costs than integration. This result explains why rent-seeking activities within firms are likely to be more costly than those between firms, and offers a formal justification for the ``costs of bureaucracy" in Williamson (1985).

Why Does Authority Mitigate Disputes over Trade Value?

TCE implicitly assumes that authority within organizations is effective and subordinates always obey their boss's orders. However, it is often pointed out that TCE has not provided any formal justification for the assumption (e.g., Hart, 1995). Chapter 3 thus formally explores why authority within firms helps trading parties immediately settle ex post surplus split despite the possibility of a subordinate's disobedience to the orders of his boss.

To achieve this, we employ three crucial behavioral assumptions: reference-dependent preference, self-serving bias, and shading. Reference-dependent preference reflects the fact that people's assessments of an outcome depend not only on the outcome itself but on its contrast with some yardsticks, which are called reference points (e.g., Kahneman and Tversky, 1979). Self-serving bias is the tendency for individuals to interpret facts in their favor (e.g., Babcock and Loewenstein, 1997). Shading is one interpretation of other-regarding preference and can be considered punishment for unfair treatments (e.g., Hart and Moore, 2008).

We point out that the choice of governance structure affects trading parties' expectations about the outcome of the surplus split, which serve as their reference points, and show that a subordinate is likely to obey orders of his boss because he is expected to do so. Nevertheless, we also point out that such a positive aspect of authority comes with the subordinate's psychological disutility.

Why Do Some Firms Choose Non-Integration and Integration Alternately?

Firms sometimes choose discrete institutional arrangements (e.g., ``non-integration or integration'' and ``centralization or decentralization'') alternately. To explain such wavering behavior, we need to analyze institutional changes dynamically, but some studies point out that TCE focuses on static assignment of transaction-cost-minimizing institution to each transaction, and thus does not fit to address dynamic problems (e.g., Dow, 1987 and Langlois and Robertson, 1995).

In Chapter 4, we address a question why firm boundaries sometimes waver by developing a multi-generation model, in which each generation chooses either non-integration or integration without knowing the reasons for predecessors' choices.

We show that under the assumption of unrecorded reasons for predecessors' governance choices, each generation's experimentation causes wavering between non-integration and integration in equilibrium. Given that the reason for each generation's choice of governance structure may not be transferred between generations, the level of relationship-specificity, from which each generation infers which governance structure is optimal, plays an important role. If the level of relationship-specificity is high (resp. low) enough, non-integration is likely to be more (resp. less) costly than integration, and hence each generation (if rational) chooses integration (resp. non-integration), which achieves actual transaction-cost minimization with high probability. However, if the level of relationship-specificity is intermediate, the governance choice which expectedly minimizes transaction costs is likely to fail in actual transaction-cost minimization. Hence, an effort to find out which governance structure is optimal is required and leads to wavering between non-integration and integration. Our model provides formal explanations for why organizational changes often follow management turnovers and why it is hard for some integrated firms to disintegrate even if integration is not optimal.

The Organization of This Dissertation

This dissertation proceeds as follows. Chapters 2, 3, and 4 address the topics, ``how rent seeking and bargaining costs interact," ``why authority mitigates disputes over trade value," and ``why some firms choose non-integration and integration alternately," respectively. Chapter 5 contains concluding comments: summaries of our results and brief discussion on what we left for future research, such as inalienable relationship-specific investments and hybrid governance structures.

Chapter 2

A Formal Theory of Firm Boundaries: A Trade-Off between Rent Seeking and Bargaining Costs

1 Introduction

This chapter focuses on two sources of ex post inefficiencies (transaction costs), namely haggling (inefficient bargaining) between firms over trade value and influence activity within firms. Haggling between firms over the value is said to cause rent-seeking costs and bargaining costs due to private information. Rent seeking does not create any value, but improves the rent-seeker's bargaining power or share of surplus at the cost of precious resources (e.g., securing competent lawyers in case of litigation). If each trading party has private information (e.g., whether each party is rational or obstinate), he then has an incentive to use such information to realize individual advantage, which can lead to bargaining costs (bargaining delay or breakdowns). It is also known that rent seeking can be observed within firms: influence activity (e.g., Milgrom and Roberts, 1988). More specifically, members of an internal organization have incentives to influence the decisions of those who have decision rights in their favor at the cost of their resources (e.g., flattering the boss).

Some existing theoretical literature (reviewed in the next section) studies these inefficiencies (i.e., rent-seeking costs and the bargaining costs) separately. We contribute to this literature by providing a formal TCE model in which they are dealt with simultaneously.

In our theory, following the arguments of TCE, processes of ex post value split differ between non-integration and integration. Under non-integration, trading parties engage in bilateral bargaining, and if the bargaining is terminated without agreement, litigation takes place (a court decides how to divide trade value). Under integration, on the other hand, a third party who has authority (i.e., a boss) determines the division of the value, and thus there is no bargaining.[1] We assume that decisions of third parties (the court and the boss) are affected by each party's rent seeking. The parties are thus eager to undertake rent seeking so as to improve their payoffs, which causes rent-seeking costs.

Furthermore, the parties are assumed to have private information about their types, which are either rational or obstinate (irrational). The obstinate type always demands a large specific share of the value denoted by θ, accepts any offer greater than or equal to that share, and rejects all smaller offers. The rational type then has an incentive to mimic the obstinate type in an attempt to obtain a larger share of the value, which leads to bargaining costs.

Our theory points out an important trade-off between rent seeking and bargaining costs: ex post inefficient bargaining, which takes place only under non-integration, can cause bargaining costs, which never occur under integration, but lowers each party's rent-seeking incentive. There are two reasons why rent-seeking incentives are lower under non-integration than under integration. First, rent seeking between firms indirectly affects rent-seekers' payoffs by improving their threat points (their expected litigation payoffs), while rent seeking within firms (influence activity) affects payoffs directly. Thus, when the aggregate litigation payoff must be smaller than the original trade value (e.g., because of time-consuming litigation), the parties' incentives for rent seeking under non-integration become smaller than those under integration. Second, the bargaining provides parties with opportunities to concede (i.e., to let their partners obtain a large share of the value). When each party becomes obstinate with high probability, any behavior other than concession is likely to delay agreement, and hence the rational type can optimally concede. Since concession terminates the game, in which

[1] Chapter 3 offers a formal explanation as to why integrated firms can avoid costly ex post bargaining by employing behavioral assumptions.

case no litigation takes place, the rational type, expecting this outcome, chooses a low level of rent seeking.

Our results explain why rent seeking within firms (influence activity) is likely to be more costly than rent seeking between firms, and provide a formal justification for the ``costs of bureaucracy" in Williamson (1985). Williamson (1985, pp. 151-152) reasons that internal operating is more subject to politicization due to the tendency of internal organizations toward reciprocity between their members, which can lead to managerial back-scratching. Our results are consistent with his argument.[2]

The rest of this chapter is organized as follows. Section 2 relates our theory to existing literature. In Section 3, we present two simple models that focus on rent-seeking costs and highlight why rent seeking between firms is likely to be less costly than rent seeking within firms (influence activity). In Section 4, by constructing a more general model, we examine both rent-seeking costs and bargaining costs, show the trade-off between them, and discuss some extensions. Section 5 contains concluding comments.

2 Related Literature

This chapter studies ex post inefficiencies by combining the rent-seeking model and the non-cooperative bargaining model in the bargaining and reputation literature. We then review, in order, the literature on rent seeking both between and within firms, bargaining and reputation, bargaining with endogenous outside options, and ex post inefficiencies.

Rent Seeking: Tullock (1980) develops a basic model of rent seeking in the context of lottery purchase, which Gibbons (2005) extends to study firm boundaries (i.e., to analyze haggling). In Gibbons (2005), two symmetric parties undertake rent seeking,

---

[2] The aim of this chapter is to formalize Williamson's informal arguments on firm boundaries and we do not intend to assert that rent seeking under non-integration is always less costly than rent seeking under integration. In fact, this remark on rent-seeking costs between and within firms needs to be examined empirically. However, to our knowledge, there is no empirical study that compares rent-seeking costs before and after integration and we still await one.

each hoping to obtain a larger portion of trade value. Gibbons shows that larger trade value makes non-integration more costly, which is consistent with the assertion of TCE.

Milgrom and Roberts (1988) and Meyer, Milgrom, and Roberts (1992) develop formal models of influence activity. In these studies, a principal requires information that is valuable for efficient decision making but is possessed by agents; this information asymmetry provides agents with incentives to manipulate the information in order to influence the decision in their favor. Milgrom and Roberts (1988) examine how organizational design (structures and policies) should respond to these incentives in the context of job assignment. Meyer, Milgrom, and Roberts (1992) explain why divestitures of divisions with poor growth prospects are more common.

We apply Tullock's rent-seeking model to rent seeking both between and within firms. Some readers might think that the boss in our model is unreasonably naive in the sense that he never ignores employees' influence activities (i.e., he never forms any institutional arrangement to avoid influence activities). However, applying Tullock's model to rent seeking both between and within firms is reasonable for three reasons. First, as Meyer, Milgrom, and Roberts (1992) discuss, influence activity is the private sector analog of rent seeking. Second, in our theory, the boss's decision only determines the division of fixed-size trade value and does not affect ex post efficiency (the size of the value), and hence he has no incentive to introduce an arrangement to prevent rent seeking. Lastly, and most importantly, in this setting, we can deal with rent-seeking costs both between and within firms in a unified and comparable way, which is consistent with the following statement by Williamson (1996, p. 228): ``One of the tasks of transaction cost economics is to assess purported bureaucratic failures in comparative institutional terms."

Bargaining and Reputation: To examine bargaining costs due to private information (each party's type), we borrow the setting and results from Abreu and Gul (2000) and Compte and Jehiel (2002). Abreu and Gul (2000) analyze a bargaining game with two-sided player-type uncertainty. More specifically, they introduce the obstinate ``irrational type,'' who always demands a fixed share θ, accepts any offer greater than or equal to that share, and rejects all smaller offers. They show that the presence of such an irrational type provides rational type with an incentive to build a reputation for obstinacy, which leads to bargaining delay.

Compte and Jehiel (2002) introduce exogenous outside options into Abreu and Gul's (2000) model. They show that when players have access to stationary outside options that yield shares larger than $1-\theta$, these outside options may cancel out the effect of obstinacy; that is, each player reveals himself as rational as soon as possible.

We adopt the symmetric version of their approaches and results to examine bargaining delay and breakdown due to private information. Nevertheless, as we will show in Section 4.5, our results hold under an asymmetric setting.

Endogenous Outside Option: As we will show in the following sections, decisions of the third parties (the court under non-integration and the boss under integration) endogenously determine trading parties' outside options. While we assume that the parties' outside options are determined by their rent seeking, there are several other approaches.

Atakan and Ekmekci (2010) and Özyurt (2010) develop the bargaining game in a searching market, which serves as an endogenous outside option. Unlike them, we consider a situation in which the parties are locked in and cannot search for other possible partners.

Lee and Liu (2010) assume that if parties cannot reach agreement in voluntary bargaining, a third party is called upon to determine how much one party pays to the other. While the third party in their model is unbiased, the court and the boss in our models can be biased (their decision is affected by rent seeking).

Ex Post Inefficiencies: Some studies have focused on ex post inefficiencies using approaches other than TCE, including the property-rights theory and the ``contracts as reference points'' approach. However, few efforts to formalize the arguments of TCE can be found.

Matouschek (2004) analyzes the optimal ownership structure that minimizes ex post inefficiency due to too much or too little trade. He develops a formal model following the property-rights theory, in which disagreement payoffs depend on the ownership structure, and shows the following results. When the expected gain from trade is large (resp. small) relative to the aggregate disagreement payoff, joint ownership (resp. either non-integration or integration) that minimizes (resp. maximizes) the aggregate disagreement payoff is optimal. His results follow from the fact that the smaller the aggregate disagreement payoff becomes, the more likely the players realize trade (including inefficient trade). In contrast to Matouschek (2004), we assume that the aggregate disagreement payoff is zero irrespective of governance structures, and hence ownership structure has no effect on the aggregate disagreement payoff. In our models, the choice of governance structure only affects how ex post value split is implemented (i.e., the way in which the trade value is distributed).

Hart and Moore (2008) and Hart (2009) develop the ``contracts as reference points'' approach to analyze inefficiencies due to ex post adaptation and present implications for firm boundaries. In their studies, a contract negotiated under ex ante competitive conditions provides players with reference points for ex post entitlement. More specifically, each player interprets the contract in a way that is most favorable to him. When he does not obtain the most favored outcome within the contract, he engages in shading (that is, he performs in a perfunctory fashion, which reduces his partner's payoff). This setting leads to the following trade-off: the more flexible the ex ante contract becomes, the easier the ex post adaptation will be, but the more likely it is that shading will take place. Hart and Moore (2008) explain why employment contracts can be optimal, and Hart (2009) examines how incentives to engage in hold-up can be reduced. Both their studies and ours are concerned with how ex post efficiencies affect firm boundaries. However, while they focus on the inefficiencies that occur after contract renegotiation (i.e., shadings), we focus on the inefficiencies that arise during renegotiation (i.e., rent-seeking costs and bargaining costs).

Bajari and Tadelis (2001) focus on construction procurement and compare two forms of contracts: fixed-price contracts, which they interpret as market transactions, and cost-plus contracts, which can be considered integration. They show a trade-off between cost-reducing efforts and ex post inefficiencies due to maladaptation. That is, while fixed-price contracts lead to high seller incentive for cost-reducing efforts, their inflexibility prevents efficient adaptation. Tadelis (2002) extends their model to address firm boundaries and show that more complex products are more likely to be internally procured under low cost-reducing incentives, while more simple products are more likely to be procured through the market under high cost-reducing incentives. Unlike these papers, we do not focus on ex ante incentives, and analyze bargaining costs rather than maladaptations as ex post inefficiencies.

Wernerfelt (2011) examines efficient mechanisms for labor procurement and points out a trade-off between specialization and bargaining cost (cost of information gathering). In his model, each buyer needs a sequence of different tasks, each of which can be supplied by any seller. In his market mechanism, while a buyer can hire the most suitable seller to each task, duplicated investment occurs because each seller has to incur some buyer-specific costs. In bilateral relationships (sequential contracting and employment), on the other hand, since the relationship between a buyer and a seller is fixed, the seller incurs the specific investment only once, but the buyer cannot hire the most suitable seller for each task. Furthermore, while trading parties are matched up through auctions in the market, they are randomly matched up in the bilateral relationships, which makes parties in the bilateral relationships willing to pay to observe their partners' private information (e.g., the buyer's value and the seller's cost). There are some differences between his study and ours. In his model, duplicated specific investment is the only downside of the market and bargaining costs are those of information gathering. On the other hand, we do not deal with ex ante investment and bargaining costs are delay in reaching agreement and bargaining breakdown. Furthermore, while Wernerfelt (2011) does not necessarily deal with bilateral monopoly (i.e., bilateral monopoly does not arise in market mechanism), we focus on transactions between firms and within a firm under bilateral monopoly due to relationship-specific investment.

Zhu (2009) attempts to develop a formal model of TCE and compares spot contracting, long-term contracting, and vertical integration, focusing on ex ante specific investment, productive action, and asset maintenance as well as bargaining friction. While both his model and ours deal with bargaining delay, the sources of the delay are different. In Zhu (2009), bargaining delay stems from the strategic choice of the timing of a contract offer and random delay in offer transmission. In this chapter, on the other hand, delay is caused by the opportunistic use of private information and there is no random delay.

3 The Model

This section introduces two simple models which explain why rent seeking under non-integration is less costly than rent seeking under integration (influence activity). There are two factors which lead to rent-seeking reduction under non-integration. One model points out that rent seeking between firms affects each party's payoff less directly than rent seeking within firms, and the other shows that only non-integration provides an opportunity for each party to concede (i.e., to let his partner obtain a large share of trade value). While this section deals with rent-seeking costs only and examines each factor separately, the next section analyzes both rent-seeking costs and bargaining costs and focuses on both factors by introducing a more general framework (the third model). For explanatory convenience, we call the model introduced in Section 3.1 (resp. Section 3.2) to examine the first (resp. second) factor Model 1 (resp. Model 2) and the general model presented in the next section Model 3.

3.1 Model 1: Indirect Effects of Rent Seeking between Firms on Payoffs

In this subsection, we point out that rent seeking under non-integration is less costly than rent seeking under integration (influence activity) because the former affects each party's payoff less directly than the latter.

There are two risk-neutral symmetric trading parties (parties 1 and 2) who are locked in due to relationship-specific investment (there is no other possible trading partner). (An asymmetric case will be discussed in Section 4.5.) These parties engage in ex post division of trade value V. Such an ex post value split is required because ex ante contract cannot be complete due to bounded rationality or other reasons.

Note that we focus on ex post inefficiencies, and thus assume that there is no ex ante inefficiency such as under-investment problems, which have been extensively analyzed in the literature on the property-rights theory. Specifically, we assume that the relationship-specific investment has been efficiently sunk and our theory does not include ex ante investment stage.

The game proceeds as follows. First, a governance structure is chosen (whether to integrate or not) to minimize ex post inefficiencies. Second, the parties simultaneously choose their levels of rent seeking, and a value split is then initiated. After the value split, the trade occurs. Figure 2.1 summarizes how the value V is divided between the parties under each governance structure.[3]

The processes of the value split depend on the governance structure chosen at the beginning. Under non-integration, the parties engage in bilateral bargaining; if the bargaining is terminated without agreement, litigation takes place (i.e., a court decides how to divide the value). If disagreement occurs, the aggregate litigation payoff shrinks to δV where $\delta \in (0,1)$ denotes a common discount factor.[4] Intuitively, litigation requires cumbersome processes that block immediate settlement. Nevertheless, in Model 1, we assume that the parties agree to the Nash bargaining solution, and hence ex post bargaining that takes place only under non-integration is efficient and there is no litigation. Under integration, on the other hand, there is no bargaining between the

[3] Figure 2.1 is based on Figure 1 in Tadelis and Williamson (2012).

[4] We can instead assume that the aggregate litigation payoff is $V-K$ $(K>0)$ without changing our main result.

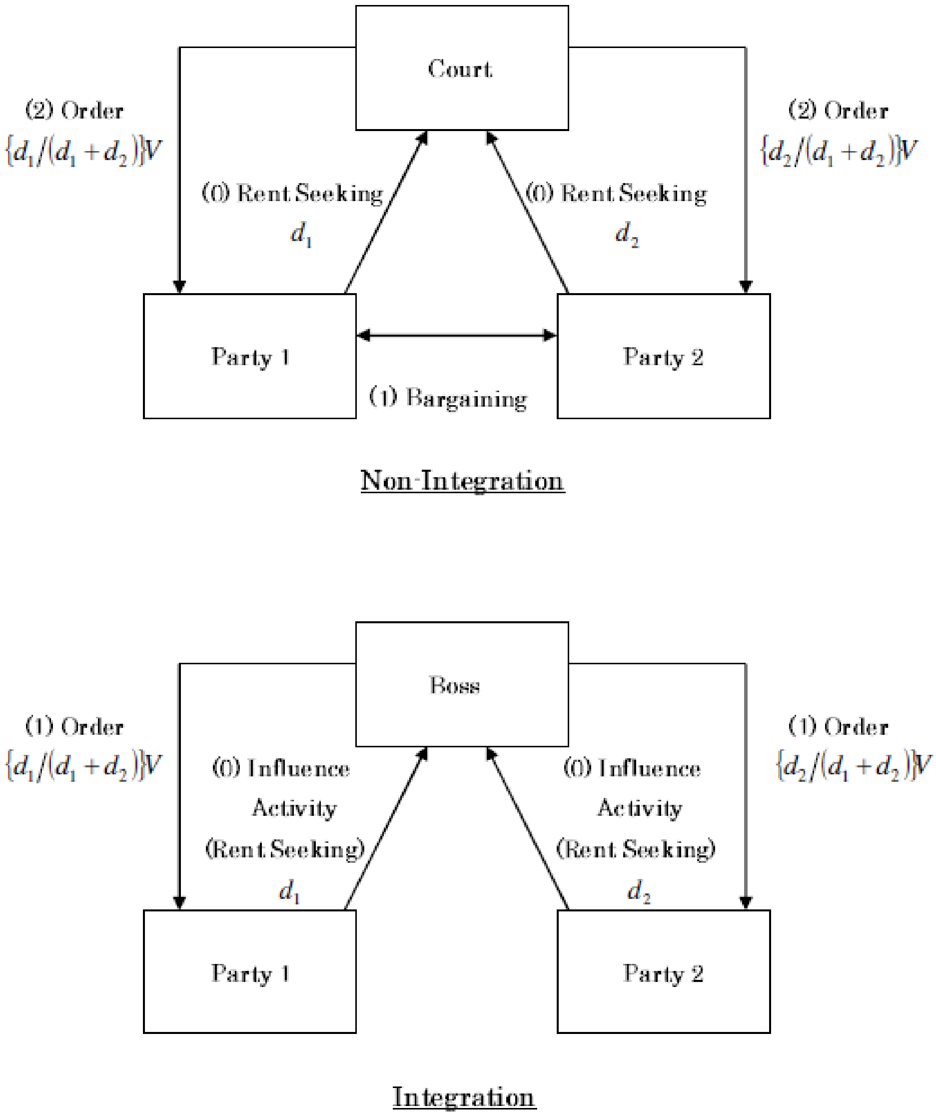

Figure 2.1: Processes to Divide Trade Value V

Under non-integration, court ordering is required only if the bargaining is terminated without agreement.

parties, and the division of V is determined by the third party who has authority (the boss). For a formal justification for the assumption that integrated firms can avoid costly bargaining, see Chapter 3.

The value split by the third party (the court's or the boss's decision making) is assumed to be affected by each party's rent seeking. Such rent seeking includes securing competent lawyers to obtain an advantage over the other party in litigation and flattering the boss. Under non-integration, party i's rent seeking increases his bargaining power by raising his expected litigation payoff, which serves as his endogenous outside option.[5] Under integration, on the other hand, party i's rent seeking increases his expected share of V by influencing the decision of the boss, and hence we interpret it as influence activity according to Milgrom and Roberts (1988).

We formalize rent seeking both between and within firms by employing Tullock's (1980) rent-seeking model. $d_i \in \Re_+$ denotes the level of party i's rent seeking ($i=1,2$) and is unobservable to the trading partner. When party i (resp. party j) provides the level of rent seeking d_i (resp. d_j), a third party distributes a share $d_i/(d_i+d_j)$ to him.[6] If neither party provides rent seeking ($d_1 = d_2 = 0$), each party receives half of V. Party i incurs rent-seeking cost $C(d_i) = kd_i$ where k is a positive constant.

We then examine each party's optimal rent-seeking level under non-integration. Party i can improve his payoff by increasing his threat point payoff (his litigation payoff), which is increasing in his rent-seeking level, d_i. Hence, i's optimal rent-seeking level d_i^* solves the following problem:

$$\max_{d_i} \frac{d_i}{d_i + d_j}\delta V + \frac{1}{2}(1-\delta)V - kd_i.$$

Note that the parties agree to the Nash bargaining solution. The first term represents i's threat point payoff, the second term denotes his share of the remaining surplus $(1-\delta)V$, and the last term is his rent-seeking cost. From symmetry assumption, we obtain $d_1^* = d_2^* = \delta V / 4k \equiv d^*$.

Under integration, on the other hand, the parties affect their payoffs by undertaking

5 It is worth noting that our result continues to hold even if rent seeking is undertaken after bargaining breaks down. It follows because disagreement never occurs and rent seeking is completely avoided under non-integration. Since this discussion is somewhat trivial, we do not deal with this case.

6 Note that the parties choose who is to be rent-sought by choosing governance structure (the court or the boss). In our models, the third parties are not players of the game, and hence we ignore their welfare.

influence activities. Let d_i^{**} denote party i's optimal influence level. d_i^{**} thus solves the following problem:

$$\max_{d_i} \frac{d_i}{d_i + d_j} V - k d_i.$$

We then find that $d_1^{**} = d_2^{**} = V / 4k \equiv d^{**} (> d^*)$. Since rent-seeking cost $C(d) = kd$ is increasing in d, we can determine that non-integration incurs lower aggregate rent-seeking cost than integration (i.e., $2C(d^*) < 2C(d^{**})$).

Model 1 presents the following observation: rent seeking between firms indirectly affects a rent seeker's payoff by increasing his threat point payoff. Thus, when the value V shrinks due to litigation, each party's incentive to provide rent seeking under non-integration becomes smaller than rent seeking under integration. In other words, when the aggregate threat point payoff must be smaller than the original V, non-integration can feature lower rent-seeking costs than integration.

3.2 Model 2: Opportunities to Concede

In the last subsection, we pointed out that indirect effect of rent seeking between firms on rent seeker's payoff makes each party less eager to engage in rent-seeking activities. Nevertheless, if $\delta = 1$ holds, the result of Model 1 fails: if litigation triggers no shrinkage in the trade value, the choice of governance structure does not affect rent-seeking costs.

This subsection introduces the second model (Model 2) and shows that the presence of private information (each party's type) makes each party less willing to engage in rent seeking under non-integration even if $\delta = 1$ holds. This result stems from the fact that ex post inefficient bargaining, which occurs only under non-integration, provides each party with an opportunity to concede (i.e., to let his partner obtain a large share of the trade value). When each party becomes obstinate with high probability, any behavior other than concession is likely to delay agreement, and hence the rational type can optimally concede. Since concession leads to no litigation, the rational type, expecting this outcome, chooses a low level of rent seeking.

There are some differences between Models 1 and 2: bargaining procedure and cost of delay. First, while the parties agree to the Nash bargaining solution in Model 1, the ex post bargaining in Model 2 is assumed to be a take-it-or-leave-it offer game. More specifically, in Model 2, either party sends an offer $x \in (0,1)$, which denotes his demanded share of the value V, and the other party decides whether to accept it.[7] The right to make the offer is assigned to each party with equal probability at the beginning of the bargaining stage. If they reach agreement, the game ends. Otherwise, litigation takes place. Second, unlike Model 1, we assume that there is no cost of delay (i.e., disagreement does not prevent immediate settlement: $\delta = 1$ holds) in Model 2. As mentioned above, in Model 1, if $\delta = 1$ holds, the result fails. Thus, Model 2 is more than just an extension of Model 1 to non-cooperative bargaining and offers a completely different insight into how non-integration economizes rent-seeking costs.

Furthermore, in Model 2, to focus on the effect of private information on each party's rent-seeking behavior, we assume that the parties may be obstinate with probability $\varepsilon \in (0,1)$ (and rational with probability $1-\varepsilon$). This probability of being obstinate is common knowledge. The obstinate type always demands a share $\theta(> 1/(1+\delta))$ for himself and never accepts any offer or the division specified by the third party unless he can obtain at least θ of V.[8] The rational type, on the other hand, accepts any division larger than or equal to 0, but can strategically mimic the obstinate type. (Since the parties do not have time to build a reputation for obstinacy, reputation effect plays little role in this model. We will deal with the reputation effect in Model 3.) As mentioned previously, the parties are symmetric, and hence share the parameters θ and ε.

The parties are uncertain about their own types before the value split is initiated.

---

[7] We refer to the proposer as ``he'' and the responder as ``she'' for the purpose of identification only.

[8] Existing literature typically assumes that θ is larger than the equilibrium share of a complete information Rubinstein offers game. $1/(1+\delta)$ is the equilibrium share of an infinite-horizon, symmetric offers game. Although Model 2 deals with one-period bargaining, the assumption $\theta > 1/(1+\delta)$ does not affect our main result.

That is, they behave rationally in the rent-seeking stage, although with probability ε they can be obstinate in the stages following the rent-seeking stage (e.g., the bargaining stage). Intuitively, once each party faces his opponent (i.e., his partner), he can lose control of himself.[9]

We adopt the same setting for rent seeking as in Model 1 and focus on symmetric rent-seeking equilibrium. Given symmetric rent-seeking behavior, the third party (the court or the boss) determines the equal division of the value V, and hence the rational type obtains the expected payoff $(1-\varepsilon)V/2$ from the third party's division. Note that the obstinate type rejects the division specified by the third party and terminates the relationship because $1/2<\theta$.

In order to show our result clearly, we make the following assumption in this subsection:

$$2\theta - 1 \leq \varepsilon < \theta. \tag{1}$$

The first inequality implies that $(1-\theta)V \geq (1-\varepsilon)V/2$, which means the rational responder prefers to accept the offer $x=\theta$ rather than reject it, given that both parties choose the same rent-seeking level. By the second inequality, which can be rewritten as $1-\theta<1-\varepsilon$, the parties prefer litigation to concession if they can obtain the whole value V in litigation against their rational partners. ``Concession" means a party either accepts $(1-\theta)V$ for herself or offers $x=1-\theta$.

We begin in Section 3.2.1 by specifying each party's optimal offer and acceptance decision in the bargaining stage. Section 3.2.2 then determines each party's optimal rent-seeking level, given the optimal behavior in the bargaining stage. In Section 3.2.3, we show the result that non-integration features lower rent-seeking costs than

[9] For the case in which the obstinate type is assumed to behave obstinately throughout the game (e.g., the obstinate type chooses irrationally high rent-seeking level which the rational type cannot match), see Section 4.5.

integration and explain its intuition.

3.2.1 The Bargaining Stage

We here examine the bargaining stage, which takes place only under non-integration. Since the obstinate type behaves mechanically in the bargaining stage, we must only specify the behavior of the rational type. Furthermore, for simplicity, we focus on pure strategies and do not consider mixed strategies. There are two cases to be analyzed separately.

Case 1. We first analyze the case in which the rational proposer concedes even if his rational partner concedes; that is, the case in which the following condition holds:

$$(1-\theta)V \geq (1-\varepsilon)\theta V. \tag{2}$$

The right-hand side of the condition is the proposer's expected payoff when he mimics the obstinate type (offers $x=\theta$) and his rational partner concedes. Intuitively, when ε is sufficiently high, his inflexible offer $x=\theta$ is likely to be rejected and lead to trade termination. Hence, even though $x=\theta$ is accepted by the rational responder, the rational proposer voluntarily concedes.

We then study the acceptance decision by the rational responder. The rational responder accepts the offer $x=\theta$ because $x=\theta$ means the proposer is obstinate given the equilibrium offer of the rational proposer. Any offer other than $x=\theta$ reveals the proposer as rational, and thus the rational responder obtains $(1/2)V$ in litigation. Hence, the rational responder accepts any offer $x \leq 1/2$ and $x=\theta$ and rejects any offer $x > 1/2$ and $x \neq \theta$.

Case 2. Suppose condition (2) does not hold. The rational proposer then optimally offers $x=\theta$. The acceptance decision by the rational responder, on the other hand, is the same as in Case 1 because condition (1) holds, namely $(1-\theta)V \geq (1-\varepsilon)V/2$. That is, she accepts any offer $x \leq 1/2$ and $x=\theta$ and rejects any offer $x > 1/2$ and $x \neq \theta$.

3.2.2 The Rent-Seeking Stage

Non-Integration

We now determine each party's optimal rent-seeking level in Cases 1 and 2 given the behavior in the bargaining stage specified above. Note that both parties choose their rent-seeking levels rationally in the situation in which each party receives the right to make an offer with equal probability and becomes obstinate with probability ε in the bargaining stage. As mentioned above, we focus on symmetric rent-seeking equilibrium.

Section 3.2.1 implies that the game ends with either concession by the rational type or termination by the obstinate type. However, this does not imply that the parties have no incentive to undertake rent seeking. Suppose party i undertakes small but positive rent seeking but party j does not. When party i is the proposer in the bargaining, i offers $x=1$ because i prefers litigation (to obtain the whole V) to concession from condition (1), $1-\varepsilon > 1-\theta$. When party i becomes the responder in the bargaining stage, on the other hand, i rejects j's offer $x=1-\theta$ because it reveals party j as rational and hence party i can obtain the whole value V in litigation.

Case 1: Let d_1^* represent the optimal rent-seeking level in Case 1. The equilibrium payoff to party i, denoted by u_i, is then given by

$$
\begin{aligned}
u_i =& \frac{1}{2}[(1-\varepsilon)(1-\theta)V + \varepsilon(1-\varepsilon)\theta V] \\
&+ \frac{1}{2}[(1-\varepsilon)\{(1-\varepsilon)\theta V + \varepsilon(1-\theta)V\} + \varepsilon(1-\varepsilon)\theta V] - kd_1^*.
\end{aligned}
$$

The first line (resp. second line) represents i's expected payoff when i is the proposer (resp. the responder) given that each party can be obstinate with probability ε in the bargaining.

In Case 1, there are two possible deviations: (i) a party chooses high rent-seeking level and triggers litigation (i.e., offers $x=1$) if he becomes the rational proposer in the bargaining stage or (ii) a party provides high rent-seeking level, rejects the rational proposer's equilibrium offer $x=1-\theta$, and goes to court when she becomes the rational responder. Let $d_{(i)}$ (resp. $d_{(ii)}$) denote the rent-seeking level that prevents deviation

(i) (resp. deviation (ii)). Since the parties are uncertain whether they will be the proposer or the responder in the bargaining stage, they choose the rent-seeking level that prevents the deviations, no matter what role they play in the bargaining. That is, each party provides $d_1^* = \max\left[d_{(i)}, d_{(ii)}\right]$. We can easily determine that $d_{(i)} > d_{(ii)}$, since the smaller the payoff party i wants party j to accept, the more i has to undertake rent seeking to prevent j's deviation. We thus find that $d_1^* = d_{(i)}$ and both deviations are prevented.

Consider rational party i's deviation (i): i chooses $d_1^* + e^*$ instead of d_1^* and, if i becomes the rational proposer, offers $x = 1$ to trigger litigation. Such e^* solves

$$\max_{e} \frac{1}{2}(1-\varepsilon)\frac{(1-\varepsilon)(d_1^* + e)}{d_1^* + (d_1^* + e)}V - k(d_1^* + e).$$

Note that this deviation occurs when party i is the rational proposer, which occurs with probability $(1-\varepsilon)/2$, and the trade is not terminated when i's partner is rational, which occurs with probability $1-\varepsilon$. If i deviates, i's expected payoff u_i' is given by

$$\begin{aligned} u_i' =& \frac{1}{2}\left[(1-\varepsilon)\frac{(1-\varepsilon)(d_1^* + e^*)}{d_1^* + (d_1^* + e^*)}V + \varepsilon(1-\varepsilon)\theta V\right] \\ &+ \frac{1}{2}[(1-\varepsilon)\{(1-\varepsilon)\theta V + \varepsilon(1-\theta)V\} + \varepsilon(1-\varepsilon)\theta V] - k(d_1^* + e^*). \end{aligned}$$

In order to prevent such a deviation, d_1^* must keep party i indifferent about whether to deviate in the situation in which i is uncertain about his type and role in the bargaining.[10] That is, d_1^* satisfies $u_i = u_i'$. We thus obtain

$$d_1^* = \frac{(1-\varepsilon)\left\{\sqrt{1-\varepsilon} - \sqrt{1-2\theta+\varepsilon}\right\}^2}{8k}V.$$

Case 2: We next derive each party's optimal rent-seeking level in Case 2, d_2^*. The expected equilibrium payoff to each party j is given by:

10 We assume that in such a situation, the parties choose not to deviate.

$$u_j = \frac{1}{2}[(1-\varepsilon)^2\theta V + \varepsilon(1-\varepsilon)\theta V] + \frac{1}{2}(1-\varepsilon)(1-\theta)V - kd_2^*.$$

The first term (resp. second term) represents j's expected payoff when j is the proposer (resp. the responder). Note that each party can be obstinate with probability ε and becomes the proposer with equal probability in the bargaining.

As in Case 1, there are two possible deviations: (iii) a party chooses high rent-seeking level and triggers litigation ($x=1$) when he becomes the rational proposer or (iv) a party provides high rent-seeking level and rejects the proposer's equilibrium offer $x=\theta$ if she becomes the rational responder. Let $d_{(iii)}$ (resp. $d_{(iv)}$) denote the rent-seeking level that prevents deviation (iii) (resp. deviation (iv)). Since the equilibrium payoff of the rational responder is smaller than that of the rational proposer (i.e., condition (2) does not hold), we obtain $d_{(iv)} > d_{(iii)}$. The parties thus choose $d_2^* = d_{(iv)}$ to prevent both deviations no matter what role they play in the bargaining.

Consider party j's deviation (iv): j chooses $d_2^* + e'$ and, if j becomes the rational responder, rejects $x=\theta$ to trigger litigation, where e' solves

$$\max_e \frac{1}{2}(1-\varepsilon)\frac{(1-\varepsilon)(d_2^*+e)}{d_2^*+(d_2^*+e)}V - k(d_2^*+e).$$

Note that the deviation occurs when party j becomes the rational responder with probability $(1-\varepsilon)/2$ and the probability with which the trade is not terminated (namely, j's partner is rational) is $1-\varepsilon$. j's expected payoff from deviation, defined as u_j', is given by

$$\begin{aligned} u_j' =& \frac{1}{2}\left[(1-\varepsilon)^2\theta V + \varepsilon(1-\varepsilon)\theta V\right] \\ &+ \frac{1}{2}(1-\varepsilon)\frac{(1-\varepsilon)(d_2^*+e')}{d_2^*+(d_2^*+e')}V - k(d_2^*+e'). \end{aligned}$$

As in Case 1, d_2^* must satisfy $u_j = u_j'$, and thus

$$d_2^* = \frac{(1-\varepsilon)\left\{\sqrt{1-\varepsilon}-\sqrt{1-2\theta+\varepsilon}\right\}^2}{8k}V = d_1^*.$$

Integration

Since there is no bargaining under integration, the parties only undertake influence activities to improve the final division in their favor. That is, party i solves the following problem:

$$\max_{d_i} \frac{(1-\varepsilon)^2 d_i}{d_i + d_j} V - k d_i.$$

In equilibrium, the boss distributes V equally to each party and the obstinate type terminates the relationship, and thus the trade takes place if both parties are rational, which occurs with probability $(1-\varepsilon)^2$. We then find that each party chooses d_I^*:

$$d_I^* = \frac{(1-\varepsilon)^2}{4k} V.$$

3.2.3 Rent-Seeking Reduction under Non-Integration

We can determine that $d_I^* > d_1^*$ and $d_I^* > d_2^*$. Since $C(d) = kd$ is increasing in d, this implies that integration features higher rent-seeking costs than non-integration (i.e., $C(d_I^*) > C(d_1^*)$ and $C(d_I^*) > C(d_2^*)$). This result stems from the presence of ex post inefficient bargaining. That is, the bargaining stage provides the parties with opportunities to concede.

The intuition of the result is as follows. When the parties are obstinate with high probability (ε is high), the rational type's litigation payoff $(1-\varepsilon)V/2$ is likely to be smaller than the concession payoff $(1-\theta)V$. Given that ε is high, the rational type thus prefers to concede rather than behave obstinately. Since concession terminates the game and litigation never takes place, the parties, expecting this outcome, choose low rent-seeking levels. As discussed, under non-integration, the parties provide the minimum rent-seeking level, which prevents their partners' deviations (if rational) no matter what roles they play in the bargaining.

3.3 Interim Summary

In this section, we presented two reasons why rent seeking between firms is less prevalent than rent seeking within firms (influence activity). First, rent seeking between firms affects the parties' payoffs indirectly, while rent seeking within firms (influence activity) affects them directly. Second, ex post bargaining, which occurs only under non-integration, provides the parties with opportunities to concede.

The analyses in this section offer some important implications for the theory of firm boundaries. First, larger trade value V makes both non-integration and integration more costly. Models 1 and 2 showed that rent-seeking costs under non-integration costs are increasing in the size of V. This corresponds to the main prediction of TCE: larger trade value makes non-integration more costly. Furthermore, we can show that influence costs are also increasing in V. This observation is consistent with Williamson (1973), who argues, ``Substantially the same factors that are ultimately responsible for market failures also explain failures of internal organization" (p. 316).

Second, rent seeking within firms (influence activity) is likely to be more costly than rent seeking between firms. As discussed above, rent-seeking costs under integration (influence costs) are always higher than rent-seeking costs under non-integration. This result offers a formal justification for the ``costs of bureaucracy" in Williamson (1985, Chapter 6). Williamson (1985) submits that internal operating is more subject to politicization, which means internal organizations tend more toward reciprocity than the market. Such reciprocity can include reciprocal managerial back-scratching, which makes integration more costly. Our result is consistent with his argument.

In this section, we analyzed two factors, which make rent seeking under non-integration less costly than rent seeking under integration, separately to show their effects starkly. The next section presents Model 3, in which both factors are at work and not only rent seeking but also bargaining costs (delay and breakdown) affect firm boundaries.

4 The Trade-off between Rent Seeking and Bargaining Costs

This section presents a general model (Model 3) in which (i) both of the previously discussed factors leading to more prevalent rent seeking within firms co-exist and (ii) bargaining costs (delay and breakdown) are introduced. We show that there is a trade-off between rent seeking and bargaining costs by applying the results of Abreu and Gul (2000) and Compte and Jehiel (2002).

In Model 3, unlike Models 1 and 2, the bargaining stage is assumed to be an infinite-horizon, alternating-offers bargaining game with private information (each party's type), and hence the reputation effect plays a central role. That is, the rational type has an incentive and an opportunity to build a reputation for obstinacy, which leads to bargaining costs.

The modified bargaining stage proceeds as follows. At the beginning of the stage, the right to make the first offer is assigned to each party with equal probability. Consider period t in which party i is the proposer ($t = 0,1,2,\ldots$). Party i either takes legal steps or makes party j an offer $x_i^t \in (0,1)$, which denotes his demanded share of the trade value V. If party i takes legal action, litigation occurs in period $t+1$ and the court specifies the division of V.[11] If party i makes an offer x_i^t, party j either accepts it, rejects it (and postpones the negotiation), or takes legal action. If party j accepts the offer, the game ends. When party j rejects the offer, the game continues and j makes the next offer in period $t+1$. If party j takes legal action, litigation takes place and the court determines the division of V in period $t+1$. The game continues unless the parties can reach agreement or one takes legal steps. Party i's payoff when the parties reach agreement in period t is given by $\delta^t \alpha_i V$, where δ denotes a common discount factor and α_i is his share specified by the accepted offer or the third party (the court or the boss).

As in the previous section, litigation endogenously determines the parties' outside

[11] If litigation occurs without such a time lag, the parties take legal steps immediately, which means the choice of governance structure does not matter.

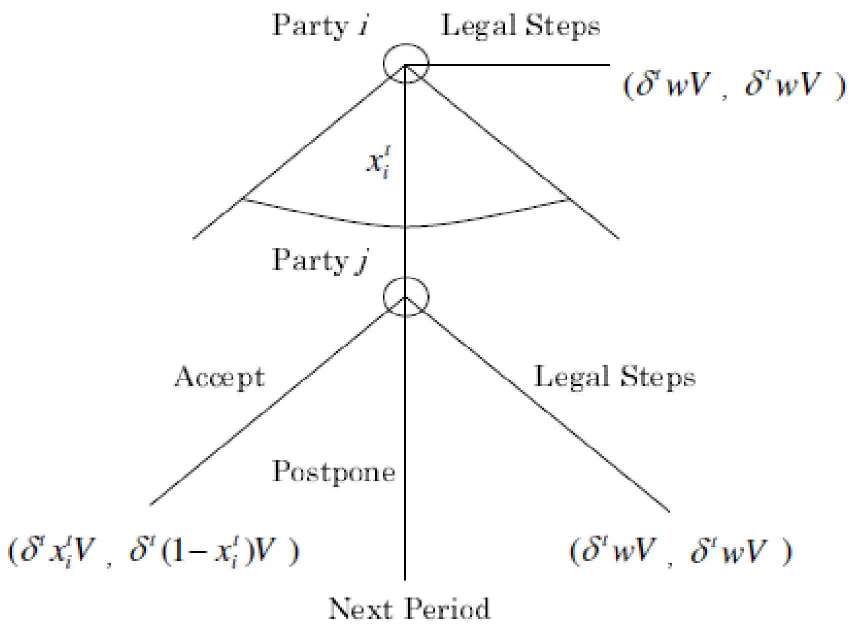

Figure 2.2: The Modified Bargaining Game

Party i's payoff is listed first and party j's second.

options. Since we continue to focus on a symmetric rent-seeking equilibrium, each party's litigation payoff when one of the parties takes legal steps in period t is given by $(\delta^{t+1}/2)V$. For notational convenience, we define $w \equiv \delta/2$. Figure 2.2 summarizes the modified bargaining stage.[12]

We further make the following five additional assumptions. First, the obstinate type never takes legal action, which means perpetual disagreement (bargaining breakdown) occurs if both parties are obstinate. Second, the obstinate type accepts any division determined by the third party.[13] Intuitively, the obstinate parties behave obstinately against people of equal rank (their partners), but reconcile to the third parties in authority (the court and the boss). These two assumptions imply that, when both

12 Figure 2.2 is based on Figure 1 in Atakan and Ekmekci (2010).

13 Although our result would continue to hold without this assumption, the analyses become a bit messy. We discuss the issue in Section 4.5.

parties are obstinate, while an agreement cannot be reached under non-integration, it is guaranteed under integration. Third, δ is sufficiently close to 1. Specifically, $\delta > v^*(=1/(1+\delta))$ holds, which means each party does not accept the equilibrium share of a complete-information, symmetric Rubinstein offers game, v^*, if he can obtain the whole value V in litigation. Fourth, mixed strategies are available to the parties. Finally, as in Compte and Jehiel (2002), $(1-\varepsilon)v^*V+\varepsilon\delta wV > wV$ holds. This implies that each party prefers to obtain the litigation payoff in period $t+2$ with probability ε (the probability of being obstinate) and the Rubinstein equilibrium share in period t with probability $1-\varepsilon$ rather than take legal action in period t when both parties choose the same rent-seeking level.

Note that two factors we presented in the previous section are included in the model: (i) litigation loss (the aggregate litigation payoff is smaller than V) and (ii) private information. In addition, there can be bargaining delay due to reputation building and bargaining breakdown.

Section 4.1 shows that our bargaining stage has two structures similar to those developed in Abreu and Gul (2000) and Compte and Jehiel (2002). Sections 4.2 and 4.3 analyze the bargaining stage and the rent-seeking stage respectively. In Section 4.4, we explore the trade-off between rent seeking and bargaining costs and present a comparative static analysis of the result. Section 4.5 briefly discusses two extensions: asymmetric parties and strong obstinacy.

4.1 Two Structures in the Bargaining Stage

Our bargaining model has two game structures developed in previous studies: one corresponds to the structure of Abreu and Gul (2000) and the other is similar to the model of Compte and Jehiel (2002). We hereafter refer to these respectively as the *AG* structure and the *CJ* structure.

AG Structure: The *AG* structure describes the situation in which the rational type prefers to concede rather than take legal steps (i.e., $wV \leq (1-\theta)V$ holds, where θ

denotes the obstinate type's inflexible demand). Hence, no litigation takes place in equilibrium. Given that each party's litigation payoff serves as his outside option, our bargaining stage corresponds to the bargaining game developed in Abreu and Gul (2000): an infinite-horizon, alternating-offers bargaining *without* outside options.

CJ Structure: The second structure considers the situation in which $wV > (1-\theta)V$ holds. Since the rational type prefers litigation to concession, he is willing to take legal action when his partner is obstinate with high probability. Thus, our bargaining game is equivalent to an infinite-horizon, alternating-offers bargaining game *with* outside options (i.e., the bargaining game that Compte and Jehiel (2002) analyze).

4.2 The Bargaining Stage

We here study the bargaining stage that occurs only under non-integration. Since the obstinate type behaves mechanically, we focus on the rational type's behavior under each structure.

AG structure ($wV \leq (1-\theta)V$): The AG structure corresponds to the symmetric version of the game developed in Abreu and Gul (2000). Hence, we can apply their result.

Lemma 1 (The Symmetric Version of Abreu and Gul's (2000) Proposition 4 and Compte and Jehiel's (2002) Proposition 3): *Consider the symmetric bargaining game described above and the case in which the rational type prefers concession to litigation (i.e., $wV \leq (1-\theta)V$ holds). The equilibrium payoff of the rational type converges to $(1-\theta)V$ as δ goes to 1 in any Perfect Bayesian Equilibrium of the game.*

Proof: See Abreu and Gul's (2000) Proposition 4 or Compte and Jehiel's (2002) Proposition 3. □

Under the AG structure, the rational type tries to build a reputation for obstinacy because if his partner (if rational) concedes, he can obtain a large share θ. However, he prefers concession if his partner never concedes (namely, if his partner is obstinate). He then concedes only at the constant rate that keeps his partner (if rational) indifferent between revealing himself as rational and mimicking the obstinate type, which causes delay in equilibrium.

As Abreu and Gul (2000) and Compte and Jehiel (2002) note, the delay emerges clearly in the symmetric case because ``parties are equally strong (weak), and thus no party is prepared to give in first with a significant probability" (Compte and Jehiel, 2002, p.1486). To see this, it is worth discussing the asymmetric case in terms of θ, ε, and δ. In the asymmetric case, one of the parties (e.g., party i) needs more time to build a reputation for obstinacy than the other (party j). That is, $T_i < T_j$, where T_i denotes the period in which party i's belief about party j's obstinacy reaches 1. Hence, in order that both parties will be known to be obstinate by the same time $T \equiv \min[T_i, T_j]$, the weaker party i has to reveal himself as rational (i.e., concede) immediately with positive probability, which is denoted by π.[14] Party j then does not concede immediately because he has the chance to get θ. Since only party i immediately concedes with probability π and the rational type randomizes his behavior after time 0, as δ goes to 1, the equilibrium payoff of the rational party i (resp. party j), which is denoted by u_i (resp. u_j), converges to $u_i = (1-\theta)V$ (resp. $u_j = \pi\theta V + (1-\pi)(1-\theta)V$).

In the symmetric case, however, both parties will be known to be obstinate by the same time without such an immediate concession ($T = T_i = T_j$). Since no immediate concession occurs ($\pi = 0$), the expected payoff of the rational type becomes $(1-\theta)V$ when δ is close to 1.

---

[14] See Abreu and Gul (2000) or Compte and Jehiel (2002) for detailed descriptions of T_i (T^j in Abreu and Gul (2000) and ϕ_i in Compte and Jehiel (2002)) and π ($\hat{F}^i(0)$ in Abreu and Gul (2000)).

CJ Structure ($wV > (1-\theta)V$): Under the *CJ* structure, the bargaining stage is equivalent to the game developed by Compte and Jehiel (2002). Hence we can apply the symmetric version of Compte and Jehiel's (2002) Proposition 5 to our bargaining stage.

Lemma 2 (The Symmetric Version of Compte and Jehiel's (2002) Proposition 5): *Consider the case in which the rational type prefers litigation to concession (i.e., $wV > (1-\theta)V$). The game then has a unique Perfect Bayesian Equilibrium. Let μ_i^h denote the current equilibrium probability that party $i(=1,2)$ is obstinate given history h. Whatever history h, $\mu_i^h \in \{0, \varepsilon, 1\}$.*

(i) If both parties are known to be rational (i.e., $\mu_i^h = \mu_j^h = 0$), they behave as in the complete information strategy profile in a symmetric alternating-offers game. That is, in each period, the proposer (e.g., party i) offers $x_i^t = v^*$ and the responder accepts an offer x_i^t iff $x_i^t \leq v^*$.

(ii) Consider a period t with history h in which party i is the proposer. If $\mu_j^h = \varepsilon$, party i (if rational) offers $x_i^t = v^*$ to party j. If $\mu_j^h = 1$, i takes legal steps.

(iii) Consider a period t in which party j is the proposer. Party i (if rational) accepts any offer $x_j^t \leq v^*$, rejects any offer greater than v^*, and takes legal action if i receives $x_j^t = \theta$.

Proof: See Proposition 5 in Compte and Jehiel (2002) □

This lemma suggests that the rational type reveals himself as rational immediately. In equilibrium, if party i makes an offer $x_i^t = \theta$, his partner j (if rational) believes that i is obstinate with probability 1, and thus takes legal steps because j prefers litigation to concession (i.e., $wV > (1-\theta)V$ holds). Party i (if rational) thus obtains only $\delta^t wV$ by mimicking the obstinate type. Since $wV < v^* V$, i has no incentive to mimic the obstinate type.

4.3 The Rent-Seeking Stage

We next analyze the rent-seeking stage and examine each party's optimal rent-seeking level under each governance structure given the equilibrium behavior in the bargaining stage. As mentioned earlier, we continue to focus on a symmetric rent-seeking equilibrium.

Non-Integration: AG Structure

Lemma 1 implies that no litigation occurs in equilibrium (the game ends with concession or perpetual disagreement). Nevertheless each party must undertake rent seeking because if party i does not engage in rent seeking (i.e., $d_i = 0$), his partner j chooses a low but positive rent-seeking level and immediately takes legal steps, which yields i nothing. Thus, in equilibrium, each party's litigation payoff must be smaller than or equal to his concession payoff $(1-\theta)V$.

Given that only the rational type triggers litigation, litigation is prevented if each party's choice of rent-seeking level d_{AG} satisfies the following condition:

$$(1-\varepsilon)(1-\theta)V - kd_{AG} = (1-\varepsilon)\frac{\delta(d_{AG}+e_{AG})}{d_{AG}+(d_{AG}+e_{AG})}V - k(d_{AG}+e_{AG})$$

where e_{AG} solves

$$\max_{e}(1-\varepsilon)\frac{\delta(d_{AG}+e)}{d_{AG}+(d_{AG}+e)}V - k(d_{AG}+e).$$

This condition suggests that party i's choice d_{AG} makes his partner j indifferent about whether to choose d_{AG} (to play the equilibrium strategy in the bargaining) or $d_{AG}+e_{AG}$ (to deviate from the equilibrium behavior in the bargaining stage). d_{AG} is thus given by

$$d_{AG} = \frac{(1-\varepsilon)\left\{\sqrt{\delta} - \sqrt{2(1-\theta)-\delta}\right\}^2}{4k}V.$$

Non-Integration: CJ Structure

Lemma 2 suggests that litigation takes place if one party is rational but the other is not. Suppose that each party provides symmetric rent-seeking level d. Party i's expected payoff is then given by

$$u_i = \frac{1}{2}\left[(1-\varepsilon)\left\{(1-\varepsilon)v^*V + \varepsilon\delta wV\right\} + \varepsilon(1-\varepsilon)wV\right]$$
$$+ \frac{1}{2}\left[(1-\varepsilon)\left\{(1-\varepsilon)(1-v^*)V + \varepsilon wV\right\} + \varepsilon(1-\varepsilon)\delta wV\right] - kd.$$

The first line (resp. second line) represents i's expected payoff when i is the first proposer (resp. the first responder) given that each party can be obstinate with probability ε in the bargaining.

We then specify the optimal rent-seeking level, d_{CJ}. There are two possible deviations: (v) a party chooses high rent-seeking level in the rent-seeking stage and takes legal action immediately if he becomes rational and the first proposer in the bargaining stage or (vi) a party provides high rent-seeking level and immediately sues the proposer when she becomes the rational responder in period 0. Let $d_{(v)}$ (resp. $d_{(vi)}$) represent the rent-seeking level that prevents deviation (v) (resp. deviation (vi)).

To prevent deviation (v), each party's rent-seeking level $d_{(v)}$ must keep his partner indifferent about whether to deviate, which means it must satisfy the following conditions:

$$\frac{1-\varepsilon}{2}\{(1-\varepsilon)v^*V+\varepsilon\delta wV\}+\frac{\varepsilon(1-\varepsilon)}{2}wV+\frac{(1-\varepsilon)\varepsilon}{2}wV+\frac{\varepsilon(1-\varepsilon)}{2}\delta wV-kd_{(v)}$$
$$=\left\{\frac{1-\varepsilon}{2}+\frac{\varepsilon(1-\varepsilon)}{2}+\frac{(1-\varepsilon)\varepsilon}{2}\right\}\frac{\delta(d_{(v)}+e_{(v)})}{d_{(v)}+(d_{(v)}+e_{(v)})}V+\frac{\varepsilon(1-\varepsilon)}{2}\frac{\delta^2(d_{(v)}+e_{(v)})}{d_{(v)}+(d_{(v)}+e_{(v)})}V$$
$$-k(d_{(v)}+e_{(v)})$$

where $e_{(v)}$ satisfies

$$\max_e\left\{\frac{1-\varepsilon}{2}+\frac{\varepsilon(1-\varepsilon)}{2}+\frac{(1-\varepsilon)\varepsilon}{2}\right\}\frac{\delta(d_{(v)}+e_{(v)})}{d_{(v)}+(d_{(v)}+e_{(v)})}V+\frac{\varepsilon(1-\varepsilon)}{2}\frac{\delta^2(d_{(v)}+e_{(v)})}{d_{(v)}+(d_{(v)}+e_{(v)})}V$$
$$-k(d_{(v)}+e_{(v)}).$$

Suppose party i chooses the rent-seeking level $d_{(v)}+e_{(v)}$ instead of $d_{(v)}$. Such a deviation improves i's litigation payoff, which is exercised in the following four cases.

First, if i is rational and becomes the first proposer, he immediately takes legal action (probability $(1-\varepsilon)/2$). Second, if i becomes obstinate and sends the first offer, his rational partner immediately sues him (probability $\varepsilon(1-\varepsilon)/2$). Third, if i becomes rational and receives the first offer, she sues her obstinate partner immediately (probability $(1-\varepsilon)\varepsilon/2$). Lastly, if i is obstinate and receives the first offer, her rational partner takes legal action in period 1 (probability $\varepsilon(1-\varepsilon)/2$).

Similarly, to prevent deviation (vi), each party's rent-seeking level $d_{(vi)}$ must make his partner indifferent about whether to deviate (namely, to choose high rent-seeking level and sues him immediately). That is, $d_{(vi)}$ satisfies

$$\frac{(1-\varepsilon)\varepsilon}{2}\delta wV+\frac{\varepsilon(1-\varepsilon)}{2}wV+\frac{(1-\varepsilon)}{2}\{(1-\varepsilon)(1-v^*)V+\varepsilon wV\}+\frac{\varepsilon(1-\varepsilon)}{2}\delta wV-kd_{(vi)}$$
$$=\left\{\frac{1-\varepsilon}{2}+\frac{\varepsilon(1-\varepsilon)}{2}\right\}\frac{\delta(d_{(vi)}+e_{(vi)})}{d_{(vi)}+(d_{(vi)}+e_{(vi)})}V$$
$$+\left\{\frac{(1-\varepsilon)\varepsilon}{2}+\frac{\varepsilon(1-\varepsilon)}{2}\right\}\frac{\delta^2(d_{(vi)}+e_{(vi)})}{d_{(vi)}+(d_{(vi)}+e_{(vi)})}V-k(d_{(vi)}+e_{(vi)})$$

where $e_{(vi)}$ satisfies

$$\max_{e}\left\{\frac{1-\varepsilon}{2}+\frac{\varepsilon(1-\varepsilon)}{2}\right\}\frac{\delta(d_{(vi)}+e_{(vi)})}{d_{(vi)}+(d_{(vi)}+e_{(vi)})}V$$
$$+\left\{\frac{(1-\varepsilon)\varepsilon}{2}+\frac{\varepsilon(1-\varepsilon)}{2}\right\}\frac{\delta^2(d_{(vi)}+e_{(vi)})}{d_{(vi)}+(d_{(vi)}+e_{(vi)})}V-k(d_{(vi)}+e_{(vi)}).$$

We thus obtain

$$d_{(v)}=\frac{(1-\varepsilon)\left[\sqrt{\delta(1+2\varepsilon+\delta\varepsilon)}-\sqrt{\delta(1+2\varepsilon+\delta\varepsilon)-2\{\delta+\delta\varepsilon-(1-\varepsilon)v^*\}}\right]^2}{8k}V$$

and

$$d_{(vi)}=\frac{(1-\varepsilon)\left[\sqrt{\delta(1+\varepsilon+2\delta\varepsilon)}-\sqrt{\delta(1+\varepsilon+2\delta\varepsilon)-2\{\delta+\delta^2\varepsilon-(1-\varepsilon)(1-v^*)\}}\right]^2}{8k}V.$$

Since the parties are uncertain what role they will play in period 0, they choose $\max\left[d_{(v)},d_{(vi)}\right]$ to prevent every possible deviation.

We next examine the rent-seeking level each party provides given that no one deviates. Let d_n denote such a level. d_n maximizes party i's expected payoff, and thus solves

$$\max_{d_i}\frac{1}{2}\left[(1-\varepsilon)\left\{(1-\varepsilon)v^*V+\varepsilon\delta\frac{\delta d_i}{d_i+d_j}V\right\}+\varepsilon(1-\varepsilon)\frac{\delta d_i}{d_i+d_j}V\right]$$
$$+\frac{1}{2}\left[(1-\varepsilon)\left\{(1-\varepsilon)(1-v^*)V+\varepsilon\frac{\delta d_i}{d_i+d_j}V\right\}+\varepsilon(1-\varepsilon)\delta\frac{\delta d_i}{d_i+d_j}V\right]-kd_i.$$

From player symmetry, we obtain

$$d_n=\frac{(1-\varepsilon)\varepsilon(1+\delta)\delta}{4k}V.$$

Since each party is uncertain whether he can send the first offer in the bargaining stage, he provides rent-seeking level $d_{CJ}=\max\left[d_{(v)},d_{(vi)},d_n\right]$.

Integration

The process of the value split under integration is the same as in Model 1. Hence, party i chooses rent-seeking level d_I, which solves the following problem:

$$\max_{d_i}\frac{d_i}{d_i+d_j}V-kd_i.$$

From symmetry assumption, we obtain $d_I=V/4k$.

4.4 Markets versus Hierarchies: A Comparison of Transaction Costs

We first focus on rent-seeking costs. From the discussion above, we can derive the following fact: $d_I>d_{AG}$ and $d_I>d_{CJ}$. Since $C(d)=kd$ is increasing in d, this implies $C(d_I)>C(d_{AG})$ and $C(d_I)>C(d_{CJ})$, which suggests that non-integration always incurs lower rent-seeking cost than integration. We thus find that the results shown in Section 3 continue to hold in Model 3. Let Δd_{AG} (resp. Δd_{CJ}) represent an excess of the aggregate influence cost over the aggregate rent-seeking cost under the AG

structure (resp. the CJ structure); that is,

$$\Delta d_{AG} \equiv 2C(d_I) - 2C(d_{AG}) = 2k(d_I - d_{AG}) > 0$$

and

$$\Delta d_{CJ} \equiv 2C(d_I) - 2C(d_{CJ}) = 2k(d_I - d_{CJ}) > 0.$$

We next analyze bargaining costs. From the existing literature (e.g., Kambe, 1999), under the AG structure, each party's expected payoff, denoted by u^{AG}, is approximately given by

$$u^{AG} = (1-\theta)V - \varepsilon^2\delta^T(1-\theta)V = (1-\varepsilon^2\delta^T)(1-\theta)V.$$

Since no one concedes immediately and the rational type employs the mixed strategy, each party expects a payoff $(1-\theta)V$ (Lemma 1). However, if both parties are obstinate (with probability ε^2), perpetual disagreement occurs and each party loses the chance to obtain $\delta^T(1-\theta)V$. Let Δb_{AG} represent the total bargaining cost under the AG structure. Δb_{AG} is then given by

$$\Delta b_{AG} = V - 2u^{AG} = \{1 - 2(1-\varepsilon^2\delta^T)(1-\theta)\}V > 0.$$

Under the CJ structure, on the other hand, no rational party has an incentive to build a reputation for obstinacy (Lemma 2). Nevertheless, bargaining costs occur if either or both parties are obstinate. There are three cases in which bargaining costs arise. First, if the first proposer is obstinate but the responder is not, which occurs with probability $(1-\varepsilon)\varepsilon$, the game ends with litigation in period 1 (the rational party takes legal steps in period 0). Second, if the first proposer is rational but the responder is not, litigation takes place in period 2 because the rational party takes legal steps in period 1. The probability with which such a case occurs is $\varepsilon(1-\varepsilon)$. Lastly, if both parties are obstinate, which arises with probability ε^2, perpetual disagreement occurs. Thus the expected payoff to each party, defined as u^{CJ}, is given by

$$
\begin{aligned}
u^{CJ} =&(1-\varepsilon)\left[(1-\varepsilon)\left\{\frac{1}{2}v^{*}V+\frac{1}{2}(1-v^{*})V\right\}+\varepsilon\left\{\frac{1}{2}\delta wV+\frac{1}{2}wV\right\}\right] \\
&+\varepsilon\left[(1-\varepsilon)\left\{\frac{1}{2}wV+\frac{1}{2}\delta wV\right\}+\varepsilon 0\right].
\end{aligned}
$$

The first line (resp. the second line) represents each party's expected payoff when he is rational (resp. obstinate). Notice that each party becomes the first proposer with equal probability and obstinate with probability ε. Hence, the total bargaining cost, denoted by Δb_{CJ}, is given by

$$\Delta b_{CJ} = V - 2u^{CJ} = [1-(1-\varepsilon)\{(1-\varepsilon)+\varepsilon\delta(1+\delta)\}]V > 0.$$

We then have the following proposition:

Proposition: *The optimal governance structure is summarized as follows.*

(i) When the litigation payoff is smaller than or equal to the concession payoff,

$$\begin{cases} \textit{Non-integration} & \textit{if } \Delta d_{AG} \geq \Delta b_{AG}, \\ \textit{Integration} & \textit{otherwise.} \end{cases}$$

(ii) When the litigation payoff is larger than the concession payoff,

$$\begin{cases} \textit{Non-integration} & \textit{if } \Delta d_{CJ} \geq \Delta b_{CJ}, \\ \textit{Integration} & \textit{otherwise.} \end{cases}$$

This proposition highlights an important trade-off which has never been focused on: while non-integration incurs lower rent-seeking costs than integration, it suffers from bargaining delay and breakdown that never occur under integration. In other words, the presence of inefficient bargaining can create a trade-off between rent-seeking costs and bargaining costs.

We now conduct comparative static analysis under each structure.

AG Structure Under the *AG* structure, non-integration is chosen if the following condition holds:

$$\Delta d_{AG} \geq \Delta b_{AG}$$

$$\Leftrightarrow \quad 2k\left[\frac{1}{4k} - \frac{(1-\varepsilon)\{\sqrt{\delta} - \sqrt{2(1-\theta)-\delta}\}^2}{4k}\right]V \geq \{1-2(1-\varepsilon^2\delta^T)(1-\theta)\}V.$$

We obtain the following results with respect to δ, θ, and ε. First, higher δ makes integration more likely to be chosen. There are two reasons for this. First, higher δ makes litigation loss $(1-\delta)V$, which leads to low rent-seeking levels, smaller, and hence the rent-seeking reduction also becomes smaller. Second, higher δ makes bargaining breakdown more costly ($2\varepsilon^2\delta^T(1-\theta)V$).

Second, larger θ makes non-integration less likely to be chosen. When θ is large, parties are apt to prefer litigation to concession under non-integration, and hence a high rent-seeking level is required to prevent deviations. In addition, larger θ leads to higher incentives to build a reputation for obstinacy because each party enjoys a share θ if his partner concedes.

Lastly, as ε decreases, both Δb_{AG} and Δd_{AG} decrease. The effect on Δb_{AG} is intuitive. When both parties are obstinate, while an agreement cannot be reached under non-integration (i.e., perpetual disagreement occurs), it is guaranteed under integration, which is the benefit of integration. As ε decreases, each party is less likely to be obstinate, and hence the benefit of integration becomes less significant. Nevertheless, lower ε also makes rent seeking under non-integration more costly. Under the *AG* structure, the only purpose of rent seeking between firms is to prevent deviations by the rational type. Thus, the lower ε becomes, the more likely each party is to be rational, and hence the more careful he must be about his rational partner's deviation.

CJ Structure Under the *CJ* structure, on the other hand, if non-integration is chosen, then the following condition must hold:

$$\Delta d_{CJ} \geq \Delta b_{CJ} \quad \Leftrightarrow \quad 2k\left(\frac{1}{4k}V - d_{CJ}\right) \geq [1-(1-\varepsilon)\{(1-\varepsilon)+\varepsilon\delta(1+\delta)\}]V$$

where

$$d_{CJ} = \max[d_{(v)}, d_{(vi)}, d_n],$$

$$d_{(v)} = \frac{(1-\varepsilon)\left[\sqrt{\delta(1+2\varepsilon+\delta\varepsilon)} - \sqrt{\delta(1+2\varepsilon+\delta\varepsilon) - 2\{\delta+\delta\varepsilon-(1-\varepsilon)v^*\}}\right]^2}{8k} V,$$

$$d_{(vi)} = \frac{(1-\varepsilon)\left[\sqrt{\delta(1+\varepsilon+2\delta\varepsilon)} - \sqrt{\delta(1+\varepsilon+2\delta\varepsilon) - 2\{\delta+\delta^2\varepsilon-(1-\varepsilon)(1-v^*)\}}\right]^2}{8k} V,$$

and

$$d_n = \frac{(1-\varepsilon)\varepsilon(1+\delta)\delta}{4k} V.$$

We obtain the following comparative static results with respect to δ and ε. First, as δ increases, both Δd_{CJ} and Δb_{CJ} decrease. The higher δ becomes, the more directly rent seeking between firms affects the rent seeker's payoff, and hence the more eager each party becomes to engage in it (Δd_{CJ} decreases). Furthermore, higher δ makes loss due to bargaining delay smaller (Δb_{CJ} decreases).

Second, while Δb_{CJ} is increasing in ε, Δd_{CJ} is non-monotonic. The effect on Δb_{CJ} is straightforward. That is, if ε is high, the case in which both parties are obstinate occurs with high probability, and hence integration is likely to be chosen to avoid perpetual disagreement. The effect of ε on Δd_{CJ} is illustrated in Figure 2.3, which describes the case in which $\delta = 4/5$ and $k = 1$, and the upper envelope curve represents d_{CJ}. (Since d_I does not depend on ε, we only need to focus on the effect on d_{CJ}.) When ε is low, since the parties become rational with high probability, litigation is less likely to occur in equilibrium. Hence, if no one deviates, the parties have low incentives to undertake rent seeking (d_n is low). However, low ε makes deviations by the rational type more likely. Since the equilibrium payoff of a rational responder $\{(1-\varepsilon)(1-v^*)+\varepsilon w\}V$ is smaller than that of a rational proposer $\{(1-\varepsilon)v^* + \varepsilon\delta w\}V$ when ε is low, the rational responder is more eager to deviate than the rational proposer. Thus, $d_{(vi)}$ is high and $d_{CJ} = d_{(vi)}$ holds when ε is low. When ε is intermediate, the parties are equally likely to become either rational or obstinate. Hence, situations in which one party is rational but the other is not are likely to occur.

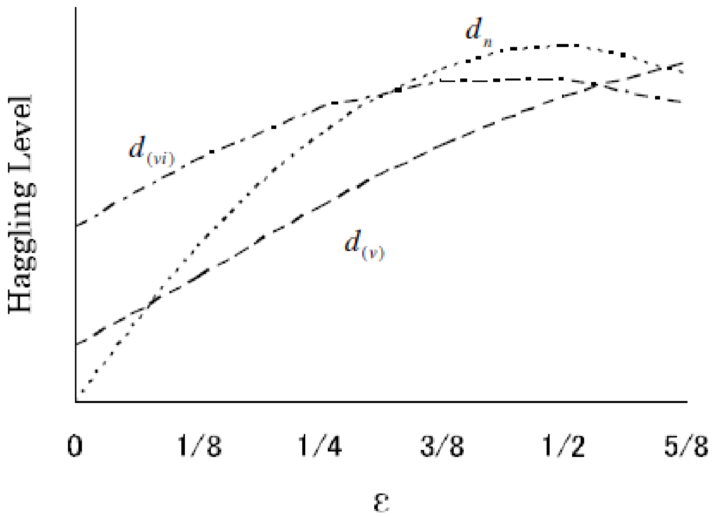

Figure 2.3: The Effect of ε on d_{CJ} ($\delta = 4/5$ and $k = 1$)

In such situations, the game ends with litigation (i.e., the rational type takes legal action in period 0 or 1). Thus, d_n is high and $d_{CJ} = d_n$ holds. If ε is high (i.e., each party is very likely to be obstinate), the rational type's equilibrium offer $x_i^0 = v^*$ is likely to be rejected, which makes the rational proposer prefer to deviate (namely, choose high rent-seeking level and take legal action immediately). To prevent such a deviation, $d_{(v)}$ becomes high and $d_{CJ} = d_{(v)}$ holds if ε is high.

It is worth noting that d_{AG}, d_{CJ}, d_I, Δb_{AG}, and Δb_{CJ} are all increasing in the size of V. This implies that larger trade value makes both non-integration and integration more costly, which is consistent with the assertion of Williamson (1973).

4.5 Extensions: An Asymmetric Case and Strong Obstinacy

In concluding this section, we examine two extensions: an asymmetric case and strong obstinacy. Although these extensions are important, they are beyond the scope of this dissertation. Hence, we make brief comments on them and leave further analysis for future research. In both extensions, the trade-off between rent seeking and bargaining costs would continue to occur.

First, extending our model to the asymmetric case in terms of θ, ε, and δ is straightforward. Abreu and Gul (2000) and Compte and Jehiel (2002) whose frameworks we have employed analyze the asymmetric game, and hence we can extend our model and results to the asymmetric case. When each party's litigation payoff is incompatible with his obstinate partner's demand (i.e., when the case that corresponds to the *CJ* structure arises), we can employ Proposition 5 in Compte and Jehiel (2002). When the rational type has no incentive to take legal steps (namely, when the case equivalent to the *AG* structure occurs), on the other hand, we can apply Abreu and Gul's (2000) Proposition 4 or Compte and Jehiel's (2002) Proposition 3.

In the asymmetric case, the third game structure arises. This structure is characterized as a one-sided outside-option case, in which only one party has the litigation payoff incompatible with his obstinate partner's inflexible demand. Under this structure, we can apply the result of Atakan and Ekmekci (2010).[15]

Lemma 3 (Atakan and Ekmekci's (2010) Lemma 1): *Consider the asymmetric version of the bargaining game and the situation in which* $w_i > (1-\theta_j)V$ *and* $w_j \leq (1-\theta_i)V$ *hold. Then (i) party* i *always demands* θ_i*, (ii) party* j *reveals himself as rational in period 0 or 1, and (iii)* i *'s share (resp.* j *'s share) conditional on facing the rational type is approximately* θ_i *(resp.* $1-\theta_i$*) in any Perfect Bayesian Equilibrium.*

Proof: See Atakan and Ekmekci's (2010) Lemma 1. □

Since party i prefers litigation to concession, party j cannot improve his payoff by mimicking the obstinate type. Thus, j reveals his rationality as soon as possible. Once j reveals himself as rational, the bargaining game with one-sided uncertainty emerges and party i obtains a payoff close to θ_i if δ_i (party i's discount factor) and δ_j are close to 1 (Myerson, 1991, Theorem 8.4).

---

15 Atakan and Ekmekci (2010) allow that $v^* > \theta$.

From this lemma, it is enough for party i to undertake rent seeking to prevent his rational partner's deviation from the equilibrium behavior. Furthermore, there is a positive probability that agreement cannot be reached immediately. Thus, the trade-off between rent seeking and bargaining costs will also emerge in the one-sided outside-option case.

The second extension includes the modification of the definition of the obstinate type. In Model 3, the obstinate type demands θV in the bargaining, but accepts any division that the third parties determine. Some readers might then think that the obstinate parties should be defined as those who will not accept any offer unless they can obtain more than or equal to θV, both in the bargaining and the third-party settlement. Even if we adopt the modified definition of the obstinate type, the trade-off between rent seeking and bargaining costs occurs because the AG and CJ structures continue to emerge.

However, this extension leads to an additional bargaining structure that has not been dealt with by the existing literature. The structure is characterized as follows: while the rational type prefers litigation to concession when there is uncertainty about his partner's type (and hence we cannot apply Lemma 1), he prefers the latter to the former when he knows his partner is obstinate with probability 1 (and thus we cannot apply Lemma 2 either).[16]

5 Conclusion

We have developed a theory of firm boundaries in the spirit of Williamson's transaction cost analysis, in which the parties engage in ex post value split. We presented three results. First, when the trade value shrinks due to delay in reaching agreement,

[16] Similarly, in the case where the obstinate type is assumed to behave obstinately throughout the game (e.g., the obstinate type chooses irrationally high rent-seeking level which the rational type cannot match), while the additional bargaining structure emerges, our trade-off between rent seeking and bargaining cost continues to hold for some ε and δ.

non-integration incurs lower rent-seeking costs than integration. Second, when the parties are obstinate with high probability, the rational type voluntarily concedes in the bargaining, and hence has small incentive to undertake rent seeking under non-integration. Lastly, and most importantly, the presence of ex post inefficient bargaining creates a trade-off between rent seeking and bargaining costs (bargaining delay and breakdown). These results explain why rent seeking within firms is likely to be more costly than rent seeking between firms, and offer a formal justification for the ``costs of bureaucracy" in Williamson (1985). Furthermore, we showed that larger trade value makes both non-integration and integration more costly, which is consistent with the argument of Williamson (1973).

There are some important topics left untouched. First, our models do not explain how internal organizations avoid costly renegotiations. That is, we assumed that the boss's order is enforceable. As Van den Steen (2010) notes, however, ``Being an employee does not mean abandoning free will: the employee decides whether or not to obey the boss's directives" (p. 466). In fact, TCE does not provide any formal answer on the issue. This issue has been dealt with in the next chapter. Chapter 3 formally explores why entering authority relation helps trading parties immediately implement ex post value split (i.e., avoid bargaining costs) by employing three behavioral assumptions: reference-dependent preference, self-serving bias, and shading (punishment for unfair treatments). In the next chapter, we point out that, under integration, a subordinate expects to obey his boss's order, and hence, it is likely to be optimal for him to comply, which leads to immediate settlement of ex post value split.

Second, we did not deal with the situation in which the parties negotiate the decision right at the beginning of the game. Some existing literature on firm boundaries, including Grossman and Hart (1986), assumes that one of the parties becomes a boss (the owner of the relevant assets) under integration. Under such an assumption, the decision right of ex post value split is transferred to party 1 or 2, and thus there is no third party. The party who has authority can then observe the other party's influence level, which means that influence activities can be used as signaling tools. That is, the level of influence activity might affect the reputation of its provider (for example, what type the provider is).

Chapter 3

A Formal Behavioral Model of Firm Boundaries: Why Does Authority Relation Mitigate Disputes over Trade Value?

1 Introduction

Transaction cost economics (TCE), such as Williamson (1985, 1996), asserts that under bilateral monopoly caused by relationship-specific investment or other factors, firms are likely to choose vertical integration. This is explained as follows. Suppose that unanticipated changes in trade circumstance occur. Such changes requires ex post adaptations, which invite dispute over trade value. Under non-integration, trading parties have to engage in bargaining to settle the dispute, which leads to bargaining costs (delay in reaching agreement and bargaining breakdown). Integrated firms, on the other hand, can settle the dispute by fiat without costly bargaining. This discussion implicitly assumes that authority within organizations is effective and subordinates always obey their boss's orders. This implicit assumption has been frequently questioned (e.g., Hart, 1995), but TCE has not provided any formal justification for it. As a matter of fact, Chapter 2 faced a same problem. That is, the previous chapter developed formal models of ex post value split in the spirit of TCE, but, as in the literature on TCE, integration is assumed to avoid bargaining costs without offering a formal justification for the assumption.

This chapter develops a formal model that explores the effectiveness of authority in the context of ex post value split, which is caused by the unprogrammed adaptations. Especially, we focus on the situation where trading parties are ``locked in'' (i.e., bilateral dependency condition appears) due to unverifiable relationship-specific investment. We adopt TCE's idea that authority is the most important aspect of integration (internal organizations) and point out that it affects each party's expectation about the outcome of the value split (i.e., reference point).

There are some recent studies which point out that reference points affect ex post renegotiation, and hence, make-or-buy decisions (e.g., Hart and Moore, 2008, and Herweg and Schmidt, 2012). We also focus on how reference points affect make-or-buy decisions and employ three behavioral assumptions about how reference points affect each party's utility and how they are set: reference-dependent preference, self-serving bias, and shading. It is worth noting that these assumptions are crucial for the result.

[1] That is, relaxing any of these assumptions leads to the result that authority relationship does not affect the timing of agreement or brings the opposite result: non-integration can realize the immediate agreement more easily than integration. The evidence that supports each of these assumptions will be presented in Section 3.

Trading parties in our model have the following four characteristics. First, as in the literature on reference-dependent preference, such as Köszegi and Rabin (2006, 2007), the parties' utility is reference dependent and their reference points are given by their expectations about the relevant outcomes.

Under this assumption, since non-integration and integration employ different processes of the value split, each governance structure leads to different reference points (i.e., the process by which the value is divided affects the parties' reference points). Under non-integration, as mentioned above, ex post value split is implemented through bargaining, and hence, the parties' reference points are given by the expected outcome of bilateral bargaining. Under integration, on the other hand, ex post value split is determined by fiat. That is, a party who has decision rights (boss) unilaterally gives an order to her subordinate and he can only choose whether to obey it or not. Thus, the parties' reference points are the expected outcome of an ultimatum game (i.e., the boss takes most of the trade value). We want to emphasize that the assumption that the boss takes most of the value under integration is not crucial to our result. More

[1] What is important here is that each party cares about his partner's gain-loss. Our result thus does not change if we employ another form of other-regarding preference instead of shading, such as altruism. See Appendix D.

specifically, our result continues to hold as long as each party expects that the party who has decision rights takes more value than the party who does not.

Second, each party has a self-serving view regarding who is to incur sunk relationship-specific investment (Babcock et al., 1995). More specifically, while a party who does not invest thinks that her partner who has invested (he) is to incur the whole investment cost, he believes that his sunk investment is to be compensated. Although his view about the sunk cost might seem unreasonable, Macleod (2007) points out a concept of fairness based on the idea that parties should be compensated for their sunk investments. Such self-serving views result in the divergence of reference points between the parties, which causes delay in reaching agreement.

Third, those who obtain the payoffs that are smaller than their reference point payoffs undertake activities that lower their partners' payoffs. Such behavior can be considered punishment for unfair treatment; it is called shading in the literature on contracts as reference points, such as Hart and Moore (2008), Hart (2009), and Hart and Holmstrom (2010). Since shading can be considered one interpretation of other-regarding preference, we can easily extend our model to analyze another form of other-regarding preference, namely altruism, which is discussed in Appendix D.

Fourth, while the value shrinks because of delay in reaching agreement, each party does not care about the cost of delay (behaves as if there were no discounting). This assumption does not only simplify our analysis substantially, but reflects the experimental fact of Binmore, Swierzbinski, and Tomlinson (2007). The case where the parties do care about discounting will be dealt with in Appendix C.

Some readers might suspect that such behavioral aspects matter at the level of individuals, but not at the level of organizations (i.e., make-or-buy decisions). Nevertheless, we believe that these aspects affect organizational-level decisions. For example, some literature points out the presence of ``boundary-role person'' (Adams, 1976) who performs ``The specialized class of roles that carry out the function of interaction between the organization and various elements in its environment'' (Perry and Angle, 1979, p. 489). This implies that since decisions at the level of organizations are made by an individual (boundary-role person), they can be affected by these behavioral aspects.

We show that integration indeed achieves immediate agreement on the division of the value more easily than non-integration despite the possibility of a subordinate's disobedience to the order of his boss. There are two reasons behind the result. First, disobedience to an order under integration provokes severer punishment than rejection of an offer under non-integration.[2] Under non-integration, trading parties are autonomous, and hence, they are entitled to reject any offer that their partners make as they please (namely, their reference point payoffs are balanced). Thus, the rejection of an offer does not cause a proposer a huge amount of feeling of loss (anger) under non-integration. Under integration, on the other hand, ex post value split is determined by fiat. That is, a boss determines how to divide the trade value, and a subordinate is supposed to obey her orders. The boss's reference point payoff is thus quite large. However, if a subordinate disobeys the boss's order, as Barnard (1938) points out, the authority relationship between the parties is terminated, and hence, the outcome of the value split is determined as if they are autonomous parties (i.e., their payoffs are balanced). This means that if the order is rejected, the boss is compelled to obtain a far smaller payoff than her reference point payoff, which provokes a huge amount of anger. Since the boss's anger leads to severe retaliation against the subordinate, he is less willing to reject the order.

The second reason is that under integration, the utility improvement for a subordinate from disobedience is not sufficient to offset damage from the severe punishment. As mentioned above, the parties' reference points under integration are the expected outcome of an ultimatum game, and hence, the subordinate expects a

[2] To facilitate the comparison between non-integration and integration, we assume that under integration, a boss does not fire a subordinate who disobeys her order. Intuitively, this assumption suggests that dismissal is not always costless: a fired employee can engage in actions that inflict damage on his ex-boss in revenge (e.g., sabotage, leakage, and theft).

small payoff. Thus, he can enjoy a large payoff improvement from rejecting the order, but such a payoff improvement is ``too much'' for him (i.e., disobedience does not lead to a large utility improvement), which makes him less eager to reject the order.

We use this result to analyze firm boundaries and point out a trade-off between immediate agreement and the aggregate sense of loss. That is, the expectation that the boss takes the entire surplus under integration makes her subordinate less willing to reject her order than under non-integration, but also leads to his larger psychological disutility than under non-integration. The reason for this is explained as follows. As mentioned above, the party who invests believes that his sunk investment will be compensated regardless of the choice of the governance structure. Nevertheless, under non-integration, each party expects a positive share of a trade surplus (namely, the trade value minus the investment cost) from bargaining, and thus, the party who invests expects to incur some portion of the investment cost. Under integration, on the other hand, a party who receives an order from the boss expects that the whole surplus will be taken by the boss, and hence, if the party who invests does not have decision rights, he does not take the investment costs into account when he sets his reference point. This discussion suggests that the divergence between the parties' reference points because of the self-serving view regarding who is to incur the investment costs is larger under integration than under non-integration. This makes the aggregate sense of loss and shading costs under integration larger than those under non-integration.

The rest of the chapter proceeds as follows. The next section relates this chapter to the existing literature. Section 3 introduces the model and Section 4 examines which governance structure achieves immediate agreement on the division of the value more easily. Section 5 presents a reduced form analysis of firm boundaries and shows the trade-off between immediate agreement and the aggregate sense of loss. Section 6 contains concluding comments. Furthermore, Appendix A shows that the three behavioral assumptions (reference-dependent preference, self-serving bias, and shading) are all crucial to our result: integration achieves immediate settlement of the division of the value more easily than non-integration. Appendix B examines the case in which the parties are risk-averse. Appendix C assumes that the parties care about discounting and checks the robustness of our result. Appendix D extends our model to analyze altruism.

2 Related Literature

This chapter employs the approach that a contractual arrangement, namely the choice of governance structure (the presence of authority), determines each party's reference point, which is influenced by self-serving bias. Hence, we first relate our study to Hart's approach, which points out that contracts serve as reference points. We then review some existing studies that share similar interests to ours.

The models of ``contracts as reference points" are presented in Hart and Moore (2008), Hart (2009), and Hart and Holmstrom (2010). These studies employ two important assumptions. First, ``each party feels entitled to the best outcome permitted by the contract" (Hart and Moore, 2008, p. 33). Second, those who obtain less than their reference points undertake retaliation against their trading parties. Such retaliation is called shading.

Our study is deeply related to contracts-as-reference-points approach in the sense that contractual arrangements affect each party's reference point and each party can engage in shading. Nevertheless, in our study, while each party's reference point is influenced by self-serving bias, he is not naive enough to believe that he is entitled to the best outcome permitted by the contract. That is, all trading parties set their reference points with the same rule, which helps their reference points converge, but cannot share the same reference point due to each party's self-serving view about who is to incur a sunk investment.

It is worth noting that our approach is quite different from that of Köszegi and Rabin (2006, 2007) in the following senses. First, while reference points are endogenously determined in their approach, they are exogenously given in ours. Second, punishment for unfair treatment (shading) plays an important role in our study, but it is not considered in their studies. Nevertheless we borrow Köszegi and Rabin's assumption that each party's reference point is his ``expectations about the relevant outcome'' (Köszegi and Rabin, 2007, p. 1051) and their utility function.

We next relate this chapter to the existing studies that share similar interests to ours: Gallice (2009), Van den Steen (2010), Akerlof (2010), and Herweg and Schmidt (2012). Gallice (2009) develops a model of Köszegi and Rabin's reference-dependent preferences with self-serving bias. However, Gallice (2009) is silent about how and what bias affects each party's reference point. As mentioned above, we assume that parties' self-serving views regarding the sunk investment result in the divergence of their reference points even if they share views on how each party sets his reference point.

Van den Steen (2010) develops a theory of interpersonal authority. He shows that it is costly for employees to disobey orders (and to get fired) because concentrating asset ownership into employer's hands (i.e., integration) improves her outside option and lowers their outside options. While Van den Steen (2010) focuses on ownership structure, it is not central to our study (e.g., the assets, which transaction in question requires, are inalienable). In this chapter, the choice of governance structure only affects the process of value split and each party's reference point.

Akerlof (2010) presents a formal model of compliance, norms (senses of duty to comply), and punishment. In his model, a failure in compliance (failure in following norms) provokes anger that leads to punishment. He points out that norms are contextual: self-interest behavior is viewed as fair in market contexts, but not within an organization. Our model also assumes that unfair treatments provoke anger and what is fair depends on the process of value split: bilateral bargaining (non-integration) or fiat (integration).[3]

Herweg and Schmidt (2012) explore how loss aversion affects the outcome of ex post

---

[3] A similar discussion can be found in Hart and Moore (2008, p. 35).

contract renegotiation and show that loss aversion interrupts efficient renegotiation. Both their study and ours assume that contractual arrangements affect reference points and point out that loss aversion matters. However, there are some differences between their study and ours. First, self-serving bias is not considered in Herweg and Schmidt (2012), but it plays an important role in our study. Second, while Herweg and Schmidt (2012) focus on inefficiencies due to maladaptation, our study focuses on delay in reaching agreement on the division of the value and shading cost (i.e., deadweight loss caused by shading).

3 The Model

This section presents the model that examines which governance structure realizes immediate agreement on ex post dispute over trade value (i.e., ex post value split) between two trading parties. We compare two polar governance structures (non-integration and integration) by employing three behavioral assumptions: reference-dependent utility, self-serving bias, and shading. We first present an overview of the model and then introduce some behavioral assumptions.

Two risk-neutral trading parties (parties 1 and 2) trade one unit of a good and are to engage in ex post value split, which is invited by unprogrammed adaptation.[4] The trade requires party 2's unverifiable relationship-specific investment I (party 1 does not invest) and creates value π. We assume that the trade is efficient and the parties cannot earn anything outside the current trade relationship. More specifically, the condition $\pi/2-I>0$ holds, which means that the Nash bargaining solution yields a positive payoff even to a party who incurs the whole sunk investment. In order to focus on ex post inefficiency, we assume that ex ante investment I is efficiently sunk (i.e., no ex ante inefficiencies).

The game proceeds as follows. First, a governance structure is chosen (non-integration or integration) to maximize the sum of the two parties' utility. Second,

[4] We refer to party 1 as ``she'' and party 2 as ``he'' for the purpose of identification only.

unanticipated changes in trade circumstances occur and trigger ex post value split. Third, the parties set their reference points regarding how the value will be divided. A process to divide the value is then initiated. We assume that under integration, party 1 (resp. party 2) becomes a boss (resp. a subordinate).[5] Some readers might wonder why the parties separate their negotiation into two phases (i.e., ex ante choice of governance structure and ex post division of the value), but this setting is appropriate to formalize TCE's arguments. That is, as the literature on TCE pointed out, ex ante contracts cannot be complete due to bounded rationality, and hence, ex post adaptations to unanticipated changes in trade circumstances are required. Such unprogrammed adaptations then invite dispute over the value.

The process of the value split consists of party 1's division offer $x=(x_1,x_2)$, where x_i represents party i's share of the value, and party 2's acceptance decision. If party 2 accepts the offer, the value is divided as the accepted offer specifies; otherwise, the game continues. This process does not necessarily mean that party 1 makes a take-or-leave-it offer. Since we focus on which governance structure realizes immediate agreement, we only need to examine whether the first offer is accepted. Thus, we can interpret this process to capture the first period of an infinite-horizon alternating-offers bargaining. The assumption of the common process of the value split between non-integration and integration is employed only to facilitate the comparison between the two structures.[6] In our model, the only difference between non-integration and integration is the presence of authority, which affects each party's expectation about the outcome of the value split. We will explain how each party's expectation is determined in the next

[5] This assumption implies that the party who has decision rights and the one who is to make the investment are different (e.g., a buyer firm merges with a seller firm which possesses a specific asset to produce a required input). We believe that this assumption is appropriate because ``the literature typically reserves the expression `make or buy' to contexts where firms integrate backward" (Lafontaine and Slade, 2007, p. 631, n. 5). If party 2 has decision rights under integration, integration should always be chosen as the optimal governance structure. See also footnote 15.

[6] We can instead assume that under non-integration, the right to send the offer is assigned to each party with equal probability without changing our result.

subsection.

For simplicity, we assume that each party does not care about discounting (the cost of delay in reaching agreement). Note that this assumption does not mean that there is no discounting. Namely, while the value actually shrinks because of delay in reaching agreement, each party ignores discounting (behaves as if there were no discounting). This assumption does not only simplify our analysis substantially, but also reflects the discussion in Binmore, Swierzbinski, and Tomlinson (2007). They conduct an experiment of Rubinstein's bargaining and point out that ``Much preliminary effort was devoted to trying to present the shrinking of the cake....But subjects then largely ignored the discounting altogether" (p. 10, n. 4). We will study the case where parties do care about discounting and generalize our main result in Appendix C.

Behavioral Assumptions

This subsection introduces three behavioral assumptions, namely reference-dependent utility, self-serving bias, and shading (other-regarding preference), and presents evidence that supports them.[7] We emphasize that these assumptions are all crucial to our result: integration can realize immediate agreement more easily than non-integration. In Appendices A and B, we show that our result does not hold if any of these assumptions is relaxed. Appendix A shows that no reference-dependence, no self-serving bias, or no shading leads to the result that the choice of the governance structure does not matter. Appendix B focuses on the case in which the parties are risk-averse and have no reference-dependent preference, and shows that such a change leads to the opposite result: non-integration achieves immediate agreement more easily than integration. Furthermore, Appendix D shows that our result holds even if we employ another form of other-regarding preference instead of shading: altruism.

Party i's utility is assumed to be reference-dependent and affected by party j's

[7] While we understand that it is important to explore whether these three behavioral assumptions can coexist, it is beyond the scope of this dissertation, and hence, we leave it for future research.

shading. More specifically, we combine Köszegi and Rabin's reference-dependent utility and the utility function of the contracts-as-reference-points approach. Let $r_i = (r_{ii}, r_{ij})$ denote party i's reference point (r_{ij} represents i's view about party j's reference point payoff). Party i's utility when an outcome of the value split is $y = (y_i, y_j)$ is thus given by

$$U_i(y \mid r_i, r_j) = y_i + n(y_i \mid r_{ii}) + \theta \min\{n(y_j \mid r_{jj}), 0\}$$

where

$$n(y_i \mid r_{ii}) = \begin{cases} \eta(y_i - r_{ii}) & \text{if } y_i \geq r_{ii} \\ \eta\lambda(y_i - r_{ii}) & \text{if } y_i < r_{ii}. \end{cases}$$

The first term of the utility function denotes party i's intrinsic payoff, the second term, $n(\cdot)$, represents his gain-loss utility (η represents weight on gain-loss payoff and $\lambda > 1$ is sensitivity of loss aversion), and the third term is the loss caused by party j's shading ($\theta > 0$ denotes an exogenous common punishment intensity, namely shading parameter). We assume that $\theta \leq (1+\eta\lambda)/\eta\lambda$, which means that each party does not have an incentive to accept a payoff which is smaller than his reference point payoff to avoid his partner's shading. Since we want to show clearly the crucial effect of loss aversion on our result, our gain-loss function $n(\cdot)$ rules out diminishing sensitivity, which is one of the features of gain-loss utility.

Shading can be interpreted as a punishment for unfair treatment. (We can extend our model to consider altruism, which will be dealt with in Appendix D.) That is, when party i obtains a payoff smaller than his reference point payoff, he experiences a sense of loss, which provokes anger and drives him to punish his partner (i.e., to engage in shading). Thus, if he obtains a payoff greater than or equal to his reference point payoff (i.e., if he does not incur any loss), he does not undertake any shading ($\theta \min\{n(y_i \mid r_{ii}), 0\} = 0$

when $y_i \geq r_{ii}$).[8] As in the contracts-as-reference-points approach, we assume that shading behavior does not inflict any cost on those who shade. Intuitively, shading makes people who are treated unfairly believe that justice has been done, and hence, brings them private benefit large enough to offset the cost of shading. Note that we use the term ``shading costs'' as deadweight loss due to shading.

It is worth noting that the first and second terms (resp. third terms) of the utility function constitute a utility function that corresponds to the utility function of Köszegi and Rabin's approach (resp. the contracts-as-reference-points approach). In other words, we introduce shading into Köszegi and Rabin's utility function. We believe that such formalization is plausible because it is well known that the threat of punishment affects people's behavior substantially. For example, the laboratory results of ultimatum games are contrary to the theoretical prediction. That is, while theory predicts that the proposer gives the receiver the smallest monetary unit possible and the receiver accepts, subjects playing the role of receiver often reject small but positive offers in ultimatum experiments. Bolton and Zwick (1995) conduct an ultimatum experiment and show that punishment for unfair treatment explains more of the deviation from the theoretical prediction in ultimatum games than the obtrusive effects of experimenter observation.

As in Köszegi and Rabin's approach, each party's reference point in our model is his expectation about the relevant outcome. However, while Köszegi and Rabin's approach assumes rational expectations, our model assumes that each party expects the relevant outcome in a biased way. More specifically, the parties correctly infer how their partners set their reference points, but perceive the game structure self-servingly.

We assume that each party has a self-serving view regarding the sunk investment I. That is, while party 1, who does not invest, thinks that party 2, who is supposed to

---

[8] The literature on contracts as reference points does not deal with gain-loss utility. Hence, shading in the literature on contracts as reference points depends not on gain-loss utility but on the difference between a party's payoff and his reference point payoff (i.e., the shading term in the literature on contracts as reference points is given by $\theta \min(y_i - r_{ii}, 0)$.

invest, is to incur his sunk investment, party 2 believes that his sunk cost is to be compensated. In other words, party 1 (resp. party 2) believes that the parties are to divide a gross value π (resp. a net value $\pi - I$). Party 2's view regarding the sunk cost might seem implausible. However, Macleod (2007, p.187) suggests that ``one can develop a concept of fairness based on the idea that it is optimal to reward sunk investment, and, hence, `fair' bargains should take this into account." Formally, party 1 believes that each party's outside option is given by

$$w_1 = (w_{11}, w_{12}) = (0, -I),$$

where w_{ij} denotes party i's view about party j's outside option. Note that each party cannot obtain anything outside the current relationship. Party 2, on the other hand, is confident that the parties' outside options are

$$w_2 = (w_{21}, w_{22}) = (0, 0).$$

This assumption reflects the fact that each party's role (in this case, whether a party has invested or not) affects his expectation in a self-serving way even if the same information is shared (Babcock et al., 1995). We further assume that each party believes that his partner shares the same view about the outside options. That is, party 1 (resp. party 2) believes that party 2's (resp. party 1's) view about the outside option is $(0, -I)$ (resp. $(0,0)$).

The ways in which parties set their reference points are assumed to be different under each governance structure; this stems from the difference in processes of the value division between non-integration and integration. Under non-integration, as the literature on TCE points out, ex post value split is determined through haggling (i.e., bargaining), and hence, each party's expectation regarding the outcome of the bilateral bargaining serves as his reference point. We thus assume that each party uses the Nash bargaining solution as his reference point; this is common knowledge.

Under integration, on the other hand, ex post dispute over the value is settled by fiat. In other words, the person who has decision rights (boss) can order any division to her subordinate (he) and he can only decide whether to accept the order or not. That is, ex post value split proceeds something like an ultimatum game, and hence, each party expects that the boss obtains most of the value (i.e., the equilibrium outcome of the ultimatum game is used as his reference point).

From these assumptions, party i's reference point under governance structure g, which is denoted by r_i^g, is given as follows: under non-integration,

$$r_1^m = (r_{11}^m, r_{12}^m) = \left(\frac{\pi}{2}, \frac{\pi}{2} - I\right) \qquad r_2^m = (r_{21}^m, r_{22}^m) = \left(\frac{\pi - I}{2}, \frac{\pi - I}{2}\right),$$

and under integration,[9]

$$r_1^h = (r_{11}^h, r_{12}^h) = (\pi, -I) \qquad r_2^h = (r_{21}^h, r_{22}^h) = (\pi - I, 0).$$

Party 1's (resp. party 2's) payoff is listed first (resp. second). Since each party believes that his partner has the same view about the outside options, he does not know that his partner has a different reference point. For example, under non-integration, party 1 (resp. party 2) believes that both parties share the same reference point r_1^m (resp. r_2^m).

Some readers might think that it is inappropriate to assume that while the parties minimize ex post inefficiencies (i.e., they recognize the presence of self-serving bias) in the stage where they choose the governance structure, they do not take into account such a bias when they construct their reference points. Nevertheless, this assumption is reasonable because even if people learn about the bias, it does not cause them to modify their expectations. As Babcock and Loewenstein (1997, p. 115) note, ``When they learned about the bias, subjects apparently assumed that the other person would succumb to it, but did not think it applied to themselves."

We then explain what will happen if party 2 rejects party 1's offer/order. We assume that each party's expectation about his/her continuation payoff, which he/she obtains if party 2 rejects party 1's offer, does not depend on the governance structure chosen at the

---

[9] What is important here is that party 1 is expected to obtain a larger payoff under integration than non-integration due to her authority. Thus, the assumption that the equilibrium outcome of the ultimatum game serves as reference points under integration is not crucial to our result. See also Section 4.2.

beginning. Some readers might wonder why this assumption is appropriate while the parties' reference points are employer-favored under integration. This assumption stems from Barnard's (1938) arguments about authority. Barnard (1938, p. 163) asserts, ``Disobedience of such a communication [directive communication] is a denial of its authority for him. Therefore, under this definition the decision as to whether an order has authority or not lies with the persons to whom it is addressed and does not reside in `persons of authority' or those who issue these orders.'' This suggests that a subordinate's rejection of an order terminates the authority relationship and each party becomes autonomous. Hence, after party 1's order is rejected, the process of the value division becomes the same under non-integration and integration, which leads to the same continuation outcome between the two governance structures. More specifically, each party's view about continuation outcome becomes the same as his reference point under non-integration (i.e., his expectation about the outcome when each party is autonomous).[10]

In the next section, it will turn out that party 1 optimally offers/orders what her reference point specifies. Party 2 thus infers party 1's true reference point (i.e., r_1^m or r_1^h) from 1's offer and modifies his view about continuation outcome. We define P as party 2's modified view about his continuation payoff and satisfies[11][12]

$$\frac{\pi}{2} - I < P \leq \frac{\pi - I}{2}.$$

This is explained as follows. Party 2 infers from 1's offer that each party's view about the continuation outcome is different, and hence, comes to believe that an actual

[10] Some readers might suspect that in reality, party 2's continuation payoff under integration is smaller than that under non-integration even after authority relationship is terminated. We employ this assumption only for simplicity, and hence, our result holds as long as party 2's expectation about his own continuation payoff under integration is larger than 0.
[11] Including $P > (\pi - I)/2$ does not change our result.
[12] This setting does not rule out party 1's view modification. For example, party 2's counter offer, which is not modeled, might help her modify her view about continuation outcome. However, since we focus on whether the first offer is accepted, such modification does not matter.

continuation outcome after 2's rejection is specified somewhere between r_1^m and r_2^m (through a negotiation, for example).

4 Which Governance Structure Achieves Immediate Agreement?

This section explores how the choice of the governance structure affects the timing of the settlement of ex post value split and shows that integration realizes immediate agreement more easily than non-integration despite the possibility of subordinates' disobedience to their boss's orders. This result can be intuitively explained by the following two discussions. First, a subordinate (party 2) believes that his disobedience to an order provokes severe punishment from his boss (party 1). Second, since the subordinate does not expect a large payoff from the outset, he is not so interested in payoff improvement from disobedience.

This section proceeds as follows. Subsection 4.1 studies each party's optimal behavior and examines when immediate agreement is realized under each governance structure. Subsection 4.2 then compares two governance structures and presents our main result and its intuition.

4.1 Each Party's Optimal Behavior

This subsection analyzes party 1's optimal offer/order, which is studied in Subsection 4.1.1, and party 2's optimal acceptance/compliance decision, which is examined in Subsection 4.1.2.

4.1.1 Party 1's Offer/Order

We first examine party 1's optimal offer/order and show that she optimally offers/orders what her reference point specifies. Note that party 1 believes that both parties share the same reference point, namely r_1^m under non-integration and r_1^h under integration, and the same view about the continuation outcome r_1^m (i.e., her view about the outcome when each party is autonomous).

Since $\theta \le (1+\eta\lambda)/\eta\lambda$ holds, any offer/order $x_1 < r_{11}^m$ or $x_1 < r_{11}^h$ is not optimal for party 1 (such an offer only leads to her loss). Hence, we must examine $x_1 \ge r_{11}^m$ under non-integration and $x_1 \ge r_{11}^h$ under integration. Furthermore, under integration, party 1's optimal order is equivalent to her reference point because there is no room for her to demand more ($r_{11}^h = \pi$). We then only need to study the optimal offering strategy under non-integration such that $x_1 = r_{11}^m + \Delta$ $(\Delta \ge 0)$.

Suppose party 1 offers $x_1 = r_{11}^m + \Delta$ under non-integration. If party 2 accepts such an offer, party 1's utility is given by

$$U_1^m(x \mid r_1^m, r_1^m) = r_{11}^m + \Delta + n(r_{11}^m + \Delta \mid r_{11}^m) + \theta n(r_{12}^m - \Delta \mid r_{12}^m)$$

$$= r_{11}^m + \Delta + \eta\Delta - \theta\eta\lambda\Delta.$$

Note that party 1 believes that party 2 also has the reference point $r_1^m = (r_{11}^m, r_{12}^m)$. If party 2 accepts the offer, party 1 obtains a payoff $r_{11}^m + \Delta$. Furthermore, since her payoff $r_{11}^m + \Delta$ is larger than her reference point payoff (r_{11}^m), she enjoys the gain $\eta\{(r_{11}^m + \Delta) - r_{11}^m\} = \eta\Delta$. However, since the offer $x_1 = r_{11}^m + \Delta$ forces party 2 to obtain $r_{12}^m - \Delta$, which is smaller than party 1's view about 2's reference point payoff (r_{12}^m), party 1 expects him to shade by $\theta\eta\lambda\{(r_{12}^m - \Delta) - r_{12}^m\} = \theta\eta\lambda\Delta$. Thus, party 1 offers $x_1 = r_{11}^m + \Delta$ instead of $x_1 = r_{11}^m$ if the following condition holds:

$$\theta \le \frac{1+\eta}{\eta\lambda}. \tag{1}$$

If this condition holds and party 2's acceptance is guaranteed, it is optimal for party 1 to choose $x_1 = \pi$, namely, she demands the whole surplus.

However, even if condition (1) holds, since party 1 believes that the parties share the same reference point r_1^m, she expects that an offer $x_1 > r_{11}^m$ will be rejected (and obtain continuation payoff r_{11}^m). Given this, making an offer $x_1 > r_{11}^m$ only delays agreement, and hence, party 1 offers $x_1 = r_{11}^m$ under non-integration.[13] If condition (1) does not

[13] We assume that when the parties face choices that yield them the same expected payoffs, they prefer the choice that achieves faster agreement.

hold, it is obviously optimal for party 1 to offer $x_1 = r_{11}^m$. We thus find that it is optimal for party 1 to offer/order what her reference point specifies. Let $x^m = r_1^m = (r_{11}^m, r_{12}^m)$ (resp. $x^h = r_1^h = (r_{11}^h, r_{12}^h)$) denote party 1's optimal offer under non-integration (resp. integration). Note that party 1 does not have an incentive to offer strategically in an attempt to affect party 2's inference about 1's reference point in her favor. It is because party 1 believes that both parties share the same reference point r_1^m or r_1^h, and thus, she does not know that her offer affects 2's view about her reference point and the continuation payoff.

4.1.2 Party 2's Acceptance/Compliance Decision

We then study party 2's acceptance/compliance decision given party 1's optimal offer $x^m = (\pi/2, \pi/2 - I)$ under non-integration and order $x^h = (\pi, -I)$ under integration. Note that party 2's reference point is $r_2^m = \{(\pi - I)/2, (\pi - I)/2\}$ under non-integration and $r_2^h = (\pi - I, 0)$ under integration.

We first study party 2's optimal acceptance strategy under non-integration. If party 2 accepts the offer $x^m = (\pi/2, \pi/2 - I)$, his utility is

$$U_2(x^m \mid r_1^m, r_2^m) = \frac{\pi}{2} - I + n\left(\frac{\pi}{2} - I \mid \frac{\pi - I}{2}\right) + \theta n\left(\frac{\pi}{2} \mid \frac{\pi}{2}\right) = \frac{\pi}{2} - I - \frac{\eta\lambda}{2} I \equiv U_2^m.$$

Note that party 2 can infer party 1's reference point r_1^m from 1's offer. If he rejects the offer, on the other hand, his utility is

$$\begin{aligned} U_2((\pi - I - P, P) \mid r_1^m, r_2^m) &= P + n\left(P \mid \frac{\pi - I}{2}\right) + \theta n\left(\pi - I - P \mid \frac{\pi}{2}\right) \\ &= P - \eta\lambda\left(\frac{\pi - I}{2} - P\right) - \theta\eta\lambda\left\{\frac{\pi}{2} - (\pi - I - P)\right\} \\ &\equiv U_2^{m\prime}. \end{aligned}$$

Party 2 then accepts the offer if

$$U_2^m \geq U_2^{m\prime} \quad \Leftrightarrow \quad \theta \geq 1 + \frac{1}{\eta\lambda} \equiv \theta_m.$$

We next analyze party 2's compliance strategy under integration. Notice that party

1's optimal order, which is equal to her reference point, is given by $x^h = r_1^h = (\pi, -I)$.

If party 2 accepts the order $(\pi, -I)$, he obtains

$$U_2(x^h \mid r_1^h, r_2^h) = -I + n(-I \mid 0) + \theta n(\pi \mid \pi) = -(1+\eta\lambda)I \equiv U_2^h.$$

If party 2 rejects the order, his utility is given by

$$\begin{aligned} U_2((\pi - I - P, P) \mid r_1^h, r_2^h) &= P + n(P \mid 0) + \theta n(\pi - I - P \mid \pi) \\ &= (1+\eta)P - \theta\eta\lambda\{\pi - (\pi - I - P)\} \equiv U_2^{h'}. \end{aligned}$$

Thus, party 2 (the subordinate) does not reject the order if the following condition holds:

$$U_2^h \geq U_2^{h'} \quad \Leftrightarrow \quad \theta \geq \frac{(1+\eta)P + (1+\eta\lambda)I}{\eta\lambda(P+I)} \equiv \theta_h.$$

4.2 Immediate Agreement and Governance Structures

This subsection derives our main result that integration is more likely to realize immediate agreement than non-integration based on the discussions in the previous subsection.

We can determine that $\theta_h < \theta_m$, which means that non-integration requires severer punishment than integration for party 2's rejection to realize immediate agreement. There are two reasons for this. First, party 2's rejection under integration provokes party 1 to greater anger than that under non-integration. Since party 1 offers/orders what her reference point specifies, party 2's rejection results in party 1's aggrievement. Furthermore, because party 1's reference point payoff under integration ($r_{11}^h = \pi$) is much larger than that under non-integration ($r_{11}^m = \pi/2$) and party 2's expectation about his continuation payoff P is independent of the choice of the governance structure, party 2 expects that his disobedience leads to party 1's larger sense of aggrievement under integration ($\eta\lambda\{\pi - (\pi - I - P)\} = \eta\lambda(P+I)$) than under non-integration ($\eta\lambda\{\pi/2 - (\pi - I - P)\} = \eta\lambda(P + I - \pi/2)$). Party 1's larger aggrievement results in severer punishment for party 2, which makes him less willing to disobey the order.

Second, while party 2's disobedience under integration leads to a larger payoff improvement than under non-integration, the former has less impact on his utility than the latter because of loss aversion. Under integration, if party 2 rejects party 1's order, he can enjoy his payoff improvement $P-(-I)=P+I$. Since party 2's reference point payoff is 0, his payoff improvement leads to gain P and reduction in loss I. Party 2's utility improvement from rejecting the order is then $\eta P+\eta\lambda I$ ($\lambda>1$). Under non-integration, on the other hand, party 2's payoff improvement $P-(\pi/2-I)$ leads to loss reduction only, and hence he enjoys the utility improvement $\eta\lambda\{P-(\pi/2-I)\}$. Intuitively, under integration, party 2 does not expect a large payoff, and hence, his payoff improvement from rejecting the order is ``too much" for him and does not lead to a large utility improvement. Such an insignificant utility improvement is not enough to offset the huge cost of the rejection discussed above (i.e., party 1's shading), and thus, party 2 is less eager to disobey the order.

The second reason suggests that each party's expectation that party 1 takes the whole surplus under integration is not critical to our result. That is, integration realizes immediate agreement more easily than non-integration as long as the following conditions hold:

$$r_{12}^m < P < r_{22}^m \text{ and } r_{12}^h < r_{22}^h < P.$$

These conditions imply that while the continuation payoff (P) does not contribute to party 2's utility improvement substantially under integration, it does so under non-integration.

We then have the following proposition:

Proposition 1: *Integration achieves immediate agreement more easily than non-integration. That is, non-integration requires severer punishment for party 2's rejection than integration to realize immediate agreement:* $\theta_h < \theta_m$ *. Thus, the governance structure that achieves faster agreement is summarized as follows:*

$$\begin{cases} \textit{Non-Integration or Integration} & \textit{if } \theta < \theta_h \textit{ or } \theta_m \leq \theta, \\ \textit{Integration} & \textit{if } \theta_h \leq \theta < \theta_m. \end{cases}$$

This proposition implies that there are three cases. The first case is that both governance structures fail in reaching immediate agreement (i.e., the case in which $\theta < \theta_h$ holds). The second case is that only integration realizes immediate agreement (namely, the case in which $\theta_h \leq \theta < \theta_m$ holds). The last case is that both governance structures achieve immediate agreement (that is, the case in which $\theta_m \leq \theta$ holds). The next section analyzes these cases separately, and hence, for convenience, we call these Cases 1, 2, and 3, respectively.

This proposition also suggests that integration can never do worse than non-integration with respect to the timing of agreement, but the choice of the governance structure does not matter when the punishment for party 2's rejection is sufficiently severe or mild (i.e., θ is either sufficiently high or low). This is quite intuitive. If the punishment for rejection is too severe (namely, θ is sufficiently high), such severe punishment makes party 2 unwilling to reject the offer/order regardless of the choice of the governance structure. If the punishment for rejection is too mild (θ is sufficiently low), on the other hand, party 2 does not care about such a negligible threat of punishment and rejects the offer/order as long as he can improve his payoff by doing so.

This result explains how integration facilitates immediate settlement in ex post dispute over trade value and presents a formal justification for the implicit assumption of TCE: integration can avoid costly ex post bargaining. Hart (1995) observes ``If there is less haggling and hold-up behaviour in a merged firm, it is important to know *why*. Transaction cost theory, as it stands, does not provide the answer'' (Hart, 1995, p. 28). Our result suggests that integration can avoid costly renegotiation because each party's expectation of the relevant outcome is different between the two governance structures due to the difference in the processes of the value split between them.

This section focused on immediate agreement ignoring transaction cost-minimization (i.e., minimizing ex post inefficiencies such as the costs of delay, the sense of loss, and shading costs). We examine these inefficiencies and study firm boundaries in the next section.

5 Which Governance Structure Minimizes Transaction Cost?

This section presents a reduced-form analysis of firm boundaries. Specifically, we examine the costs of delay, the sense of loss, and shading costs under each governance structure and study which governance structure minimizes these inefficiencies in Cases 1, 2, and 3. We then point out a trade-off between immediate agreement and the aggregate sense of loss (shading costs).

As mentioned previously, while the value actually shrinks because of bargaining delay, the parties ignore discounting. Specifically, although the parties behave as if there were no discounting, the surplus shrinks to $\delta\pi - I$ because of delay in reaching agreement, where δ is a source of the cost of delay and can be interpreted as a discount factor. (We discuss the case in which the parties care about discounting in Appendix C.)

Case 1 ($\theta < \theta_h$): In this case, the parties cannot reach agreement immediately regardless of the choice of the governance structure (the cost of delay is the same between the two governance structures). Hence, we need to examine the sense of loss and shading costs.

As mentioned previously, the continuation outcome after party 2's rejection is determined to be somewhere between r_1^m and r_2^m, and thus, under non-integration, the negotiation after party 1's offer is rejected can be seen as the division of the aggregate loss $\eta\lambda(r_{11}^m - r_{21}^m) = \eta\lambda(r_{22}^m - r_{12}^m) = (\eta\lambda/2)I$ between the parties. Hence, the aggregate shading cost (i.e., the sum of each party's shading) is $\theta(\eta\lambda/2)I$.

Under integration, on the other hand, given party 2's disobedience, he obtains at least $r_{12}^m = \pi/2 - I$, and hence, enjoys gain at least $\eta\{(\pi/2 - I) - 0\} = \eta(\pi/2 - I)$. However, party 1 experiences a loss larger than $\eta\lambda[\pi - \{(\pi - I) - (\pi/2 - I)\}] = (\eta\lambda/2)\pi$ because she believes that she can obtain π, but party 2's disobedience forces her to receive at most $(\pi - I) - (\pi/2 - I)$. Thus, under integration, the aggregate loss is equal to or greater than $(\eta\lambda/2)\pi - \eta(\pi/2 - I)$ and the aggregate shading cost is at least $\theta(\eta\lambda/2)\pi$.

This discussion implies that in Case 1 there is no reason to choose integration because integration does not facilitate agreement and incurs a larger sense of loss and shading cost than non-integration.

Case 2 ($\theta_h \le \theta < \theta_m$): Unlike Case 1, only integration can realize immediate agreement. In other words, integration can save the cost of delay $(1-\delta)\pi$ that non-integration cannot avoid.

While integration can avoid the cost of delay, it suffers from a larger loss and shading cost than non-integration. As shown in Case 1, since the offer is rejected, non-integration incurs the aggregate loss $(\eta\lambda/2)I$ and the aggregate shading cost $\theta(\eta\lambda/2)I$. Under integration, on the other hand, party 1's order, which is equal to her reference point, is accepted, and hence, only party 2 experiences loss $\eta\lambda(0-(-I)) = \eta\lambda I$ and engages in shading $\theta\eta\lambda I$.

Thus, integration should be chosen if the cost of delay under non-integration is larger than the excess of the aggregate loss and shading cost under integration over those under non-integration. That is, the optimal governance structure is summarized as follows:

$$\begin{cases} \text{Non-integration} & \text{if} \quad \theta \geq \max[\theta_h, \theta_2] \\ \text{Integration} & \text{otherwise,} \end{cases}$$

where

$$\theta_2 \equiv \frac{2(1-\delta)\pi}{\eta\lambda I} - 1.$$

θ_2 equalizes the cost of delay with the excess of the aggregate loss and shading cost under integration over those under non-integration.

Case 2 is the case where $\theta_h \le \theta < \theta_m$ holds. Hence, if $\theta_2 < \theta_h$ holds, integration should not be chosen. That is, if integration can be the optimal governance structure, the following condition must hold in addition to the condition above:

$$\theta_2 \ge \theta_h \quad \Leftrightarrow \quad 1-\delta \ge \frac{\{(1+\eta+\eta\lambda)P+(1+2\eta\lambda)I\}I}{2(P+I)\pi}.$$

Case 3 ($\theta_m \le \theta$): Case 3 is similar to Case 1 in that the choice of the governance structure does not affect the timing of agreement (namely, immediate agreement is reached regardless of the choice of the governance structure). Hence, we again need to focus on the sense of loss and shading costs, as in Case 1.

Under non-integration, party 2 accepts the offer, and hence, only party 2 experiences loss $\eta\lambda\{(\pi-I)/2-(\pi/2-I)\}=(\eta\lambda/2)I$ and undertakes shading $\theta(\eta\lambda/2)I$. Under integration, on the other hand, as in Case 2, immediate agreement is reached, and thus, only party 2 feels aggrievement $\eta\lambda I$ and shades by $\theta\eta\lambda I$.

The above discussion suggests that non-integration should be chosen in Case 3, as in Case 1.

From Cases 1, 2, and 3, we have the following proposition:

Proposition 2: *Integration should be chosen as the optimal governance structure (that minimizes the transaction costs) if and only if the following conditions hold:*

$$1-\delta \ge \frac{\{(1+\eta+\eta\lambda)P+(1+2\eta\lambda)I\}I}{2(P+I)\pi} \tag{2}$$

and

$$\theta_h \le \theta < \theta_2,$$

where

$$\theta_2 \equiv \frac{2(1-\delta)\pi}{\eta\lambda I}-1.$$

This result implies that integration should be chosen when the punishment for party 2's rejection (θ) is intermediate and the cost of delay is larger than the sense of loss and shading cost. The explanation as to why integration should be chosen when θ is intermediate has been presented in the intuition of Proposition 1. Furthermore, even if only integration can realize immediate agreement (i.e., θ is intermediate), it should not be chosen when the cost of delay is insignificant (namely, δ is sufficiently close to 1) and the excess of loss and shading costs under integration over those under non-integration are quite large (i.e., either η or λ or both are large). This is what condition (2) means.

The right-hand side of condition (2) (resp. θ_2) is decreasing (resp. increasing) in π. This implies that larger trade value makes integration more likely to be chosen, which is consistent with the main assertion of TCE. Furthermore, this observation is also consistent with empirical studies on TCE, such as Monteverde and Teece (1982), Masten (1984), and Joskow (1988) (see Lafontaine and Slade (2007) for the review of these studies). These empirical studies provide support for the hypothesis that the more relationship-specific a trade becomes, the larger quasi-rent gets, and hence, the more likely it is that integration should be chosen.

A Trade-Off between Immediate Agreement and Shading Costs

The above discussions suggest that integration always suffers larger shading costs and sense of loss than non-integration. This stems from the fact that the level of divergence between two parties' reference points under integration is larger than under non-integration. That is, while the divergence between r_{12}^m and r_{22}^m is $I/2$, the difference between r_{12}^h and r_{22}^h is I. This can be explained by the fact that under integration, party 2 sets his reference point without internalizing investment cost I.

Under either governance structure, party 2 believes that his investment cost I is to be compensated. Nevertheless, under non-integration, party 2 somewhat internalizes the investment cost when he sets his reference point because he obtains a positive share of the surplus $\pi - I$ from ex post bargaining. Under integration, on the other hand, party 2 expects that he cannot obtain any portion of the surplus (i.e., $r_{22}^{h} = 0$), and hence, there is no room for him to internalize the investment cost I.

This implies that there is a trade-off between immediate agreement and the aggregate sense of loss. That is, the expectation that party 1 (boss) takes the entire surplus under integration makes party 2 less willing to reject her order than under non-integration (see Section 4), but also makes him set his reference point without internalizing the investment cost, which leads to larger aggregate loss and shading costs than under non-integration.[14][15]

6 Conclusion

This chapter examined the question of why authority (integration) helps ex post value split to be settled immediately. We showed that, despite the possibility of subordinates' disobedience to their boss's orders, integration can realize immediate settlement of value split because each party's reference point under integration is employer-favored due to the process of the value split under integration. This employer-favored reference

---

[14] Even if party 2 obtains some portion of the surplus under integration, this trade-off continues to emerge as long as the following conditions hold: $r_{12}^{m} < P < r_{22}^{m}$ and $r_{12}^{h} < r_{22}^{h} < P$.

[15] As mentioned in footnote 5, if party 2 becomes the boss under integration, integration dominates non-integration. This is because in such a case, both parties share the same reference point under integration: $r_1^{h'} = r_2^{h'} = (0, \pi - I) \equiv r^{h'}$. Since the same reference point is shared between the parties, party 1, who is now the subordinate, accepts party 2's order, which is equal to $r^{h'}$, without incurring any sense of loss. Hence, integration completely avoids ex post inefficiencies (delay in reaching agreement, sense of loss, and shading costs).

point makes a subordinate less eager to reject his boss's order for the following two reasons. First, it is very costly for the subordinate to reject the order from his boss because disobedience to the order results in the boss's huge amount of anger and severe punishment. Second, it is not so rewarding for the subordinate to reject the order because he does not expect a large payoff from the outset.

We further showed that integration incurs larger aggregate loss and shading cost than non-integration. This follows because, under integration, the expectation that party 2 cannot obtain any portion of the surplus makes him set his reference point without internalizing the investment cost. These discussions suggest that the employer-favored reference points create a trade-off between immediate agreement and shading costs.

In conclusion, we make a brief comment on some extensions: asymmetric shading parameters, endogenous reference points, and the limit of firm scope. First, we discuss the case in which the parties have different shading parameters. While our model assumes that the parties share the same shading parameter θ, asymmetric shading does not affect our result because party 2's shading does not matter. Hence, any change in either party's shading parameter does not substantially affect our analysis and results.

We next discuss endogenous reference points. Our model takes each party's reference point as exogenous. Nevertheless, we can extend our model to deal with endogenous reference points by employing the assumption of imperfect recall, which can be found in Bénabou and Tirole (2004). For example, suppose party 1 is completely rational, but party 2 forgets that he can be biased and set his reference point self-servingly with positive probability. Since party 1 is rational, she takes party 2's bias into account when she sets her reference point. In such a case, as in Köszegi and Rabin's approach, party 1's reference point is given by her probabilistic belief concerning the relevant outcome.

Finally, we can extend our model to analyze the limit of firm scope. Suppose party 1 faces some other transactions similar to the trade in which parties 1 and 2 engage and that θ is decreasing in the number of transactions she conducts: $\theta'(n) \leq 0$, where n represents the number of transactions she handles. The intuition of the latter assumption is that the more transactions party 1 conducts, the smaller effort and the less time she can provide to each transaction (i.e., the harder it is for her to punish those who disobey her orders). Under these assumptions, an integrated firm can become larger as long as $\theta_h \leq \theta(n)$ and condition (2) hold (see Proposition 2). That is, party 1 can acquire at most n^* trading partners where n^* satisfies $\theta(n^*+1) < \theta_h \leq \theta(n^*)$. This discussion is consistent with diminishing returns to management (e.g., Coase, 1937).

Appendix A: Relaxing Three Behavioral Assumptions

This appendix shows that three behavioral assumptions (reference-dependent utility, self-serving bias, and shading) are all crucial to our result: integration realizes immediate agreement more easily than non-integration. Sections A.1, A.2, and A.3 examine the no-reference-dependence case, the no-self-serving bias case, and the no-shading case, respectively. All these cases yield the same result: the choice of the governance structure does not affect the timing of agreement.

A.1 No Reference-Dependence

We first explore the no reference-dependence case. Suppose the outcome of the value split is $y=(y_i, y_j)$. In the case where there is no reference-dependence, the utility of party i who has a reference point r_i is given by

$$U_i(y \mid r_i, r_j) = y_i + \theta \min(y_j - r_{jj}, 0).$$

Since there is no reference-dependence, each party's utility function does not include a gain-loss term and each party's shading depends on the difference between his payoff and his reference point payoff (namely, the shading term does not include η, which denotes weight on gain-loss payoff, and λ, which represents the sensitivity of loss aversion). In other words, the utility function above is similar to that of contracts as reference points. Since parameters η and λ are not used, we assume that $\theta \leq 1$, which means that each party does not have an incentive to give up any payoff to avoid his partner's shading and corresponds to the assumption $\theta \leq (1+\eta\lambda)/\eta\lambda$ in the main model.

Note that the optimal offer/order of party 1 does not change. We thus need to examine party 2's optimal acceptance/compliance decision only. Under non-integration, while party 2's acceptance payoff is given by

$$U_2(x^m \mid r_1^m, r_2^m) = \frac{\pi}{2} - I - \theta \cdot 0 = \frac{\pi}{2} - I \equiv U_{NRD}^m,$$

his rejection payoff is

$$U_2((\pi - I - P, P) \mid r_1^m, r_2^m) = P - \theta \left\{ \frac{\pi}{2} - (\pi - I - P) \right\} \equiv U_{NRD}^{m'}.$$

Note that party 1 optimally offers $(\pi/2, \pi/2 - I)$, party 2's reference point is $((\pi - I)/2, (\pi - I)/2)$, and party 2's expectation about his continuation payoff is P. Comparing U_{NRD}^m and $U_{NRD}^{m'}$ implies that party 2 does not reject the offer if $\theta \geq 1$.

Under integration, on the other hand, if party 2 accepts the order, he obtains

$$U_2(x^h \mid r_1^h, r_2^h) = -I - \theta \cdot 0 = -I \equiv U_{NRD}^h.$$

Note that party 1's optimal order is $(\pi, -I)$ and party 2's reference point is $(\pi - I, 0)$. If party 2 rejects the order, his utility is given by

$$U_2((\pi - I - P, P) \mid r_1^h, r_2^h) = P - \theta\{\pi - (\pi - I - P)\} = P - \theta(P + I) \equiv U_{NRD}^{h'}.$$

We find that when $\theta \geq 1$, party 2 does not reject the order under integration.

This discussion implies that if there is no reference dependence, the choice of the governance structure does not matter (i.e., does not affect party 2's acceptance/compliance decision). In the no-reference-dependence case, the marginal benefit from payoff improvement is 1 and its marginal cost is θ, and hence, party 2 rejects the offer/order as long as the former is larger than the latter: $\theta < 1$. As mentioned in Section 4, one of the reasons why integration achieves immediate agreement more easily than non-integration is that while the utility improvement from rejection under non-integration consists of loss reduction only, that under integration includes not only loss reduction but also gain. No reference dependence (no loss aversion) makes both gain and loss equally important for both parties and eliminates the difference between the effects of gains and losses on each party's utility.

A.2 No Self-Serving Bias

We next study what will happen if there is no self-serving bias. As in the previous subsection, party 1's optimal offer/order does not change, and thus, we focus on party 2's optimal behavior.

Suppose both parties share the same view regarding each party's outside option: $w_1' = w_2' = (0, -I)$.[16] Both parties then share the same reference point. That is, under non-integration, their reference points are

$$r_1^m = r_2^m = \left(\frac{\pi}{2}, \frac{\pi}{2} - I\right) \equiv r_{NSSB}^m,$$

and, under integration,

$$r_1^h = r_2^h = (\pi, -I) \equiv r_{NSSB}^h.$$

Party 2's acceptance payoff under non-integration is thus

$$U_2(x^m \mid r_{NSSB}^m, r_{NSSB}^m) = \frac{\pi}{2} - I + n\left(\frac{\pi}{2} - I \mid \frac{\pi}{2} - I\right) + \theta n\left(\frac{\pi}{2} \mid \frac{\pi}{2}\right) = \frac{\pi}{2} - I \equiv U_{NSSB}^m.$$

Since party 2 has the same reference point as party 1, accepting the offer leads to no aggrievement. If he rejects the offer, he obtains

$$\begin{aligned} U_2((\pi - I - P, P) \mid r_{NSSB}^m, r_{NSSB}^m) &= P + n\left(P \mid \frac{\pi}{2} - I\right) + \theta n\left(\pi - I - P \mid \frac{\pi}{2}\right) \\ &= P + \eta(1 - \theta\lambda)\left\{P - \left(\frac{\pi}{2} - I\right)\right\} \equiv U_{NSSB}^{m'}. \end{aligned}$$

[16] Assuming $w_1' = w_2' = (0,0)$ does not affect the result.

The comparison between U^{m}_{NSSB} and $U^{m'}_{NSSB}$ suggests that party 2 does not reject the offer if $\theta \geq (1+\eta)/\eta\lambda$ holds. Similarly, under integration, if party 2 accepts the order, his utility is

$$U_2(x^h \mid r^h_{NSSB}, r^h_{NSSB}) = -I + n(-I \mid -I) + \theta n(\pi \mid \pi) = -I \equiv U^h_{NSSB}.$$

If party 2 rejects the order, he obtains:

$$\begin{aligned} U_2((\pi - I - P, P) \mid r^h_{NSSB}, r^h_{NSSB}) &= P + n(P \mid -I) + \theta n(\pi - I - P \mid \pi) \\ &= P + \eta(1 - \theta\lambda)(P + I) \equiv U^{h'}_{NSSB}. \end{aligned}$$

Hence, we find that party 2 does not reject the order if $\theta \geq (1+\eta)/\eta\lambda$ holds. These discussions imply that the choice of the governance structure does not affect the timing of the agreement when there is no self-serving bias.

This result is explained as follows. Without self-serving bias, both parties share the same reference point, and hence, $|r^g_{NSSB1} - (\pi - I - P)|$ and $|r^g_{NSSB2} - P|$ become the same, where r^g_{NSSBi} represents party i's reference point payoff under governance structure g. Party 2 thus rejects the offer/order if the marginal benefit from rejecting the offer/order (i.e., $1+\eta$) is larger than or equal to the marginal cost from doing so (namely, $\theta\eta\lambda$).

A.3 No Shading

Lastly, we examine the case in which there is no shading. This case corresponds to the one in which there is no punishment for rejecting an offer/order and party i's utility function is characterized as follows:

$$U_i(y \mid r_i) = y_i + n(y_i \mid r_{ii}),$$

where

$$n(y_i \mid r_{ii}) = \begin{cases} \eta(y_i - r_{ii}) & \text{if } y_i \geq r_{ii} \\ \eta\lambda(y_i - r_{ii}) & \text{if } y_i < r_{ii}. \end{cases}$$

Since there is no shading, party i's utility does not depend on his partner's reference point. This formulation corresponds to the simple version of Köszegi and Rabin's reference-dependent utility function.

Since there is no punishment for party 2's rejection of an offer/order, he rejects any offer/order that yields him a smaller payoff than his continuation payoff. Given that party 1 believes that the continuation payoff is given by r_1^m (her reference point when the parties are autonomous), she optimally offers $x_{NS} = (\pi/2, \pi/2 - I)$ under both non-integration and integration.

Under non-integration, if party 2 accepts the offer, he receives

$$U_2(x_{NS} \mid r_2^m) = \frac{\pi}{2} - I + n\left(\frac{\pi}{2} - I \mid \frac{\pi - I}{2}\right) = \frac{\pi}{2} - I - \frac{\eta\lambda}{2} I \equiv U_{NS}^m,$$

and if he rejects it, his utility is

$$U_2((\pi - I - P, P) \mid r_2^m) = P + n\left(P \mid \frac{\pi - I}{2}\right) = P - \eta\lambda\left(\frac{\pi - I}{2} - P\right) \equiv U_{NS}^{m'}.$$

Note that there is no shading even if the offer that corresponds to party 1's reference point is rejected. By assumption $P > \pi/2 - I$, U_{NS}^m is smaller than $U_{NS}^{m'}$, which means that party 2 always rejects the offer.

Under integration, on the other hand, if party 2 accepts the order, his utility is given by[17]

$$U_2(x_{NS} \mid r_2^h) = \frac{\pi}{2} - I + n\left(\frac{\pi}{2} - I \mid 0\right) = (1 + \eta)\left(\frac{\pi}{2} - I\right) \equiv U_{NS}^h.$$

If he rejects the order, he enjoys

$$U_2((\pi - I - P, P) \mid r_2^h) = P + n(P \mid 0) = (1 + \eta)P \equiv U_{NS}^{h'}.$$

Since $P > \pi/2 - I$, $U_{NS}^h < U_{NS}^{h'}$ always holds. That is, under integration, party 2 rejects the order for certain.

The above discussion implies that the governance structure does not matter if there is

[17] From the discussion of the optimal ordering strategy, some readers might suspect that without shading, each party's reference point under integration becomes the same as that under non-integration. Nevertheless, such a change does not affect our discussion.

no shading. This is quite intuitive: party 2's rejection cannot be prevented without any punishment for it.

Appendix B: Risk-Averse Parties

We here examine a different type of no reference dependence. More specifically, in this appendix, we assume that the parties are risk-averse instead of assuming that they have reference-dependent preferences and are risk-neutral.

Suppose that each party i has concave utility function $m(x)$, which is twice differentiable ($m'(\cdot) > 0$ and $m''(\cdot) < 0$), and his overall utility is

$$U_i(x = (x_i, x_j) \mid r_i, r_j) = m(x_i) + \theta \min\{m(x_j) - m(r_{jj}), 0\}.$$

This utility function is similar to that of contracts as reference points. This change in the assumption does not affect party 1's optimal offer, and thus, we need to analyze party 2's behavior only.

Under non-integration, party 2's acceptance utility is

$$U_2(x^m = r_1^m \mid r_1^m, r_2^m) = m(r_{12}^m).$$

Note that party 1's optimal offer is equivalent to her reference point, $x^m = r_1^m = (r_{11}^m, r_{12}^m)$, and party 2 has a reference point $r_2^m = (r_{21}^m, r_{22}^m)$. If party 2 rejects the offer, his utility is

$$U_2((\pi - I - P, P) \mid r_1^m, r_2^m) = m(P) - \theta\{m(r_{11}^m) - m(\pi - I - P)\}.$$

Party 2 does not reject the offer under non-integration if his acceptance utility is larger than or equal to his expected continuation utility:

$$\begin{aligned} m(r_{12}^m) &\geq m(P) - \theta\{m(r_{11}^m) - m(\pi - I - P)\} \\ \Leftrightarrow \qquad \theta &\geq \frac{m(P) - m(r_{12}^m)}{m(r_{11}^m) - m(\pi - I - P)} \equiv \theta'_m. \end{aligned}$$

Under integration, party 2's compliance utility is

$$U_2(x^h = r_1^h \mid r_1^h, r_2^h) = m(r_{12}^h),$$

and his rejection utility is

$$U_2((\pi - I - P, P) \mid r_1^h, r_2^h) = m(P) - \theta\{m(r_{11}^h) - m(\pi - I - P)\}.$$

Hence, party 2 does not reject the order if the following condition holds:

$$\begin{aligned} m(r_{12}^h) &\geq m(P) - \theta\{m(r_{11}^h) - m(\pi - I - P)\} \\ \Leftrightarrow \qquad \theta &\geq \frac{m(P) - m(r_{12}^h)}{m(r_{11}^h) - m(\pi - I - P)} \equiv \theta'_h. \end{aligned}$$

We thus determine

$$\theta'_m < \theta'_h$$

because $m(\cdot)$ is concave and the following relationships hold:

$$r_{12}^h < r_{12}^m < P \leq \pi - I - P < r_{11}^m < r_{11}^h.$$

This discussion implies that non-integration achieves immediate agreement more easily than integration, which means that our main result cannot be obtained by assuming risk-averse parties.

In the main model, party 1's punishment for rejecting the order under integration is severer than her shading under non-integration because both parties' reference points are employer-favored under integration. In the risk-averse case, however, the same factor leads to the opposite result. This is illustrated in Figure A. Since the parties have concave utility functions, the same amount of payoff increase/decrease affects their utility differently. Under integration, the amount of party 2's payoff improvement from rejecting the order ($P-(-I)=P+I$) is the same as that of party 1's payoff decrease ($\pi-(\pi-I-P)=P+I$). Nevertheless, the amount of party 2's utility improvement from his rejection, which corresponds to (b) in Figure A, is far larger than that of party 1's utility decrease from it, which is denoted by (a) in Figure A. Under non-integration, on the other hand, party 1's utility decrease from party 2's rejection ($\pi/2-(\pi-I-P)=P+I-\pi/2$), which is denoted by (c) in Figure A, is not so small compared with party 2's utility improvement from it ($P-(\pi/2-I)=P+I-\pi/2$), which corresponds to (d) in Figure A. Hence, integrated firms require much severer punishment for party 2's rejection to offset party 2's benefit from it than autonomous trading parties do.

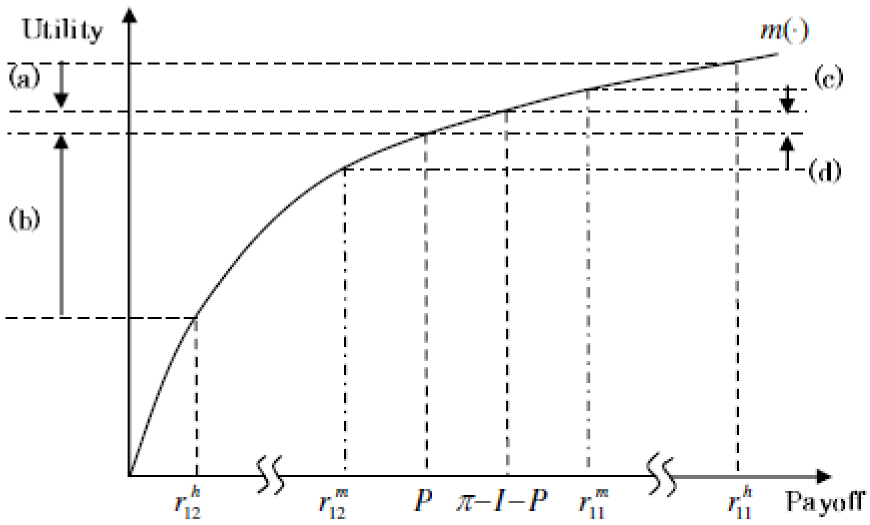

Figure A: Each Party's Utility Improvement/Decrease from Party 2's Rejection

(a) (resp. (b)) represents party 1's utility decrease (resp. party 2's utility improvement) under integration. (c) (resp. (d)) denotes 1's utility decrease (resp. 2's utility improvement) under non-integration.

Appendix C: Parties Who Care about Discounting

This section studies the case in which the parties care about discounting and checks the robustness of our result. To achieve this, we change the setting in the following way (the rest of the settings are the same as in the main model). First, the parties do care about discounting. That is, they share a common discount factor δ and their payoffs are discounted if they cannot reach agreement immediately; this is common knowledge. Second, we assume that the following condition holds:[18]

$$\frac{\delta\pi}{2} - I < \delta P \leq \frac{\delta\pi - I}{2}.$$

[18] We continue to assume that the following condition holds:

$$\pi / 2 - I < P \leq (\pi - I)/2 .$$

The first inequality implies that party 2 has an incentive to reject party 1's offer/order that corresponds to party 1's reference point (each party's reference point will be specified below). The second inequality means that party 2 does not expect more than what he thinks he is entitled to obtain (namely, his reference point payoff). This condition also implies that $I/(\pi - 2P) \leq \delta(\leq 1)$.

This appendix proceeds as follows. Section C.1 specifies each party's reference point and party 1's optimal offer/order under each governance structure. Section C.2 studies party 2's optimal acceptance/compliance decision under each governance structure. Section C.3 presents the result, which is a modified version of Proposition 1.

C.1 Reference Points and Party 1's Optimal Offer/Order

We first specify each party's reference point and party 1's optimal offer/order under each governance structure. It is common knowledge that the parties care about discounting, and hence, their reference points are different from those in the main model. Since both parties expect that party 1 sends the offer which makes party 2 indifferent about whether he accepts it and party 2 accepts such an offer, party 1's reference point under non-integration is

$$r_1^{m*} = (r_{11}^{m*}, r_{12}^{m*}) = \left(\pi - I - \left(\frac{\delta\pi}{2} - I\right), \frac{\delta\pi}{2} - I\right).$$

Note that the expected bargaining outcome is given by the Nash bargaining solution and party 1 believes that party 2 is to incur his sunk investment (i.e., she believes that the parties' outside options are $w_1 = (w_{11}, w_{12}) = (0, -I)$).

As mentioned in the main model, since party 1 believes that both parties share the same reference point r_1^{m*}, it is optimal for her to offer what her reference point specifies. Thus, her optimal offer is given by

$$x^{m*} = (x_1^{m*}, x_2^{m*}) = \left(\pi - I - \left(\frac{\delta\pi}{2} - I\right), \frac{\delta\pi}{2} - I\right) = r_1^{m*}.$$

Under integration, on the other hand, party 1's reference point is

$r_1^{h*} = (r_{11}^{h*}, r_{12}^{h*}) = (\pi, -I)$, which is the same as in the main model, because there is no room for her to demand more. The optimal order, which is equal to party 1's reference point, is thus given by

$$x^{h*} = (x_1^{h*}, x_2^{h*}) = (\pi, -I) = r_1^{h*}.$$

We then determine party 2's reference point. Party 2 infers that party 1's offer makes him indifferent about whether he accepts it. However, he believes that the parties' outside options are $w_2 = (w_{21}, w_{22}) = (0,0)$. Thus, his reference point under non-integration is given by

$$r_2^{m*} = (r_{21}^{m*}, r_{22}^{m*}) = \left(\pi - I - \left(\frac{\delta\pi - I}{2}\right), \frac{\delta\pi - I}{2}\right).$$

Party 2's reference point under integration, on the other hand, is

$$r_2^{h*} = (r_{21}^{h*}, r_{22}^{h*}) = (\pi - I, 0).$$

C.2 Party 2's Acceptance/Compliance

We first study party 2's optimal acceptance decision under non-integration given party 1's optimal offer x^{m*} and the parties' reference points, r_1^{m*} and r_2^{m*}. If he accepts the offer, he obtains payoff $\delta\pi/2 - I$, which leads to his sense of loss $\eta\lambda\{(\delta\pi - I)/2 - (\delta\pi/2 - I)\} = \eta\lambda I/2$, and incurs no shading from party 1. Party 2's utility is thus given by

$$U_2(x^{m*} \mid r_1^{m*}, r_2^{m*}) = \frac{\delta\pi}{2} - I + n\left(\frac{\delta\pi}{2} - I \mid \frac{\delta\pi - I}{2}\right) + \theta \cdot 0 = \frac{\delta\pi}{2} - I - \frac{\eta\lambda}{2} I \equiv U_{Dis}^m.$$

If he rejects the offer, on the other hand, his utility is

$$\begin{aligned}
& U_2((\delta\pi - I - \delta P, \delta P) \mid r_1^{m*}, r_2^{m*}) \\
= \; & \delta P + n\left(\delta P \mid \frac{\delta\pi - I}{2}\right) + \theta n\left(\delta\pi - I - \delta P \mid \pi - \frac{\delta\pi}{2}\right) \\
= \; & \delta P - \eta\lambda\left(\frac{\delta\pi - I}{2} - \delta P\right) - \theta\eta\lambda\left\{\left(1 - \frac{3\delta}{2}\right)\pi + I + \delta P\right\} \equiv U_{Dis}^{m'}.
\end{aligned}$$

Note that party 2's expectation about the continuation outcome is discounted since the parties care about discounting. We thus find that party 2 does not reject the offer if the following condition holds:

$$U_{Dis}^{m} \geq U_{Dis}^{m'} \quad \Leftrightarrow \quad \theta \geq \frac{(1+\eta\lambda)\left\{\delta P - \left(\frac{\delta}{2}\pi - I\right)\right\}}{\eta\lambda\left\{\delta P + \left(1 - \frac{3\delta}{2}\right)\pi + I\right\}} \equiv \theta_m^*.$$

We next examine party 2's compliance decision under integration. If he accepts the optimal order x^{h*}, his utility is

$$U_2(x^{h*} \mid r_1^{h*}, r_2^{h*}) = -I + n(-I \mid 0) + \theta \cdot 0 = -(1+\eta\lambda)I \equiv U_{Dis}^{h}.$$

If he rejects it, on the other hand,

$$\begin{aligned} U_2((\delta\pi - I - \delta P, \delta P) \mid r_1^{h*}, r_2^{h*}) &= \delta P + n(\delta P \mid 0) + \theta n(\delta\pi - I - \delta P \mid \pi) \\ &= (1+\eta)\delta P - \theta\eta\lambda\{\pi - (\delta\pi - I - \delta P)\} \equiv U_{Dis}^{h'}. \end{aligned}$$

Party 2 thus does not reject the order if

$$U_{Dis}^{h} \geq U_{Dis}^{h'} \quad \Leftrightarrow \quad \theta \geq \frac{(1+\eta)\delta P + (1+\eta\lambda)I}{\eta\lambda\{\pi - (\delta\pi - I - \delta P)\}} \equiv \theta_h^*.$$

C.3 Immediate Agreement and Governance Structures

Comparing θ_m^* and θ_h^* leads to the following result:

$$\begin{cases} \theta_m^* \geq \theta_h^* & \text{if} \quad \delta \geq \dfrac{\frac{1+\eta\lambda}{2}\pi^2 - \eta(\lambda-1)P(\pi+I)}{\frac{1+\eta\lambda}{2}\pi^2 - \eta(\lambda-1)P\left(\frac{3}{2}\pi - P\right)} \equiv \delta^* \\ \theta_m^* < \theta_h^* & \text{otherwise.} \end{cases}$$

Since $\delta^* < 1$ holds, the case in which $\theta_m^* \geq \theta_h^*$ does exist. We thus have the following proposition:

Proposition 3: *When the parties care about discounting, integration achieves immediate agreement more easily than non-integration ($\theta_h^* \leq \theta_m^*$) if and only if the following condition holds:*

$$\delta \geq \max\left[\delta^*, \frac{I}{\pi - 2P}\right],$$

where

$$\delta^* \equiv \frac{\frac{1+\eta\lambda}{2}\pi^2 - \eta(\lambda-1)P(\pi+I)}{\frac{1+\eta\lambda}{2}\pi^2 - \eta(\lambda-1)P\left(\frac{3}{2}\pi - P\right)}.$$

This implies that when the cost of delay is not so large, integration achieves immediate agreement more easily than non-integration. When the parties care about discounting, party 2 faces two costs from rejecting the offer/order: punishment for rejection (party 1's shading) and the cost of delay. As discussed in the main model, the punishment under integration is much severer than that under non-integration. This implies that the cost of delay has an insignificant effect on party 2's utility compared to party 1's shading under integration. Hence, if integration achieves faster agreement than non-integration, the cost of delay must be small enough to have little effect on the parties' utility under either governance structure (i.e., δ is close enough to 1).

Appendix D: Altruism

In this appendix, we examine the case in which the parties are altruistic. That is, each party i considers party j's gain and loss. (In the main model, party i does not care j's gain.) In such a case, each party i's utility is given by

$$U_i(y \mid r_i, r_j) = y_i + n_i^i(y_i \mid r_{ii}) + \theta n_j^i(y_j \mid r_{jj})$$

where

$$n_j^i(y_j \mid r_{jj}) = \begin{cases} \eta(y_j - r_{jj}) & \text{if } y_j \geq r_{jj} \\ \eta\lambda_{ij}(y_j - r_{jj}) & \text{if } y_j < r_{jj}. \end{cases}$$

$\lambda_{ij}(>1)$ represents party i's sensitivity to j's loss and we assume that $\lambda_{ii} > \lambda_{ij}$ and $\lambda_{11} = \lambda_{22}$ for simplicity. We further assume that $\theta \leq (1+\eta\lambda_{ii})/\eta\lambda_{ij}$, which corresponds to the assumption $\theta \le (1+\eta\lambda)/\eta\lambda$ in the main model. In this setting, θ can be considered each party's level of altruism.

We can easily check that such a change in each party's utility function does not change party 1's optimal offer/order (i.e., she offers what her reference point specifies). Hence, we only need to examine party 2's acceptance/compliance decision given that party 1 offers $x^m = (\pi/2, \pi/2 - I)$ under non-integration and order $x^h = (\pi, -I)$ under integration.

Under non-integration, if party 2 accepts x^m, then his utility is given by

$$U_2(x^m \mid r_1^m, r_2^m) = \frac{\pi}{2} - I + n\left(\frac{\pi}{2} - I \mid \frac{\pi - I}{2}\right) + \theta n\left(\frac{\pi}{2} \mid \frac{\pi}{2}\right) = \frac{\pi}{2} - I - \frac{\eta\lambda_{22}}{2} I.$$

Note that party 2's reference point is $r_2^m = \{(\pi - I)/2, (\pi - I)/2\}$. If he rejects the offer, on the other hand, his utility is

$$\begin{aligned} U_2((\pi - I - P, P) \mid r_1^m, r_2^m) &= P + n\left(P \mid \frac{\pi - I}{2}\right) + \theta n\left(\pi - I - P \mid \frac{\pi}{2}\right) \\ &= P - \eta\lambda_{22}\left(\frac{\pi - I}{2} - P\right) - \theta\eta\lambda_{21}\left\{\frac{\pi}{2} - (\pi - I - P)\right\}. \end{aligned}$$

Thus, we can determine that party 2 accepts the offer if the following condition holds:

$$\theta \ge \frac{1 + \eta\lambda_{22}}{\eta\lambda_{21}} \equiv \theta_m^A.$$

Note that θ_m^A corresponds to θ_m in our main model.

We then analyze party 2's compliance strategy under integration given that party 1's order is $x^h = r_1^h = (\pi, -I)$ and party 2's reference point is $r_2^h = (\pi - I, 0)$. If he accepts the order, his utility is

$$U_2(x^h \mid r_1^h, r_2^h) = -I + n(-I \mid 0) + \theta n(\pi \mid \pi) = -(1 + \eta\lambda_{22})I.$$

If he rejects the order, on the other hand, his utility is given by

$$\begin{aligned} U_2((\pi - I - P, P) \mid r_1^h, r_2^h) &= P + n(P \mid 0) + \theta n(\pi - I - P \mid \pi) \\ &= (1 + \eta)P - \theta\eta\lambda_{21}\{\pi - (\pi - I - P)\}. \end{aligned}$$

Hence, party 2 accepts the order if

$$\theta \geq \frac{(1+\eta)P + (1+\eta\lambda_{22})I}{\eta\lambda_{21}(P+I)} \equiv \theta_h^A.$$

θ_h^A corresponds to θ_h in our main model.

We can easily determine that $\theta_m^A > \theta_h^A$. This implies that immediate settlement under non-integration requires party 2 to be more altruistic than immediate agreement under integration. We thus find that our main message (i.e., integration can achieve immediate agreement more easily than non-integration) continues to emerge under the altruism case.

Chapter 4

A Dynamic Model of Firm Boundaries: Why Do Firm Boundaries Waver?

1 Introduction

Some studies point out that TCE, which focuses on static assignment of transaction-cost-minimizing governance structure to each transaction, cannot address dynamic problems (e.g., Dow, 1987 and Langlois and Robertson, 1995). In fact, the models we presented in the previous chapters focus on static analysis of make-or-buy decisions and do not fit to deal with dynamic problems.

In this chapter, we analyze dynamic changes in firm boundaries by providing a multi-generation model. Especially, we address a question why some firms waver their boundaries between non-integration and integration (i.e., choose non-integration and integration alternately), which TCE has had trouble explaining.

Our model formalizes the situation in which each generation chooses either non-integration or integration to govern a transaction. The game consists of decision and transition phases: each generation t chooses either non-integration or integration in the decision phase, and then receives his payoff, which is observed by generation $t+1$, and exits in the transition phase. Each generation t (if rational) chooses governance structure so as to minimize governance costs given that generation $t-1$'s payoff, the history of predecessors' choices of governance structure, and the level of relationship-specificity, which is chosen by generation 0.

Non-integration and integration come with different governance costs. That is, non-integration may suffer from hold-up costs, which are increasing in the level of relationship-specificity but do not necessarily emerges. Integration, on the other hand, never incurs hold-up costs but cannot avoid bureaucratic costs (e.g., incentive degradation and logrolling).

The key assumption of our model is that some generations choose governance structure without knowing the reasons for predecessors' choices of governance structure. More specifically, each generation t whose predecessor $t-1$ chose integration cannot infer why $t-1$ chose integration. This assumption is explained as follows. If generation $t-1$ chose integration, $t-1$ never experiences hold up, and thus generation t cannot infer from generation $t-1$'s payoff whether hold up occurs. This unobservability makes generation t uncertain about whether generation $t-1$ chose integration rationally or not (e.g., hold up did not occur, but generation $t-1$ simply sought to leave his distinctive mark by changing his predecessor $t-2$'s organizational choice).

This assumption of the unrecorded reasons stems from the impression that some firms choose distinct structural alternatives alternately as if they did not know or care about the problems which triggered structural changes before. For example, Nickerson and Zenger (2002) and Boumgarden, Nickerson, and Zenger (2012) report that Hewlett-Packard chose centralization and decentralization alternately over 25 years (pre-1982 to 2008). According to their studies, while all the changes from decentralization to centralization were triggered by low coordination, which was the cost of decentralization, all the changes from centralization to decentralization were caused by inflexibility and low innovation, which were the cost of centralization. Nickerson and Zenger (2002) and Boumgarden, Nickerson, and Zenger (2012) assert that such ``organizational vacillation" is an effort to achieve dynamic balance between coordination, innovation, and flexibility (we will review their studies in the next section). Nevertheless, we wonder why no attempt to achieve first-best solution, namely static balance of them (e.g., an effort to form a hybrid structure between centralization and decentralization), is observed in their case study of HP.[1] We then hit on an idea that

[1] Nickerson and Zenger (2002) point out that ``complementarities logic" makes the choice of hybrid structures difficult. That is, some elements and decision variables need to be consistently determined and inconsistent combinations suffer from poor performance or simply die out. See Nickerson and Zenger (2002) for the detail

each manager did not know the problems which triggered structural changes before and such wavering behavior is caused by experimentation (an effort to figure out which choice triggers what problems).

We show that under the assumption of unrecorded reasons for predecessors' governance choice, non-integration and integration can be optimally chosen alternately in equilibrium. Given that the reason for each generation's choice of governance structure may not be transferred between generations, the level of relationship-specificity, from which each generation infers which governance structure is optimal, plays an important role. If the level of relationship-specificity is high (resp. low) enough, non-integration is likely to be more (resp. less) costly than integration, and hence each generation (if rational) chooses integration (resp. non-integration), which achieves actual transaction-cost minimization with high probability. However, if the level of relationship-specificity is intermediate, the governance choice which expectedly minimizes transaction costs is likely to fail in actual transaction-cost minimization. Hence, an effort to find out which governance structure is optimal is required and leads to wavering between non-integration and integration. Our model provides formal explanations for why organizational changes often follow management turnovers and why it is hard for some integrated firms to disintegrate even if integration is not optimal.

The rest of the chapter is organized as follows. Section 2 relates this chapter to existing literature. In Sections 3, we introduce our model. Section 4 (resp. Section 5) examines the dynamic changes in firm boundaries when relationship-specificity is constant (resp. when relationship-specificity changes). Section 6 applies our results to the choice between centralization and decentralization and provides an illustration of our results by focusing on the history of Hewlett-Packard's organizational changes, which is based on Boumgarden, Nickerson, and Zenger (2012). Section 7 concludes the chapter.

---

explanation for it.

2 Related Literature: Organizational Vacillation

The question why some firms choose distinct institutional arrangements, such as non-integration or integration and centralization or decentralization, alternately is relatively unexplored. The literature on ``organizational vacillation," including Nickerson and Zenger (2002) and Boumgarden, Nickerson, and Zenger (2012), focuses on the choice between centralization and decentralization and considers such vacillation an effort to achieve balance between exploration, which ``includes things captured by terms such as search, variation, risk taking, experimentation, play, flexibility, discovery, innovation" (March, 1991, p.71), and exploitation, which ``includes such things as refinement, choice, production, efficiency, selection, implementation, execution" (p.71).

The problem with balancing exploration and exploitation is that they work as complements in improving performance but require quite opposite organizational designs (i.e., centralization promotes exploitation, but decentralization promotes exploration). They assert that ``Organizational vacillation is a dynamic approach to achieving high performance through simultaneously high levels of exploration and exploitation" (Boumgarden, Nickerson, and Zenger, 2012, p.591). That is, firms can dynamically achieve ``high levels of both exploration and exploitation by temporally and sequentially alternating between organizational structures that promote either exploration or exploitation, respectively" (Boumgarden, Nickerson, and Zenger, 2012, p.588).

Such dynamic balance in exploration and exploitation is achievable because while formal organizational structures change discretely (e.g., decentralization or centralization), informal organizations, which the levels of exploration and exploitation reflect, ``adjust more continuously in response to structural shifts" (Boumgarden, Nickerson, and Zenger, 2012, p.591). Such informal organization includes ``the routines, decision-making processes, and knowledge flows within the organization, as well as the general behaviors, decisions, and actions of individuals within the organization"

(Boumgarden, Nickerson, and Zenger, 2012, p.591). Thus ``Repeated modulation ... produces brief periods of dual capability" (Boumgarden, Nickerson, and Zenger, 2012, p.592).

Although the aims of their approach (i.e., examine the choice between centralization and decentralization by focusing on ``inertia in the informal organization") and ours (i.e., analyze dynamic changes in firm boundaries by focusing on unrecorded reason for predecessors' choices), we believe that our analysis may complement their explanation for organizational vacillation, which will be illustrated in Section 6 by focusing on Hewlett-Packard case.

3 The Model

In this section, we introduce a multi-generation model in which each generation chooses how to govern a trade: either non-integration or integration. The key assumption of our model is that the reason for each generation's choice of governance structure is not recorded, and hence successors may not be able to understand why his predecessors chose non-integration or integration. We show that such unrecorded reasons for the choices of governance structure may lead to experimentation which triggers swings between non-integration and integration in equilibrium. Our model provides formal explanations for why organizational changes often follow management turnovers and why inefficient choice of integration can persist.

Our multi-generation model consists of decision and transition phases. Each generation t chooses either non-integration or integration in the decision phase, and then receives his payoff, which is observed by generation $t+1$, and exits in the transition phase. Specifically, the game proceeds as follows: [D. t] represents generation t's decision phase and [T. $t-(t+1)$] denotes transition phase from generation t to $t+1$,

[D. 0] Given non-integration, generation 0 chooses the level of relationship-specificity of the trade.[2]

[T. 0-1] Generation 0 receives his payoff and exits, and generation 1 observes 0's payoff.

[D. 1] Given generation 0's payoff and the level of relationship-specificity, generation 1 chooses whether to integrate or not.

[T. 1-2] Generation 1 receives his payoff and exits, and generation 2 observes 1's payoff.

[D. 2] Given generation 1's payoff, the predecessor's choice (non-integration or integration), and the level of relationship-specificity, generation 2 chooses whether to integrate or not.

[T. 2-3] Generation 2 receives his payoff and exits, and generation 3 observes 2's payoff.

$\vdots$

[D. n] Given generation $n-1$'s payoff, the history of the predecessors' choices (non-integration or integration), and the level of relationship-specificity, generation n chooses whether to integrate or not.

[T. n-(n+1)] Generation n receives his payoff and exits, and generation $n+1$ observes n's payoff.

$\vdots$

Let $k \in \Re_+$ be the index of relationship-specificity and $k = 0$ corresponds to the case where the trade is not relationship-specific at all. Intuitively, the level of relationship-specificity is determined by how specific generation 0's ex ante investments, which is required for a trade in question, are. We want to emphasize that under-investment problems are not dealt with.

[2] Following TCE's classic assumption that ``in the beginning there were markets" (Williamson, 1975, p.20), we assume that the game starts with non-integration. Nevertheless, our result does not change qualitatively if the game starts with integration.

Each generation's payoff is given by governance costs, which depend on the choice of governance structure. Following Williamson (1996, ch.3), $M(k)$ denotes the governance costs of markets and $B(k)$ represents those of hierarchies, and $M(0) < B(0)$ and $M'(k) < B'(k)$ hold. Under non-integration, as the literature on TCE points out, when trading parties are locked in (e.g., due to relationship-specific assets), ex post contract renegotiation may invite hold-up behavior.[3] We thus assume that market governance cost is $M(k)$ if hold up occurs and 0 otherwise. We further assume that whether hold up occurs is determined exogenously when generation 0's payoff is realized (i.e., [T. 0-1]), and if hold up occurs (resp. did not occur) in [T. 0-1], then future generations who choose non-integration suffer (resp. do not suffer) from hold up unless the level of relationship-specificity k changes.

Under integration, on the other hand, ex post adaptations are implemented by fiat, and thus no hold up occurs. Nevertheless, as Williamson (1985, ch.6) points out, internal organizations suffer from the costs of bureaucracy (e.g., incentive degradation and logrolling). Hence, we assume that such bureaucratic costs $B(k)$ cannot be avoided as long as integration is chosen.

Each generation becomes either rational type (with probability $1-q$) or appeal type (with probability q). The probability of being appeal type q is common knowledge. While the rational type generations are assumed to maximize their (expected) payoffs (i.e., minimize governance costs), appeal type generation t never chooses the same governance structure as generation $t-1$ (i.e., if generation $t-1$ chose non-integration, t chooses integration, and vice versa). Intuitively, this assumption reflects the possibility that ``[organizational] vacillation may reflect management turnover--new managers seeking to leave their distinctive mark by initiating organizational change" (Nickerson and Zenger, 2002, p.548). What is crucial to our result is that the changes from non-integration to integration do not necessarily take

[3] Note that, in this chapter, the term ``hold up" does not mean ex ante under-investment problem.

place for rational reasons. Thus, our result continues to hold if we introduce imperialist type, who always chooses integration, instead of the appeal type.

We assume that if generation $t-1$ chose integration, generation t cannot infer whether hold up occurs from $t-1$'s payoff. In other words, generation t cannot infer whether generation $t-1$ chose integration rationally. It follows because integrated firms do not suffer from hold up and generation $t-1$'s payoff does not include any information about the occurrence of hold up.[4] Information available to each generation t thus depends on generation $t-1$'s choice of governance structure. If generation $t-1$ chooses non-integration, the following information is available to generation t: (a) the level of relationship-specificity k, (b) $M(k)$ and $B(k)$, (c) the history of the predecessors' choices of governance structure, (d) generation $t-1$'s payoff (payoffs of generations 0,1,..., and $t-2$ are not available), (e) the probability with which each generation becomes appeal type q, (f) common prior belief about how likely hold up is to occur given k, which is denoted by $p(k)$ ($p'(k)>0$), and (g) whether hold up occurs or not.[5] If generation $t-1$ chooses integration, on the other hand, information (a)-(f) is available to generation t, but (g) is not. Which information is available to each generation is common knowledge. Let μ_t denote generation t's posterior belief about how likely hold up is to occur. μ_t is determined by Bayes' rule.

To focus on governance-cost minimization, we assume that generation 0 is indifferent among his choices of k: $V(k)-p(k)M(k)=V(k')-p(k')M(k')$ holds for all $k,k'\in\Re_+$ where $V(k)$ denotes the trade value given the relationship-specificity k and $V'(k)>0$ holds. Note that generation 0 chooses k given non-integration. We

[4] Some public information made by predecessors, such as annual reports, might be available to successors and help them infer the reason for predecessors' choice of governance structures. However, such information is not necessarily reliable: ``Particular care must be exercised in the interpretation of annual reports that are biased in presentation" (Boumgarden, Nickerson, and Zenger, 2012, p.607).

[5] Some accounting information of generations 0, 1,..., and $t-2$, such as income statements, might be available. Nevertheless, such information may reflect not only the occurrence of hold up but other factors (e.g., some exogenous shocks). Thus, it is hard for generation t to infer whether earlier generations experienced hold up from such information.

assume that generation 0 is benevolent in the sense that he chooses k so as to minimize inefficiencies due to his successors' maladaptation (i.e., minimize the probability of each generation's (if rational) wrong choice of governance structure).

Analyses proceed as follows. Section 4 focuses on the case where the level of relationship-specificity is constant and shows our main result. Section 5 then analyzes the situation where relationship-specificity changes and extends the result of Section 4. Section 6 applies our results to analyze the choice between centralization and decentralization and briefly illustrates our results.

4 Analysis 1: Relationship-Specificity k Is Constant

We here focus on the situation where the level of relationship-specificity k is constant. Since the appeal type behaves mechanically, we only need to examine the optimal behavior of rational generation t. The behavior of rational generation t depends on generation $t-1$'s choice of governance structure. Hence, we first analyze the case where generation $t-1$ chose non-integration, and then the case where generation $t-1$ chose integration.

4.1 The Case Where Generation $t-1$ Chose Non-Integration

When generation $t-1$ chose non-integration, generation t can understand whether hold up occurs by observing $t-1$'s payoff: $-M(k)$ or 0. That is, $-M(k)$ (resp. 0) implies that hold up occurs (resp. no hold up occurs). Thus, the rational choice of generation t is as follows:

$$\begin{cases} \text{Non-Integration} & \text{if } -M(k) \geq -B(k) \text{ or no hold up occurs,} \\ \text{Integration} & \text{if } -M(k) < -B(k). \end{cases}$$

4.2 The Case Where Generation $t-1$ Chose Integration

If generation $t-1$ chose integration, he does not suffer from hold up and obtains

$-B(k)$. Since generation t observes $-B(k)$, which does not tell him anything about whether hold up occurs (i.e., whether market governance cost is $-M(k)$ or 0), he cannot understand whether hold up triggered generation $t-1$'s choice of integration.

This case includes three subcases: (1) non-integration was chosen for several generations in a row before generation t, (2) not non-integration but integration was chosen for several generations in a row before generation t, and (3) non-integration and integration were chosen alternately, namely each generation i changed generation $i-1$'s governance structure ($i<t$).

4.2.1 Subcase (1): non-integration was chosen for several generations in a row before generation t

Suppose that generations $j-1$ and j chose non-integration ($j<t$). Since the appeal type always changes governance structure, generation t can infer that generation j's choice was rational. Given that generation j's choice of non-integration was rational, generation t also infers that either of the following two conditions must hold:

$$\begin{cases} \text{(1) no hold up occurs} \quad \text{or} \\ \text{(2) } -M(k) \geq -B(k) \text{ holds.} \end{cases}$$

In either case, it is optimal for generation t (if rational) to choose non-integration.

4.2.2 Subcase (2): not non-integration but integration was chosen for several generations in a row before generation t

Suppose generations $j-1$ and j chose integration ($j<t$). Since the appeal type always changes governance structure, generation t can infer that generation j's choice was rational. Formally, the following condition holds:

$$-\mu_j M(k) < -B(k).$$

This suggests that observed changes in governance structure from integration to non-integration were not rational. Hence, generation t ignores such changes and focuses on how many times changes from non-integration to integration occurred before generation t to update his belief. If changes in governance structure from non-integration to integration occurred n times before generation t, generation t's rational choice is given by

$$\begin{cases} \text{Non-Integration} & \text{if } -\mu_t M(k) \geq -B(k) \\ \text{Integration} & \text{if } -\mu_t M(k) < -B(k) \end{cases}$$

where

$$\mu_t = \frac{p(k)}{p(k) + \{1 - p(k)\}q^n}.$$

Given that $-\mu_j M(k) < -B(k)$ holds, since $\mu_t > \mu_i$ holds for all $i < t$, $-\mu_t M(k) < -B(k)$ also holds. Thus, generation t (if rational) optimally chooses integration.

4.2.3 Subcase (3): each generation i changed generation $i-1$'s governance structure ($i < t$)

Generation t knows the history of predecessors' choices of governance structure, q, and $p(k)$, and hence infers generation $t-2$'s belief μ_{t-2} correctly. Generation t can then evaluate whether generation $t-2$'s choice of non-integration was rational:

$$\begin{cases} \text{Rational} & \text{if } -\mu_{t-2} M(k) \geq -B(k) \\ \text{Irrational} & \text{if } -\mu_{t-2} M(k) < -B(k). \end{cases}$$

Since the changes from integration to non-integration does not provide any additional information about the occurrence of hold up, $\mu_{t-2} = \mu_{t-1}$ holds regardless of generation $t-2$'s choice and μ_t is given by

$$\mu_t = \frac{\mu_{t-1}}{\mu_{t-1} + (1 - \mu_{t-1})q} (> \mu_{t-2}).$$

Thus, if $-\mu_{t-2} M(k) < -B(k)$ holds, since $\mu_t > \mu_{t-2}$, $-\mu_t M(k) < -B(k)$ also holds and generation t (if rational) optimally chooses integration. If $-\mu_{t-2}M(k) \geq -B(k)$ holds, on the other hand, generation t's rational choice is given by

$$\begin{cases} \text{Non-Integration} & \text{if } -\mu_t M(k) \geq -B(k) \\ \text{Integration} & \text{if } -\mu_t M(k) < -B(k). \end{cases}$$

From the discussions above, the following proposition summarizes the optimal choice of rational generation t:

Proposition 1: *Suppose the level of relationship-specificity k is constant.*

(a) If generation $t-1$ chose non-integration, generation t (if rational) chooses

$$\begin{cases} \textit{Non-Integration} & \textit{if } -M(k) \geq -B(k) \textit{ or no hold up occurs,} \\ \textit{Integration} & \textit{if } -M(k) < -B(k). \end{cases}$$

(b) If generation $t-1$ chose integration and generations $j-1$ and j chose non-integration ($j < t$), generation t (if rational) chooses non-integration.

(c) If generation $t-1$ chose integration and there is no $j(< t)$ such that generations $j-1$ and j chose non-integration, generation t (if rational) chooses

$$\begin{cases} \textit{Non-Integration} & \textit{if } -\mu_t M(k) \geq -B(k) \\ \textit{Integration} & \textit{if } -\mu_t M(k) < -B(k) \end{cases}$$

where

$$\mu_t = \frac{p(k)}{p(k) + \{1 - p(k)\}q^n}$$

and n denotes how many times the changes from non-integration to integration occurred before generation t.

This proposition implies that the lower (resp. higher) relationship-specificity k becomes, the less (resp. more) likely integration is chosen. This follows because $M(k)$,

$B(k)$, $p(k)$, and μ_t are all increasing in k . This is consistent with the main assertion of TCE and a number of empirical studies (see Lafontaine and Slade, 2007 for a survey of these empirical studies).

Furthermore, this proposition suggests that, if k is intermediate (i.e., k satisfies $-M(k) < -B(k)$ and $-\mu_2 M(k) \geq -B(k)$) and hold up occurs under non-integration, non-integration and integration are chosen alternately before certain generation t^* . This is explained as follows. Generation 1, who knows that hold up occurs, chooses integration since $-M(k) < -B(k)$. Generation 2 then cannot infer from generation 1's payoff whether hold up occurs and chooses non-integration since $-\mu_2 M(k) \geq -B(k)$ holds. Generation 3 chooses integration again because he knows that hold up occurs and $-M(k) < -B(k)$ holds.... This cycle continues until generation t^* , which satisfies $-\mu_{t*-1} M(k) \geq -B(k)$ and $-\mu_{t*} M(k) < -B(k)$, and each generation $i(\geq t^*)$ (if rational) chooses integration. This wavering behavior is caused by each generation's experimentation, namely an effort to figure out which structure triggers what problems. The assumption of unrecorded rationality for governance choice is thus crucial to this wavering behavior. If each generation knows whether the choice of non-integration triggers hold up, there is no need for such experimentation and each generation (if rational) chooses the same governance structure (i.e., generation t (if rational) chooses non-integration if $-M(k) \geq -B(k)$ or no hold up occurs, and integration if $-M(k) < -B(k)$). Our result also presents a formal explanation for why organizational changes often follow the changes in top managements. That is, new top management cannot completely understand the intention of his/her predecessors' choices of governance structure, and hence experimentation takes place and firm boundaries waver.

We can derive several implications from this proposition. First, when the choice of governance structure is discrete (i.e., no hybrid structure between non-integration and integration is available), intermediate level of relationship-specificity degrades performance (i.e., governance-cost minimization is not achieved when the level of relationship-specificity is intermediate), and hence the level of relationship-specificity should be either high or low. This observation is consistent with the argument about complementarity in Milgrom and Roberts (1990) and suggests that there is a complementarity between the level of relationship-specificity and the level of conscious coordination. Milgrom and Roberts (1990) points out that ``The defining characteristic of these groups of complements is that if the levels of any subset of the activities are increased, then the marginal return to increases in any or all of the remaining activities rises"(p.514). As mentioned above, each generation (if rational) chooses non-integration (resp. integration) if k is sufficiently low (resp. high), which leads to governance-cost minimization. However, if k is intermediate (i.e., k satisfies $-M(k) < -B(k)$ and $-\mu_2 M(k) \geq -B(k)$) and hold up occurs under non-integration, integration should be always chosen but several generations (including rational ones) choose non-integration before t^* (i.e., failure in governance-cost minimization occurs). This is consistent with Roberts (2004): ``when the variables are complements, it is quite possible that changing any one of them alone would worsen performance" (pp. 37-38).

Second, even if integration is not optimal, each generation can rationally continue inefficient integration. Suppose that generation 1 is appeal type and relationship-specificity k is high enough to satisfy $-\mu_2 M(k) < -B(k)$, but hold up does not occur. Appeal type generation 1 chooses integration in an attempt to leave his distinctive mark and since $-\mu_2 M(k) < -B(k)$ holds, the successors (if rational) continue to choose inefficient integration, which continues until the next appeal type generation appears. This explains why it is hard for some integrated firms to disintegrate even if integration is not optimal.

Third, our result implies that demands for organizational changes triggered by poor performance may make wavering behavior last longer. Poor performance often triggers organizational changes. In such cases, the working generation t observes his predecessor $t-1$'s low payoff which makes generation t more likely to think that generation $t-1$ was appeal type and his choice of governance structure was irrational (i.e., q is large). The larger q is, the smaller μ_t becomes, and hence the longer t^* gets.

5 Analysis 2: Relationship-Specificity k Changes

The previous section focused on the situation where the level of relationship-specificity is constant. However, in reality, unforeseen changes in relationship-specificity occur. For example, in Klein, Crawford, and Alchian's (1978) famous case of General Motors-Fisher Body relationship, dramatic decline in relationship specificity is illustrated. Klein, Crawford, and Alchian (1978) report that ``in 1919 General Motors entered a ten-year contractual agreement with Fisher Body for the supply of closed auto bodies" (p.308) to encourage Fisher Body's investment in specific stamping machines, which were important for metal closed body construction. Such a contractual arrangement was required because ``The original production process for automobiles consisted of individually constructed open, largely wooden, bodies" (p.308). However, ``The demand conditions facing General Motors and Fisher Body changed dramatically over the next few years. There was a large increase in the demand for automobiles and a significant shift away from open bodies to the closed body styles supplied by Fisher" (p.309). This section thus examines the situation where the level of relationship-specificity changes and extends Proposition 1.

The timing of the game changes as follows: suppose that the level of relationship-specificity changes from k to k' in generation n, then

[D. n] Given generation $n-1$'s payoff, the history of predecessors' choices (non-integration or integration), and the level of relationship-specificity k, generation n chooses whether to integrate or not.

——————Change in Relationship-Specificity : $k \to k'$——————

[T. $n-(n+1)$] Generation n receives his payoff under relationship-specificity k' and exits, and generation $n+1$ observes n 's payoff.

[D. $n+1$] Given generation n 's payoff, the history of predecessors' choices (non-integration or integration), and the history of relationship-specificity (i.e., k and k'), generation $n+1$ chooses whether to integrate or not.

[T. $(n+1)-(n+2)$] Generation $n+1$ receives his payoff and exits, and generation $n+2$ observes $n+1$'s payoff.

$\vdots$

We assume that whether hold up occurs under relationship-specificity k does not affect whether hold up occurs under k'. That is, generation $n+1$'s belief about whether hold up occurs is equal to the prior belief about how likely hold up is to occur under k', namely $p(k')$. Note that we employ such an assumption only for simplicity, and thus our result does not change if generation $n+1$'s belief about how likely hold up is to occur under k' is given by $p(k',\mu_t)$ where μ_t denotes posterior belief just before the change in relationship-specificity.

Since there is no change in the analysis, we can extend Proposition 1 to the case where relationship-specificity changes with slight modification:

Proposition 2: *Suppose relationship-specificity changes from k to k' in generation i. Each generation's governance choice before generation i is described in Proposition 1. Generation t's rational governance choice after generation i is summarized as follows:*

(a) If generation $t-1$ chose non-integration, generation t (if rational) chooses

$$\begin{cases} \textit{Non-Integration} & \textit{if } -M(k') \geq -B(k') \textit{ or no hold up occurs,} \\ \textit{Integration} & \textit{if } -M(k') < -B(k'). \end{cases}$$

(b) If generation $t-1$ chose integration and generation $j-1$ and j chose non-integration ($j \in \{i+1, i+2, \ldots, t-1\}$), generation t (if rational) chooses non-integration.

(c) If generation $t-1$ chose integration and there is no l such that generations

$l-1$ and l chose non-integration ($l \in \{i+1, i+2, \ldots, t-1\}$), generation t (if rational) chooses

$$\begin{cases} \textit{Non-Integration} & \textit{if } -\mu_t' M(k') \geq -B(k') \\ \textit{Integration} & \textit{if } -\mu_t' M(k') < -B(k') \end{cases}$$

where

$$\mu_t' = \frac{p(k')}{p(k') + \{1 - p(k')\} q^{n'}}$$

and n' denotes how many times the changes from non-integration to integration occurred between generations i and t.

Note that properties of Proposition 1 are preserved in Proposition 2 (e.g., the higher relationship-specificity becomes, the more likely integration is to be chosen). Propositions 1 and 2 imply that changes in relationship-specificity alter the pattern of transition between non-integration and integration, which is often observed in practice. The next section applies our results to the choice between centralization and decentralization and presents an example of such change in the patterns.

6 Application to Centralization and Decentralization: Hewlett-Packard

Our model presented in the previous sections is originally developed to analyze dynamic changes in firm boundaries, but we here show that it can be applied to other problems of institutional arrangements, namely the choice between centralization and decentralization.

To apply our results to the decisions whether firms should be centralized or decentralized, we need to reinterpret k (i.e., the level of relationship-specificity in the main model), $M(k)$ (i.e., costs of hold up in the main model), and $B(k)$ (i.e., bureaucratic costs in the main model). We can now consider k ``relative importance of coordination over flexibility or innovation," and $M(k)$ and $B(k)$, ``inefficiencies due to coordination failure" (i.e., costs of decentralization) and ``inefficiencies due to inflexibility" (i.e., costs of centralization), respectively. This interpretation reflects the fact that ``Centralization promotes coordination and specialization within functions, but mutes incentives for flexibility and localized innovation. Decentralization, by contrast, promotes innovation and flexibility but impedes coordination" (Nickerson and Zenger, 2002, p.551).

Propositions 1 and 2 then implies that

(i) each generation (if rational) chooses decentralization if coordination does not matter (i.e., k is low),

(ii) decentralization and centralization are alternately chosen before generation t^* and each generation $t(\geq t^*)$ (if rational) chooses centralization if coordination and flexibility are equally important (i.e., k is intermediate),

(iii) each generation (if rational) chooses centralization if coordination is highly important (i.e., k is high),

and changes in k may alter the pattern of institutional choices.

Given that the reason for each generation's choice of governance structure may not be transferred between generations, the level of relationship-specificity, from which each generation infers which governance structure is optimal, plays an important role. If the level of relationship-specificity is high (resp. low) enough, each generation (if rational) chooses integration (resp. non-integration) without wavering. However, if the level of relationship-specificity is intermediate, each generation (if rational) is then uncertain whether the choice of integration is optimal, and thus chooses non-integration and integration alternately to find out which governance structure is optimal. Our model explained why organizational changes often follow management turnovers and why inefficient choice of integration can persist.

To illustrate the claim above, we focus on Hewlett-Packard case. Our discussion here is based on Boumgarden, Nickerson, and Zenger (2012) who present a detailed illustration of HP.

According to Boumgarden, Nickerson, and Zenger (2012), HP case can be briefly described as follows. Prior to early 1980s, decentralized structure served well. However, ``as HP's product portfolio increasingly drifted away from instruments and into computing and software" (p.597), which triggered customers' demands for integrated solutions rather than stand-alone components, HP's organizational structure began to vacillate. Such vacillation emerged as follows. The demands for integrated solution made coordination important and autonomy problematic. Managements then chose centralization to eliminate coordination problems. However, centralized structure eventually became bureaucratic and suffered from bureaucratic costs (e.g., loss of flexibility), managements thus chose decentralization to revive flexibility and innovativeness. Nevertheless, such decentralized structure eventually spoiled coordination again. HP's organizational choices over 25 years (pre-1982 to 2008) are summarized as follows:

Decentralization: pre-1982, 1988 to 1994, 1998 to 1999, and 2005 to 2008

Centralization: 1982 to 1988, 1995 to 1998, and 1999 to 2005.

Boumgarden, Nickerson, and Zenger (2012) explain this ``organizational vacillation" as follows. Managements want to achieve balance between exploration and exploitation to improve performance, but they requires quite opposite organizational designs. Firms then try to dynamically achieve ``high levels of both exploration and exploitation by temporally and sequentially alternating between organizational structures that promote either exploration or exploitation, respectively" (Boumgarden, Nickerson, and Zenger, 2012, p.588). Such dynamic balance between exploration and exploitation is achievable because while formal organizational structures change discretely, informal organizations, such as routines, ``adjust more continuously in response to structural shifts" (Boumgarden, Nickerson, and Zenger, 2012, p.591).

This chapter, on the other hand, tries to explain the transition of HP's organizational structure by unrecorded rationality for predecessors' organizational choices and resulting experimentation. According to Boumgarden, Nickerson, and Zenger (2012), analysts and HP executives initially viewed each change in organizational structure positively and believed that it resolved problems, which the previous structure could not avoid, and balance between exploration and exploitation was achieved. However, such a belief eventually turned out to be wrong. This repeated process of initial applause for the organizational changes and disillusionment seems to suggest that people focus on the fact that the previous structure did not work well, but do not know or care about the problems which previous similar changes triggered. Our assumption that each generation may not understand the reason for his predecessors' choices reflects this interpretation.

Our results (Propositions 1 and 2) then explain the transition of HP's organizational structure as follows. In pre-1982, ``HP had developed a leading position in test and measurement equipment, designing and producing largely stand-alone products'' (Boumgarden, Nickerson, and Zenger, 2012, p.596) and coordination was not so important relative to innovativeness and flexibility (i.e., k is low). Each generation thus chose ``decentralization'' (i.e., pattern (i) arose), which corresponds to the fact that HP was well known for its highly decentralized structure by the early 1980s. However, customers ``increasingly demanded integrated solutions rather than stand-alone components'' (p.597) and ``For the first time in its history, HP faced strong pressure to coordinate designs, products, and marketing efforts across divisions'' (p.597), which made innovativeness and coordination equally important (i.e., k becomes intermediate). Each generation who faced intermediate k thus chose decentralization and centralization alternately, which corresponds to pattern (ii).

Although our explanation above is different from that of Boumgarden, Nickerson, and Zenger (2012), we have no intention of saying that our approach replaces the approach of Boumgarden, Nickerson, and Zenger (2012) or our model completely explains HP's choices between centralization and decentralization. In fact, their focus and aim are different from ours and our assumption that each manager (generation) chooses organizational structure only once contradicts the observation that some organizational changes were initiated by the same CEO (e.g., Lewis Platt, who was HP's CEO from 1992 to 1999). Thus, the most that this section tells us is that there can be some complement explanations for transition of HP's organizational structure other than the one presented in Boumgarden, Nickerson, and Zenger (2012).

7 Conclusion

This chapter addressed a question why firm boundaries sometimes waver by developing a multi-generation model. The key assumption of our model is that each generation chooses either non-integration or integration to govern a transaction without knowing the reasons for predecessors' choices of governance structure.

We show that under the assumption of unrecorded reasons for predecessors' governance choices, each generation's experimentation causes wavering between non-integration and integration in equilibrium, which TCE has had trouble explaining. Given that the reason for each generation's choice of governance structure may not be transferred between generations, the level of relationship-specificity affects governance choices. If the level of relationship-specificity is high (resp. low) enough, non-integration is likely to be more (resp. less) costly than integration, and hence each generation (if rational) chooses integration (resp. non-integration), which achieves actual transaction-cost minimization with high probability. However, if the level of relationship-specificity is intermediate, the governance choice which expectedly minimizes transaction costs is likely to fail in actual transaction-cost minimization. Hence, an effort to find out which governance structure is optimal is required and the wavering between non-integration and integration occurs. Our model provides formal explanations for why organizational changes often follow management turnovers and why it is hard for some integrated firms to disintegrate even if integration is not optimal. We then applied our result to the choice between centralization and decentralization and offered an explanation for the changes in HP's organizational structures.

There are some extensions to be done. For example, we did not analyze what would happen if each generation can expect the changes in relationship-specificity or if each generation t whose predecessor $t-1$ chose non-integration do not know whether bureaucratic costs occur, and how our result would change if single decision maker with imperfect recall of the reason for his previous choices repeatedly chooses governance structure. Although these extensions are important, they are beyond the scope of the dissertation, and hence we leave further analysis for future research.

Chapter 5
Conclusion

In this dissertation, we developed formal models of firm boundaries in the spirit of transaction cost analysis (i.e., examined firm boundaries by focusing on ex post inefficiencies). Specifically, we addressed the following three topics: ``how rent seeking and bargaining costs interact,'' ``why authority mitigates disputes over trade value,'' and ``why some firms choose non-integration and integration alternately.''

How Do Rent Seeking and Bargaining Costs Interact? We showed that there is a trade-off between rent seeking, which is undertaken regardless of the choice of governance structure, and bargaining costs, which occur only under non-integration. More specifically, while non-integration suffers from bargaining delay and breakdown, which never occur under integration, it incurs lower rent-seeking costs than integration.

There are two reasons for the result. First, rent seeking under non-integration indirectly affects rent-seekers' payoffs by improving their outside options, while rent seeking under integration affects their payoffs directly. Thus, when the aggregate outside option must be smaller than the original trade value (e.g., because of time-consuming litigation), the parties' incentives for rent seeking under non-integration become smaller than those under integration. Second, the bargaining provides parties with opportunities to concede (i.e., to let their partners obtain a large share of the value). When each party behaves obstinately in the bargaining with high probability, any behavior other than concession is likely to delay agreement, and hence parties (if rational) can optimally concede. Since concession terminates the game, in which case outside option is not exercised, each party, expecting this outcome, becomes less eager to engage in rent seeking to improve his outside option.

Why Does Authority Mitigate Disputes over Trade Value? We pointed out that a subordinate is likely to obey orders of his boss because he is expected to do so. This stems from the discussion that the choice of governance structure affects trading parties' expectations about how they should behave. Under non-integration, each party is autonomous, and hence he is entitled to reject any offer that his partner makes as he pleases. Under integration, on the other hand, an authority relationship exists, and hence a subordinate is supposed to obey orders from his boss. Hence, the subordinate's disobedience to his boss provokes his sense of guilt or his fear of his boss's punishment, which makes him likely to obey. Such obedience makes immediate settlement of ex post value split more likely to be achieved, but comes with subordinate's psychological disutility.

Why Do Some Firms Choose Non-Integration and Integration Alternately? We pointed out that unrecorded reason for previous choices of governance structures leads to an effort to figure out which structure triggers what problems (i.e., experimentation), and hence makes firms choose non-integration and integration alternately.

Our multi-generation model showed that under the assumption of unrecorded reasons for predecessors' governance choices, each generation's experimentation causes wavering between non-integration and integration in equilibrium, which TCE has had trouble explaining. Given that the reason for each generation's choice of governance structure may not be transferred between generations, the level of relationship-specificity, from which each generation infers which governance structure is optimal, plays an important role. If the level of relationship-specificity is high (resp. low) enough, non-integration is likely to be more (resp. less) costly than integration, and hence each generation (if rational) chooses integration (resp. non-integration), which achieves actual transaction-cost minimization with high probability. However, if the level of relationship-specificity is intermediate, the governance choice which expectedly minimizes transaction costs is likely to fail in actual transaction-cost minimization. Hence, an effort to find out which governance structure is optimal is required and leads to wavering between non-integration and integration. Our model explained why organizational changes often follow management turnovers and why inefficient choice of integration can persist.

Although our analyses in this dissertation have shortcomings and limitations (some of these will be briefly discussed in the next subsection), we hope that these findings stimulate progress in formalizing TCE and help us improve our understanding of make-or-buy decisions.

Important Topics Left Untouched

In concluding this dissertation, we briefly discuss some topics which are important but left untouched: inalienable relationship-specific assets and hybrid governance structures.

Inalienable Relationship-Specific Assets

Throughout this dissertation (except Chapter 3, which implicitly assumed inalienable assets), we assumed that relationship-specific assets which led to bilateral dependency between trading parties were alienable (e.g., specific physical assets). Nevertheless, in reality, a number of inalienable relationship-specific assets can be found (e.g., some specific knowledge and skills). Furthermore, it is known that there is a serious gap between theoretical hypothesis and empirical findings regarding how such inalienable assets affect firm boundaries. More specifically, while theoretical studies, including Gibbons (2005), assert that vertical integration may not stop opportunistic behavior (i.e., haggling) caused by inalienable relationship-specific assets, most empirical studies show that the more relationship-specific inalienable assets in question become, the more likely firms are to choose vertical integration (see Lafontaine and Slade, 2007 for a survey of these studies).

The negative theoretical hypothesis about inalienable relationship-specific assets stems from the view that integration (only) affects control rights. Theoretical studies assert that vertical integration can reduce or avoid haggling caused by alienable relationship-specific assets because it can remove the relevant control rights from hagglers. In other words, ``the most that integration can do is to unify the alienable control rights" (Gibbons, 2005, p. 205). Thus, if the assets are inalienable, such a mechanism fails and integration cannot stop haggling.

We believe that Klein's (1988) informal approach is the key to the gap filling between theoretical and empirical views regarding how inalienable specific assets affect firm boundaries. Klein (1988) asserts that ``Vertical integration may solve a hold-up potential even when it hinges on human capital" (p.207), which is consistent with the empirical studies. This assertion stems from the following arguments. Consider the situation where a buyer firm wants a good whose production requires an inalienable relationship-specific asset and there is only one seller firm which possesses such an asset. This means that the buyer firm and the seller firm are locked in, and hence market transaction suffers from inefficiencies due to hold up. Nevertheless, integration reduces the possibility of such small numbers relation. It follows because, even if there is only one seller firm which possesses the asset, it is natural that it is shared among employees in the seller firm (Klein refers to these employees as ``key employees"). Hence, the buyer firm can obtain all those key employees by acquiring the seller firm. Holding up the buyer under integration then requires that all the key employees simultaneously threaten the buyer firm. Such coordination between the key employees is difficult to achieve if there are many key employees. Thus, ``with many key individuals involved, the organization will generally be secure" (Klein, 1988, p. 208).

While Klein's (1988) approach has not been used for theoretical analysis on firm boundaries, it presents an important idea that the number of key employees, which plays little role in the cases where the specific assets are alienable, does matter when the assets are inalienable. In other words, his idea illustrates a fundamental difference between hazards triggered by alienable specific assets and those caused by inalienable specific assets.

Hybrid Structures and Integrative Theory of Firm Boundaries

Throughout this dissertation, we focused on how ex post inefficiencies affected the choice between two polar governance structures: non-integration and integration. However, in reality, we observe governance structures that are located between markets and hierarchies (e.g., long-term contracting and franchising). Such governance structures are called hybrid governance structures. By definition, hybrid governance structures include various kinds of institutional arrangements, and hence it is difficult to define these in a way that we can compare them with non-integration and integration. We believe that such a problem might be solved by defining each governance structure based on the combination of ex ante and ex post features (e.g., who makes ex ante investments and how trading parties handle ex post governance problems), which have been dealt with separately in existing literature.

This implies that an integrative theory which can deal with both ex ante and ex post problems in a unified way is required. However, since there is much to be done to completely formalize how ex post aspects determine the choice between the two polar governance structures, we need more time to be ready to develop such an integrative theory. Nevertheless, we believe that it will not take long to achieve the satisfactory formalization of ex post approaches (especially, TCE). Recently, it seems that ex post problems receive theoretical attention (e.g., Bajari and Tadelis, 2001 and Matoushek, 2004). Furthermore, recent remarkable progress in behavioral economics and its application to the theory of the firm (e.g., Hart and Moore, 2008 and Herweg and Schmidt, 2012) will help us reach the next stage of the studies on firm boundaries.

Bibliographically

Abreu, Dilip, and Faruk Gul. 2000. ``Bargaining and Reputation." *Econometrica*, 68(1): 85-117.

Adams, J. S. 1976. ``The Structure and Dynamics of Behavior in Organizational Boundary Roles." in *Handbook of Industrial and Organizational Psychology*, ed. Marvin. D. Dunnette, 1175-1199. Chicago: Rand McNally.

Akerlof, Robert. 2010. ``Punishment, Compliance, and Anger in Equilibrium." Job Market Paper, MIT, Sloan School.
http://academia.edu/228988/Punishment_Compliance_and_Anger_in_Equilibrium_JOB_MARKET_PAPER_

Atakan, Alp E., and Mehmet Ekmekci. 2010. ``Bargaining and Reputation in Search Markets."
http://www.kellogg.northwestern.edu/faculty/ekmekci/personal/assets/MatchingwithIrrattionalTypes.pdf

Babcock, Linda, George Loewenstein, Samuel Issacharoff, and Colin Camerer. 1995. ``Biased Judgments of Fairness in Bargaining." *American Economic Review*, 85(5): 1337-1343.

Babcock, Linda, and George Loewenstein. 1997. ``Explaining Bargaining Impasse: The Role of Self-Serving Biases." *Journal of Economic Perspectives*, 11(1): 109-126.

Bajari, Patrick, and Steven Tadelis. 2001. ``Incentives versus Transaction Costs: A Theory of Procurement Contracts." *Rand Journal of Economics*, 32(3): 387-407.

Barnard, Chester I. 1938, *The Functions of the Executive*. Cambridge, MA: Harvard University Press.

Bénabou, Roland and Jean Tirole. 2004. ``Willpower and Personal Rules." *Journal of Political Economy*, 112(4): 848-886.

Binmore, Ken, Joseph Swierzbinski, and Chris Tomlinson. 2007. ``An Experimental Test of Rubinstein's Bargaining Model." ELSE Working Papers #260, ESRC Centre for Economic Learning and Social Evolution, London, UK.

Bolton, Gary and Rami Zwick. 1995. ``Anonymity versus Punishment in Ultimatum Bargaining." *Games and Economic Behavior*, 10(1): 95-121.

Boumgarden, Peter, Jackson Nickerson, and Todd R. Zenger. 2012. ``Sailing into the Wind: Exploring the Relationships among Ambidexterity, Vacillation, and Organizational Performance." *Strategic Management Journal*, 33(6): 587-610.

Coase, Ronald. 1937. ``The Nature of the Firm." *Economica*, 4(16): 386-405.

Compte, Olivier, and Philippe Jehiel. 2002. ``On the Role of Outside Options in Bargaining with Obstinate Parties." *Econometrica*, 70(4): 1477-1517.

Dow, Gregory K. 1987. ``The Function of Authority in Transaction Cost Economics." *Journal of Economic Behavior and Organization*, 8(1): 13-38.

Gallice, Andrea. 2009. ``Self-Serving Biased Reference Points." University of Torino. http://ideas.repec.org/p/usi/depfid/0909.html

Gibbons, Robert. 2005. ``Four Formal(izable) Theories of the Firm?" *Journal of Economic Behavior and Organization*, 58(2): 200-245.

Grossman, Sanford, and Oliver Hart. 1986. ``The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy*, 94(4): 691-719.

Hart, Oliver. 1995. *Firms, Contracts, and Financial Structure*. Oxford University Press.

Hart, Oliver. 2009. ``Hold-Up, Asset Ownership, and Reference Points." *Quarterly Journal of Economics*, 124(1): 267-300.

Hart, Oliver and Bengt Holmstrom. 2010. ``A Theory of Firm Scope." *Quarterly Journal of Economics*, 125(2): 483-513.

Hart, Oliver, and John Moore. 1990. ``Property Rights and the Nature of the Firm." *Journal of Political Economy*, 98(6): 1119-1158.

Hart, Oliver, and John Moore. 2008. ``Contracts as Reference Points." *Quarterly Journal of Economics*, 123(1): 1-48.

Herweg, Fabian, and Klaus M. Schmidt. 2012. ``Loss Aversion and Ex Post Inefficient

Renegotiation." University of Munich.
http://www.am.vwl.uni-muenchen.de/forschung/publikationen/renegotiation_loss_aversion.pdf

Holmstrom, Bengt, and Paul Milgrom. 1991. ``Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, and Organization*, 7: 24-52.

Joskow, Paul L. 1988. ``Asset Specificity and the Structure of Vertical Relationships: Empirical Evidence." *Journal of Law, Economics, and Organization*, 4(1): 95-117.

Kahneman, Daniel, and Amos Tversky. 1979. ``Prospect Theory: An Analysis of Decision under Risk." *Econometrica*, 47(2): 263-292.

Kambe, Shinsuke. 1999. ``Bargaining with Imperfect Commitment," *Games and Economic Behavior*, 28(2): 217-237.

Klein, Benjamin, Robert Crawford, and Armen Alchian. 1978. ``Vertical Integration, Appropriable Rents and the Competitive Contracting Process." *Journal of Law and Economics*, 21(2): 297-326.

Köszegi, Botond, and Matthew Rabin. 2006. ``A Model of Reference-Dependent Preferences." *Quarterly Journal of Economics*, 121(4): 1133-1165.

Köszegi, Botond, and Matthew Rabin. 2007. ``Reference-Dependent Risk Attitudes." *American Economic Review*, 97(4): 1047-1073.

Lafontaine, Francine, and Margaret Slade. 2007. ``Vertical Integration and Firm Boundaries: The Evidence." *Journal of Economic Literature*, 45(3): 629-685.

Langlois, Richard N., and Paul L. Robertson. 1995. *Firms, Markets, and Economic Change: A Dynamic Theory of Business Institutions*. Routledge.

Lee, Jihong, and Qingmin Liu. 2010. ``Gambling Reputation: Repeated Bargaining with Outside Options."
http://www.econ.upenn.edu/~qingmin/LeeLiuMar10.pdf

MacLeod, W. Bentley. 2007. ``Can Contract Theory Explain Social Preferences?" *American Economic Review*, 97(2): 187-192.

March, James G. 1991. ``Exploration and Exploitation in Organizational Learning." *Organizational Science*, 2(1): 71-87.

Masten, Scott E. 1984. ``The Organization of Production: Evidence from the Aerospace Industry." *Journal of Law and Economics*, 27(2): 403-417.

Masten, Scott E., James W. Meehan Jr., and Edward A. Snyder. 1989. ``Vertical Integration in the U.S. Auto Industry: A Note on the Influence of Transaction Specific Assets." *Journal of Economic Behavior and Organization*, 12(2): 265-273.

Matouschek, Niko. 2004. `` Ex Post Inefficiencies in a Property Rights Theory of the Firm." *Journal of Law, Economics, and Organization*, 20(1): 125-147.

Meyer, Margaret, Paul Milgrom, and John Roberts. 1992. ``Organizational Prospects, Influence Costs, and Ownership Changes." *Journal of Economics and Management Strategy*, 1(1): 9-35.

Milgrom, Paul, and John Roberts. 1988. ``An Economic Approach to Influence Activities in Organizations." *American Journal of Sociology*, 94, Supplement: Organizations and Institutions: Sociological and Economic Approaches to the Analysis of Social Structure (1988): S154-S179.

Milgrom, Paul, and John Roberts. 1990. ``The Economics of Modern Manufacturing: Technology, Strategy, and Organization." *American Economic Review*, 80(3): 511-528.

Monteverde, Kirk, and David Teece. 1982. ``Supplier Switching Costs and Vertical Integration in the Automobile Industry." *Bell Journal of Economics*, 13(1): 206-213.

Myerson, Roger. 1991. *Game Theory: Analysis of Conflict*. Cambridge, Massachusetts: Harvard University Press.

Nickerson, Jack A., and Todd R. Zenger. 2002. ``Being Efficiently Fickle: A Dynamic Theory of Organizational Choice." *Organizational Science*, 13(5): 547-566.

Özyurt, Selçuk. 2010. ``Searching a Bargain: `Play it Cool' or Haggle." http://research.sabanciuniv.edu/14597/

Perry, James L., and Harold L. Angle. 1979. ``The Politics of Organizational Boundary Roles in Collective Bargaining." *Academy of Management Review*, 4(4): 487-495.

Riordan, Michael H., and Oliver E. Williamson. 1985. ``Asset Specificity and Economic Organization." *International Journal of Industrial Organization*, 3: 365-378.

Roberts, John. 2004. *The Modern Firm: Organizational Design for Performance and Growth*. New York: Oxford University Press.

Tadelis, Steven. 2002. ``Complexity, Flexibility, and the Make-or-Buy Decision." *American Economic Review*, 92(2): 433-437.

Tadelis, Steven, and Oliver E. Williamson. 2012. ``Transaction Cost Economics." *The Handbook of Organizational Economics*, ed. Robert Gibbons and John Roberts, Princeton University Press.

Tullock, Gordon. 1980. ``Efficient Rent Seeking." in *Toward a Theory of the Rent-Seeking Society*, ed. James M. Buchanan, Robert D. Tollison, and Gordon Tullock, 97-112. College Station, TX: Texas A&M University Press.

Van den Steen, Eric. 2010. ``Interpersonal Authority in a Theory of the Firm." *American Economic Review*, 100(1): 466-490.

Wernerfelt, Birger. 2011. ``Efficient Adaptation Versus Gains from Specialization: Procuring Labor Services."
Available at SSRN: http://ssrn.com/abstract=1783722 or http://dx.doi.org/10.2139/ssrn.1783722

Williamson, Oliver E. 1971. ``The Vertical Integration of Production: Market Failure Considerations." *American Economic Review*, 61(2): 112-123.

Williamson, Oliver E. 1973. ``Markets and Hierarchies: Some Elementary Considerations." *American Economic Review*, 63(2): 316-325.

Williamson, Oliver E. 1975. *Markets and Hierarchies: Analysis and Antitrust Implications*. New York: Free Press.

Williamson, Oliver E. 1985. *The Economic Institutions of Capitalism*. New York: Free Press.

Williamson, Oliver E. 1996. *The Mechanisms of Governance*. New York: Oxford University Press.

Zhu, Tian. 2009. ``Towards a Formal Model of the Transaction Cost Theory of the Firm."
Available at SSRN: http://ssrn.com/abstract=1486464 or
http://dx.doi.org/10.2139/ssrn.1486464