A Formal Behavioral Model of Firm Boundaries: Why Does Authority Relation Mitigate Ex Post Adaptation Problems?*

Yusuke Mori†

This version: December 25, 2012

Abstract

We explore why authority within firms helps trading parties immediately settle ex post adaptation problems despite the possibility of a subordinate's disobedience to the orders of his boss. By employing three crucial behavioral assumptions (reference-dependent preference, self-serving bias, and shading), we point out that the choice of governance structure affects trading parties' expectations about outcome of ex post adaptations and show that a subordinate is likely to obey orders of his boss because he is expected to do so. Nevertheless, our study also points out that such a positive aspect of authority comes with subordinate's psychological disutility.

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, Junichiro Ishida, Akifumi Ishihara, Shinsuke Kambe, Shintaro Miura, Tomoharu Mori, Fumitoshi Moriya, Takeshi Murooka, Sadao Nagaoka, and Yasuhiro Shirata 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, 2012 Japanese Economic Association Autumn Meeting, the 6th Annual Meeting of the Association of Behavioral Economics and Finance, and the seminar at Hitotsubashi University for their helpful comments. An earlier version of this paper entitled "How Can Integration Reduce Inefficiencies Due to Ex Post Adaptation?" received the Osaka University Institute of Social and Economic Research/Moriguchi Prize. 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

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. It follows because while non-integrated parties have to engage in costly negotiations for the ex post adaptations to unanticipated disturbances, 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 is supported by a number of empirical studies (see Lafontaine and Slade, 2007 for a survey of these studies).

The discussion above 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.

This paper develops a formal model that explores the effectiveness of authority in the context of ex post adaptations. Especially, we focus on the situation where trading parties in bilateral monopoly due to relationship-specific investment engage in the division of trade value, which is most likely to cause conflicts between them. We show that authority helps trading parties reach agreement on the division of the value immediately despite the possibility of a subordinate's disobedience to the order of his boss (i.e., integration achieves immediate agreement more easily than non-integration).

There is 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 our study employs 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 adaptation processes, each governance structure leads to different reference points (i.e., the process by which adaptation outcomes are determined affects the parties' reference points). Under non-integration, as mentioned above, ex post adaptation 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 adaptation is implemented 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).

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 belief about the sunk cost might seem unreasonable, Macleod (2007) points out a concept of fairness based on the idea that par-

[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.

ties 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). It is worth noting that our model can be easily extended 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 reflect the experimental fact of Binmore, Swierzbinski, and Tomlinson (2007), but also substantially simplifies our analysis. The case where the parties do care about discounting will be dealt with in Appendix C.

Some reader 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 those who make decisions at the level of organizations is an individual (boundary-role person), his decisions can be affected by these behavioral aspects.

Our model presents two reasons why integration achieves immediate agreement on the division of the value more easily than non-integration despite the possibility of the disobedience to orders. First, disobedience to an order under integration provokes severer punishment than rejec-

tion 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 adaptation is implemented 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 adaptation outcome 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 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, while integration can economize inefficiencies

[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).

due to delay in reaching agreement, it incurs larger shading costs (i.e., deadweight loss caused by shading) than non-integration. The reason for this is 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 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 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 paper 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 contracts-as-reference-points approach, which points out that contracts serve as reference points. 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 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 belief 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 our paper 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 an 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. In our study, the choice of governance structure only affects the ex post adaptation process 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 adaptation process: bilateral bargaining (non-integration) or fiat (integration).[3]

Herweg and Schmidt (2012) explore how loss aversion affects the outcome of ex post contract renegotiation and show that loss aversion interrupts efficient renegotiation. Both their study and

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

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).

We lastly review some approaches to firm boundaries: TCE and the property-rights theory. While the former approach, as in Coase (1937) and Williamson (1996), focuses on authority, the latter approach, as in Grossman and Hart (1986), Hart and Moore (1990), and Hart (1995), stresses the choice of ownership structures.

TCE asserts that authority helps integrated firms to 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 takes place only under non-integration, creates a trade-off between rent seeking and bargaining costs. In Mori (2011), however, as in the literature on TCE, integration is assumed to avoid bargaining costs without offering a formal justification for the assumption. This study adopts TCE's idea that authority is the most important aspect of integration (internal organizations) and complements its arguments by showing that the presence of authority (i.e., the choice of governance structure) affects each party's expectation, and hence, the timing of agreement.

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

non-integration or integration maximizes the aggregate disagreement payoff, joint ownership minimizes it). Our study, on the other hand, assumes that ownership structure does not affect parties' 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 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 examines which governance structure realizes immediate agreement on 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 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 the division of trade value.[4] The trade requires party 2's 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, the parties set their reference points regarding how the value will be divided. A process to divide the value is then initiated.

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

We assume that under integration, party 1 (resp. party 2) becomes a boss (resp. a subordinate).[5]

The process of the value division 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.[6] If party 2 accepts the offer, the surplus 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.

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.

[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 14.

[6]The assumption that party 1 has the right to send an offer 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 the offer is assigned to each party with equal probability without changing our result.

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 shading. That is, 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 belief about party j's reference point payoff). Party 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

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

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

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

[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 $\theta \min(y_i - r_{ii}, 0)$).

(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 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 belief about his 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 belief 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).

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 Williamson (1996) notes, "the autonomous stages would need to bargain these [adaptations to unanticipated disturbances] through to agreement" (p. 17), 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, "the unified firm can implement adaptations to unanticipated disturbances by fiat" (Williamson, 1996, p. 17). 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 adaptation 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). We assume that each party knows about self-serving bias. That is, each party knows that his partner has a different reference point (e.g., under non-integration, party 1 knows that party 2's reference point is 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. For simplicity, we assume that after party 2 rejects party 1's offer/order, each party obtains a continuation payoff. As mentioned above, since each party is affected by self-serving bias, party 1 believes that party 2's continuation payoff cannot be larger than $r_{12}^m = \pi/2 - I$, which is her belief about party 2's reference point payoff when the trading parties are autonomous. Party 2 is also influenced by self-serving bias (i.e., before observing party 1's offer/order, party 2 believes that his continuation payoff cannot be smaller than $r_{22}^m = (\pi - I)/2$), but he can update his belief about his continuation payoff by observing party 1's offer/order. That is, party 2 infers what party 1 seriously believes about continuation outcome from her offer/order, and expects that real continuation out-

[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.

come is to be specified somewhere between r_1^m and r_2^m (through a negotiation, for example). Hence, party 2's belief about his continuation payoff after observing party 1's offer/order, P, satisfies[10][11]

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

It will turn out that party 1 optimally offers what her reference point specifies, and hence, this assumption about party 2's continuation payoff P implies that he has an incentive to reject party 1's optimal offer regardless of the choice of governance structure.

We assume that each party's belief about party 2's continuation payoff does not depend on the governance structure chosen at the 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. 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 belief about each party's continuation payoff between the two governance structures.

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

[11] This setting does not rule out party 1's belief update. For example, party 2's counter offer, which is not modeled, might help her modify her belief about continuation outcome. However, since we focus on whether the first offer is accepted, such update does not matter.

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 (the division of the trade value) 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 party 2's continuation payoff is given by r_{12}^m (i.e., her belief about what he is entitled to obtain as an autonomous party).

Since $\theta \leq (1+\eta\lambda)/\eta\lambda$ 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 $x_1 \geq r_{11}^m$ under non-integration and $x_1 \geq r_{11}^h$ 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 $x_1 = r_{11}^m + \Delta$ ($\Delta \geq 0$).

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

$$\begin{aligned} U_1^m(x \mid r_1^m, r_2^m) =& r_{11}^m + \Delta + n(r_{11}^m + \Delta \mid r_{11}^m) + \theta n(r_{12}^m - \Delta \mid r_{22}^m) \\ =& r_{11}^m + \Delta + \eta\Delta - \theta\eta\lambda\left(\frac{I}{2} + \Delta\right). \end{aligned}$$

Note that party 1 knows party 2's reference point $r_2^m = (r_{21}^m, r_{22}^m)$. 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 his reference point payoff (r_{22}^m), party 1 expects him to shade by $\theta\eta\lambda\{(r_{12}^m - \Delta) - r_{22}^m\} = \theta\eta\lambda\{(I/2) + \Delta\}$. Thus, party 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 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 2 knows party 1's 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.[12] If condition (1) does not 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

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

non-integration (resp. integration).

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 knows party 1's reference point r_1^m. If he rejects the offer, on the other hand, his utility is

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

Party 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 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

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

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 belief 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 improve-

ment $P-(-I)=P+I$. 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 $\eta P+\eta\lambda I$ (λ>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 belief 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:*

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

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 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 party's expectation of the relevant outcome is different between the two governance structures due to 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 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. We thus assume that 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 (θ<θ_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 ($\theta_h \leq \theta < \theta_m$): 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:

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

where

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

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 ($\theta_m \leq \theta$): 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:*

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

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 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 belief 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.[13] [14]

6 Conclusion

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

[13] 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 $r_{12}^h < r_{22}^h < P$.

[14] 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).

reject the order because he does not expect a large adaptation 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 sets 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 adaptation outcome 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

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

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 belief 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: $w_1' = w_2' = (0, -I)$.[15] 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

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

Since party 2 has the same reference point as party 1, accepting the offer leads to no aggrieve-

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

ment. If he rejects the offer, he 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 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

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

If party 2 rejects the order, he 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 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, $\left|r^g_{NSSB1} - (\pi - I - P)\right|$ and $\left|r^g_{NSSB2} - P\right|$ 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 party 2's 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[16]

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

[16]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.

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 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 belief about the 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.

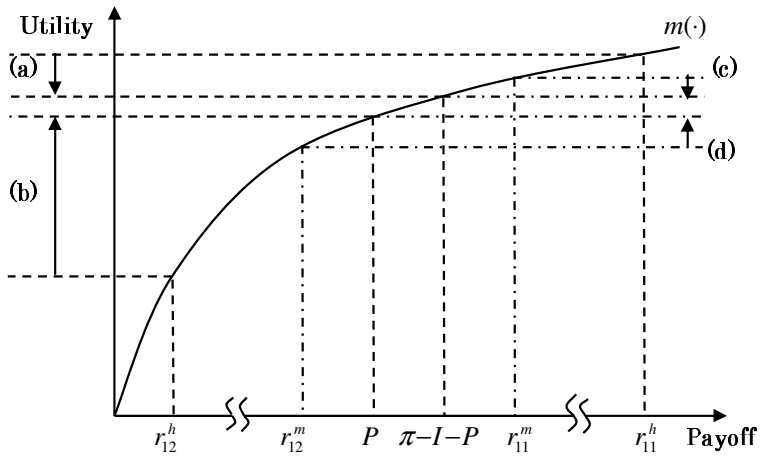

Figure A: Each Party's Utility Improvement/Decrease from Party 2's Rejection

(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.

In the model in Section 4, 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.

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:[17]

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

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.

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

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

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, it is optimal for party 1 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 belief 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,

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

Party 2 thus does not reject the order if

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

C.3 Immediate Agreement and Governance Structures

Comparing θ^*_m and θ^*_h leads to the following result:

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

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 $\theta \leq (1 + \eta\lambda)/\eta\lambda$ 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:

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

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

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

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.

References

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://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 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.

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.

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

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

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

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.

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

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

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

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

Mori, Yusuke. 2011. “A Formal Theory of Firm Boundaries: A Trade-Off between Rent Seeking and Bargaining Costs.” Hitotsubashi University.
Available at SSRN: http://ssrn.com/abstract=1975624 or
http://dx.doi.org/10.2139/ssrn.1975624

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.

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

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.