How Can Integration Reduce Inefficiencies Due to *Ex Post* Adaptation?

Yusuke Mori
（Graduate School of Commerce and Management
Hitotsubashi University）

April 2012

No.142

How Can Integration Reduce Inefficiencies Due to *Ex Post* Adaptation?*

Yusuke Mori†

April 25, 2012

Abstract

How can integrated firms immediately settle *ex post* adaptations to unanticipated disturbances? While this question is crucial to the understanding of transaction cost economics (TCE), TCE has not provided any formal answer. This paper develops a model that explores this question by employing three behavioral assumptions: reference-dependent preference, self-serving bias, and shading. We present two reasons why integration can avoid costly renegotiations and realize immediate adaptations; these stem from the fact that while non-integrated parties have to engage in negotiations for the adaptations, integrated firms can implement these by fiat. First, punishments for rejection of an order under integration are severer than those for rejection of an offer under non-integration. Second, under integration, the utility improvement for a subordinate from rejecting an order is not sufficient to offset the loss from a severe punishment. Furthermore, we point out a trade-off between immediate agreement and the aggregate sense of loss.

Keywords: Reference-dependent preference; self-serving bias; contracts as reference points; transaction cost; *ex post* adaptation

JEL Classification: D23; L22

*I am grateful to Reiko Aoki, Kohei Daido, Shinsuke Kambe, Fumitoshi Moriya, Takeshi Murooka, and Sadao Nagaoka for their beneficial comments, and especially to Hideshi Itoh for his suggestions and encouragement. I also thank the participants at Contract Theory Workshop, Contract Theory Workshop East, and the seminar at Hitotsubashi University for their helpful comments. Any remaining errors are my own.

†Graduate School of Commerce and Management, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8601, Japan. E-mail: cd101008@g.hit-u.ac.jp

1 Introduction

How can integrated firms immediately settle *ex post* adaptations to unanticipated disturbances? Although this question is crucial to the understanding of transaction cost economics (TCE) and has been frequently asked (e.g., Hart (1995)), TCE has not provided any formal answer. This paper formally shows that integration indeed realizes immediate agreement more easily than non-integration.

The above question is critical to TCE because its main prediction is derived from the implicit assumption that integration can implement the adaptations immediately. TCE (e.g., Williamson (1996)) asserts that under bilateral monopoly caused by relationship-specific investment or other factors, while non-integrated parties have to engage in costly negotiations for the *ex post* adaptations, which leads to bargaining inefficiencies (delay in reaching agreement and bargaining breakdown), integrated firms can implement these by fiat without such costly negotiations. This assertion leads to TCE's main prediction: when a bilateral monopoly arises, firms are likely to choose vertical integration; this is supported by a number of empirical studies (see Lafontaine and Slade (2007) for a survey of these studies).

In this paper, we develop a model that explores how integration reduces inefficiencies due to *ex post* adaptation (especially the division of trade value) and show that integration indeed settles the adaptation more easily than non-integration by employing three behavioral approaches: reference-dependent preference, self-serving bias, and shading (contracts as reference points). We emphasize that these behavioral assumptions are all crucial to our result. That is, relaxing any of these assumptions leads to the result that the choice of governance structure does not affect the timing of agreement or brings the opposite result: non-integration can realize the immediate settlement of *ex post* adaptation more easily than integration. The evidence that supports each of these assumptions will be presented in Section 3.

Players in our model have the following three characteristics. First, as in the literature on reference-dependent preference, such as Köszegi and Rabin (2006, 2007), the players' 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 adaptation processes, each governance structure leads to different reference points. That is, the process by which adaptation outcomes are determined matters for the players' reference points. Under non-integration, as mentioned above, *ex post* adaptation is implemented through bargaining, and hence, the players' reference points are given by the expected outcome of bilateral bargaining. Under integration, on the other hand, *ex post* adaptation is implemented by fiat, and thus, the players' reference points are the expected outcome of an ultimatum game (i.e., a player who has a decision right takes most of the trade value).

Second, the players' reference points are biased because each player has a self-serving view regarding who is to incur sunk relationship-specific investment. More specifically, while a player who does not invest thinks that his partner who has invested ("she") is to incur the whole investment cost, she believes that her sunk investment is to be compensated. This assumption reflects the fact that each player's role (in this case, whether a player has invested) affects his expectation in a self-serving way even if the same information is shared (Babcock et al., 1995). Such self-serving views result in the divergence of reference points between the players.

Third, those who do not obtain the payoffs that are smaller than their reference point payoffs engage in 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).

Our model presents two reasons why integration achieves immediate settlement of *ex post* adaptation more easily than non-integration. First, a rejection of an order under integration provokes a severer punishment than a rejection of an offer under non-integration.[1] As mentioned above, under integration, *ex post* adaptation is implemented by fiat. That is, a person who has authority (a boss) determines how to divide the trade value, and a subordinate is supposed to obey the boss's orders. The boss's reference point payoff is thus quite large. However, if a subordinate rejects the boss's order, as Barnard (1938) points out, the authority relationship between them is terminated, and hence, the adaptation outcome is determined as if they are autonomous parties (i.e., their adaptation payoffs are balanced). This means that if the order is rejected, the boss is compelled to obtain a far smaller payoff than his reference point payoff, which provokes a huge amount of anger. Since the boss's anger leads to severe retaliation against the subordinate, the subordinate is less willing to reject the order. Under non-integration, on the

[1] To facilitate the comparison between governance structures, we assume that under integration, an employer does not fire an employee who disobeys an 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).

other hand, 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 result in a huge amount of sender's anger under non-integration.

The second reason is that under integration, the utility improvement for a subordinate from rejecting an order is not sufficient to offset the loss from the severe punishment. As mentioned above, the players' reference points under integration are the expected outcome of an ultimatum game, and hence, the subordinate who has no decision right does not expect a large adaptation payoff. Given that the subordinate expects a small adaptation payoff, the payoff improvement from rejecting the order is "too much" for him (i.e., the rejection of the order 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, while integration can economize inefficiencies that are due to delay in reaching agreement, it incurs larger shading costs than non-integration. The reason for this is as follows. As mentioned above, the player who invests believes that her sunk investment will be compensated regardless of the choice of the governance structure. Nevertheless, under non-integration, each player receives a positive share of a trade surplus (namely, the value the trade creates minus the investment cost) from bargaining, and thus, the player who invests expects to incur some portion of the investment cost. Under integration, on the other hand, a player who receives an order from his boss expects that the entire surplus will be taken by the boss, and hence, if he invests, he does not take the investment costs into account when he sets his reference point. This discussion suggests that the divergence between the players' reference points because of the self-serving view regarding who is to incur the investment costs is larger under integration than under non-integration, which makes the aggregate sense of loss and shading costs under integration larger than those under non-integration.

The rest of the paper proceeds as follows. The next section relates our study to the existing literature. Section 3 introduces the model and Section 4 examines which governance structure achieves immediate settlement of *ex post* adaptation. 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 I 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 *ex post* adaptation more easily than non-integration. Appendix II examines the case in which the players are risk-averse. Appendix III assumes that the players care about discounting and checks the robustness of our result.

2 Related Literature

This section reviews the related literature. This paper employs the approach that each player's reference point is assumed to be set as his expectation of the relevant outcome under the influence of self-serving bias. Hence, we first relate our study to the existing literature in which expectations are considered reference points: Köszegi and Rabin's rational expectation approach and Hart's contracts-as-reference-points approach. We then review some existing studies that share similar interests to ours. Lastly, since this paper derives implications for firm boundaries, some approaches to them are reviewed.

The literature on reference-dependent preference, such as Köszegi and Rabin (2006) and Köszegi and Rabin (2007), assumes that "a person's reference point is the rational expectations about the relevant outcome" (Köszegi and Rabin, 2007, p. 1051). Since the expectations are formed rationally, each person's reference point is not affected by self-serving bias. Furthermore, their assumption of rationality makes it possible to derive each player's reference point endogenously.

The models of contracts as reference points are presented in Hart and Holmstrom (2008), Hart (2009), and Hart and Holmstrom (2010). Unlike Köszegi and Rabin's approach, this approach assumes that "each party feels entitled to the best outcome permitted by the contract" (Hart and Moore, 2008, p. 33). This assumption implies that each player's reference point is completely biased.

This paper adopts a middle ground approach between these two approaches: the players in our model are neither completely rational nor completely naive. That is, while they are not completely rational in the sense that they are affected by a self-serving view regarding who is to incur the investment cost, they are not naive enough to believe that they are entitled to the best outcome permitted by the contract.

We next relate our paper to the existing studies that share similar interests to ours: Gallice (2009), Van den Steen (2010), and Akerlof (2010). 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 player's reference point. As mentioned above, we assume that players' self-serving views regarding the sunk investment result in the divergence of their reference points even if they share views on how each player 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 because concentrating asset ownership into an employer's hands 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. In our study, the choice of governance structure only affects the *ex post* adaptation process and each player'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 that what is fair depends on the adaptation process chosen: bilateral bargaining (non-integration) or fiat (integration).[2]

We next review some approaches to firm boundaries: TCE and the property rights theory. While the former approach, as in Coase (1937) and Williamson (1975 and 1985), focuses on authority, the latter approach, as in Grossman and Hart (1986), Hart and Moore (1992), and Hart (1995), emphasizes ownership structure.

TCE asserts that authority helps integration avoid costly *ex post* renegotiation, but does not explain how it does this. Mori (2011), for example, develops formal models of *ex post* adaptation in the spirit of TCE and shows that inefficient *ex post* bargaining, which only takes place under non-integration, creates a trade-off between rent seeking and bargaining losses. In Mori (2011), however, as in the literature on TCE, integration is assumed to avoid bargaining losses without offering a formal justification for the assumption. This study complements TCE's arguments by presenting the idea that authority works because the choice of governance structure affects each player's expectation.

Our study is quite different from the existing studies on the property rights theory with respect to how ownership structures affect players' outside options. Matouschek (2004), for example, develops a formal model following property rights theory and examines the optimal ownership structure that minimizes *ex post* inefficiency caused by too much or too little trade.

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

In Matouschek (2004), disagreement payoffs depend on the ownership structure (namely, while non-integration or integration maximize the aggregate disagreement payoff, joint ownership minimizes it). Our study, on the other hand, assumes that ownership structure does not affect players' outside options. Furthermore, while the property rights theory has often been employed to examine *ex ante* inefficiency (underinvestment problems), our study assumes that there is no such *ex ante* inefficiency (namely, the investment has been efficiently sunk) and focuses on *ex post* inefficiencies.

3 The Model

This section presents the model that focuses on which governance structure realizes immediate settlement of *ex post* adaptation, particularly the division of trade value 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 present an overview of the model and then introduce three assumptions.

Two risk-neutral trading parties (players 1 and 2) trade one unit of a good and are to engage in the division of trade value.[3] The trade requires player 2's specific investment I in an alienable relationship-specific asset (player 1 does not invest) and creates value π. We assume that the trade is efficient and that the players 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 player who incurs the whole sunk investment. In order to focus on *ex post* inefficiency, we assume that *ex ante* investment I is efficiently sunk and that there are 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 players' utility. Second, the players set their reference points regarding the outcome of the *ex post* adaptation. An adaptation process is then initiated. We assume that under integration, player 1 (resp. player 2) becomes a boss (resp. a subordinate).[4]

The *ex post* adaptation process consists of player 1's division offer $x = (x_1, x_2)$, where x_i

[3] We refer to player 1 as "he" and player 2 as "she" for the purpose of identification only.

[4] If we assume that player 2 has decision rights under integration, integration should always be chosen as the optimal governance structure. We will discuss this topic in Section 5.

represents player i's share of the trade surplus and player 2's acceptance decision.[5] If player 2 accepts the offer, the surplus is divided as the accepted offer specifies; otherwise, the adaptation continues. Since we do not specify what will happen after player 2 rejects the offer, the adaptation process does not necessarily mean that player 1 makes a take-or-leave-it offer. Since we here focus on which governance structure realizes immediate agreement, we only need to examine whether the first offer is accepted. Thus, we think that this model captures the first period of infinite-horizon alternating-offers bargaining.

For simplicity, we assume that each player 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 player ignores discounting (behaves as if there were no discounting). This assumption 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 players do care about discounting and generalize our main result in Appendix III.

Behavioral Assumptions

Our model employs three behavioral assumptions: reference-dependent utility, self-serving bias, and shading. This subsection introduces these three assumptions and presents evidence that supports them. We emphasize that these assumptions are all crucial to our result: integration can realize immediate agreement more easily than non-integration. In Appendices I and II, we show that our result does not hold if any of these assumptions is relaxed. Appendix I 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 II focuses on the case in which the players 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.

[5] The assumption that player 1 has the right to send an offer/order under both governance structures is employed only to facilitate the comparison between non-integration and integration. Thus, we can instead assume that under non-integration, the right to send an offer is assigned to each player with equal probability without changing our result.

We first present each player's utility function. Player i's utility is assumed to be reference-dependent and affected by player j's shading. Formally, 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 player i's reference point (r_{ij} represents i's belief about player j's reference point payoff). Player i's utility when an adaptation outcome 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

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

$n(\cdot)$ is the gain-loss function, η represents weight on gain-loss payoff, $\lambda(>1)$ is sensitivity of loss aversion, and $\theta(>0)$ denotes an exogenous common punishment intensity (i.e., the shading parameter). The first term of the utility function denotes player i's intrinsic payoff, the second term represents his gain-loss utility, and the third is the loss caused by player j's shading. It is worth noting that the first and second terms (resp. the third term) 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). This formalization suggests that our gain-loss function $n(\cdot)$ captures loss aversion but rules out diminishing sensitivity, which is one of the features of gain-loss utility. We adopt such an assumption because we want to show clearly the crucial effect of loss aversion on our result.

Shading can be interpreted as a punishment for unfair treatment. That is, when player i obtains a payoff less 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 $y_i \geq r_{ii}$).[6] 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

[6] 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 player's payoff and his reference point payoff (i.e., the shading term in the literature on contracts as reference points is given by $\theta \min(y_i - r_{ii}, 0)$).

of shading.[7]

As mentioned above, we introduce shading into Köszegi and Rabin's reference-dependent preference. 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 (1993) 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.

We next explain how each player's reference point is defined in our model. As in Köszegi and Rabin's approach, we assume that each player's reference point is his expectation about the relevant outcome. However, while Köszegi and Rabin's approach assumes rational expectations, our model assumes that each player expects the relevant outcome in a boundedly rational way. More specifically, the players correctly infer how their partners set their reference points, but perceive the game structure self-servingly.

Our model assumes that each player has a self-serving view regarding the sunk investment I. That is, while the player who does not invest (player 1) thinks that his partner, who is supposed to invest (player 2), is to incur her sunk investment, his partner (player 2) believes that her sunk cost is to be compensated. In other words, player 1 (resp. player 2) believes that the players are to divide a gross value π (resp. a net value $\pi - I$). Formally, player 1 (who does not invest) believes that each player's outside option is given by

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

where w_{1i} denotes player 1's belief about player i's outside option. Note that each player can earn 0 outside the current trade relationship. Player 2, on the other hand, is confident that the players' outside options are

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

This assumption reflects the fact that each player's role (in this case, whether he has invested or not) affects his expectation in a self-serving way even if the same information is shared (Babcock et al., 1995).

[7] Note that we use the term "shading costs" as inefficiencies (deadweight loss) due to shading.

Some readers might think that it is inappropriate to assume that while the players 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 (Babcock and Loewenstein, 1997). 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."

The ways in which players set their reference points are assumed to be different under each governance structure; this stems from the difference in adaptation processes between non-integration and integration. Under non-integration, as in Williamson (1996), "the autonomous stages would need to bargain these [adaptations to unanticipated disturbances] through to agreement" (p. 17), and hence each player's expectation regarding the outcome of the bilateral bargaining serves as his reference point. We thus assume that each player uses the Nash bargaining solution as his reference point; this is common knowledge.

Under integration, on the other hand, "the unified firm can implement adaptations to unanticipated disturbances by fiat" (Williamson, 1996, p. 17). In other words, the person who has decision rights can order any division to his subordinate (she) and his order will be accepted as long as it yields her a payoff larger than or equal to her outside option. That is, *ex post* adaptation proceeds something like an ultimatum game, and hence each player's reference point is given by his expectation regarding the outcome of the game (i.e., a player with a right to send the offer obtains most of the surplus).

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

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

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

and under integration,

$$r_1^h = (r_{11}^h, r_{12}^h) = (\pi, -I)$$

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

Player 1's (resp. player 2's) payoff is listed first (resp. second).

We next explain what will happen if player 2 rejects the offer/order from player 1. In the next section, it will turn out that player 1 optimally offers/orders what his reference point specifies. Player 2 thus infers what player 1's real reference point is when she receives the offer/order. To focus on the situation in which player 2 has an incentive to reject such an offer (i.e., the situation in which immediate agreement is not necessarily realized), if player 2 rejects the offer/order, she is assumed to believe that she can obtain a continuation payoff P that satisfies[8]

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

The first inequation implies that player 2 has a belief that she can obtain more than the offer. The second inequation suggests that player 2 believes that she cannot obtain more than what she is entitled to (namely, her reference point payoff when the trading parties are autonomous). Since P is player 2's belief about the continuation payoff, player 1 does not know it and believes that player 2's continuation payoff cannot be larger than r_{12}^m, which is his belief about player 2's reference point payoff when the trading parties are autonomous. Since each player's belief about the continuation payoff is different, a real continuation outcome after player 2's rejection is specified somewhere between r_1^m and r_2^m (through a negotiation, for example).

We assume that player 2's belief about her continuation payoff (P) does not depend on the governance structure chosen at the beginning. Some readers might wonder why this assumption is appropriate under integration while the players' reference points are employer-favored. 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 an employee's rejection of an order terminates the authority relationship. Hence, after player 1's order is rejected, the negotiation process becomes the same between non-integration and integration, which leads to the same continuation payoff between the two governance structures.

[8]Including $P > (\pi - I)/2$ does not change our result.

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* adaptation and shows that integration realizes immediate agreement more easily than non-integration. This result can be intuitively explained by the following two discussions. First, since an employee (player 2) is supposed to obey an order from her boss (player 1), she believes that her rejection of an order provokes severe punishment from player 1. Second, since an employee does not expect a large adaptation payoff from the outset, she is not so interested in payoff improvement from rejecting the order.

This section proceeds as follows. Subsection 4.1 studies the optimal behavior of players 1 and 2 and examines when immediate agreement is realized under each governance structure. Subsection 4.2 then compares two governance structures, presents our main result, and explains the intuition of the result.

4.1 Each Player's Optimal Behavior

This subsection analyzes the optimal offering/ordering strategy of player 1, which is studied in Subsection 4.1.1, and the optimal acceptance/compliance decision of player 2, which is examined in Subsection 4.1.2.

4.1.1 Player 1's Offer/Order

We first examine player 1's optimal offering/ordering strategy and show that he optimally offers or orders what his reference point specifies. Note that player 1 believes that both players share the same reference point (r_1^m under non-integration and r_1^h under integration) and also believes that player 2's continuation payoff is given by r_{12}^m (the payoff that player 2 is entitled to obtain as an autonomous player).

Obviously, any offer/order $x_1 < r_{11}^m$ or $x_1 < r_{11}^h$ is not optimal because such an offer would be accepted by player 2 but leads to player 1's loss. Hence, we must examine $x_1 \geq r_{11}^m$ under non-integration and $x_1 \geq r_{11}^h$ under integration. Under integration, player 1's optimal order is equivalent to his reference point because $r_{11}^h = \pi$ and there is no room for player 1 to demand more. We then only need to study the optimal offering strategy under non-integration such that $x_1 = r_{11}^m + \Delta$ ($\Delta \geq 0$).

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

$$\begin{aligned}U_1^m(x \mid r_1^m, r_1^m) =& r_{11}^m + \Delta + n(r_{11}^m + \Delta \mid r_{11}^m) + \theta n(r_{12}^m - \Delta \mid r_{12}^m) \\ =& r_{11}^m + \Delta + \eta\Delta - \theta\eta\lambda\Delta.\end{aligned}$$

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

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

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

However, since player 1 believes that the players share the same reference point r_1^m, he expects that an offer $x_1 > r_{11}^m$ will be rejected and the continuation outcome is to be that each player obtains what he/she is entitled to obtain (i.e., r_1^m). Given this, making an offer $x_1 > r_{11}^m$ only delays agreement, and hence player 1 offers $x_1 = r_{11}^m$ under non-integration.[9] If condition (1) does not hold, it is obviously optimal for player 1 to offer $x_1 = r_{11}^m$. We thus find that it is optimal for player 1 to offer/order what his 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 player 1's optimal offer under non-integration (resp. integration).

4.1.2 Player 2's Acceptance/Compliance Decision

We next study player 2's acceptance decision under each governance structure given player 1's optimal offer $x^m = (\pi/2, \pi/2 - I)$ under non-integration and order $x^h = (\pi, -I)$ under integration. Note that player 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.

Under non-integration, if player 2 accepts the offer $x^m = (\pi/2, \pi/2 - I)$, her utility is given

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

by

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

The shading term suggests that player 2 knows player 1's reference point r_1^m. This follows because player 2 rationally expects that the sender optimally offers what his reference point specifies. If she rejects the offer, on the other hand, her utility is

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

Player 2 then accepts the offer if

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

We next analyze player 2's compliance strategy under integration. Notice that player 1's reference point, which is equal to his optimal order, is given by $r_1^h = x^h = (\pi, -I)$.

If player 2 accepts the order $(\pi, -I)$, she obtains

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

If player 2 rejects the order, her utility is given by

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

Thus, the employee (player 2) 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.

From the previous discussions, we can determine that $\theta_h < \theta_m$. This suggests that non-integration requires a severer punishment than integration for player 2's rejection to realize immediate agreement. There are two reasons for this. First, player 2's rejection under integration provokes player 1 to greater anger than that under non-integration. Since player 1 offers/orders what his reference point specifies, player 2's rejection results in player 1's aggrievement. Furthermore, because player 2 believes player 1's reference point payoff under integration ($r_{21}^h = \pi - I$) to be much larger than that under non-integration ($r_{21}^m = (\pi - I)/2$) and player 2's belief about her continuation payoff P is independent of the choice of the governance structure, player 2 believes that her rejection makes player 1 experience a larger sense of aggrievement under integration ($\eta\lambda \{\pi - (\pi - I - P)\} = \eta\lambda(P + I)$) than under non-integration ($\eta\lambda \{\pi/2 - (\pi - I - P)\} = \eta\lambda(P+I-\pi/2)$). Larger aggrievement results in severer punishment for player 2, which makes her less willing to reject the order.

Second, while player 2's rejection under integration leads to a larger payoff improvement than under non-integration, the former has less impact on her utility than the latter because of loss aversion. Under integration, if player 2 rejects player 1's offer, she can enjoy her payoff improvement $P-(-I) = P+I$. Since player 2's reference point payoff is 0, her payoff improvement leads to gain P and reduction in loss I. Player 2's utility improvement from rejecting the order is then $\eta P + \eta\lambda I$ ($\lambda > 1$). Under non-integration, on the other hand, player 2's payoff improvement $P - (\pi/2 - I)$ leads to loss reduction only, and hence she enjoys the utility improvement $\eta\lambda\{P - (\pi/2 - I)\}$. Intuitively, under integration, player 2 does not expect a large payoff, and hence her payoff improvement from rejecting the order is "too much" for her 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., player 1's shading), and thus the employee (player 2) is less eager to reject the order.

The second reason suggests that each player's belief that player 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 player 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 player 2's rejection than integration:* θ_h

$< \theta_m$. *Thus, the governance structure that achieves faster agreement is summarized as follows:*

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

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 reaching agreement, but the choice of the governance structure does not matter when the punishment for player 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 a severe punishment makes player 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, player 2 does not care about such a negligible threat of punishment and rejects the offer/order as long as she can improve her payoff by doing so.

This result explains how integration facilitates immediate agreement 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 player's expectation of the relevant outcome is different between the two governance structures because of the difference in the adaptation processes 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 Achieves Transaction Cost Minimization?

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.

As mentioned above, while the value actually shrinks because of bargaining delay, the players ignore discounting. We thus assume that although the players behave as if there were no discounting, the surplus shrinks to $\delta\pi - I$ because of bargaining delay, where δ is a source of the cost of delay and can be interpreted as a discount factor.

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

As mentioned previously, the continuation outcome after player 2's rejection is determined to be somewhere between r_1^m and r_2^m, and thus, under non-integration, the negotiation after player 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 players. Hence, the aggregate shading cost (i.e., the sum of each player's shading) is $\theta(\eta\lambda/2)I$.

Under integration, on the other hand, given player 2's rejection, player 2 obtains at least $r_{12}^m = \pi/2 - I$, and hence she enjoys gain at least $\eta\{(\pi/2 - I) - 0\} = \eta(\pi/2 - I)$. However, player 1 experiences a loss larger than $\eta\lambda[\pi - \{(\pi - I) - (\pi/2 - I)\}]$ because he believes that he can obtain π, but player 2's rejection forces him 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: 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 a larger loss and shading cost than non-integration, as in Case 1. 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, player 1's order equal to his reference point is accepted, and hence only player 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:

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

where

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

Such a 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 $\theta_h \leq \theta < \theta_m$. 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:

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

Case 3: 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, player 2 accepts the offer, and hence only player 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 player 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:*

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

and

$$\theta_h \leq \theta < \theta_2,$$

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 player 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.

Condition (2) can be rewritten as follows:

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

Similarly, θ_2 can be rewritten as

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

The right-hand side of condition (2) (resp. θ_2) is decreasing (resp. increasing) in π/I. This implies that larger trade surplus 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. The 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), provide support for the hypothesis that the more relationship-specific a trade becomes, the larger quasi-rents get, 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 players' reference points under integration is larger than under non-integration. That is, while the divergence between r^m_{12} and r^m_{22} is $I/2$, the difference between r^h_{12} and r^h_{22} is I. This can be explained by the fact that under integration, player 2 sets her reference point without internalizing investment cost I.

As mentioned previously, under either governance structure, player 2 believes that her investment cost I is to be compensated and $w_2 = (0, 0)$. Nevertheless, under non-integration, player 2 somewhat internalizes the investment cost when she sets her reference point because she obtains a positive share of the surplus $(\pi - I)$ from *ex post* bargaining. Under integration, on the other hand, player 2 believes that she cannot obtain any portion of the surplus (i.e., $r^h_{22} = 0$), and hence, there is no room for her 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 belief that player 1 takes the entire surplus under integration makes player 2 less willing to reject player 1's offer than under non-integration (see Section 4), but also makes her set her reference point without internalizing the investment cost, which leads to larger aggregate loss and shading costs than under non-integration.[10]

It is worth noting that if player 2 becomes a boss under integration, the trade-off above is eliminated and integration dominates non-integration. This is because in such a case, both players share the same reference point under integration: $r^{h'}_1 = r^{h'}_2 = (0, \pi - I) \equiv r^{h'}$. Since the same reference point is shared between the players, player 1, who is now the subordinate, accepts player 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).

6 Conclusion

This paper examined the question of why integration can avoid costly *ex post* renegotiations and provided an answer. We showed that integration can realize immediate settlement of the adaptation because each player's reference point under integration is employer-favored because of the *ex post* adaptation process under integration. This employer-favored reference point makes

[10]Even if player 2 obtains some portion of the surplus under integration, this trade-off continues to emerge as long as the following conditions hold:

$$r^m_{12} < P < r^m_{22} \text{ and } r^h_{12} < r^h_{22} < P.$$

an employee less eager to reject his boss's order for the following two reasons. First, it is very costly for an employee to reject the order from her boss because the rejection results in a huge amount of anger from the boss and severe punishment. Second, it is not so rewarding for the employee to reject the order because she does not expect a large adaptation payoff from the outset.

We further showed that integration incurs larger aggregate loss and sense of loss than non-integration. This follows because the expectation that player 2 cannot obtain any portion of the surplus makes her set her 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 players have different shading parameters. While our model assumes that the players share the same shading parameter θ, asymmetric shading does not affect our result because player 2's shading does not matter. Hence, any change in either player's shading parameter does not substantially affect our analysis and results.

We next discuss endogenous reference points. Our model takes each player'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 player 1 is completely rational, but player 2 forgets that she can be biased and sets her reference point self-servingly with positive probability. Since player 1 is rational, he takes player 2's bias into account when he sets his reference point. In such a case, as in Köszegi and Rabin's approach, player 1's reference point is given by his probabilistic belief concerning the relevant outcome.

Finally, we can extend our model to analyze the limit of firm scope. Suppose player 1 faces some other transactions similar to the trade in which player 1 and player 2 engage and that θ is decreasing in the number of transactions he conducts: $\theta'(n) \leq 0$, where n represents the number of transactions he handles. The intuition of the latter assumption is that the more transactions player 1 conducts, the smaller effort and the less time he can provide to each transaction (i.e., the harder it is for him to punish those who disobey his 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, player 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 I: 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 AI.1, AI.2, and AI.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.

AI.1 No Reference-Dependence

We first explore how our result is affected when there is no reference-dependence. Note that the optimal offer/order of player 1 does not change in the no-reference-dependence case. We thus need to examine player 2's optimal acceptance/compliance decision only.

Suppose the adaptation outcome is $y = (y_i, y_j)$. In the case where there is no reference-dependence, the utility of player i who has a reference point payoff 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 player's utility function does not include a gain-loss term and, as in the literature on contracts as reference points, each player's shading depends on the difference between a player's 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.

Under non-integration, while player 2's acceptance payoff is given by

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

her rejection payoff is

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

Note that player 1 optimally offers $(\pi/2, \pi/2 - I)$ and that player 2's reference point is $((\pi - I)/2, (\pi - I)/2)$. Comparing U^m_{NRD} and $U^{m'}_{NRD}$ implies that player 2 does not reject the offer if $\theta \geq 1$.

Under integration, on the other hand, if player 2 accepts the order, she obtains

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

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

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

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

The above discussion implies that if there is no reference dependence, the choice of the governance structure does not matter; the governance structure does not affect player 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, player 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 the players and eliminates the difference between the effects of gains and losses on each player's utility.

AI.2 No Self-Serving Bias

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

Suppose both players share the same view regarding each player's outside option: $w_1' = w_2' = (0, -I)$.[11] Hence, both players share the same reference point. That is, under non-integration, their reference points are

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

and, under integration,

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

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

Player 2's acceptance payoff under non-integration is thus given by

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

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

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

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

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

If player 2 rejects the order, she obtains:

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

Hence, we find that player 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 can be explained as follows. Without self-serving bias, both players share the same reference point, and hence, the size of player 2's payoff increase from rejection and that of player 1's payoff decrease from it are exactly the same. Player 2 thus rejects the offer/order if the marginal benefit from rejecting the offer/order (i.e., $1+\eta$) is equal to or smaller than the marginal cost from doing so (namely, $\theta\eta\lambda$).

AI.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 player 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, player 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 player 2's rejection of an offer/order, player 2 rejects any offer/order that yields her a smaller payoff than her continuation payoff. Given that player 1 believes that the continuation payoff is given by r_1^m (his reference point when the players are autonomous), player 1 then optimally offers $x_{NS} = (\pi/2, \pi/2 - I)$ under both non-integration and integration.

Under non-integration, if player 2 accepts the offer, she 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 she rejects it, her 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 player 1's reference point is rejected. By assumption $P > \pi/2 - I$, U_{NS}^m is smaller than $U_{NS}^{m'}$, which means that player 2 always rejects the offer.

Under integration, on the other hand, if player 2 accepts the order, her utility is given by[12]

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

If she rejects the order, she 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, player 2 rejects the order for certain.

The above discussion implies that the governance structure does not matter if there is no shading. This is quite intuitive: player 2's rejection cannot be prevented without any punishment for it.

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

Appendix II: Risk-Averse Players

We here examine a different type of no reference dependence. The utility function of our model exhibits concavity, which is the characteristic of risk-averse preferences. This appendix thus assumes that the players do not have reference-dependent preferences and are risk-averse instead of assuming that they have reference-dependent preferences and are risk-neutral.

Since the players are risk-averse, they have concave utility functions $m(x)$, which is twice differentiable, $m'(\cdot) > 0$ and $m''(\cdot) < 0$. Suppose that each player i has the following overall utility function:

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

This utility function corresponds to that of contracts as reference points. This change in the assumption does not affect player 1's optimal offering strategy, and thus we need to analyze player 2's behavior only.

Under non-integration, player 2's acceptance utility is given by

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

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

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

The discussion suggests that player 2 does not reject the offer under non-integration if her acceptance utility is larger than or equal to her belief about the continuation utility:

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

Under integration, player 2's compliance utility is

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

and her 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, player 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 the following relationships hold:

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

The above 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 players.

In the model in Section 4, player 1's punishment for rejecting the order under integration is severer than his shading under non-integration because both players' reference points are employer-favored under integration. In the concave utility function case, however, the same factor leads to the opposite result. This is illustrated in Figure 1. Since the players have concave utility functions, the same amount of payoff increase/decrease affects their utility differently. Under integration, the amount of player 2's payoff improvement from rejecting the order ($P - (-I) = P + I$) is the same as that of player 1's payoff decrease ($\pi - (\pi - I - P) = P + I$). Nevertheless, the amount of player 2's utility improvement from her rejection, which corresponds to (b) in Figure 1, is far larger than that of player 1's utility decrease from it, which is denoted by (a) in Figure 1. Under non-integration, on the other hand, player 1's utility decrease from player 2's rejection, which is denoted by (c) in Figure 1, is not so small compared with player 2's utility improvement from it, which corresponds to (d) in Figure 1. Hence, integrated firms require much severer punishment for player 2's rejection to offset player 2's benefit from it than autonomous trading parties do.

Appendix III: Players Who Care about Discounting

This section studies the case in which the players care about discounting and check the robustness of our result. To achieve this, we change the setting in the following way (the rest of the settings

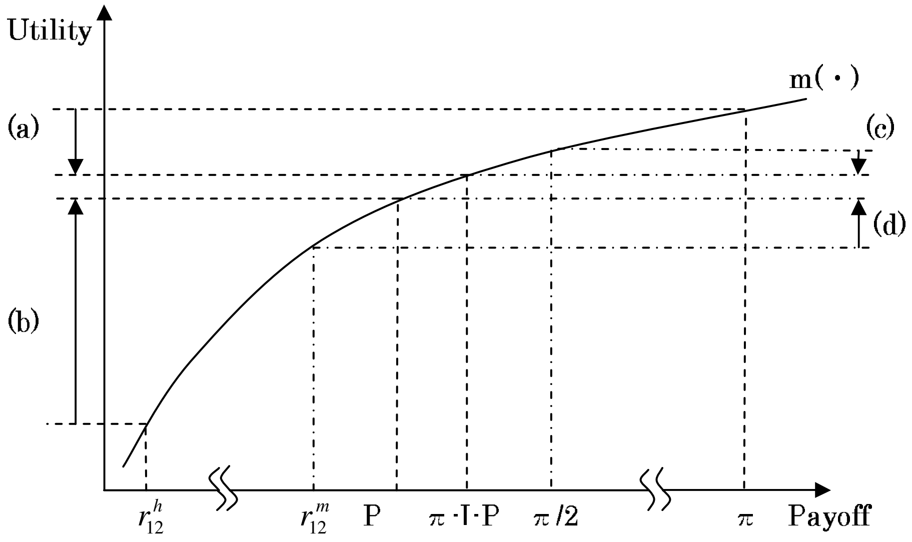

Figure 1: Each Player's Utility Improvement/Decrease from Player 2's Rejection
(a) represents player 1's utility decrease under integration, (b) is player 2's utility improvement under integration, (c) denotes player 1's utility decrease under non-integration, and (d) represents player 2's utility improvement under non-integration.

are the same as in the main model). First, the players 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 conditions hold:

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

and

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

The second condition is the same as in the main model. The first inequation of the first condition implies that player 2 has an incentive to reject player 1's offer/order that corresponds to player 1's reference point (which will be specified below). The second inequation of the first condition means that player 2 does not expect more than what she thinks she is entitled to obtain (namely, her reference point payoff, which will be determined below). Furthermore, the first condition implies that $I/(\pi - 2P) \leq \delta < 2I/(\pi - 2P)$.

This appendix proceeds as follows. Section AIII.1 examines each player's reference point and player 1's optimal offer/order under each governance structure. Sections AIII.2 and AIII.3 study player 2's optimal acceptance/compliance decision under each governance structure. Section AIII.4 presents the result, which is a modified version of Proposition 1.

AIII.1 Reference Points and Player 1's Optimal Offer/Order

We first specify each player's reference point and player 1's optimal offer/order under each governance structure. It is common knowledge that players care about discounting, and hence, their reference points are different from those in our main model. Since both players expect that player 1 sends the offer that makes player 2 indifferent about whether she accepts it and that player 2 accepts such an offer, player 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 player 1 believes that player 2 is to incur her sunk investment (i.e., he believes that the players' outside options are $w_1 = (w_{11}, w_{12}) = (0, -I)$) and that the expected bargaining outcome is assumed to be given by the Nash bargaining solution.

As mentioned in the main model, it is optimal for player 1 to offer what his reference point specifies given that player 1 believes that both players share the same reference point. Thus, his 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, player 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 player 1 to demand more. The optimal order, which is equal to player 1's reference point, is thus given by

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

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

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

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

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

AIII.2 Player 2's Acceptance: Non-integration

This subsection studies player 2's optimal acceptance decision under non-integration given player 1's optimal offer x^{m*} and the players' reference points, r_1^{m*} and r_2^{m*}. If player 2 accepts the offer, she obtains payoff $\delta\pi/2 - I$, which leads to her sense of loss $\eta\lambda\{(\delta\pi - I)/2 - (\delta\pi/2 - I)\} = \eta\lambda I/2$, and incurs no shading from player 1. Player 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 she rejects the offer, on the other hand, her 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 player 2's belief about the continuation outcome is discounted since the players care about discounting. We thus find that player 2 does not reject the offer if the following condition holds:

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

AIII.3 Player 2's Compliance Decision: Integration

We next examine player 2's compliance decision under integration. If player 2 accepts the optimal order x^{h*}, her 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 she rejects it, on the other hand,

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

Player 2 thus does not reject the order if

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

AIII.4 Immediate Agreement and Governance Structures

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

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

Since $\delta^* \leq 2I/(\pi - P)$ holds, the case in which $\theta_m^* \geq \theta_h^*$ does exist. We thus have the following proposition:

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

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

where

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

This implies that when the cost of delay is not so large, integration achieves immediate agreement more easily than non-integration. When the players care about discounting, player 2 faces two costs from rejecting the offer/order: punishment for rejection (player 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 player 2's utility compared to player 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 players' utility under either governance structure (i.e., δ is close enough to 1).

References

Akerlof, Robert. 2010. "Punishment, Compliance, and Anger in Equilibrium." Job Market Paper, MIT, Sloan School. http://mit.academia.edu/RobertAkerlof/Papers/163148/Punishment_Compliance_and_Anger_in_Equilibrium_JOB_MARKET_PAPER_

Babcock, Linda, George Loewenstein, Samuel Issacharoff, and Colin Camerer. 1995. "Biased

Judgements 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.

Barnard, Chester I. 1938, *The Functions of the Executive.* 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 discussion paper, University College, London.

Bolton, Gary and Rami Zwick. 1993. "Anonymity versus Punishment in Ultimatum Bargaining." *Games and Economic Behavior*, 10(1): 95-121.

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

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

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

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

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

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*, XLV: 629-685.

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

Matouschek, Nico. 2004. "Ex Post Inefficiencies in a Property Rights Theory of the Firm."

Journal of Law, Economics, and Organization, 20: 125-147.

Monteverde, Kirk, and David Teece. 1982. "Supplier Switching Costs and Vertical Integration in the Automobile Industry." *Bell Journal of Economics*, 13: 206-12.

Mori, Yusuke. 2011. "A Formal Theory of Ex Post Adaptation: A Trade-Off between Rent Seeking and Bargaining Losses."
Available at SSRN: http://ssrn.com/abstract=1975624 or
http://dx.doi.org/10.2139/ssrn.1975624

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

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