Essays on Relational Contracts

Akifumi Ishihara

A thesis submitted to the Department of Economics of the London School of Economics for the degree of Doctor of Philosophy, London, December 2011

Declaration

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

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

Abstract

This dissertation contains three essays on self-enforcing implicit contracts in economic transactions and politics.

Chapter 2 studies a repeated agency model with two tasks where the agent has private information on the first task and there is no verifiable performance signal for the second task. The equilibrium level of the first task is determined so as to guarantee the credibility of the relational contracts to provide incentives for the second task. It implies interesting economic results including non-monotonic relation between the discount factor and the total surplus, social desirability of unverifiability, and implications for organization design.

Chapter 3 studies a model of political contribution of dynamic common agency where state-contingent agreements must be self-enforced. First, we investigate the punishment strategy for supporting the self-enforcing mechanism. The most severe punishment strategy on the principals takes the form of a two-phase scheme in general. Second, we characterize the payoff set of the equilibria on which the same decision is chosen by the agent through implicit agreements and examine whether it can achieve the same payoff as in the standard static menu auction model. It implies that there could be an equilibrium outcome in a static menu auction that cannot be supported in our model for any discount factor.

Chapter 4 studies repeated political competition with policy-motivated citizen candidates. The dynamic relationship could cause strategic candidacy in two-candidate competition, such as in circumstances where two candidates stand for election and one of them has no chance to win. The candidate can choose her implementing policy depending

on the set of the rival candidates in the election and the rival candidate actually has an incentive to stand even with no chance to win since it can induce policy compromises from the winning candidate.

Contents

Acknowledgement 9

1 Introduction **11**

2 The Interaction of Formal and Implicit Contracts with Adverse Selection **15**
2.1 Introduction 15
2.2 Related Literature 21
2.3 The Model 24
2.3.1 Environment 24
2.3.2 Economic Interpretation of the Setting 27
2.3.3 Strategy and Equilibrium 28
2.4 Analysis 32
2.5 Social Desirability of Unverifiability 39
2.6 Task Assignment Problem 43
2.6.1 Centralization 44
2.6.2 Task Separation 45
2.6.3 Implications 46
2.7 Extension and Robustness 50
2.7.1 Extension to a General Setting 50
2.7.2 Assumptions on the Behaviour 52
2.8 Conclusion 58
2.9 Appendix: Proofs 60
2.9.1 Proof of Proposition 2.1 60
2.9.2 Proof of Lemma 2.1and 2.2 72
2.9.3 Proof of Corollary 2.1 78
2.9.4 Proof of Proposition 2.3 78
2.9.5 Proof of Lemma 2.3 78
2.9.6 Proof of Lemma 2.4 79
2.9.7 Proof of Proposition 2.5 80
2.10 Appendix: The Optimal Contract without Assumption of Limited Commitment Ability 80

3 Relational Political Contribution under Common Agency **87**
3.1 Introduction 87
3.2 The Model 93
3.2.1 Environment 93
3.2.2 Simple Strategy Representation 94
3.3 Decision-Stationary Equilibria 98
3.4 The Optimal Penal Code 101
3.4.1 The Optimal Penal Code on the Agent 102

3.4.2 The Optimal Penal Code on the Principals . . . 102
3.5 Validity of Menu Auction for Political Contribution . . . 112
3.5.1 BW's Model . . . 112
3.5.2 Comparison between SMA and RPC . . . 113
3.5.3 Example . . . 118
3.6 Discussion . . . 119
3.6.1 Approximation . . . 119
3.6.2 Conflict Makes the SMA Vulnerable . . . 120
3.6.3 Other SMA Equilibria . . . 121
3.6.4 SMA with Caps on Transfer . . . 121
3.7 Conclusion . . . 123
3.8 Appendix: Proofs . . . 124
3.8.1 Proof of Proposition 3.1 . . . 124
3.8.2 Proof of Proposition 3.2 . . . 126
3.8.3 Proof of Proposition 3.3 . . . 130
3.8.4 Proof of Lemma 3.4 . . . 132
3.8.5 Proof of Proposition 3.5 . . . 133
3.8.6 Proof of Corollary 3.1 . . . 135
3.8.7 Proof of Proposition 3.6 . . . 135
3.8.8 Proof of Lemma 3.5 . . . 136
3.8.9 Proof of Proposition 3.7 . . . 136
3.8.10 Proof of Proposition 3.8 . . . 136
3.8.11 Proof of Lemma 3.6 . . . 137
3.8.12 Proof of Proposition 3.10 . . . 138

4 Strategic Candidacy *via* Endogenous Commitment **139**
4.1 Introduction . . . 139
4.2 One-Shot Political Competition . . . 145
4.2.1 Environment . . . 145
4.2.2 Political Equilibria . . . 146
4.2.3 Impossibility of Strategic Candidacy with Two Candidates . . . 148
4.3 Repeated Competition . . . 149
4.4 Strategic Candidacy in Two-Candidate Competition . . . 151
4.4.1 Definition . . . 151
4.4.2 Example: A Spatial Model with Five Citizens . . . 151
4.4.3 The Necessary and Sufficient Condition . . . 155
4.5 Is Candidacy Necessary for Political Compromises? . . . 158
4.5.1 Example Revisited . . . 159
4.5.2 Deterrence of the Median's Candidacy . . . 161
4.6 Concluding Remarks . . . 165
4.7 Appendix: Proofs . . . 167
4.7.1 Proof of Proposition 4.2 . . . 167
4.7.2 Proof of Proposition 4.3, 4.4, and 4.5and Corollary 4.1 . . . 167
4.7.3 Proof of Proposition 4.6and 4.7 . . . 174
4.8 Appendix: Characterization of Weakly Dominated Voting . . . 178

List of Figures

2.1 Timing within a Period 24
2.2 Equilibrium Constraints 33
2.3 Total Surplus and P's payoff in the OPPE 38
2.4 Comparison between the Verifiable and Unverifiable Cases 40
2.5 Comparison among (B), (C), and (S) 47

3.1 In Case of Exclusion-type 108
3.2 In Case of Sanction-type 109

4.1 Consequent Policy 153
4.2 Policies by citizen 2 and 3 161
4.3 When $k' < M$ 163
4.4 When $k' \geq M$ 164

List of Tables

3.1 Example 1 109
3.2 Example 2 118

Acknowledgement

.

There are so many people to support my academic life throughout the graduate programme so far and it would be impossible to achieve the completion of this dissertation without them. Nevertheless expressing my gratitude to all of them would make a too big book. So I want to apologize to acknowledge very limited persons.

I am greatly indebted to my advisors Andrea Prat and Leonardo Felli. Andrea continuously gave me insightful suggestions which often made a breakthrough in my research. He also often told me the importance of economic insights from the model. It is an invaluable lesson for me who sometimes falls to do just a mathematical exercise in my research. Leonardo gave many specific and careful comments those were definitely helpful for improving the dissertation. His suggestion on my writing and presentation has also been useful for expressing my research in an effective way.

This dissertation substantially benefited by comments from and discussions with Simon Board, Kyoichi Eguchi, Susanne Goldlücke, Makoto Hanazono, Daisuke Hirata, Hideshi Itoh, Andrea Mattozzi, Gerard Padró i Miquel, Volker Nocke, Martin Peitz, Michele Piccione, Ronny Razin, Tadashi Sekiguchi, Konrad Stahl, Balazs Szentes, Yuichi Yamamoto and anonymous referees. Kara Contreary and Daniel Vernazza kindly proofread my drafts. I also had exciting discussions on economics with Bonsoo Koo, Prakarsh Singh, Noriyuki Yanagawa, and Yukihiro Yazaki. Fortunately I had many opportunities to present my research for participants in various seminars and conferences. I am grateful to all of these people.

Finally and very personally, I would like to express my sincere gratitude to my farther and mother, Toshio Ishihara and Keiko Ishihara, for their understanding, patience, and support of my graduate study.

Chapter 1

Introduction

Contracts and agreements have played a major role in a broad range of economic, political, and social activities. One of the important aspects of contracts and agreements is that these are typically concluded *ex ante* for stipulating the *ex post* behaviour. It means that contracts and agreements have little influence if the *ex post* enforcement is not guaranteed.

Whereas recent development of the theory of contract and mechanism design has provided a huge amount of insights on information and incentives, it has less emphasized its importance of the role of enforcement of contracts and agreements. The standard agency model and mechanism design problems typically assume that a contract or mechanism offered by a principal or mechanism designer is enforced by exogenous third party such as courts (or able to be committed by them). However this is often too ideal for the following reasons at least.

Illegal Contract: the law may stipulate to prohibit (illegal) agreements being enforced.

Imperfect Law System: the law system may not be matured enough to guarantee the enforcement.

Lack of Information: the court (or another third party who is considered the enforcing party) may not be able to judge whether an *ex post* situation is what was stipulated in the contract.

Nevertheless we have observed that contracts and agreements have positive impacts in many cases even if there are some enforcement problems due to the above reasons. It indicates a next important direction of incentive theory to consider the enforcement problem explicitly to obtain additional insights on information and incentive problems.

Motivated by the above discussion, this dissertation theoretically investigates incentive problems in economics and politics and discusses the impact of agreements when the enforcement is not perfectly guaranteed. Especially, we focus on situations in which self-enforcement agreements are supported by ongoing relationships between the involved parties. If an agreement in each period is not guaranteed to be enforced exogenously, then each of the involved parties has an option to deviate from the agreement. In the ongoing relationships, however, such a deviation might be deterred since the deviation could induce the loss from punishment for the future relationship.

In the following three chapters, we consider incentive problems of multitasking agency, lobbying activity *via* political contribution, and political competition in elections respectively. Each of them has a common question; how to deal with the self-enforcing agreements in the ongoing relationship.

Chapter 2 studies a relational contracting problem where there exist adverse selection and multitasking problems. Specifically we consider a repeated agency model with two tasks where the agent has private information on the first task and there is no verifiable performance signal for the second task. The principal is faced with inefficiency generated by information rent from the first task and the problem of designing self-enforcing contracts for the second task. Under the assumption that the principal has full bargaining power and imperfect commitment ability, the equilibrium level of the first task is determined so as to guarantee the credibility of the relational contracts to provide incentives for the second task. This interaction provides interesting results as follows. First, the total surplus is not monotonically increasing in the discount factor. In particular, the total surplus for intermediate values of the discount factor is greater than for high values since the principal must care about sustaining relational contracts rather than distorting the first task to reduce the information rent to the agent. Second, for intermediate values of the discount factor, the equilibrium is more socially efficient than when both the first and second tasks are verifiable. Interestingly it implies that to assure enforcement of agreements sometimes induces socially worse outcomes. Furthermore, in order to obtain insights for organizational design, we consider a job design problem of whom the principal allocates the second task to. If the second task can be allocated to the principal herself, then the total surplus could be lower than in allocating the second task to the agent. Since the information rent can be an incentive device for the unverifiable task, the principal never exercises her option to allocate the second task to a different agent. From a perspective of organizational design, these results imply the benefit of delegation and bundling of tasks due to the effect of information rent on the other task through the self-enforcement mechanism.

Chapter 3 studies self-enforcing contracts of political contribution. While the literature typically assumes that a payment plan offered by lobbyists is treated as a binding contract enforced by the court after the agent makes a decision, it is not the case in many countries. From the concern of the enforcement problems of political contribution, we consider a model of dynamic common agency where state-contingent agreements must be self-enforced. First, we investigate the punishment strategy for supporting the self-enforcing mechanism. The most severe punishment strategy on the principals takes the form of a two-phase scheme in general. More specifically, it is either an "Exclusion-type" where a decision undesirable for the deviating principal is chosen over time, or a "Sanction-type" where once the undesirable decision is chosen and the deviating principal contributes an additional amount of the payment, it is cancelled. Second, we characterize the payoff set of the equilibria on which the same decision is chosen by the agent through implicit agreements and examine whether it can achieve the same payoff as in the standard static menu auction model. It implies that there could be an equilibrium outcome in a static menu auction that cannot be supported in our model for any discount factor. This result has an important meaning since a political outcome predicted by a static menu auction model might not be justified by ongoing relationships even when the political players are assumed to be patient.

Chapter 4 studies repeated political competition with policy-motivated citizen candidates. While an election promise cannot be a binding agreement in practice, the dynamic relationship can guarantee an ability to commit to a policy that is different from the ideal policy for the fear of worse political process after implementing her ideal policy. It causes strategic candidacy in two-candidate competition, such as in circumstances where two candidates stand for election and one of them has no chance to win which never emerges when there is no commitment ability. The candidate can choose her implementing policy depending on the set of the rival candidates in the election and the rival candidate actually has an incentive to stand even with no chance to win since it can induce policy compromises from the winning candidate. We furthermore show that in some situations, a political compromise is induced with strategic candidacy which would be impossible without the strategic candidacy.

Chapter 2

The Interaction of Formal and Implicit Contracts with Adverse Selection

2.1 Introduction

Problems of incentive provision are often observed in economic relationships such as procurement transaction, organization, and regulation of industry. The incentive problem often becomes more serious when there are multiple dimensions of tasks, all of which the right incentives must be provided for. As Holmström and Milgrom (1991) demonstrate in the principal-agent relationship, if the agent is faced with multi-dimensional tasks, then the principal must be concerned with balancing incentives among the tasks as well as the incentive provision. In the extreme case where some tasks do not generate any objective performance measure, legal contracts can never compensate him for the effort devoted to these tasks and then the agent would devote his effort only to a subset of the tasks. Thus the multitask property has suggested additional insights for contract design. It has also stimulated some issues in the theory of the firm. In particular, the task assignment problem to decide who has the responsibility of the task is definitely motivated by the multitasking incentive problem.

The aim of this chapter is to find new insight into the multitasking incentive problem caused by *ex ante* asymmetric information and unverifiability of performance by courts.

We argue that the interaction of information rent and reneging temptation of relational contracts causes several interesting results. Specifically, raising the discount factor and/or making the signal on the performance verifiable do not necessarily improve economic efficiency. This interaction provides new incentive perspective for the job design problem. The deriving force of these results are the information rent caused by *ex ante* asymmetric information, which affects the incentive provision in the task without verifiable signals.

In what follows, we analyse a repeated principal-agent relationship with two tasks. Each task can lead to a potential inefficiency. On the first task, the agent has *ex ante* private information on his cost in each period. On the second task, the cost is technologically independent of his private information. However the performance measure is unverifiable so that it cannot be assessed by a third party and formal contracts contingent on it are not enforced by the court. We assume that both tasks are essential for the parties' payoff so that the principal must deal with these two incentive problems simultaneously.

One way to create incentives related to the second task is the use of relational contracts based on the unverifiable measurement. More specifically, the principal uses both formal contracts contingent only on verifiable signals and informal agreements contingent on unverifiable signals as well as verifiable ones. This is consistent with evidence showing that both formal and informal contracts are used as incentive devices.[1] In our model, the principal optimizes hew own payoff every period subject to two kinds of constraint. First, the principal must deal with the information rent caused by the agent's private information. Second, since informal agreements cannot be enforced by the court exogenously, both parties must be concerned with ensuring to honour the informal agreements

[1]Gibbons (1998, 2005b) describes a number of cases in which bonus plans based on subjective assessments together with formal contracts were used within the firm. Asanuma (1989) investigates and concludes that in the Japanese automobile and electric machinery industries, the written contracts between a manufacturer and a supplier were relatively coarse, and that well-established informal practices supported their economic performances.

voluntarily.

It is well known that there are many equilibria in repeated games even if they are restricted to the public perfect equilibria those are tractable in repeated games with public monitoring.[2] In order to obtain the prediction from our model, we make the additional assumption (Assumption 2.4) which requires that in each period the principal maximizes her own payoff as long as the players are on the equilibrium path. This restriction is analogous to the standard principal-agent models which presume the principal to make a take-or-leave-it offer. It also means the principal's imperfect ability at commitment. It implies that she cannot commit to the strategy which is not maximizing her own payoff in some period even if it can induce a preferable outcome *ex ante*.

The above assumption implies that the equilibrium is stationary; in equilibrium, the principal offers the same contract and the agent decision depends only on the current type.[3] Together with an additional assumption[4] on the equilibrium behaviour, this stationarity further implies that the equilibrium is characterized by the principal's static optimization problem as in the standard adverse selection model with an additional constraint requiring the players to honour the informal agreements. This additional constraint determines the scope of relational contracts and it is satisfied if the sum of the discounted future total surplus between the principal and the agent exceeds the agent's current deviation benefit on the second task. Thus, if this additional constraint is relevant, then the principal's concern of incentive provision in unverifiable tasks can alter the design of formal contracts from the optimal contract without the additional constraint. This effect causes the following results.

[2] The definition of the public perfect equilibrium is provided by Fudenberg et al. (1994).

[3] Without Assumption 2.4, under the strategy which maximizes the principal's payoff, she can exploit the agent in the first period and compensate for it in the future by maximizing the future aggregate benefit. See Section 2.7 for details.

[4] It is corresponding to Assumption 2.3, which we make later.

It is the comparative static analysis of our equilibrium characterization with respect to the discount factor that gives us our first main result. Clearly, the discount factor determines the scope of relational contracts since it affects the future discounted total surplus. It is important to distinguish three cases: high, intermediate, or low discount factor. If the discount factor is low, then relational contracts do not work at all and there is no value of the relationship between the principal and the agent since there is no way to provide the agent incentives for the second task. Conversely, if the discount factor is high, then relational contracts work well enough to provide the right incentive with respect to the second task. Nevertheless the principal is still faced with hidden information with respect to the first task and hence she attempts to decrease the information rent to the agent. As in the standard adverse selection problem, this concern leads to the inefficient outcome with respect to the first task.[5]

When the discount factor is intermediate, the same equilibrium outcome as in the case of high discount factor cannot be supported since it does not create a sum of the discounted future total surplus that is large enough to sustain the relational contracts. Nevertheless the principal can increase the total surplus by mitigating the inefficiency caused by the information rent problem in the first task. In order to ensure the success of the relational contracts, she improves the efficiency relative to the case of a high discount factor. In other words, by decreasing the discount factor from a high level to an intermediate level, the principal's objective is altered from purely reducing the information rent to ensuring the functioning of the relational contract, which leads to a more efficient outcome.

The second main result shows that unverifiability could be socially desirable. Specifically, we compare our equilibrium characterization described above with one in which the performance measure on the second task is also verifiable under Assumption 2.4. In

[5]Specifically, the information rent problem causes the downward distortion on the first task.

the latter, the principal's concern for ensuring the efficacy of relational contracts vanishes and then, regardless of the discount factor, the outcome is the same as the former for high discount factors.[6] It implies that, for intermediate discount factors, the aggregate benefit between the two parties is greater when the second task is unverifiable than when it is verifiable.[7] When the second task is also verifiable, the inefficiency caused by her concern about the information rent still remains even if the discount factor is not high.

In order to obtain insights for organizational design, we next consider the following task assignment problem. Before the game starts, the principal is allowed to allocate the second task to the agent or another party. We investigate three possibilities as follows: centralization where the principal performs the second task, task bundling where both tasks are assigned to a single agent, and task separation where the second task is allocated to a different agent . Clearly, task bundling is the same as the setting explained above.

The equilibrium under centralization is exactly the same as the one where the second task is verifiable. Under centralization, there is no agency problem in the second task and the principal's only concern is to reduce the information rent. Then as in the discussion of verifiability, when the discount factor is intermediate, centralization generates lower social surplus than task bundling. Note that in contrast with centralization, task bundling is interpreted as delegation of the decision right to the agent. Thus it advocates the delegation of authority to the informed party; if the decision right can be sold to the agent, then the principal might prefer delegation since it can work as a commitment device to produce a more efficient outcome.

When only task bundling and task separation are available, the former is preferred to

[6]Under Assumption 2.4, when the performance measure on the second task is also verifiable, the equilibrium results in the repetition of the equilibrium in the one shot game. The same outcome is implemented when the performance measure on the second task is unverifiable and the discount factor is high.

[7]The principal's payoff is greater when the second task is verifiable than when it is unverifiable.

the latter from the social perspective. Intuitively, by bundling the tasks, the information rent obtained from the first task in the future can be an incentive device for the second task. By contrast, in task separation, the agent working on the second task no longer has opportunity to obtain the future information rent. Furthermore, the principal no longer has an incentive to improve the aggregate benefit among all the parties since doing so just increases the information rent for agent 1 and does not help to relax the self-enforcement constraint at all. These effects imply that task separation is undesirable not only from the social perspective but also the principal's perspective.

It should be emphasized that while the literature has shown the results similar to our main results, the logic behind our results is different from them. On desirability of unverifiability, the literature typically has argued that unverifiability can make a stronger punishment for deviation which can reinforce incentive provision through relational contracts. The literature of the job design problem in multitasking agency problems mainly has discussed complementarity of tasks and identified its conditions in terms of technological aspects such as cost structure and observability and verifiability of signals. Both of the literatures usually abstract *ex ante* asymmetric information which causes the information rent problem. By contrast, this chapter argues that the information rent problem causes the similar phenomena. Our explanation for desirability of unverifiability is that since the principal cares about the information rent, unverifiability can alter the her objective from her own payoff to total surplus for sustaining relational contracts, which can improve the total surplus. The task assignment problem in our model suggests that since the information rent could be a useful instrument for the incentive provision on tasks without verifiable signals, tasks with hidden information and unverifiable measurement should be treated as complementary tasks from an incentive perspectives and allocated to a single agent.

The next section reviews the related literature and clarifies the contribution of this chapter. Section 2.3 describes the main model and defines the strategy and equilibrium used in the analysis. Section 2.4 characterizes the equilibrium conditions and provides comparative statistic analyses. Section 2.5 points out the possibility that unverifiability is socially desirable. Section 2.6 considers the task assignment problem. Section 2.7 discusses the robustness of the results. The final section concludes. The appendix contains proofs for some propositions and lemmas.

2.2 Related Literature

This chapter is related to several strands of the literature on contract theory and the theory of the firm. In order to clarify the relation to the literature, we classify the related papers into the following three categories, relational contracts and screening, verifiability of performance measures, and multitasking incentive agency and job design.[8]

Relational Contracts and Screening Several papers study the design of relational contracts for screening. However none of them include multitasking problems. Levin (2003) develops the design of relational contracts with asymmetric information.[9] Athey and Miller (2007) study repeated bilateral trading with a budget balancing constraint and two-sided private information. Alonso and Matouschek (2007) study a repeated coordination problem where monetary transfer is prohibited.

The following papers also study repeated principal-agent relationships with screening and are related to our results. Calzolari and Spagnolo (2010) study competitive bidding

[8]MacLeod (2007) and Malcomson (2010) survey the various topics of self-enforcing relational contracts. Dewatripont et al. (2000) briefly summarize the topics of multitasking incentive and job design problems.

[9]In his model the agent obtains private information after contracting while in our model he obtains private information before contracting.

in procurement and suggest the importance of limiting the competition since it maintains information rent to the agent which is necessary for implementing unverifiable quality.[10] We partly share this logic and argue for a similar role of information rent in a bilateral relationship. Wolitzky (2010) considers a dynamic monopolist problem with price discrimination and unverifiable delivery decisions. Setting a price corresponds to our first task in the sense that it is related to private information about consumer value and the delivery decision corresponds to our second task in the sense that it is unverifiable. In our terminology, delivery is implemented by the principal and then his model is similar to centralization in our task assignment problem. He investigates both non-durable and durable goods monopolists and demonstrates that the so-called Coase conjecture, due to the ratchet effect in durable goods markets, is not observed when an unverifiable delivery cost is introduced. While his analysis of durable goods includes the dynamics of market rationing, our model essentially rules out such a dynamic path by excluding persistent state variables and then there is no issue of learning over the periods.[11] Dynamic flow of information in repeated principal-agent relationships will be left for future research.[12]

Verifiable and Unverifiable Performance Measure As long as unverifiable performance is more informative on the task than verifiable performance, an interaction of formal and informal contracts could emerge. Baker et al. (1994), Pearce and Stacchetti (1998) and Itoh and Morita (2010) study this topic in an environment with biased measurement, risk-sharing, and relation-specific investments, respectively. However these abstract from multitasking issues. The interaction of formal and informal contracts in multitasking

[10]Board (2010) demonstrates the similar outcome in repeated competitive procurement. However in his model, the agents have no private information and then the principal does not need to attempt to screen the agents' types.

[11]We briefly discuss the case of serially correlated type in Section 2.7.

[12]Halac (2009) studies repeated principal-agent relationships where each party holds private information on her/his own outside value. She assumes that the outside value is completely persistent over periods, which induces dynamics of learning on the other's private information.

environment is studied by Schmidt and Schnitzer (1995), Daido (2006), Schöttner (2008), and Iossa and Spagnolo (2009).[13] Neither of them includes hidden information.

Several papers have already found the possibility of welfare improvement by making a task unverifiable (Baker et al., 1994; Schmidt and Schnitzer, 1995; Bernheim and Whinston, 1998; Kovrijnykh, 2010). Their idea is basically that unverifiability can make a stronger punishment for deviation and then provide a larger incentive *via* relational contracts, which improves efficiency. As argued in Section 2.1, our idea is different from theirs.[14]

Kvaløy and Olsen (2009) introduce the principal's endogenous verification investment – interpreted as a cost of writing effective formal contracts – which stochastically determines the verifiability. They show that the discount factor and verifiability could have a negative impact on the quality of the transaction.[15] This chapter derives this phenomenon caused by hidden information instead of an endogenous verification decision.

Multitasking Incentive Agency and Job Design Multitasking incentive problems and job design have been mainly discussed in environments without relational contracts. One of the main strands in the literature studies the environment with hidden action and identifies conditions on the verifiable measurement under which task bundling is preferable to task separation between multiple agents. (Holmström and Milgrom, 1991; Itoh, 1994; Meyer et al., 1996).[16] In the environment with hidden information, Jackson and Sonnenschein (2007) and Matsushima et al. (2010) investigate the linking mechanism in which many identical tasks are allocated to a single agent and the agent's messages

[13] Dewatripont et al. (1999) study a multitasking problem with implicit contracts. They focus on an agent with career concerns and the contract design problem is different from our model.

[14] We discuss the detail of the difference in Section 2.5.

[15] Their interest is in how a change of such factors affects the quality level implemented in the transaction and they do not mention the overall economic efficiency including the cost of the investment for verification. Actually, in their model, the overall economic efficiency is weakly increasing in the discount factor.

[16] A number of papers study the vertical job design problem, i.e. centralization or delegation. See Mookherjee (2006) for a survey of this literature.

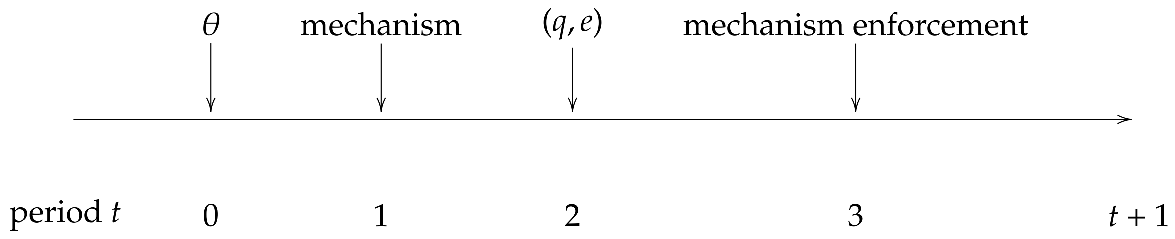

Figure 2.1: Timing within a Period

on his preference are rationed so as to approximate the *ex ante* belief on the preference. Both of their papers demonstrate that when the number of the tasks are sufficiently large, the linking mechanism can almost beat the incentive problem without any transfer. The research interest in this chapter is different from them in the sense that we consider the effect of a small number of the tasks with hidden information on the multitasking incentive problem.

Schöttner (2008) is one of the most related papers to this chapter. She studies the effect of job design on relational contracts and points out that task bundling is often preferred to task separation. In her paper, however, hidden information is excluded. The implication for the job design problem with hidden information is one of our contributions to this literature.

2.3 The Model

2.3.1 Environment

There are two parties, a principal (denoted by P) and an agent (denoted by A). Both live in periods $t = 0, 1, \ldots$ until infinity. Their common discount factor is $\delta \in [0, 1)$. Assume that they are risk-neutral. In each period t, they have a business opportunity and in period t the game proceeds as in Figure 2.1. At stage 0, A privately observes the cost parameter θ_t

drawn from $[\underline{\theta}, \overline{\theta}]$ according to the cumulative distribution function $F(\theta)$ with the density function $f(\theta)$. Assume that θ_t is independently drawn in each period. At stage 1, P offers A a mechanism, which is to be defined formally later. Assume that P can choose to abstain from offering mechanisms. If P does not offer a mechanism or A rejects the mechanism, then the period ends and both players obtain zero payoff. If A accepts the mechanism, then A works on two tasks at stage 2. In the first task, he chooses $q_t \in [0, \overline{q}] \equiv Q$ where $\overline{q} > 0$ and in the second one he faces a binary choice $e_t \in \{0, \overline{e}\}$ where $\overline{e} > 0$. These decisions yield P the benefit $y(q_t, e_t)$ and A the cost $c(q_t, \theta_t) + e_t$. Finally, at stage 3 the parties make a decision on the enforcement of the mechanism. Later we will explain this stage precisely.

In period t, given θ_t, q_t, e_t, and a monetary transfer w_t from P to A, P and A obtain the (*ex post*) payoff $y(q_t, e_t) - w_t$ and $w_t - c(q_t, \theta_t) - e_t$ respectively. Denote the aggregated surplus by $s(q, e, \theta) \equiv y(q, e) - c(q, \theta) - e$ and the first best decision by $(q^{FB}(\theta), e^{FB}(\theta)) \in \arg\max_{q,e} s(q, e, \theta)$. We make the following assumptions.

Assumption 2.1 *For all $q \in Q$, $e \in \{0, \overline{e}\}$, and $\theta \in [\underline{\theta}, \overline{\theta}]$, the following conditions are satisfied;*

1. *$c(q, \theta)$ is three-times differentiable and bounded in each component, $c_q(q, \theta) > 0$, $c_\theta(q, \theta) > 0$, $c_{q\theta}(q, \theta) > 0$ and $c(0, \underline{\theta}) > 0$.*

2. *$s(q, e, \theta)$ is twice differentiable in q, bounded in each component, $s(q, 0, \theta) < 0$, $s(0, e, \theta) < 0$, $s_{qq}(q, \overline{e}, \theta) < 0$ and there exists $q' \in (0, \overline{q})$ such that $s_q(q', \overline{e}, \theta) = 0$.*

3. *Let $J(q, \overline{e}, \theta) \equiv s(q, \overline{e}, \theta) - c_\theta(q, \theta)F(\theta)/f(\theta)$. Then $J_{qq}(q, \overline{e}, \theta) < 0$, $J_{q\theta}(q, \overline{e}, \theta) < 0$, and $J(q^{FB}(\theta), \overline{e}, \theta) > 0$.*

If the second task e is absent, implying that it is a single task model, then these assumptions would be fairly standard in the adverse selection literature.[17] Nevertheless some remarks

[17]The assumption that $c_\theta(0, \theta) > 0$ rules out a case where $c(q, \theta) \equiv \theta q + K$ for some $K > 0$ since it is not satisfied at $q = 0$. This assumption is only used for simplifying the proof of Lemma 2.1 and it can be shown even if $c_\theta(0, \theta) = 0$.

are useful here. Part 2 of Assumption 2.1 implies that both tasks are strict complements in the sense that to implement $e = \bar{e}$ and $q > 0$ is essential for the value of the relationship. If it is impossible to induce positive amounts of (q, e), then the business is less valued than the outside option.

$J(q, e, \theta)$ is known as the virtual-surplus, which becomes the consequent objective function for P in a typical adverse selection model. Part 3 guarantees that for each θ, $\arg\max_{q,e} J(q, e, \theta)$ is unique. Hereafter denote the maximand of $J(q, e, \theta)$ by $(e^{SB}(\theta), q^{SB}(\theta))$. It can be seen that $e^{SB}(\theta) = \bar{e}$ and $J_q(q^{SB}(\theta), \bar{e}, \theta) = 0$[18] for each $\theta \in [\underline{\theta}, \overline{\theta}]$ and $q^{SB}(\theta)$ is decreasing in θ. Part 2 and Part 3 further guarantee that $e^{FB}(\theta) = \bar{e}$[19] and $s_q(q^{FB}(\theta), \bar{e}, \theta) = 0$ for each $\theta \in [\underline{\theta}, \overline{\theta}]$, $q^{FB}(\theta)$ is decreasing in θ,[20] and $q^{FB}(\theta) \geq q^{SB}(\theta)$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$ with equality holding only if $\theta = \underline{\theta}$.

The assumption on the observability and verifiability is as follows. All variables except for θ are observable by both parties. P cannot observe θ and she believes that it is distributed according to $F(\theta)$ and $f(\theta)$. Both q_t and e_t are describable at stage 1 in period t whereas unless otherwise stated, only the first task q_t is verifiable.

A mechanism is defined as a pair of a formal contract and an informal agreement. A formal contract is a transfer schedule $p_t(q_t)$ from P to A contingent on q_t and enforced by the court.[21] We assume that $p_t(q_t)$ is continuous in q_t.[22] An informal agreement $b_t(q_t, e_t)$ is a transfer schedule from P to A which can be contingent on both q_t and e_t and must be enforced by themselves.[23] Both $p_t(q_t)$ and $b_t(q_t, e_t)$ can be positive or negative. Enforcement

[18] Note that $\max_q J(q, 0, \theta) < 0$ for any q and θ by Part 2 of Assumption 2.1.

[19] It is because $\max_q s(q, \bar{e}, \theta) > \max_q J(q, \bar{e}, \theta) \geq J(q^{FB}(\theta), \bar{e}, \theta) > 0 > \max_q s(q, 0, \theta)$ for any q and θ.

[20] Note that $s_{q\theta}(q, e, \theta) < 0$ for all q and θ.

[21] Here we implicitly assume that monetary transfer is also verifiable.

[22] The assumption of continuity of $p_t(q_t)$ is required only for the proof of Lemma 2.6, which shows the existence of an equilibrium strategy used for the punishment. It means that this technical assumption is only for the construction of the off-path outcome and innocuous in the other parts of the analysis. Thus except for Lemma 2.6, we will not explicitly mention this assumption. See the proof of Lemma 2.6 and footnote 44 there for the detail.

[23] The assumption on the describability of q_t and e_t implies that neither p_t nor b_t can be contingent on q_τ

of informal agreements is defined as follows. At stage 3, each party chooses $I_t^i \in \Im \equiv \{H, R\}$ for $i = P, A$ where H denotes honouring the informal agreement, and R reneging. The informal agreement $b_t(q_t, e_t)$ is enforced if and only if both parties chose to honour it, i.e., $I_t \equiv (I_t^P, I_t^A) = (H, H)$. If it is not enforced, i.e., $I_t \neq (H, H)$, then no transfer is made by the informal agreements.

We later discuss the effect of verifiability of the second task e. We define that the second task e is verifiable if an informal agreement $b_t(q, e)$ is also enforced by the court for each t. It implies that when the second task is verifiable, $b_t(q, e)$ is successfully transferred regardless of I_t.

2.3.2 Economic Interpretation of the Setting

Throughout this chapter, we treat this environment as a manufacturer-supplier relationship. The manufacturer offers a menu contract $p(q)$ contingent on units of the products q from the supplier. However this product is valueless without the supplier's effort which can be interpreted as improvement of quality, careful delivery, or regular maintenance. For simplicity, in what follows, we use the term "quantity" for the first task and "effort" for the second task.

This framework can be applied to the analysis of other situations. On management of a production line of automobiles or or electric machineries, the product consists of hard parts q assembled in the expertized line and software e for controlling the system the performance of which is hard to be evaluated from the outside. In an relationship between a firm owner and a CEO, q might be interpreted as the size of a project and e as the effort to find and tailor the project for successful implementation. In the context of

and e_τ for $\tau > t$.

regulation, the regulation authority asks the regulated firm to meet several environmental standards and some of the standards might be hard to measure objectively. The analysis below would be applicable to these cases.

2.3.3 Strategy and Equilibrium

A pure strategy in the repeated game is defined as a mapping from a privately observable history to a decision variable. Formally, let $D \equiv Q \times \{0, \bar{e}\}$ be the set of the task levels and $\mathfrak{W} \equiv (\{(p(\cdot), b(\cdot,\cdot)) \mid p : Q \to \mathbb{R}, b : D \to \mathbb{R}\})$ be the set of feasible mechanisms. Denote $\overline{\mathfrak{W}} \equiv \mathfrak{W} \cup \{\phi\}$ where ϕ denotes no mechanism being offered. Denote also the decision of A's rejection of the mechanism by ω. Then the public history, a sequence of publicly observable variables[24], up to period $t \geq 1$ is defined as $h^t \equiv (h_\tau)_{\tau=0}^{t-1}$ where for each $\tau = 0, \ldots, t-1$, $h_\tau \in \{\phi\} \cup (\mathfrak{W} \times (\{\omega\} \cup (D \times \mathfrak{I}^2)))$.[25] Let h^0 be the null history and $\mathcal{H}$ be the set of the public histories. A's private history, the sequence of A's observable variables, up to period t is defined as $h^{A0} = h^0$ and for $t \geq 1$, $h^{At} \equiv (\theta_\tau, h_\tau)_{\tau=0}^{t-1}$. Let $\mathcal{H}^A$ be the set of A's private histories.

P's pure strategy is a pair of mappings $\sigma^P \equiv (\Gamma, \iota^P)$ where $\Gamma : \mathcal{H} \to \overline{\mathfrak{W}}$ and $\iota^P : \mathcal{H} \times \mathfrak{W} \times D \to \mathfrak{I}$. These define the mechanism offered by P and whether P enforces the informal agreement given that the mechanism is accepted, respectively. A's pure strategy is a pair of mappings $\sigma^A \equiv (\chi, \iota^A)$ where $\chi : \mathcal{H}^A \times [\underline{\theta}, \overline{\theta}] \times \mathfrak{W} \to \overline{D}$, $\iota^A : \mathcal{H}^A \times [\underline{\theta}, \overline{\theta}] \times \mathfrak{W} \times D \to \mathfrak{I}$, and $\overline{D} \equiv D \cup \{\omega\}$. χ stipulates A's response to an offered mechanism and ι^A whether to enforce the informal agreement. Denote the set of strategies by Σ^i for $i = P, A$. A pure strategy profile (σ^P, σ^A) is a perfect Bayesian equilibrium if after every history, there is no

[24]We mean by "publicly observable" and "public history" that it is observable for both P and A but not necessarily observable for others. Then note that it does not necessarily coincide with "verifiable".

[25]It is implicitly presumed that if h_τ contains an element in D, then it means that A accepts the mechanism in period τ.

incentive to deviate to improve her/his own average payoff at any stage of the period and P's belief about θ is computed by Bayes' rule whenever it is possible and consistent with the strategy profile.[26]

Since the game has a recursive structure, we can decompose strategy (σ^P, σ^A) in the following way. For any public history up to stage 0 of period t, $h^t \in \mathcal{H}$, let

$$\sigma_+^P(h^t) \begin{cases} \Gamma(h^t) \in \overline{\mathfrak{W}} \\ \iota^P(\cdot \mid h^t) : \mathfrak{W} \times D \to \mathfrak{I} \\ \sigma_+^P(h^t, \cdot) : \{\phi\} \cup (\mathfrak{W} \times (\overline{D} \cup \{\omega\})) \to \Sigma^P \end{cases}$$

where $\sigma_+^P(h^0) \equiv \sigma^P$, and for any A's private history up to stage 0 of period t, $h^{At} \in \mathcal{H}^A$,

$$\sigma_+^A(h^{At}) \begin{cases} \chi(\cdot \mid h^{At}) : [\underline{\theta}, \overline{\theta}] \times \mathfrak{W} \to \overline{D} \\ \iota^A(\cdot \mid h^{At}) : [\underline{\theta}, \overline{\theta}] \times \mathfrak{W} \times D \to \mathfrak{I} \\ \sigma_+^A(h^{At}, \cdot) : [\underline{\theta}, \overline{\theta}] \times (\{\phi\} \cup (\mathfrak{W} \times (\overline{D} \cup \{\omega\}))) \to \Sigma^A \end{cases}$$

where $\sigma_+^A(h^0) \equiv \sigma^A$. This formulation is interpreted as that the strategy stipulates players' actions in the current period and a strategy itself played in the repeated game starting from next period, that is the continuation strategy.

The repeated game typically has multiple equilibria. Thus we make some additional restrictions on the equilibrium strategy. First, we focus on the public strategy in the equilibrium.

Assumption 2.2 *σ^A satisfies that for all $t \geq 0$ and $h^{At} \in \mathcal{H}^A$, $\sigma_+^A(h^{At})$ is independent from $(\theta_0 \ldots, \theta_{t-1})$ and $\iota^A(\theta_t, W_t, d_t \mid h^{At})$ is independent from $(\theta_0 \ldots, \theta_{t-1}, \theta_t)$.*

It simply states that A's equilibrium strategy is independent from θs which are irrelevant

[26]Since the equilibrium mechanism seen later does not have bunching for the first task, the restriction to pure strategy profiles is without loss of generality for the optimum. See Strausz (2006) for details.

to his current payoff. An equilibrium of a public strategy is known as the public perfect equilibrium (PPE) and every pure strategy equilibrium payoff can be achieved by PPE.[27] Instead of $\sigma_+^A(h^{At})$, hereafter denote A's public strategy by $\sigma_+^A(h^t)$ where h^t is a public history.

The next assumptions stipulate the bargaining power and the commitment ability.

Assumption 2.3 *A public strategy σ satisfies $\sigma_+(h^t, W, \omega) = \sigma_+(h^t, \phi) = \sigma_+(h^t)$ for all $h^t \in \mathcal{H}$.*

Assumption 2.4 *Let $\hat{\mathcal{H}}(\sigma) \subset \mathcal{H}$ be the set of public histories achieved with positive probability by public strategy σ. For any $h^t \in \hat{\mathcal{H}}(\sigma)$, $\sigma_+(h^t) \equiv (\sigma_+^P(h^t), \sigma_+^A(h^t))$ must attain the maximum average payoff for P in the PPE.*

Assumption 2.3 means that the event of no contracting at the beginning of the period is not interpreted as serious misbehaviour and then the same strategy is taken in the continuation game. These could prevent the punishment for deviating party which can reinforce efficiency. Assumption 2.4 means that as long as the players are on the equilibrium path, P has absolute bargaining power at the stage of contracting in the sense that the equilibrium strategy will optimize P's utility subject to the equilibrium constraints. This is for consistency with the standard principal-agent models which presume the principal to make a take-or-leave-it offer. It also implies that P's commitment ability to the future play is not so strong that P may not voluntarily choose the worse outcome, which could lead to more efficient outcome *ex ante*.

Assumption 2.3 and 2.4 cause some loss of generality. In particular, Assumption 2.4 is crucial if the type is independently drawn each period as in our model. In Section 2.7, we discuss how much loss of generality is caused by these assumptions.

Hereafter call the equilibrium satisfying Assumption 2.2, 2.3, and 2.4 the Optimal Public Perfect Equilibrium (OPPE). The following proposition shows that it is without

[27]See Fudenberg et al. (1994) for details.

loss of generality to focus on the following simple trigger strategy for finding the OPPE. To state the proposition, let $\mathcal{H}^R \equiv \{h^t \in \mathcal{H} \mid {}^\exists\tau < t, {}^\exists i, {}^\exists(d_\tau, I_\tau) \in D \times \mathfrak{I}^2, I_\tau^i = R\}$, that is the set of public histories in which there exists a party having reneged on an informal agreement and $\mathcal{H}^H \equiv \mathcal{H} \setminus \mathcal{H}^R$, the set of public histories in which informal agreements have been honoured on.

Proposition 2.1 *Let σ be an OPPE and (π, u) be the OPPE payoff. Then there exists an OPPE σ^* such that*

- *for all $h^t \in \mathcal{H}^H$,*
  - $\Gamma^*(h^t) = W^* \in \overline{\mathfrak{W}}$,
  - *if $W^* \in \mathfrak{W}$, then $\chi^*(W^* \mid h^t, \theta_t) = d^*(\theta_t) \in \overline{D} \cup \{\omega\}$, and $\iota^{P*}(W^*, d \mid h^t) = \iota^{A*}(W^*, d \mid h^t, \theta_t) = H$,*
- *for $h^t \in \mathcal{H}^R$, $\Gamma^*(h^t) = \phi$, and*
- *the associated payoff (π^*, u^*) satisfies that $\pi^* = \pi$ and $u^* \geq u$.*

Proposition 2.1 drastically simplifies our analysis. Specifically, we obtain several notable features of σ^* that (i) on the equilibrium path, if $W^* \in \mathfrak{W}$ and $d^*(\theta_t) \in \overline{D}$, i.e., P actually offers a mechanism and A accepts it, then the informal agreement must always be honoured, (ii) P's mechanism and A's decision are independent of public histories as long as they have reneged on informal agreements, and (iii) after deviation to reneging on an informal agreement, both players obtain 0 payoff from the outside options. Furthermore if $W^* = \phi$, then it is straightforward that both parties obtain 0 as their average payoff. Thus our problem is to seek the mechanism $W^* \in \mathfrak{W}$ which maximizes P's average payoff. If there exists such a mechanism with P's average payoff $\pi^* \geq 0$, then W^* is offered on the equilibrium path and if not, then no mechanism is offered on the equilibrium path.

Hereafter we focus on the OPPE which satisfies Proposition 2.1.

2.4 Analysis

Thanks to Proposition 2.1, the OPPE can be derived by a stationary problem with some constraints. Furthermore, the following proposition shows that the OPPE is more simplified in that all types accept the equilibrium mechanism.

Lemma 2.1 *Let $W^* \in \mathfrak{W}$ be the OPPE mechanism and $d^*(\theta)$ be A's OPPE decision given that mechanism W^* is offered. Then $d^*(\theta) \in D$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$.*

Let $W^* = (p^*(q), b^*(e, q))_{q \in Q, e \in \{0, \bar{e}\}}$ and $d^*(\theta) = (q^*(\theta), e^*(\theta))$ be the equilibrium task level chosen by A given the mechanism W^*. Due to the stationarity of W^* and $d^*(\theta)$, P's average payoff is given by

$$\begin{aligned} \pi^* &= (1-\delta)\sum_{\tau=0}^{\infty} \delta^{\tau} E_{\theta}\left[y(q^*(\theta), e^*(\theta)) - w^*(q^*(\theta), e^*(\theta))\right] \\ &= E_{\theta}\left[y(q^*(\theta), e^*(\theta)) - w^*(q^*(\theta), e^*(\theta))\right] \end{aligned}$$

where $w^*(q, e) = p^*(q) + b^*(q, e)$. This is the objective to be maximized subject to the equilibrium conditions. In each period, the players are faced with four equilibrium conditions which are common among periods on the equilibrium path. Figure 2.2 describes them in order. At stage 1, the participation constraint (PC) is satisfied if and only if A with type $\theta \in [\underline{\theta}, \overline{\theta}]$ accepts W^* offered by P. At stage 2, A chooses $d^*(\theta) = (q^*(\theta), e^*(\theta))$. There are two concerns on this incentive compatibility constraint. First, the truth-telling constraint (TT) is satisfied if and only if type θ does not pretend to be any other type θ'. Second, the effort incentive constraint (EI) is satisfied if and only if type θ does not choose $e \neq e^*(\theta)$.[28]

[28]Strictly, we must also check the incentive constraint not to choose $q \notin Q^* \equiv \{q' \in Q \mid {}^{\exists}\theta \in [\underline{\theta}, \overline{\theta}], q' = q^*(\theta)\}$. However, since we can show that Q^* is an closed interval and $p^*(q)$ is non-decreasing in $q \in Q^*$, the

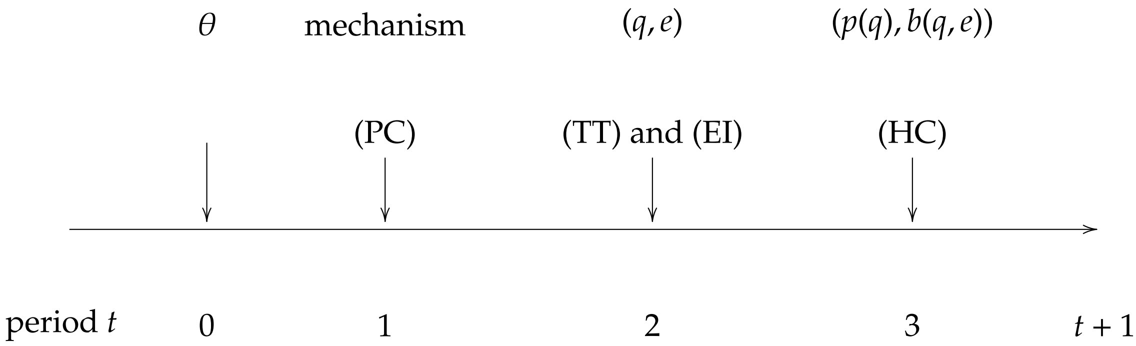

Figure 2.2: Equilibrium Constraints

Finally, at stage 3, the honouring constraint (HC) is satisfied if and only if both parties choose to honour the informal agreement.

Let

$$
\begin{aligned}
U^*(\theta) &\equiv w^*(q^*(\theta), e^*(\theta)) - c(q^*(\theta), \theta) - e^*(\theta), \\
u^* &\equiv E_\theta\left[U^*(\theta)\right], \\
s^* &\equiv u^* + \pi^*
\end{aligned}
$$

be A's *ex post* and *ex ante* payoff and the *ex ante* total surplus within a period. In general, the players are concerned about their future continuation payoff as well as the payoff in the current period. Nevertheless by construction of the strategy, the continuation payoff is not altered by any deviation on (TT), (EI), and (PC). It means that it is enough to consider these constraints without the future continuation payoff. On (HC), choosing $\iota^{i*} = H$ induces the continuation payoff, π^* and u^* for P and A respectively. Conversely, choosing $\iota^{i*} = R$ induces no mechanism in the future and then the continuation payoff is 0 for both parties. Thus, while the continuation payoff must be taken into account in (HC), the value of the

possibility to choose $q \notin Q^*$ can be excluded by constructing $p^*(q)$ such that for $q > \sup Q^*$, $p^*(q) = \sup_{q \in Q^*} p^*(q)$ and for $q < \inf Q^*$, $p^*(q)$ is sufficiently small.

continuation payoff is still history-independent. Therefore the optimization problem of characterizing the OPPE can be transformed to a static mechanism design problem.

First, look at (TT). For each $\theta, \theta \in [\underline{\theta}, \overline{\theta}]$, θ does not pretend $\theta' \in [\underline{\theta}, \overline{\theta}]$ if and only if

$$\begin{aligned} U^*(\theta) &\geq w^*(q^*(\theta'), e^*(\theta')) - c(q^*(\theta'), \theta) - e^*(\theta') \\ &= U^*(\theta') + c(q^*(\theta'), \theta') - c(q^*(\theta'), \theta). \end{aligned}$$

The envelope theorem by Milgrom and Segal (2002) implies that this is equivalent to that $q^*(\theta)$ is non-increasing in θ and

$$U^*(\theta) = U^*(\overline{\theta}) + \int_{\theta}^{\overline{\theta}} c_\theta(q^*(z), z)dz. \tag{2.1}$$

Second, consider (EI). (EI) is satisfied if for each θ, θ' and $e' \neq e^*(\theta')$, θ does not choose $(q(\theta'), e')$. It is expressed as

$$\begin{aligned} U^*(\theta) &\geq w^*(q^*(\theta'), e') - c(q^*(\theta'), \theta) - e' \\ &= U^*(\theta') - b^*(q^*(\theta'), e^*(\theta')) + b^*(q^*(\theta'), e') + c(q^*(\theta'), \theta') - c(q^*(\theta'), \theta) + e^*(\theta') - e'. \end{aligned}$$

Substituting (2.1) implies that

$$\int_{\theta}^{\theta'} c_\theta(q^*(x), x)dx - c(q^*(\theta'), \theta') + c(q^*(\theta'), \theta) \geq -b^*(q^*(\theta'), e^*(\theta')) + b^*(q^*(\theta'), e') + e^*(\theta') - e' \tag{2.2}$$

Third, (PC) means that every type accepts the mechanism. Since to reject generate zero payoff for A, the condition is described as

$$U^*(\theta) \geq 0$$

for all $\theta \in [\underline{\theta}, \overline{\theta}]$. Since $U^*(\theta)$ is non-increasing due to (2.1), it is satisfied if and only if

$$U^*(\overline{\theta}) \quad \geq \quad 0. \tag{2.3}$$

Finally, consider (HC). Given that q and e were observed and A chooses $I^A = H$, if P chooses $I^P = H$, then $b^*(q, e)$ is transferred from P to A and the continuation payoff from the next period is π^*. Conversely if P chooses $I^P = R$, then $b^*(q, e)$ is not to be transferred and the continuation payoff from the next period is 0 because she does not offer any mechanism from the next period. Thus P chooses $I^P = H$ if and only if

$$-b^*(q, e) + \frac{\delta}{1-\delta}\pi^* \geq 0.$$

Similarly, A chooses $I^A = H$ if and only if

$$b^*(q, e) + \frac{\delta}{1-\delta}u^* \geq 0.$$

Combining them implies that

$$\frac{\delta}{1-\delta}\pi^* \geq b^*(q, e) \geq -\frac{\delta}{1-\delta}u^*. \tag{2.4}$$

Plugging (2.1) into π^* and integrating by parts yield that

$$\begin{aligned} \pi^* &= E_\theta\left[y(q^*(\theta), e^*(\theta)) - c(q^*(\theta), \theta) - e^*(\theta) - U^*(\theta)\right] \\ &= \int_{\underline{\theta}}^{\overline{\theta}} J(q^*(\theta), e^*(\theta), \theta) f(\theta) d\theta - U^*(\overline{\theta}) \end{aligned} \tag{2.5}$$

where $J(q, e, \theta)$ is defined as in Assumption 2.1. In equilibrium, it is maximized by control variables $(q^*(\theta), e^*(\theta), U^*(\theta), b^*(q^*(\theta), e))$ subject to (2.2), (2.3), (2.4), and monotonicity of $q^*(\theta)$ in θ. This problem is simplified as follows.

Lemma 2.2 *The optimization problem described above is equivalent to the following;*

$$\max_{q^*(\theta)} \int_{\underline{\theta}}^{\overline{\theta}} J(q^*(\theta), \bar{e}, \theta) f(\theta) d\theta$$

$$\textit{subject to } \frac{\delta}{1-\delta} \int_{\underline{\theta}}^{\overline{\theta}} s(q^*(z), \bar{e}, z) f(z) dz \geq \bar{e} \tag{2.6}$$

and $q^*(\theta)$ is non-increasing in θ.

If constraint (2.6) is absent, then the Euler equation implies that $q^*(\theta) = q^{SB}(\theta)$ for all θ. Since $q^{SB}(\theta)$ is non-increasing in θ, it is the solution if $q^{SB}(\theta)$ satisfies (2.6), that is

$$\frac{\delta}{1-\delta} \int_{\underline{\theta}}^{\overline{\theta}} s(q^{SB}(z), \bar{e}, z) f(z) dz \geq \bar{e} \iff \delta \geq \bar{\delta} \equiv \frac{\bar{e}}{\bar{e} + \int_{\underline{\theta}}^{\overline{\theta}} s(q^{SB}(z), \bar{e}, z) f(z) dz}. \tag{2.7}$$

Suppose that (2.7) is not satisfied. Now consider the following Lagrangian with multiplier $\lambda^* \geq 0$:

$$\mathcal{L} \equiv \int_{\underline{\theta}}^{\overline{\theta}} J(q^*(\theta), \bar{e}, \theta) f(\theta) d\theta + \lambda^* \left[\frac{\delta}{1-\delta} \int_{\underline{\theta}}^{\overline{\theta}} s(q^*(\theta), \bar{e}, \theta) f(\theta) d\theta - \bar{e} \right].$$

The Euler equation implies that the solution satisfies that

$$s_q(q^*(\theta), \bar{e}, \theta) = \frac{1}{1 + \frac{\delta}{1-\delta}\lambda} \frac{F(\theta)}{f(\theta)} c_{q\theta}(q^*(\theta), \theta). \tag{2.8}$$

Recall that by definition of $q^{SB}(\theta)$, $q^*(\theta) = q^{SB}(\theta)$ if $\lambda = 0$. However, since it does not satisfy (2.6), then λ^* is positive and determined to satisfy (2.6) with equality.

Under what conditions is it feasible? It is easy to see that $q^*(\theta)$ characterized by (2.8) is decreasing in θ by Assumption 2.1. The right hand side of (2.8) is greater then 0 and $q^{SB}(\theta)$ satisfies that $J_q(q^*(\theta), \bar{e}, \theta) = 0$ or equivalently

$$s_q(q^{SB}(\theta), \bar{e}, \theta) = \frac{F(\theta)}{f(\theta)} c_{q\theta}(q^{SB}(\theta), \theta),$$

the right hand side of which is greater that of (2.8). Thus we see that for $\delta < \overline{\delta}$, $s_q(q^{FB}(\theta), \bar{e}, \theta) < s_q(q^*(\theta), \bar{e}, \theta) < s_q(q^{SB}(\theta), \bar{e}, \theta)$, which implies that by concavity of $s(q, e, \theta)$

$$q^{FB}(\theta) > q^*(\theta) > q^{SB}(\theta)$$

for all $\theta \in (\underline{\theta}, \overline{\theta})$. Furthermore, (2.8) shows that $q^*(\theta)$ is increasing in λ^* (provided that $\delta > 0$) and approaches $q^{FB}(\theta)$ as λ^* goes to infinity. Then by concavity of $s(q, \bar{e}, \theta)$ in q, s^* is increasing as λ^* goes up and the supremum of it is $\int s(q^{FB}(z), \bar{e}, z) f(z) dz$. Thus the solution obtained by (2.8) is feasible if and only if

$$\frac{\delta}{1-\delta} \int_{\underline{\theta}}^{\overline{\theta}} s(q^{FB}(z), \bar{e}, z) f(z) dz > \bar{e} \iff \delta > \underline{\delta} \equiv \frac{\bar{e}}{\bar{e} + \int_{\underline{\theta}}^{\overline{\theta}} s(q^{FB}(z), \bar{e}, z) f(z) d\theta} \tag{2.9}$$

where strictly inequality must be satisfied because of Slater constraint qualification for constrained optimization problems. When $\delta = \underline{\delta}$, it is obvious that $q^{FB}(\theta)$ for each θ is the unique feasible q and then it is the solution.

Finally, if $\delta < \underline{\delta}$, then there is no $q^*(\theta)$ which satisfies (2.6). This implies that $e^*(\theta) = 0$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$ and then $\pi^* < 0$. In this case, P strictly prefers to offer no mechanism. The summary of the above analysis is as follows.

Proposition 2.2 *The OPPE satisfies the following.*

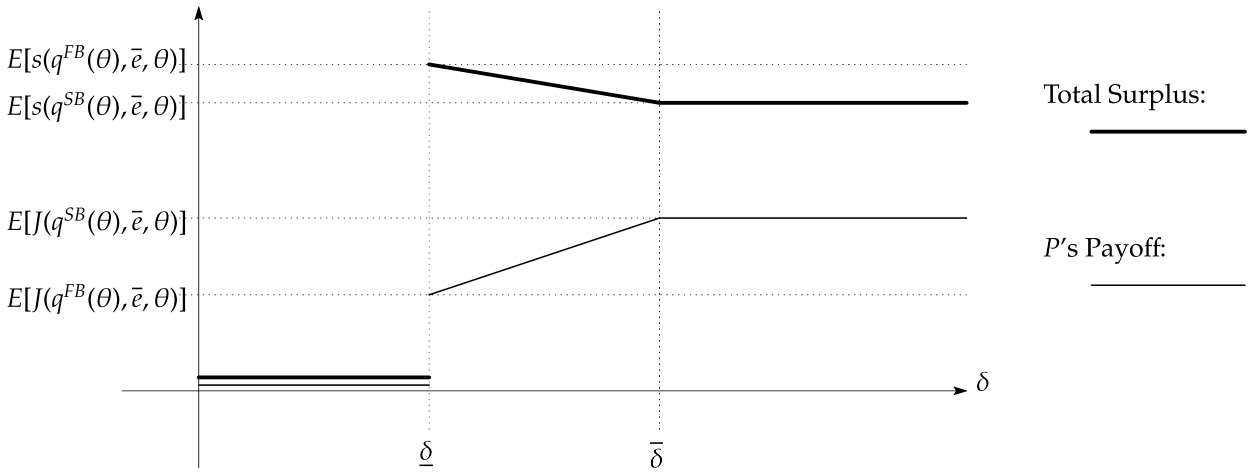

Figure 2.3: Total Surplus and P's payoff in the OPPE

1. *For $\delta \geq \overline{\delta}$, $d^*(\theta) = (q^{SB}(\theta), \bar{e})$.*

2. *For $\delta \in (\underline{\delta}, \overline{\delta})$, $d^*(\theta) = (q^*(\theta), \bar{e})$ where $q^*(\theta)$ satisfies (2.6) and (2.8) with equality for some $\lambda^* > 0$.*

3. *For $\delta = \underline{\delta}$, $d^*(\theta) = (q^{FB}(\theta), \bar{e})$.*

4. *For $\delta < \underline{\delta}$, $W^* = \phi$.*

Proposition 2.2 implies the following corollary.

Corollary 2.1 *On the OPPE, the total surplus for $\delta \in (\underline{\delta}, \overline{\delta})$ is decreasing in δ and strictly greater than $\delta \geq \overline{\delta}$. P's payoff for $\delta \in (\underline{\delta}, \overline{\delta})$ is increasing in δ and strictly less than $\delta \geq \overline{\delta}$.*

Figure 2.3 illustrates the result in Proposition 2.2 and Corollary 2.1.

The intuition of Corollary 2.1 is as follows. Since effort is unverifiable, self-enforcement contracts are required to induce positive effort from A. The self-enforcement contracts are credible if the value of the relationship in the future is large enough. This is captured by constraint (2.6). For high discount factors such that $\delta \geq \overline{\delta}$, the players value the future surplus so much that P is allowed to optimize her utility with respect to the first task. As the discount factor is lower such that $\delta \in [\underline{\delta}, \overline{\delta})$, the optimal level of the first task does not ensure the future surplus enough to sustain the self-enforcing contracts. Nevertheless, as (2.6) shows, it is possible to induce $\bar{e}$ by raising the total surplus for an intermediate value of the discount factor. Thus the total surplus is rather improved by mitigating inefficiency with respect to the first task caused by asymmetric information. In other words, P's objective is altered from optimizing her own payoff to caring about the total surplus to sustain the relational contracts.

2.5 Social Desirability of Unverifiability

Contrary to the previous assumption, assume now that e is also verifiable within a period. We show that verifiability could lead to a lower social surplus.

Even when e is verifiable, the basic procedure for deriving the equilibrium is similar to the previous section.[29] The difference is that informal contracts $b^*(q^*(\theta), e)$ do not have to satisfy (HC) which is corresponding to (2.4). Thus, when e is verifiable, the optimization problem to derive the OPPE is to maximize (2.5) by control variables $(q^*(\theta), e^*(\theta), U^*(\theta), b^*(q^*(\theta), e))$ subject to (2.2), (2.3), and monotonicity of $q^*(\theta)$ in θ.

Proposition 2.3 *When e is verifiable, the OPPE satisfies that $W^* \in \overline{\mathfrak{W}}$, $q^*(\theta) = q^{SB}(\theta)$, and $e^*(\theta) = \bar{e}$ for any $\delta \in [0, 1)$.*

Comparing it with the result in Proposition 2.2 immediately implies the following corollary.

Corollary 2.2 *For $\delta \geq \underline{\delta}$, unverifiability of e is weakly socially desirable. In particular, for $\delta \in [\underline{\delta}, \overline{\delta})$, unverifiability of e is strictly socially desirable.*

[29] The optimality of stationary contracts with all the types participating can be proven by a similar proof of Proposition 2.1 and Lemma 2.1.

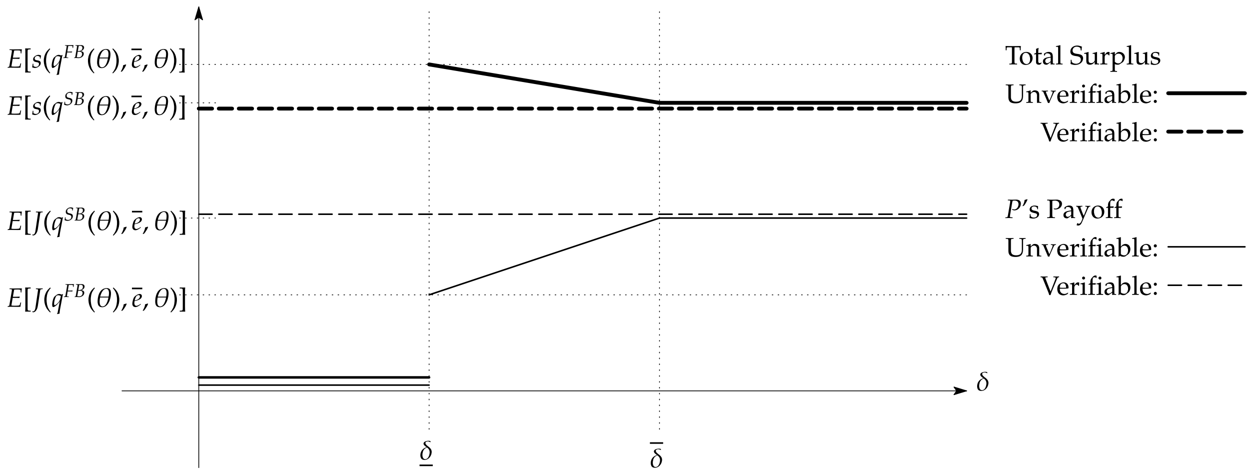

Figure 2.4: Comparison between the Verifiable and Unverifiable Cases

Figure 2.4 illustrates the comparison between the unverifiable case and the verifiable one. Actually the problem of the verifiable case is the same as that of the unverifiable case without (2.6). Thus the equilibrium achieved in the verifiable case is the same as in the unverifiable case with $\delta \geq \overline{\delta}$. However it is less socially efficient than the unverifiable case with $\delta \in [\underline{\delta}, \overline{\delta})$.

The intuition is similar to the comparison between the cases of high and intermediate discount factor in the previous section. Even for intermediate δ, if e is verifiable, then P has no problem about implementing $\bar{e}$. Unverifiability of e urges P to consider the implementation of $\bar{e}$ more seriously, which induces more social efficiency as discussed before.

Note that unverifiability is not desirable from P's view. Actually, regardless of verifiability of e, P's objective is the expected value of $J(q^*(\theta), \bar{e}, \theta)$ and the unverifiability could change the quantity level from $q^{SB}(\theta)$ to a more socially efficient level. Since $q^{SB}(\theta)$ is the optimum of $J(q^*(\theta), \bar{e}, \theta)$, the unverifiability is obviously less preferred by P. Thus P does not strategically leave the contract incomplete even if she can.[30]

[30]Nevertheless we can demonstrate the strategic contractual incompleteness. by extending the model. One way of extension is just introducing a positive cost of writing a contract on e. Another extension is

We can also obtain an economic implication for the role of courts and contract law. The standard contract theory typically tells us that if the court could void a contract, a party anticipating that action would lose the incentive to implement better economic performance due to the lack of the commitment to the *ex post* compensation. This view makes it difficult to explain why courts sometimes intervene to void contractual terms in reality. One of the explanations from our model is that when the long-term relationship is working and the party with bargaining power are faced with the party with private information, the court should sometimes void the terms of the contract. Intuitively, contract design by the party without information causes inefficiency due to the concern of the information rent. If verifiability is not assured by the court, then the contracting party cannot control the information rent and then the inefficiency caused by it can be mitigated.

Several papers in the literature point out the desirability of unverifiability. Nevertheless it should be emphasized that the logic here is quite different from that of the literature. To understand the difference, consider a strategy which implements the same decision every period in equilibrium. Then it can be a public perfect equilibrium if and only if the following inequality is satisfied:

$$\frac{\delta}{1-\delta}\left[s^{*}-\underline{s}\right] \geq g^{*} \tag{2.10}$$

where s^* denotes the equilibrium total surplus, $\underline{s}$ does the total surplus without relational contracts and g^* the agent's deviation benefit from unverifiable tasks. It expresses that the self-enforcing contract can work if and only if the future total surplus from relational con-

assuming that at the beginning of the repeated game there is an option to invest in a technology which makes the effort verifiable (e.g., introducing the objective measurement related to the effort). Then P does not necessarily choose to invest in the technology by receiving a transfer from A as a consideration for leaving the unverifiability.

tracts exceeds the one-shot benefit from deviation in the unverifiable tasks. By substituting $\underline{s} = 0$ and $g^* = \bar{e}$, it is the same as (2.6) in our model.

The typical argument is that unverifiability can make for a stronger punishment on deviation, which improves efficiency. For instance, Schmidt and Schnitzer (1995) study a relational contracting model where, as in our model, the agent's tasks are two dimensional. They show that the above inequality is necessary and sufficient for the equilibrium condition and compare the cases where both tasks are unverifiable and one of the tasks is verifiable. Then in the latter, since the verifiable task can be implemented by formal contracts, the deviation incentive g^* is smaller. At the same time, however, the total surplus without relational contracts $\underline{s}$ could be larger since formal contracts can always give incentives for the verifiable task. Overall, whether the constraint is relaxed or not is ambiguous. In particular, if the effect on $\underline{s}$ dominates that of g, then verifiability is not necessarily desirable. Here the payoff $\underline{s}$ is the total surplus after informal contracts are reneged on. Thus $\underline{s}$ expresses the degree of effectiveness of the punishment. Making one task verifiable can weaken the punishment necessary for sustaining relational contracts, which can worsen efficiency.[31]

By contrast, our argument starts with the case where both tasks are verifiable. Then the formal contract can completely eliminate the deviation incentive meaning that $g^* = 0$. Furthermore, since there is no unverifiable task, the principal will implement the decision as she likes. Now suppose that one of the tasks becomes unverifiable. Since the same strategy as in the verifiable case is most preferred to the principal, it would be implemented as long as it satisfies (2.10). If not, then the principal would attempt to improve the total surplus to satisfy (2.10). In other words, by increasing g, the principal's objective is

[31] A similar argument is developed by Baker et al. (1994), Bernheim and Whinston (1998), and Kovrijnykh (2010).

changed from maximizing her own payoff distorted from the total surplus to improving the total surplus to sustain the relational contracts. In this sense, a more important effect of unverifiability on the efficiency in our argument is increasing g rather than decreasing $\underline{s}$.

2.6 Task Assignment Problem

In the main analysis, we assume that a single agent dedicates to the two tasks. The literature on the multitasking agency problem demonstrates that the distribution of the tasks to multiple agents could mitigate or reinforce the inefficiency caused by the multitasking incentive problem. The effect of the task allocation suggests whether the tasks should be allocated to one agent as a bundle or two or more agents separately and gives us implications on organizational design.

In this section, we develop a discussion of the task assignment problem in the following manner. The first task is highly expertized so that only the original agent can work on it and he still holds private information on that technology. By contrast, the second task can be delegated to another party. However the performance of it is still unverifiable so that a self-enforcing mechanism is required.

Formally, e is unverifiable but the decision right of e is contractible for infinite periods. Thus we modify the repeated game by introducing period -1 in which P chooses a party to whom the second task is allocated. Then from period 0, they play the repeated game in which in each period the party with the decision right of e makes a decision and incurs the cost of e.

We consider the following three modes.

- Task Bundling (B): the decision right is allocated to the agent deciding q

- Centralization (C): P keeps the decision right of e
- Task Separation (S): the decision right is allocated to an agent different from one who decides q.

Recall that (B) is exactly the same as in the analysis of section 2.4. Here we investigate (C) and (S).[32]

2.6.1 Centralization

First, suppose that P keeps the decision right of e and denote the OPPE under (C) by superscript C. If she offers a mechanism, among the equilibrium conditions, the constraint (EI) can be ignored because she no longer has to care about A's incentive to exert effort. Instead, she must care about her own incentive to choose e^C.[33] The constraint would be

$$\int_{\underline{\theta}}^{\overline{\theta}} [y(q^C(\theta), e^C) - w^C(q^C(\theta), e^C)]d\theta \geq \int_{\underline{\theta}}^{\overline{\theta}} [y(q^C(\theta), e') - w^C(q^C(\theta), e')]d\theta \tag{2.11}$$

for $e^C \neq e'$. The other constraints are unchanged. Then given that P offers some mechanism $W^C \in \mathfrak{W}$, the problem of deriving the OPPE under centralization is to maximize (2.5) subject to (2.1), (2.3), (2.4), (2.11) and monotonicity of $q^C(\theta)$. The equilibrium under (C) is characterized as follows.

Lemma 2.3 *The optimization problem described above is equivalent to the following one;*

$$\max_{q^C(\theta)} \int_{\underline{\theta}}^{\overline{\theta}} J(q^C(\theta), \bar{e}, \theta) f(\theta) d\theta$$

[32] While depending on the allocation the strategy space is changed, a similar proof of Proposition 2.1 and Lemma 2.1 shows that we can still focus on the stationary strategy where all the types accept if a mechanism is offered.

[33] Here e^C must be independent from θ because we assume that P chooses e before observing A's choice of q. Nevertheless the same result is established even if P chooses e after observing A's choice of q so that e^C can be type-dependent.

subject to monotonicity of $q^C(\theta)$*.*

It is easy to see that Assumption 2.1 assures the following result.

Proposition 2.4 *For any* $\delta \in [0, 1)$*, the OPPE under (C) satisfies that* $W^C \in \overline{\mathfrak{W}}$*,* $q^C(\theta) = q^{SB}(\theta)$*, and* $e^C = \bar{e}$*.*

2.6.2 Task Separation

Next consider (S) in which P delegates the decision right of e to another agent. Denote the equilibrium by superscript S and let us call the agent to choose q $A1$ and one to choose e $A2$, respectively. The timing of the game within a period is modified as follows. At stage 0, only $A1$ knows his type θ. At stage 1, P offers a mechanism $p^S(q)$ to $A1$ and $b^S(e)$ to $A2$ simultaneously and publicly in the sense that every party observes this offer.[34] If at least either of the agents rejects the mechanism offered by P, then the period ends and all the parties obtain zero payoff as their outside option. If both accept, then at stage 2, $A1$ and $A2$ make a decision on their respective tasks simultaneously. Finally, at stage 3, P and $A2$ decide whether to enforce the informal agreement or not.

Suppose that P offers a mechanism. Since $A1$ is not involved in the second task or relational contracts, the constraints (EI) and (HC) are ignored and then the constraints on $A1$ are characterized by (2.1), (2.3), and monotonicity of $q^S(\theta)$. Instead, P must care about $A2$'s incentive to choose e^S, $A2$'s participation constraint, and the self-enforcing constraint between P and $A2$. Let $u^{2S} \equiv b^S(e^S) - e^S$ be $A2$'s payoff within a period. The constraints are respectively described as

$$u^{2S} \;\; \geq \;\; b^S(e') - e' \text{ for } e^S \neq e' \tag{2.12}$$

[34]The result is not changed even if P can offer an informal agreements to $A1$, an informal agreement can depend on q, or the offers are sequential.

$$u^{2S} \geq 0 \tag{2.13}$$

$$\frac{\delta}{1-\delta}\pi^S \geq b^S(e) \geq \frac{\delta}{1-\delta}u^{2S} \text{ for } e = \{0, \bar{e}\}. \tag{2.14}$$

Although these additional constraints must be taken into account, the procedure for deriving the optimal solution can be similarly simplified.

Lemma 2.4 *The optimization problem described above is equivalent to the following;*

$$\max_{q^S(\theta)} \int_{\underline{\theta}}^{\overline{\theta}} J(q^S(\theta), \bar{e}, \theta) f(\theta) d\theta$$

$$\textit{subject to } \frac{\delta}{1-\delta} \int_{\underline{\theta}}^{\overline{\theta}} J(q^S(z), \bar{e}, z) f(z) dz \geq \bar{e} \tag{2.15}$$

and $q^S(\theta)$ is non-increasing in θ.

By using this simplified problem, the equilibrium can be characterized as follows.

Proposition 2.5 *Let $\underline{\delta}^S \equiv \bar{e}/(\bar{e} + \int_{\underline{\theta}}^{\overline{\theta}} J(q^{SB}(z), \bar{e}, z) f(z) dz)$. The OPPE under (S) satisfies the following.*

1. *For $\delta \geq \underline{\delta}^S$, $q^S(\theta) = q^{SB}(\theta)$ and $e^S = \bar{e}$.*

2. *For $\delta < \underline{\delta}^S$, $W^S = \phi$.*

2.6.3 Implications

Note that given that P offers a mechanism in equilibrium, P's objective is $\int_{\underline{\theta}}^{\overline{\theta}} J(q(\theta), \bar{e}, \theta) f(\theta) d\theta$ which is common among (B), (C), and (S). Then, if P can choose among those, she chooses the one attaining the maximum equilibrium payoff. The corollary below directly follows from Proposition 2.2, 2.4, and 2.5.

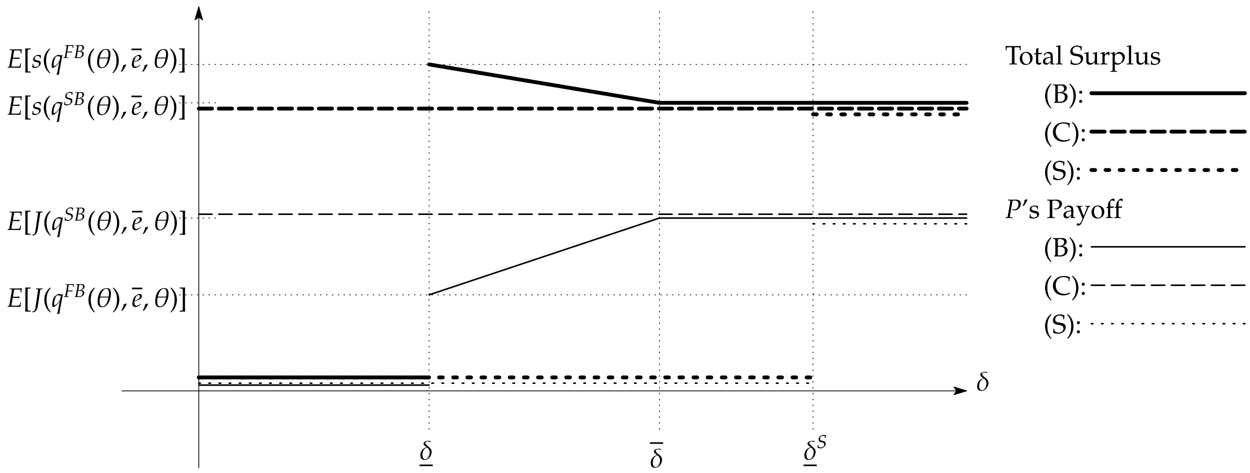

Figure 2.5: Comparison among (B), (C), and (S)

Corollary 2.3 1. *Suppose that P can choose (C). Then she weakly prefers (C) to the others. She strictly prefers (C) to (B) if* $\delta < \overline{\delta}$ *and to (S) if* $\delta < \underline{\delta}^S$.

2. *For* $\delta \in [\underline{\delta}, \overline{\delta})$*, when (B) and (C) are available, the total surplus is increased by preventing (C) from being chosen.*

3. *Suppose that P can choose only (B) and (S). Then she weakly prefers (B) to (S) for any* $\delta \in [0, 1)$ *and strictly prefers (B) for* $\delta \in [\underline{\delta}, \underline{\delta}^S)$.

4. *The total surplus in (B) is weakly greater than (S) for any* $\delta \in [0, 1)$ *and strictly greater for* $\delta \in [\underline{\delta}, \underline{\delta}^S)$.

Corollary 2.3 can be easily obtained from Figure 2.5, which illustrates the result of the three modes. Comparison between (B) and another mode gives us interesting implications. First, centralization is always preferred by P but could be less socially efficient than task bundling. Since e generates direct benefit for P, P is willing to choose $e = \bar{e}$ in (C) and then she can simply optimize $\int_{\underline{\theta}}^{\overline{\theta}} J(q, \bar{e}, \theta) f(\theta) d\theta$ by formal contracts. This is very similar to the situation in which e is verifiable as in Section 2.5. It eliminates the effect of improving social efficiency caused by the concern for implementing the unverifiable task when $\delta \in [\underline{\delta}, \overline{\delta})$.

This analysis provides a rationale for delegation in organizations. Specifically, if monetary transfer is allowed in period -1 in which the allocation of decision right on e is determined, P should "sell" the decision right to A unless δ is so small that relational contracts do not work at all. Interestingly, this is somewhat counter-intuitive in the sense that since e has a direct benefit for P, P has a right incentive to work on the unverifiable task and then the inefficiency caused by incomplete contracts disappears under centralization. Here, delegation to the party with small bargaining power can be a commitment device to avoid the inefficiency caused by the asymmetric information.

Second, task separation is generally less preferred from both P's and the social perspective. It is obvious from the comparison between the constraints of (B) and that of (S). Contrary to (2.6) in (B), the consequent constraint in (S) becomes (2.15). The key feature of (S) different from (B) is that whether relational contracts work or not is independent from $A1$'s payoff. Actually the left hand side is the aggregate surplus between P and $A2$, which does not include $A1$'s payoff. Since $s(q, \bar{e}, \theta) > J(q, \bar{e}, \theta)$ for all $q \in Q$ and $\theta \in (\underline{\theta}, \overline{\theta}]$, if $q^S(\theta)$ is feasible in the optimization problem of (S), then it is also feasible in the optimization problem of (B). Then P does not prefer (S) to (B) because of the feasibility. In addition, P does not have any incentives to mitigate the social inefficiency caused by $A1$'s information rent under (S). This implies that (S) is also socially less desirable.

Comparison between (B) and (S) suggests that while the tasks have no technological complementarity in the sense that the total costs are additively separable, they are complements due to the incentive problem. Intuitively, it is because the future information rent can be an incentive device for the task with unverifiable performance. Bundling the tasks provides information rent caused by the first task for the agent who also works on the second task and then the incentive constraint for the second task can be relaxed in comparison with task separation. In this sense, tasks with hidden information and unverifiable measurement would be complements when relational contracts are used for providing incentives.

The benefit of task bundling from an incentive perspective has been already discussed in the literature. In static settings, it has been shown that, given that tasks have no complementarity in the agent's cost, the tasks should be bundled and allocated to one agent if the difficulty of measuring performance is similar among those tasks (Holmström and Milgrom, 1991) or if there is only an aggregate measurement for the tasks available for terms in contracts (Itoh, 1994). In dynamic settings, this would also be the case if there is a component which causes a ratchet effect (Meyer et al., 1996). All of these studies rule out private information held by the agent and the possibility of relational contracting.[35]

Recently, as argued by Gibbons (2005a), there has been an increasing interest in the effect of organizational design on relational contracts. Schöttner (2008) studies a relational contracting model with multiple tasks and discusses the effect of job design on relational contracts. She focuses on moral hazard environments and shows that task bundling can reinforce relational contracts due to the malfunction of formal contracts.[36] Our model suggests another perspective for job design problems with relational contracts; task bundling is beneficial since the task with private information can provide an incentive for the unverifiable task *via* information rent.

[35]In case of hidden information with a large number of tasks, task bundling could also be beneficial by using the linking mechanism. See Jackson and Sonnenschein (2007) and Matsushima et al. (2010) for the detail.

[36]This argument is similar to the discussion of verifiability in the sense that both verifiability and job design can be instruments that change the effectiveness of punishment for cheating in relational contracts.

2.7 Extension and Robustness

2.7.1 Extension to a General Setting

For tractability, our model is simplified in some aspects which might seem not to be reasonable. Here we discuss how these assumptions are innocuous to the results derived above.

Binary Choice of Effort First, we assume that the second task e is a binary decision. This is not a crucial assumption as long as both q and e are assumed to be essential for the value of the relationship. For instance, suppose that e is chosen from interval $[0, E]$ where $E \geq \bar{e}$. Then the same result is obtained if $y(q, e)$ is a step function in e which jumps from 0 to positive value at $e = \bar{e}$. Or even if $y(q, e)$ is increasing and differentiable everywhere in e, a similar equilibrium is attained if $y_e(q, e)$ is sufficiently high. The crucial assumption for our result is that e has a large marginal effect on the aggregate benefit so that providing an incentive for e is a priority. It forces P to improve the total surplus to induce high effort. The extreme form of this idea is a binary decision.

Measurement Error Second, our model does not contain any error in measuring the outcome of the task. It can be relaxed as follows. Suppose that both q and e are unobservable for P and then unverifiable. Instead, a stochastic signal $k \in \{1, 0\}$ correlated with q is verifiable and both parties can observe an unverifiable signal $y \in \{1, 0\}$ correlated with e. Finally, P's *ex post* benefit is given by Bky where B is constant positive. In this setting, according to Levin (2003), without loss of generality, the equilibrium is still characterized by the stationary strategy without any possibilities of terminating the relationship and the result is qualitatively similar to the case without measurement errors.

While we abstract from measurement errors for focusing on the information rent problem, this extension would be valuable to investigate. One promising topic for future research is the usefulness of less informative verifiable signals as discussed in Baker et al. (1994) for single task agency. They study whether and how a noisy verifiable signal is useful when the parties intend to use relational contracts based on a more informative but unverifiable signal. In our model, the verifiable signal k does not contain any information on the unverifiable task e. By introducing correlation between k and e, we may discuss which kind of verifiable signal is desirable for relational contracts. Another topic is further investigation into task assignment problem. Our model generates no trade-off between task bundling and task separation; task bundling is weakly dominant from an incentive perspective. However, as Holmström and Milgrom (1991) point out, heterogeneity in the degree of informativeness of the signals can lead to task separation since task separation can avoid misallocation of effort across tasks and Schöttner (2008) points out that it has the opposite effect if relational contracts are available. Thus incorporating measurement noise into our model can stimulate discussion on the job design problem.

Additive Separability of Cost Function The third simplification is about additive separability of A's cost function. We should admit that this assumption is for tractability of the model. Nevertheless the result seems to be robust as long as a marginal change of θ and q has little effect on that of e.

Let us briefly review the technical difficulty. Suppose that A's cost function is generalized to $c(q, e, \theta)$ and denote $\overline{c}(q, \theta) \equiv c(q, \overline{e}, \theta)$ and $\underline{c}(q, \theta) \equiv c(q, 0, \theta)$. Assume that $\overline{c}_q(q, \theta) \geq \underline{c}_q(q, \theta)$ and $\overline{c}_\theta(q, \theta) \geq \underline{c}_\theta(q, \theta)$. When the analysis proceeds as in Section 2.4, (2.6) becomes

$$\frac{\delta}{1-\delta}\int_{\underline{\theta}}^{\overline{\theta}} s(q^*(z),\overline{e},z)f(z)dz \geq \overline{c}(q^*(\theta),\theta) - \underline{c}(q^*(\theta),\theta)$$

for all θ who accept the mechanism. This inequality says as before that the cost of the unverifiable task must be no more than the future discounted sum of the expected total surplus. Here the cost of the unverifiable task depends on the type.[37] Then the optimization problem must take into consideration the constraints for every θ. Since there are infinitely many θ, it makes the problem much more complex.[38] It further suggests that it could be the case that some types do not accept the mechanism in equilibrium even under Assumption 2.1. Thus we might have to consider the set of types which reject the mechanism. It makes the analysis more complex.

2.7.2 Assumptions on the Behaviour

Punishment upon Rejection of Mechanisms Assumption 2.3 imposes that if no mechanism is offered or an offered mechanism is rejected, then the players play the same strategy in the continuation game. It implies that if the equilibrium stipulates that P offers a mechanism in period 0, then the same mechanism must be offered in the next period if it was rejected in period 0.

For $\delta \geq \underline{\delta}$, on the OPPE the same mechanism is offered each period and all types accept it. Then rejection is an unexpected deviation in the sense that it leads off the equilibrium path. According to Abreu (1988), allowing the punishment for deviations leading off the equilibrium path could make the players better off. Thus allowing termination of the

[37] It would be the case even if $\overline{c}_\theta(q,\theta) = \underline{c}_\theta(q,\theta)$ because the type affects the optimal quantity, which further affects the effort as long as $\overline{c}_q(q,\theta) > \underline{c}_q(q,\theta)$.

[38] If we make an assumption that $\overline{c}_\theta(q,\theta) = \underline{c}_\theta(q,\theta)$, then it can be shown that the right hand side achieves the maximum at $\theta = \underline{\theta}$.

relationship after rejecting the mechanism could change the equilibrium.

In order to look at this possibility, consider the following situation. After A rejects the equilibrium mechanism, instead of the same mechanism being offered, P does not offer any mechanism and both parties obtain zero payoff. Denote this equilibrium by superscript $**$. Now consider the participation constraint for A with type θ. If he accepts the mechanism, then he obtains $U^{**}(\theta)$ in the current period and u^{**} as his continuation payoff. Conversely, If he rejects it, then he obtains 0 in the current period and also 0 as his continuation payoff. Then instead of (2.3), the participation constraint becomes

$$U^{**}(\theta) + \frac{\delta}{1-\delta}u^{**} \geq 0$$

for all $\theta \in [\underline{\theta}, \overline{\theta}]$.

By the same procedure as in Section 2.4, we see that the equilibrium satisfies that for high δ,

$$s_q(q^{**}(\theta), \bar{e}, \theta) \quad = \quad (1-\delta)c_{q\theta}(q^{**}(\theta), \theta)\frac{F(\theta)}{f(\theta)}$$

and for intermediate δ,

$$s_q(q^{**}(\theta), \bar{e}, \theta) \quad = \quad \frac{1-\delta}{1+\frac{\delta}{1-\delta}\lambda^{**}}c_{q\theta}(q^{**}(\theta), \theta)\frac{F(\theta)}{f(\theta)}$$

where $\lambda^{**} > 0$. It immediately implies that for high δ, $q^{FB}(\theta) > q^{**}(\theta) > q^{*}(\theta) = q^{SB}(\theta)$ for $\theta \in (\underline{\theta}, \overline{\theta}]$ meaning that total efficiency is improved by imposing punishment for A's rejection.[39] Intuitively, by rejecting the mechanism, A loses two benefits: the current

[39] For intermediate δ, direct comparison between $q^{*}(\theta)$ and $q^{**}(\theta)$ is impossible in general since the threshold of the discount factor and the Lagrange multiplier are different. Nevertheless the same inequality is established.

information rent and the expected payoff from future trade. It means that relative to the case without punishments for the rejection, the current payoff is less important for A and then the incentive to mimic another type to obtain the information rent is also less important. It leads to a more efficient outcome since P's concern about the reduction of information rent is less serious. Imposing punishment on A's rejection also alters A's payoff level. Specifically,

$$U^{**}(\overline{\theta}) = -\delta \int_{\underline{\theta}}^{\overline{\theta}} c_\theta(q^{**}(\theta), \theta) F(\theta) d\theta,$$

which is obviously negative for $\delta > 0$. Then, even if the current payoff is lower than his outside option, A could choose to participate because present loss can be compensated by future trade.

Whereas imposing punishment on rejection can improve efficiency, we rule out this possibility in the main analysis for two reasons. First, it is not robust to some aspects. For instance, the result that every type participates in equilibrium is due to Assumption 2.1 on $J(q, \overline{e}, \theta)$. If this assumption is relaxed and there are some types rejecting the mechanism in equilibrium, rejection is no longer an unexpected observable deviation and then the punishment is not necessarily optimal. Second, while the equilibrium is altered by imposing punishment on rejection, our main results derived above are qualitatively the same. In particular, the social desirability of unverifiability and the issues in job design can be demonstrated similarly even if we allow punishment for rejection and then our message from the analysis is basically unchanged.

Principal's Commitment Ability The important restriction made by Assumption 2.4 is that in each period the strategy must maximize the principal's payoff subject to the equilibrium conditions as long as the players are on the equilibrium path. In other words, the principal cannot commit themselves to a worse equilibrium in the continuation game. While this assumption seems to be reasonable in the case where we presume that P has always full bargaining power, we should review the case where the principal can commit themselves to a worse outcome.

If both Assumption 2.3 and 2.4 are absent, then P's optimal equilibrium strategy is no longer stationary. Nevertheless, when e is verifiable, it is still simply characterized as follows.[40]

- If $\delta \geq \underline{\delta}$, then the equilibrium decision in period t, $(q_t(\theta), e_t(\theta))$ is independent of the past history and satisfies $(q_0(\theta), e_0(\theta)) = (q^{SB}(\theta), e^{SB}(\theta))$

$$(q_t(\theta), e_t(\theta)) = \begin{cases} (q^{SB}(\theta), e^{SB}(\theta)) & \text{for } t = 0 \\ (q^{FB}(\theta), e^{FB}(\theta)) & \text{for } t \geq 1 \end{cases}$$

  and

- otherwise, the principal offers no contract every period.

When $\delta \geq \underline{\delta}$, only the decision in the first period is distorted in equilibrium. When the principal implements the first best, she must give more information rent to the agent than the second best level. While it is costly for her in the one shot relation, the dynamic structure allows that this loss for her can be compensated by transfer from the agent in the first period.

It is not hard to confirm that these results are not consistent with what we have argued so far. Specifically, raising δ and verifiability of e never worsen efficiency. Then it seems that our argument is vulnerable without Assumption 2.4. Nevertheless we argue that it

[40] The proof is in the Appendix.

is due to the assumption that θ is independently drawn over time, which we impose for tractability of the model. If θ is serial correlated, then our argument would be valid even without Assumption 2.4 while we do not formally show the result here.

The above conjecture can be obtained from the result by Battaglini (2005) who studies a model of dynamic price discrimination where the consumer has private information about the valuation of the good and the valuation evolves in a Markovian way. Whereas he considers a single task model with binary types, the first task level in the above equilibrium is very similar to his equilibrium characterization where the type is independently drawn over time. As Battaglini (2005)'s result suggests, if the type evolves in a Markovian way, then the decision optimal for the principal still remains inefficient distortion in the future. It implies that if in our two task model the type evolves in a Markovian way and Assumption 2.3 and 2.4 are absent, then the principal attempts to implement the quantity similar to Battaglini (2005)'s characterization as long as it can provide an incentive for the second task.

We have argued that when the constraint to honour relational contracts are stringent, the future inefficiency can be mitigated for ensuring the relational contract to work. In our framework with independent drawn type, however, without Assumption 2.4 the information rent problem causes the inefficiency only in the first period and then our logic is not valid. Nevertheless we guess that if the type is serially correlated which causes persistent inefficiency due to the screening problem, our argument would be restored even if Assumption 2.3 and 2.4 are dropped.

Incentives through Efficiency Wages Throughout this chapter, we assume that *ex post* transfer stipulated by informal agreements is available every period as long as the players have no incentive to deviate. Nevertheless it might be presumably infeasible from some reasons.[41] If *ex post* transfer is unavailable, the efficiency wage can be a substitute for providing incentives for unverifiable tasks. As Shapiro and Stiglitz (1984) demonstrate, in the efficiency wage scheme, the players terminate their relationship after observing that the agent makes an unexpected deviation.

Let us describe the stationary equilibrium with efficiency wages. Denote the equilibrium with efficiency wage by superscript EW. Since they do not use any informal agreements, (HC) is no longer relevant. It is straightforward that (TT) and (PC) are the same as before. The constraint is changed in (EI). If A with type θ chooses $q = q^{EW}(\theta)$ and $e = \bar{e}$, then he can obtain $U^{EW}(\theta)$ in the current period and u^{EW} as the continuation payoff. Conversely if he chooses $q = q^{EW}(\theta')$ and $e = 0$, then his payoff in the current period is

$$p^{EW}(q^{EW}(\theta')) - c(q^{EW}(\theta'), \theta) = U^{EW}(\theta') + c(q^{EW}(\theta'), \theta') - c(q^{EW}(\theta'), \theta) + \bar{e}$$

and the continuation payoff is 0 because the relationship with the principal ends. Hence the equilibrium condition is

$$U^{EW}(\theta) + \frac{\delta}{1-\delta} u^{EW} \geq U^{EW}(\theta') + c(q^{EW}(\theta'), \theta') - c(q^{EW}(\theta'), \theta) + \bar{e}.$$

As Lemma 2.2, the equilibrium is characterized by the optimization problem to maximize

[41] For instance, if we consider a sequence of short-lived principals instead of one principal playing in infinite periods, then each principal cannot commit to honour any informal agreements. Another example is a situation where there are multiple agents competing each period to get the monopolistic transaction opportunity with the principal. In this case, the principal has an opportunity to find and transact with another agent next period. If the parties outside the transaction have no way of knowing if cheating was taken place in the informal agreement, then punishing such cheating becomes impossible. It makes it impossible for the principal to credibly commit herself to informal agreements no matter how patient she is. See MacLeod and Malcomson (1989, 1998) and Calzolari and Spagnolo (2010) for an analysis of the competitive situation.

$\int_{\underline{\theta}}^{\overline{\theta}} J(q^{EW}(\theta), \bar{e}, \theta)F(\theta)d\theta$ subject to

$$\frac{\delta}{1-\delta}u^{EW} \geq \bar{e} \tag{2.16}$$

and monotonicity of $q^{EW}(\theta)$ where $u^{EW} = \int_{\underline{\theta}}^{\overline{\theta}} c_\theta(q^{EW}(\theta), \theta)F(\theta)d\theta$. It is again straightforward that if δ is so high that $q^{EW}(\theta) = q^{SB}(\theta)$ satisfies (2.16), then it is the solution. Otherwise, it is still valid to solve the Lagrangian if $c_{\theta qq} \leq 0$. In this case, $q^{EW}(\theta)$ satisfies that

$$s_q(q^{EW}(\theta), \bar{e}, \theta) = \left(1 - \frac{\delta\lambda^{EW}}{1-\delta}\right)\frac{F(\theta)}{f(\theta)}c_{\theta q}(q^{EW}(\theta), \theta)$$

with $\lambda^{EW} \geq 0$ which satisfies (2.16) with equality provided that $\lambda^{EW} \leq (1-\delta)/\delta$.[42] If these qualifications are satisfied, then a similar result is obtained to the case with informal agreements; the total surplus for intermediate discount factors is greater than for high discount factors.

If the above characterization by the Lagrangian is not valid, then the quantity could be distorted upward from $q^{FB}(\theta)$. Whereas we admit it to be an open question, our guess is reasonable since (2.16) shows that to induce $e^{EW} = \bar{e}$ is possible if A's (future) information rent is sufficiently large. Since the information rent is assured by increasing q monotonically, P would distort q upward as long as such large q is still beneficial to her.

2.8 Conclusion

This chapter studied the interaction of formal and informal contracts in multitasking agency problem with adverse selection. We investigated the multitasking incentive prob-

[42] If $\lambda^{EW} > (1-\delta)/\delta$, then $q^{EW}(\theta)$ does not satisfy monotonicity.

lem to show how the incentive problems due to hidden information and unverifiability interact. The analysis provided various insights on contract and organization design.

We have shown that unverifiability can improve social efficiency. If a performance measure of some task becomes unverifiable, then the principal needs to use relational contracts to provide incentives to undertake the task. Then the principal's main concern might shift from the reduction of the agent's rent to assuring credible relational contracts. This shift could improve social efficiency since relational contracts can be sustained by increasing the future total surplus.

Second, the task assignment problem implies that those tasks should be bundled and allocated to a single agent. Specifically, whereas leaving the decision rights on the unverifiable task in the hands of the principal seems to mitigate the agency problem, as in the discussion of the verifiable task, it negatively affects the relational contracts. Furthermore, separating those tasks between two agents is totally undesirable. This suggests that those tasks should be treated as complements from the incentive perspective.

We conclude by suggesting a future research agenda. First, our model does not include any measurement error. According to Levin (2003), an extension in this direction seems to be somewhat easy since the optimal equilibrium would still be stationary and, as mentioned in Section 2.7, it can provide additional insights on contract and organization design. Second and more broadly, we guess that a similar result can be obtained without hidden information. In particular, the key idea behind the assumption on the first task is that the decision on the first task is distorted from the efficient level due to the principal's cost of implementing the efficient decision. This phenomenon also emerges for instance in moral hazard environments with limited liability. However, if there is another task whose performance measure is unverifiable, the principal again alters her priority from incentivizing the first task according to her payoff to ensuring credible relational contracts.

Thus a formal analysis that investigates what kind of multitasking agency problems generates an outcome similar to ours is a future research topic for examining the robustness of our prediction.[43]

2.9 Appendix: Proofs

2.9.1 Proof of Proposition 2.1

First, we show two lemmas.

Lemma 2.5 *Let (π, u) be a PPE payoff. Then $\pi \geq 0$ and $u \geq 0$.*

Proof (Lemma 2.5) *P can achieve at least zero payoff by abstaining from offering mechanisms regardless of the public history. A can obtain at least zero payoff by rejecting any mechanism regardless of the public history.* □

Lemma 2.6 *There exists a PPE σ such that $\Gamma(h^t) = \phi$ for all $h^t \in \mathcal{H}$.*

Proof (Lemma 2.6) *Consider the strategy satisfying the following. For all $h^t \in \mathcal{H}$, $\Gamma(h^t) = \phi$ and for any $\tilde{W} = (\tilde{p}, \tilde{b}) \in \mathfrak{W}$ and $d \in D$*

$$
\begin{aligned}
\chi(\tilde{W} \mid h^t, \theta) &= \begin{cases} (\tilde{q}, 0) \in D & \text{if } \tilde{q} \in \arg\max_{q \in Q}[\tilde{p}(q) - c(q, \theta)] \text{ and } \tilde{p}(\tilde{q}) - c(\tilde{q}, \theta) \geq 0 \\ \omega & \text{otherwise,} \end{cases} \\
\iota^i(\tilde{W}, d \mid h^t) &= R.
\end{aligned}
$$

Namely, P does not offer any mechanism and A does not choose $e = \bar{e}$ after every history as long as it assures non-negative payoff in the current period. Given history h^t, type θ, and mechanism $\tilde{W}$,

[43] The difficulty of the model of relational contracts with limited liability is that the optimal equilibrium is in general non-stationary and then the dynamics must be taken into account. For instance, see Thomas and Worrall (1994) and Fong and Li (2010).

A has no incentive to deviate from $\chi(\tilde{W} \mid h^t, \theta)$ since deviation does not improve his current payoff nor change the continuation payoff.[44] *Given history h^t, type θ, mechanism $\tilde{W} \in \mathfrak{W}$, and decision $d \in D$, to change from $\iota^A(\tilde{W}, d \mid h^t) = R$ to H does not alter A's payoff since $\iota^H(\tilde{W}, d \mid h^t) = R$ and the continuation payoff is not changed. Thus the strategy is sequentially rational for A. Suppose that changing P's strategy alters her expected payoff. Note that due to the same argument, to change from $\iota^P(\tilde{W}, d \mid h^t) = R$ to $I^P = H$ does not alter P's payoff. Then it could be the case only when she deviates at stage 0 in some period and A must accept the mechanism. However because of the construction of A's strategy, if A accepts the mechanism, then the total benefit must be strictly lower than* 0 *by Assumption 2.1. It implies that either P or A's payoff must be strictly lower than* 0*, which contradicts Lemma 2.5. Thus changing P's strategy does not alter her expected payoff implying that P has no incentive to change her strategy. Then this strategy is a PPE.* □

Note that Lemma 2.6 implies that there is a PPE in which the expected payoff vector is $(0, 0)$.

Suppose that $\Gamma(h^0) = \phi$. Then by assumption 2.4, $\Gamma(h^t) = \phi$ for any h^t on the equilibrium path. The strategy described in the proof of Lemma 2.6 generates the equilibrium payoff $(0, 0)$. Since Lemma 2.5 assures that the OPPE payoff must be no less than 0 for both parties, the statement in Proposition 2.1 is satisfied.

[44] We have assumed that an formal contract $p(\cdot)$ must be continuous. When we allow P to offer non-continuous $p(\cdot)$, this strategy is not necessarily sequentially rational. Specifically, for some non-continuous $p(\cdot)$, $\arg\max_{q \in Q}[p(q) - c(q, \theta)]$ might be empty. It means that in the subgame after such $p(\cdot)$ is offered, there is no optimal decision for A and then this strategy does not satisfy the conditions of PPE off the equilibrium path.

In the rest of the proof, suppose that $\Gamma(h^0) \in \mathfrak{W}$. Denote the set of public strategies by $\hat{\Sigma}^i$ for $i = P, A$. As mentioned in Section 2.3.3, σ^P and σ^A can be respectively rewritten as

$$\sigma^P \begin{cases} W \in \mathfrak{W} \\ \iota^P : \mathfrak{W} \times D \to \mathfrak{I} \\ \sigma_+^P : \{\phi\} \cup (\mathfrak{W} \times ((D \times \mathfrak{I}^2) \cup \{\omega\})) \to \hat{\Sigma}^P \end{cases}$$

and

$$\sigma^A \begin{cases} \chi : [\underline{\theta}, \overline{\theta}] \times \mathfrak{W} \to \overline{D} \\ \iota^A : \mathfrak{W} \times D \to \mathfrak{I} \\ \sigma_+^A : \{\phi\} \cup (\mathfrak{W} \times ((D \times \mathfrak{I}^2) \cup \{\omega\})) \to \hat{\Sigma}^A. \end{cases}$$

For notational simplicity, let $\ell(W, d) = \mathbf{1}\{\iota^i(W, d) = H \text{ for } i = P, A\}$ and $(\pi_+(h^1), u_+(h^1))$ be the continuation payoff where the public history after period 0 was h^1. Let (π, u) be the corresponding payoff. By applying the one-shot deviation principle, it is shown that σ satisfies if the following five conditions (PPC), (IC), (PHC), (AHC), and (CE) are satisfied;

(PPC):

$$\begin{aligned}
& \int_{\theta:d(\theta,W)\in D} \Big[y(\chi(\theta, W)) - p(q(\theta, W)) - \ell(W, \chi(\theta, W))b(\chi(\theta, W)) \\
& + \frac{\delta}{1-\delta}\pi_+(W, \chi(\theta, W), \iota^P(W, \chi(\theta, W)), \iota^A(W, \chi(\theta, W))) \Big] f(\theta)d\theta + \int_{\theta:d(\theta,W)\notin D} \frac{\delta}{1-\delta}\pi f(\theta)d\theta \\
\geq \ & \max \Bigg\{ \sup_{W' \in \mathfrak{W}} \Bigg[\int_{\theta:d(\theta,W')\in D} \Big[y(\chi(\theta, W')) - p(q(\theta, W')) - \ell(W', \chi(\theta, W'))b(\chi(\theta, W')) \\
& + \frac{\delta}{1-\delta}\pi_+(W', \chi(\theta, W'), \iota^P(W', \chi(\theta, W')), \iota^A(W', \chi(\theta, W'))) \Big] f(\theta)d\theta \\
& + \int_{\theta:d(\theta,W')\notin D} \frac{\delta}{1-\delta}\pi f(\theta)d\theta \Bigg], \frac{\delta}{1-\delta}\pi \Bigg\}
\end{aligned}$$

(IC): when $\chi(\theta, W) = (q(\theta, W), e(\theta, W)) \in D$,

$$\begin{aligned}
& p(q(\theta, W)) - c(q(\theta, W), \theta) - e(\theta, W) + \ell(W, \chi(\theta, W))b(\chi(\theta, W)) \\
& + \frac{\delta}{1-\delta} u_+(W, \chi(\theta, W), \iota^P(W, \chi(\theta, W)), \iota^A(W, \chi(\theta, W))) \\
\geq \; & \max\left\{ \sup_{d' \equiv (q', e') \in D} \left[p(q') - c(q', \theta) - e' + \ell(W, d')b(d') + \frac{\delta}{1-\delta} u_+(W, d', \iota^P(W, d'), \iota^A(W, d')) \right], \right. \\
& \left. \frac{\delta}{1-\delta} u \right\}
\end{aligned}$$

and when $\chi(\theta, W) = \omega$,

$$\frac{\delta}{1-\delta} u \geq \sup_{d' \equiv (q', e') \in \in D} \left[p(q') - c(q', \theta) - e' + \ell(W, d')b(d') + \frac{\delta}{1-\delta} u_+(W, d', \iota^P(W, d'), \iota^A(W, d')) \right],$$

(PHC): when $\iota^P(W, d) = H$,

$$-b(d) + \frac{\delta}{1-\delta} \pi_+(W, d, H, \iota^A(W, d)) \geq \frac{\delta}{1-\delta} \pi_+(W, d, R, \iota^A(W, d)),$$

and when $\iota^P(W, d) = R$,

$$-b(d) + \frac{\delta}{1-\delta} \pi_+(W, d, H, \iota^A(W, d)) \leq \frac{\delta}{1-\delta} \pi_+(W, d, R, \iota^A(W, d)),$$

(AHC): when $\iota^A(W, d) = H$,

$$b(d) + \frac{\delta}{1-\delta} u_+(W, d, \iota^P(W, d), H) \geq \frac{\delta}{1-\delta} u_+(W, d, \iota^P(W, d), R),$$

and when $\iota^A(W, d) = R$,

$$b(d) + \frac{\delta}{1-\delta} u_+(W, d, \iota^P(W, d), H) \leq \frac{\delta}{1-\delta} u_+(W, d, \iota^P(W, d), R),$$

and

(CE): for all $z = \phi$ or $(W, d) \in \mathfrak{W} \times (\{\omega\} \cup (D \times \mathfrak{I}^2))$, $(\pi_+(z), u_+(z))$ is a PPE and it is also an OPPE if z is on the equilibrium path.

Let $\Sigma^* \equiv \{\sigma \in \hat{\Sigma} \mid \sigma$ satisfies (PPC), (IC), (PHC), (AHC), and (CE)$\}$. The OPPE is the strategy which maximizes P's payoff in Σ^*. Let Ψ^* be the set of payoff vectors attained by $\sigma \in \Sigma^*$.

Lemma 2.7 *If $\sigma \in \Sigma^*$, its associated mechanism in period 0 is $W \in \mathfrak{W}$, and W is accepted with positive probability, then there exists an OPPE $\tilde{\sigma}$ with its associated mechanism in period 0 is $\tilde{W}$ such that the payoff is the same as σ and the informal agreements are honoured whenever $\tilde{W}$ is offered and A accepts it.*

Proof (Lemma 2.7) *Construct strategy $\tilde{\sigma}$ as follows; for $d \in D$ such that $\iota(W, d) \neq (H, H)$,*

$$\begin{aligned} \tilde{b}(d) &= 0, \\ (\tilde{\pi}_+(\tilde{W}, d, H, H), \tilde{u}_+(\tilde{W}, d, H, H)) &= (\pi_+(W, d, \iota(W, d)), u_+(W, d, \iota(W, d))), \\ (\tilde{\pi}_+(\tilde{W}, d, I), \tilde{u}_+(\tilde{W}, d, I)) &= (0, 0) \text{ for } I \neq (H, H), \\ \tilde{\iota}^P(\tilde{W}, d) = \tilde{\iota}^A(\tilde{W}, d) &= H, \end{aligned}$$

and the others are the same as σ, i.e., $\tilde{\chi}(\theta, \tilde{W}) = \chi(\theta, W)$, $\tilde{p}(q) = p(q)$, $\tilde{b}(d) = b(q)$, $\tilde{\iota}^i(\tilde{W}, d) = \iota^i(W, d)$ for $i = P, A$, $(\tilde{\pi}_+(\tilde{W}, d, I), \tilde{u}_+(\tilde{W}, d, I)) = (\pi_+(W, d, I), u_+(W, d, I))$, and $(\tilde{\pi}_+(\tilde{W}, \omega), \tilde{u}_+(\tilde{W}, \omega)) = (\pi_+(W, \omega), u_+(W, \omega))$. Note that $(\tilde{\pi}_+(\tilde{W}, d, H, H), \tilde{u}_+(\tilde{W}, d, H, H)) \in \Psi^*$ and $(\tilde{\pi}_+(\tilde{W}, d, I), \tilde{u}_+(\tilde{W}, d, I))$ is a PPE for $I = (H, R), (R, H), (R, R)$ thanks to Lemma 2.5. Then (CE) is satisfied. By construction, (PHC) and (AHC) with $\tilde{\iota}^P(\tilde{W}, d') = \tilde{\iota}^A(\tilde{W}, d') = H$ are satisfied for any $d' \in D$ and

$$\tilde{p}(q) + \tilde{\ell}(\tilde{W}, d')b(d') + \frac{\delta}{1-\delta}\tilde{u}_+(\tilde{W}, d', H, H)$$

$$= \quad p(q) + \ell(W,d')b(d') + \frac{\delta}{1-\delta}u_+(W,d',\iota^P(W,d'),\iota^A(W,d')) \tag{2.17}$$

that assures (PPC) and (IC). Hence $\tilde{\sigma}$ is a PPE. Since $\tilde{\chi}(\tilde{W},\theta) = \chi(W,\theta)$ and (2.17) holds, A's payoff in $\tilde{\sigma}$ is the same as σ, generating the same payoff as σ. □

Then without loss of generality, (PHC) and (AHC) are written as

(PHC′): for any $d \in D$,

$$-b(d) + \frac{\delta}{1-\delta}\pi_+(W,d,H,H) \quad \geq \quad \frac{\delta}{1-\delta}\pi_+(W,d,R,H)$$

(AHC′): for any $d \in D$,

$$b(d) + \frac{\delta}{1-\delta}u_+(W,d,H,H) \quad \geq \quad \frac{\delta}{1-\delta}u_+(W,d,H,R).$$

Lemma 2.8 *Ψ^* is compact.*

Proof (Lemma 2.8) *For all $(\pi,u) \in \Psi^*$, Lemma 2.5 assures $(\pi,u) \geq (0,0)$ and since $s(q,e,\theta)$ is bounded from above, $\pi + u$ is also bounded above. Then Ψ^* is bounded.*

To show that it is closed, consider an arbitrary converging sequence $\{(\pi^n,u^n)\}_{n=0}^{\infty}$ where $(\pi^n,u^n) \in \Psi^*$ for all $n \geq 0$ and $\lim_{n\to\infty}(\pi,u)$ and let σ^n be the corresponding strategy profile. The proof is completed if $(\pi,u) \in \Psi^*$. Let $\underline{\mathfrak{W}} \equiv \{p : Q \to [-K,K], b : D \to [-K,K]\}$ where $K \equiv \sup_{q'\in Q, e'\in\{0,\bar{e}\}, \theta'\in[\underline{\theta},\bar{\theta}]} s(q',e',\theta')\delta/(1-\delta)$. Then the mechanism offered under σ^n is without loss of generality in $\underline{\mathfrak{W}}$ which is compact. Then we can construct a converging subsequence $\{\tilde{\sigma}^n\}_{n=0}^{\infty}$. Let $\tilde{\sigma}$ be its limit. Note that $\tilde{\sigma}$ can be decomposed to $\tilde{W}$, $\tilde{d}(\theta,W)$, $\tilde{\iota}^i(W,d)$, and $\tilde{\sigma}^i_+(h^1)$ for $i = P,A$. Now suppose that $(\pi,u) \notin \Psi^*$. Then either (PPC), (IC), (PHC′), (AHC′), or (CE) is violated. If

(PPC) is violated, then there exists $W' \in \mathfrak{W}$ such that

$$\begin{aligned}
&\int_{\theta:\tilde{d}(\theta,\tilde{W})\in D}\Big[y(\tilde{\chi}(\theta,\tilde{W})) - \tilde{p}(\tilde{q}(\theta,\tilde{W})) - \tilde{\ell}(\tilde{W},\tilde{\chi}(\theta,\tilde{W}))\tilde{b}(\tilde{\chi}(\theta,\tilde{W})) \\
&+\frac{\delta}{1-\delta}\tilde{\pi}_{+}(\tilde{W},\tilde{\chi}(\theta,\tilde{W}),\tilde{\iota}^{P}(\tilde{W},\tilde{\chi}(\theta,\tilde{W})),\tilde{\iota}^{A}(\tilde{W},\tilde{\chi}(\theta,\tilde{W})))\Big]f(\theta)d\theta \\
&+\frac{\delta}{1-\delta}\int_{\theta:\tilde{d}(\theta,\tilde{W})\notin D}\tilde{\pi}f(\theta)d\theta + \varepsilon \\
< \ & \max\Big\{\int_{\theta:\tilde{d}(\theta,W')\in D}\Big[y(\tilde{\chi}(\theta,W')) - p'(\tilde{q}(\theta,W')) - \tilde{\ell}(W',\tilde{\chi}(\theta,W'))b'(\tilde{\chi}(\theta,W')) \\
&+\frac{\delta}{1-\delta}\tilde{\pi}_{+}(W',\tilde{\chi}(\theta,W'),\tilde{\iota}^{P}(W',\tilde{\chi}(\theta,W')),\tilde{\iota}^{A}(W',\tilde{\chi}(\theta,W')))\Big]f(\theta)d\theta \\
&+\frac{\delta}{1-\delta}\int_{\theta:\tilde{d}(\theta,W')\notin D}\tilde{\pi}f(\theta)d\theta, \frac{\delta}{1-\delta}\tilde{\pi}\Big\}
\end{aligned}$$

for some $\varepsilon > 0$. Since $\tilde{\sigma}$ is a limit of a converging sequences $\{\tilde{\sigma}^n\}_{n=0}^{\infty}$, There exists $N > 0$ such that

$$\begin{aligned}
&\int_{\theta:d^N(\theta,W^N)\in D}\Big[y(\chi^N(\theta,W^N)) - p^N(q^N(\theta,W^N)) - \ell^N(W^N,\chi^N(\theta,W^N))b^N(\chi^N(\theta,W^N)) \\
&+\frac{\delta}{1-\delta}\pi_{+}^{N}(W^N,\chi^N(\theta,W^N),\iota^{PN}(W^N,\chi^N(\theta,W^N)),\iota^{AN}(W^N,\chi^N(\theta,W^N)))\Big]f(\theta)d\theta \\
&+\frac{\delta}{1-\delta}\int_{\theta:d^N(\theta,W^N)\notin D}\tilde{\pi}f(\theta)d\theta \\
< \ & \max\Big\{\int_{\theta:d^N(\theta,W')\in D}\Big[y(\chi^N(\theta,W')) - p'(q^N(\theta,W')) - \ell^N(W',\chi^N(\theta,W'))b'(\chi^N(\theta,W')) \\
&+\frac{\delta}{1-\delta}\pi_{+}^{N}(W',\chi^N(\theta,W'),\iota^{PN}(W',\chi^N(\theta,W')),\iota^{AN}(W',\chi^N(\theta,W')))\Big]f(\theta)d\theta \\
&+\frac{\delta}{1-\delta}\int_{\theta:d^N(\theta,W')\notin D}\pi^N f(\theta)d\theta, \frac{\delta}{1-\delta}\pi^N\Big\},
\end{aligned}$$

which contradicts that $\tilde{\sigma}^N$ is a PPE. Thus it could not be the case that (PPC) is violated. Similar arguments show that neither (IC), (PHC′), nor (AHC′) is not violated. Finally, suppose that (CE) is violated. Then there exists a strategy profile $\tilde{\sigma}_{+}(h^1)$ for some public history up to period 1, h^1, such that party i can increase his payoff by changing to another strategy $\sigma^{i\prime} \in \Sigma^i$. If $i = P$, then it states that $\pi(\tilde{\sigma}_{+}(h^1)) + \varepsilon < \pi(\sigma^{P\prime}, \tilde{\sigma}_{+}^{A}(h^1))$ for some $\varepsilon > 0$ where $\pi(\sigma)$ is P's average payoff given strategy profile σ. Note that since $\pi(\sigma)$ is a sum of the discounted payoff, if $\{\tilde{\sigma}_{+}^{n}(h^1)\}_{n=0}^{\infty}$ is *a converging sequence, then* $\{\pi(\tilde{\sigma}_+^n(h^1))\}_{n=0}^{\infty}$ *and* $\{\pi(\sigma^{P\prime}, \tilde{\sigma}_+^n(h^1))\}_{n=0}^{\infty}$ *are converging sequences too. Then there exists* $N > 0$ *such that* $\pi(\tilde{\sigma}_+^N(h^1)) < \pi(\sigma^{P\prime}, \tilde{\sigma}_+^{NA}(h^1))$ *for some* $\sigma^{P\prime} \in \Sigma^P$*. It contradicts that* $\tilde{\sigma}_+^N(h^1)$ *is a PPE. The same argument can be applied for* $i = A$. □

Suppose that σ is an OPPE. By Lemma 2.8, σ can be restricted to one such that the payoff $(\pi, u) \in \Psi^*$ satisfies that for all $(\pi', u') \in \Psi^*$, $\pi \geq \pi'$ and $u \geq u'$ if $\pi = \pi'$. That is, (π, u) is on the Pareto frontier of Ψ^* subject to π attaining the maximum. Now we construct a strategy σ^* in Ψ^* generating a payoff which weakly Pareto-dominates that of σ. For $d \in D$, let $W^* = (p^*(\cdot), b^*(\cdot))$ be such that

$$\begin{aligned} b^*(d) &= b(d) + \frac{\delta}{1-\delta}\left[u_+(W, d, H, H) - u\right] - \inf_{d \in D}\left[b(d) + \frac{\delta}{1-\delta} u_+(W, d, H, H)\right] \\ p^*(q) &= p(q) + \inf_{d \in D}\left[b(d) + \frac{\delta}{1-\delta} u_+(W, d, H, H)\right]. \end{aligned}$$

In each period t, P chooses $W_t = \phi$ if each of the players has chosen $I_s^i = R$ for some $s < t$ and $W_t = W^*$ otherwise. If W^* is offered, then A chooses $\chi^*(\theta, W^*) = \chi(\theta, W)$ for all θ.

Note that Lemma 2.6 ensures that there is a PPE such that P does not offer any mechanism. Since P offers W^* repeatedly on the equilibrium path of σ^*, if (IC), (PHC′), and (AHC′) under σ^* are satisfied, there is no incentive to deviate at any information set implying that σ^* is a PPE. Note that the continuation payoff is u^* unless the implicit contracts were reneged on.

We first confirm that (IC) is satisfied under W^*. Note that since σ is an OPPE, (IC) under σ implies that for $\chi(\theta, W) = (q(\theta, W), e(\theta, W)) \in D$,

$$\begin{aligned} &p(q(\theta, W)) - c(q(\theta, W), \theta) - e(\theta, W) + b(\chi(\theta, W)) + \frac{\delta}{1-\delta} u_+(W, \chi(\theta, W), H, H) \\ &\geq \max\left\{\sup_{d' \equiv (q', e') \in D}\left[p(q') - c(q', \theta) - e' + b(d') + \frac{\delta}{1-\delta} u_+(W, d', H, H)\right], \frac{\delta}{1-\delta} u_+(W, \omega)\right\} \end{aligned}$$

and for $\chi(\theta, W) = \omega$,

$$\frac{\delta}{1-\delta} u_+(W, \omega) \geq \sup_{d' \equiv (q', e') \in D} \left[p(q') - c(q', \theta) - e' + b(d') + \frac{\delta}{1-\delta} u_+(W, d', H, H) \right].$$

These imply (IC) under σ^* by the following argument. First, note that

$$\begin{aligned}
u^* &= (1-\delta) \int_{\theta: d(W,\theta) \in D} \left[p^*(q(\theta, W)) - c(q(\theta, W), \theta) - e(\theta, W) + b^*(\chi(\theta, W)) \right] f(\theta) d\theta \\
&\quad + \delta u^* \\
&= (1-\delta) \int_{\theta: d(W,\theta) \in D} \Big[p(q(\theta, W)) - c(q(\theta, W), \theta) - e(\theta, W) + b(\chi(\theta, W)) \\
&\quad + \frac{\delta}{1-\delta} \left[u_+(W, \chi(\theta, W), H, H) - u \right] \Big] f(\theta) d\theta + \delta u^* \\
\Longleftrightarrow u^* &= \int_{\theta: d(W,\theta) \in D} \Big[p(q(\theta, W)) - c(q(\theta, W), \theta) - e(\theta, W) + b(\chi(\theta, W)) \\
&\quad + \frac{\delta}{1-\delta} \left[u_+(W, \chi(\theta, W), H, H) - u \right] \Big] f(\theta) d\theta
\end{aligned}$$

and

$$\begin{aligned}
u &= \int_{\theta: d(W,\theta) \in D} \left[(1-\delta) \left(p(q(\theta, W)) - c(q(\theta, W), \theta) - e(\theta, W) + b(\chi(\theta, W)) \right) \right. \\
&\quad \left. + \delta u_+(W, \chi(\theta, W), H, H) \right] f(\theta) d\theta + \delta \int_{\theta: d(W,\theta) \notin D} u_+(W, \omega) f(\theta) d\theta.
\end{aligned}$$

Since (π, u) is on the Pareto frontier of Ψ^*, Assumption 2.4 ensures that $u \geq u_+(W, \omega)$. It implies that

$$\begin{aligned}
u^* &= \frac{1}{1-\delta} u - \frac{\delta}{1-\delta} \int_{\theta: d(W,\theta) \notin D} u_+(W, \omega) f(\theta) d\theta - \frac{\delta}{1-\delta} \int_{\theta: d(W,\theta) \notin D} u f(\theta) d\theta \\
&\geq \frac{1}{1-\delta} u - \frac{\delta}{1-\delta} \int_{\theta: d(W,\theta) \notin D} u f(\theta) d\theta - \frac{\delta}{1-\delta} \int_{\theta: d(W,\theta) \notin D} u f(\theta) d\theta = u
\end{aligned}$$

Then for $\chi(\theta, W) = (q(\theta, W), e(\theta, W)) \in D$,

$$\begin{aligned}
& p^*(q(\theta, W)) - c(q(\theta, W), \theta) - e(\theta, W) + b^*(\chi(\theta, W)) + \frac{\delta}{1-\delta} u^* \\
= \; & p(q(\theta, W)) - c(q(\theta, W), \theta) - e(\theta, W) + b(\chi(\theta, W)) + \frac{\delta}{1-\delta} u_+(W, \chi(\theta, W), H, H) \\
\geq \; & \max \left\{ \sup_{d' \equiv (q', e') \in D} \left[p(q') - c(q', \theta) - e' + b(d') + \frac{\delta}{1-\delta} u_+(W, d', H, H) \right], \frac{\delta}{1-\delta} u_+(W, \omega) \right\} \\
= \; & \max \left\{ \sup_{d' \equiv (q', e') \in D} \left[p^*(q') - c(q', \theta) - e' + b^*(d') + \frac{\delta}{1-\delta} u \right], \frac{\delta}{1-\delta} u_+(W, \omega) \right\} \\
= \; & \max \left\{ \sup_{d' \equiv (q', e') \in D} \left[p^*(q') - c(q', \theta) - e' + b^*(d') + \frac{\delta}{1-\delta} u^* \right], \frac{\delta}{1-\delta} u^* \right\}
\end{aligned}$$

and for $\chi(\theta, W) = \omega$,

$$\begin{aligned}
\frac{\delta}{1-\delta} u^* = \frac{\delta}{1-\delta} u_+(W, \omega) \; & \geq \; \sup_{d' \equiv (q', e') \in D} \left[p(q') - c(q', \theta) - e' + b(d') + \frac{\delta}{1-\delta} u_+(W, d', H, H) \right] \\
& = \; \sup_{d' \equiv (q', e') \in D} \left[p^*(q') - c(q', \theta) - e' + b^*(d') + \frac{\delta}{1-\delta} u \right] \\
& = \; \sup_{d' \equiv (q', e') \in D} \left[p^*(q') - c(q', \theta) - e' + b^*(d') + \frac{\delta}{1-\delta} u^* \right].
\end{aligned}$$

These inequalities are equivalent to (IC) under σ^*.

We next confirm (AHC′). By construction of $b^*(d)$,

$$\begin{aligned}
& b^*(d) + \frac{\delta}{1-\delta} u^* \\
= \; & b(d) + \frac{\delta}{1-\delta} \left[u_+(W, d, H, H) - u \right] - \inf_{d \in D} \left[b(d) + \frac{\delta}{1-\delta} u_+(W, d, H, H) \right] + \frac{\delta}{1-\delta} u^* \\
\geq \; & b(d) + \frac{\delta}{1-\delta} u_+(W, d, H, H) - \inf_{d \in D} \left[b(d) + \frac{\delta}{1-\delta} u_+(W, d, H, H) \right] \geq 0,
\end{aligned}$$

which implies (AHC′) for σ^*.

Finally, we confirm (PHC′). Note that

$$
\begin{aligned}
\pi^* &= (1-\delta)\int_{\theta:d(W,\theta)\in D}\left[y(\chi(\theta,W)) - p^*(q(\theta,W)) - b^*(\chi(\theta,W))\right] f(\theta)d\theta + \delta\pi^* \\
&= (1-\delta)\int_{\theta:d(W,\theta)\in D}\Big[y(\chi(\theta,W)) - p(q(\theta,W)) - b(\chi(\theta,W)) \\
&\quad -\frac{\delta}{1-\delta}\left[u_+(W,\chi(\theta,W),H,H) - u\right]\Big] f(\theta)d\theta + \delta\pi^* \\
\Longleftrightarrow \pi^* &= \int_{\theta:d(W,\theta)\in D}\Big[y(\chi(\theta,W)) - p(q(\theta,W)) - b(\chi(\theta,W)) \\
&\quad -\frac{\delta}{1-\delta}\left[u_+(W,\chi(\theta,W),H,H) - u\right]\Big] f(\theta)d\theta
\end{aligned}
$$

and

$$
\begin{aligned}
\pi &= \int_{\theta:d(W,\theta)\in D}\left[(1-\delta)\left(y(\chi(\theta,W)) - p(q(\theta,W)) - b(\chi(\theta,W))\right)\right. \\
&\quad \left. +\delta\pi_+(W,\chi(\theta,W),H,H)\right] f(\theta)d\theta + \delta\int_{\theta:d(W,\theta)\notin D}\pi_+(W,\omega)f(\theta)d\theta.
\end{aligned}
$$

Since Assumption 2.4 ensures that $\pi_+(W,\omega) = \pi$,

$$
\begin{aligned}
&\pi^* \\
= \;& \frac{1}{1-\delta}\pi - \int_{\theta:d(W,\theta)\in D}\frac{\delta}{1-\delta}\pi_+(W,\chi(\theta,W),H,H)f(\theta)d\theta \\
& -\frac{\delta}{1-\delta}\int_{\theta:d(W,\theta)\notin D}\pi_+(W,\omega)f(\theta)d\theta - \frac{\delta}{1-\delta}\int_{\theta:d(W,\theta)\in D}\left[u_+(W,\chi(\theta,W),H,H) - u\right] f(\theta)d\theta \\
= \;& \pi + \frac{\delta}{1-\delta}\left[\pi - \int_{\theta:d(W,\theta)\in D}\left(s_+(W,\chi(\theta,W),H,H) - u\right) f(\theta)d\theta - \int_{\theta:d(W,\theta)\notin D}\pi_+(W,\omega)f(\theta)d\theta\right] \\
= \;& \pi + \frac{\delta}{1-\delta}\left[\int_{\theta:d(W,\theta)\in D}\left(s - s_+(W,\chi(\theta,W),H,H)\right) f(\theta)d\theta\right].
\end{aligned}
$$

Note that by construction of σ, $s \geq s_+(W,\chi(\theta,W),H,H)$ for all θ. Thus we obtain $\pi^* \geq \pi$.

(PHC′) is satisfied if and only if

$$\begin{aligned}
& -b^{*}(d) + \frac{\delta}{1-\delta}\pi^{*} \\
= \quad & -b(d) - \frac{\delta}{1-\delta}\left[u_{+}(W,d,H,H) - u\right] + \inf_{d\in D}\left[b(d) + \frac{\delta}{1-\delta}u_{+}(W,d,H,H)\right] + \frac{\delta}{1-\delta}\pi^{*}
\end{aligned}$$

is no less than 0 for all d. It is equivalent to that the following is no less than 0;

$$\begin{aligned}
& -\sup_{d\in D}\left[b(d) + \frac{\delta}{1-\delta}u_{+}(W,d,H,H)\right] + \inf_{d\in D}\left[b(d) + \frac{\delta}{1-\delta}u_{+}(W,d,H,H)\right] + \frac{\delta}{1-\delta}[\pi^{*} + u] \\
= \quad & -\sup_{d\in D}\left[b(d) + \frac{\delta}{1-\delta}u_{+}(W,d,H,H)\right] + \inf_{d\in D}\left[b(d) + \frac{\delta}{1-\delta}u_{+}(W,d,H,H)\right] + \frac{\delta}{1-\delta}s^{*} \geq 0 \\
\iff \quad & \frac{\delta}{1-\delta}s^{*} \geq \sup_{d\in D}\left[b(d) + \frac{\delta}{1-\delta}u_{+}(W,d,H,H)\right] - \inf_{d\in D}\left[b(d) + \frac{\delta}{1-\delta}u_{+}(W,d,H,H)\right]. \quad (2.18)
\end{aligned}$$

Recall from (PHC′) and (AHC′) under σ that

$$b(d) + \frac{\delta}{1-\delta}u_{+}(W,d,H,H) \geq \frac{\delta}{1-\delta}u_{+}(W,d,H,R) \geq 0$$

implying that

$$-\inf_{d\in D}\left[b(d) + \frac{\delta}{1-\delta}u_{+}(W,d,H,H)\right] \leq 0$$

and

$$\begin{aligned}
\sup_{d\in D}\left[b(d) + \frac{\delta}{1-\delta}u_{+}(W,d,H,H)\right] &= \sup_{d\in D}\left[b(d) + \frac{\delta}{1-\delta}\left(s_{+}(W,d,H,H) - \pi_{+}(W,d,H,H)\right)\right] \\
&\leq \sup_{d\in D}\left[b(d) + \frac{\delta}{1-\delta}\left(s^{*} - \pi_{+}(W,d,H,H)\right)\right] \\
&= -\inf_{d\in D}\left[-b(d) + \frac{\delta}{1-\delta}\pi_{+}(W,d,H,H)\right] + \frac{\delta}{1-\delta}s^{*} \\
&\leq -\frac{\delta}{1-\delta}\pi_{+}(W,d,R,H) + \frac{\delta}{1-\delta}s^{*} \leq \frac{\delta}{1-\delta}s^{*}
\end{aligned}$$

Summing these inequalities implies (2.18).

2.9.2 Proof of Lemma 2.1 and 2.2

Let Θ^* be the set of types for which $d^*(\theta) \in D$, i.e. the set of types who accept the mechanism in the OPPE, and $\Theta^{*C} \equiv [\underline{\theta}, \overline{\theta}] \setminus \Theta^*$. We first establish Lemma 2.2 by showing that given $\Theta^* = [\underline{\theta}, \overline{\theta}]$, the equilibrium can be characterized by the optimization problem in Lemma 2.2. We next establish Lemma 2.1 by showing the equilibrium must satisfy either $\Theta^* = [\underline{\theta}, \overline{\theta}]$ or $\Theta^* = \emptyset$.

Proof of Lemma 2.2 Note that because of (2.1),

$$u^* = U^*(\overline{\theta}) + \int_{\underline{\theta}}^{\overline{\theta}} c_\theta(q^*(\theta), \theta)F(\theta)d\theta$$

and then (2.4) can be written as

$$\begin{aligned} & \frac{\delta}{1-\delta}\left[\int_{\underline{\theta}}^{\overline{\theta}} J(q^*(\theta), e^*(\theta), \theta)f(\theta)d\theta - U^*(\overline{\theta})\right] \\ \geq \;& b^*(q,e) \geq -\frac{\delta}{1-\delta}\left[U^*(\overline{\theta}) + \int_{\underline{\theta}}^{\overline{\theta}} c_\theta(q^*(\theta), \theta)F(\theta)d\theta\right]. \end{aligned} \tag{2.19}$$

Now $b^*(q, e)$ appears only in constraints (2.2) and (2.19). By observing (2.2), we see that (2.2) is relaxed by increasing $b^*(q^*(\theta'), e^*(\theta'))$ and by decreasing $b^*(q^*(\theta'), e')$. It implies that without loss of generality

$$b^*(q^*(\theta), e) = \begin{cases} \frac{\delta}{1-\delta}\left[\int_{\underline{\theta}}^{\overline{\theta}} J(q^*(z), e^*(z), z)f(z)dz - U^*(\overline{\theta})\right] & \text{if } e = e^*(\theta) \\ -\frac{\delta}{1-\delta}\left[U^*(\overline{\theta}) + \int_{\underline{\theta}}^{\overline{\theta}} c_\theta(q^*(z), z)F(z)dz\right] & \text{if } e = e' \end{cases} \tag{2.20}$$

and then (2.2) becomes

$$\begin{aligned}
&\frac{\delta}{1-\delta}\int_{\underline{\theta}}^{\overline{\theta}} s(q^*(z),e^*(z),z)f(z)dz + \int_{\theta}^{\theta'} c_\theta(q^*(x),x)dx - c(q^*(\theta'),\theta') + c(q^*(\theta'),\theta) \\
&\geq e^*(\theta') - e' \\
\iff \quad &\frac{\delta}{1-\delta}\int_{\underline{\theta}}^{\overline{\theta}} s(q^*(z),e^*(z),z)f(z)dz + \int_{\theta}^{\theta'} \left[c_\theta(q^*(x),x) - c_\theta(q^*(\theta'),x)\right] dx \\
&\geq e^*(\theta') - e'. \qquad (2.21)
\end{aligned}$$

Now (2.19) is redundant and $U^*(\theta)$ appears only in the objective and (2.3). Then (2.3) is obviously binding; $U^*(\overline{\theta}) = 0$. Thus the rest of the constraints are (2.21) and monotonicity of $q^*(\theta)$.

The objective function shows that $e^*(\theta')$ must be $\overline{e}$ as long as it is feasible. If $e^*(\theta') = \overline{e}$, then the right hand side is $\overline{e}$. However recall that given $\theta' \in [\underline{\theta},\overline{\theta}]$ fixed, (2.21) must be satisfied for any $\theta \in [\underline{\theta},\overline{\theta}]$. Now the following lemma is useful.

Lemma 2.9 *Suppose that $q^*(\theta)$ is non-increasing in θ. Then*

$$\min_{\theta,\theta' \in [\underline{\theta},\overline{\theta}]} \left[\int_{\theta}^{\theta'} \left[c_\theta(q^*(x),x) - c_\theta(q^*(\theta'),x)\right] dx\right] = 0.$$

Proof (Lemma 2.9) *Suppose that $q^*(\theta)$ is non-increasing in θ. Given arbitrary $\theta' \in [\underline{\theta},\overline{\theta}]$ fixed,*

$$\frac{d}{d\theta}\left[\int_{\theta}^{\theta'} \left[c_\theta(q^*(z),z) - c_\theta(q^*(\theta'),z)\right] dz\right] = -c_\theta(q^*(\theta),\theta) + c_\theta(q^*(\theta'),\theta) \begin{cases} \geq 0 & \text{if } \theta > \theta' \\ = 0 & \text{if } \theta = \theta' \\ \leq 0 & \text{if } \theta < \theta' \end{cases}$$

meaning that the term $\int_{\theta}^{\theta'}[c_\theta(q^*(z),z) - c_\theta(q^*(\theta'),z)]dz$ is non-increasing in $\theta < \theta'$ and non-decreasing in $\theta > \theta'$. This implies that it attains the minimum at $\theta = \theta'$ the value of which is *obviously zero.* $\square$

Lemma 2.9 implies that $e^*(\theta') = \bar{e}$ if and only if

$$\frac{\delta}{1-\delta}\int_{\underline{\theta}}^{\overline{\theta}} s(q^*(z), e^*(z), z) f(z) dz \geq \bar{e},$$

which is independent from θ'. This further implies that if $e^*(\theta) = \bar{e}$ for some θ, then $e^*(\theta) = \bar{e}$ for all θ. If $e^*(\theta) = 0$ for all θ, then Assumption 2.1 implies that $\pi^* < 0$ and P should choose to offer no mechanism. Thus the equilibrium in which some mechanism is offered must satisfy that $e^*(\theta) = \bar{e}$ for all θ and the condition becomes (2.6).

Proof of Lemma 2.1 Note that $U^*(\theta) = 0$ for all $\theta \in \Theta^{*C}$. If Θ^* has zero probability measure, then π^* is also 0. In this case, to offer no mechanism is indifferent. Thus without loss of generality, suppose that Θ^* has positive probability measure.

Furthermore, suppose that for $\theta < \theta'$, $\theta' \in \Theta^*$ and $\theta \in \Theta^{*C}$. For type θ', there is no incentive to mimic θ. It is satisfied if and only if

$$U^*(\theta') \geq 0.$$

Conversely, type θ has no incentive to mimic θ' if and only if

$$\begin{aligned} 0 &\geq w^*(q^*(\theta'), e^*(\theta')) - c(q^*(\theta'), \theta) - e^*(\theta') \\ &= U^*(\theta') + c(q^*(\theta'), \theta') - c(q^*(\theta'), \theta). \end{aligned}$$

Combining them implies that $c(q^*(\theta'), \theta') \leq c(q^*(\theta'), \theta)$, which contradicts $\theta < \theta'$. It implies that if $\theta \in \Theta^{*C}$, then $\theta' \in \Theta^{*C}$ for all $\theta' > \theta$. It also means that there exists $\hat{\theta} \in [\underline{\theta}, \overline{\theta}]$ such that $[\underline{\theta}, \hat{\theta}) \subset \Theta^*$ and $(\hat{\theta}, \overline{\theta}] \subset \Theta^{*C}$.

Given $\hat{\theta}$ fixed, the equilibrium conditions can be characterized in the same way as in the proof of Lemma 2.2 by replacing $\overline{\theta}$ with $\hat{\theta}$. Then it can be characterized by the following optimization problem;

$$\max \int_{\underline{\theta}}^{\hat{\theta}} J(q^*(z), \overline{e}, z) f(z) dz \quad \text{subject to} \quad \frac{\delta}{1-\delta} \int_{\underline{\theta}}^{\hat{\theta}} s(q^*(z), \overline{e}, z) f(z) dz \geq \overline{e}$$

and monotonicity of $q^*(\theta)$. The Euler equation provides the characterization of the equilibrium as follows.

1. If

$$\frac{\delta}{1-\delta} \int_{\underline{\theta}}^{\hat{\theta}} s(q^{SB}(z), \overline{e}, z) f(z) dz \geq \overline{e} \iff \delta \geq \overline{\delta}(\hat{\theta}) \equiv \frac{\overline{e}}{\overline{e} + \int_{\underline{\theta}}^{\hat{\theta}} s(q^{SB}(z), \overline{e}, z) f(z) dz},$$

then the equilibrium satisfies that $d^*(\theta) = (q^{SB}(\theta), \overline{e})$ for $\theta > \hat{\theta}$.

2. If $\delta < \overline{\delta}(\hat{\theta})$ and

$$\frac{\delta}{1-\delta} \int_{\underline{\theta}}^{\hat{\theta}} s(q^{FB}(z), \overline{e}, z) f(z) dz > \overline{e} \iff \delta > \underline{\delta}(\hat{\theta}) \equiv \frac{\overline{e}}{\overline{e} + \int_{\underline{\theta}}^{\hat{\theta}} s(q^{FB}(z), \overline{e}, z) f(z) dz},$$

then the equilibrium satisfies that

$$\begin{aligned} s_q(q^*(\theta), \overline{e}, \theta) &= \frac{1}{1 + \frac{\delta}{1-\delta} \lambda^*(\hat{\theta})} \frac{F(\theta)}{f(\theta)} c_{q\theta}(q^*(\theta), \theta), \\ \frac{\delta}{1-\delta} \int_{\underline{\theta}}^{\hat{\theta}} s(q^*(z), \overline{e}, z) f(z) dz &= \overline{e}, \end{aligned}$$

and $e^*(\theta) = \overline{e}$ with some $\lambda^*(\hat{\theta}) > 0$.

3. For $\delta = \underline{\delta}(\hat{\theta})$, $q^*(\theta) = q^{FB}(\theta)$ and $e^*(\theta) = \overline{e}$.

4. For $\delta < \underline{\delta}(\hat{\theta})$, $W^* = \phi$.

Note that the value of P's payoff is positive if $\delta \geq \underline{\delta}(\hat{\theta})$ while it is zero if $\delta < \underline{\delta}(\hat{\theta})$.

Now fix $\delta \in [0, 1)$ and suppose that $\hat{\theta} < \overline{\theta}$. In the following, we show that P's payoff is weakly increasing in $\hat{\theta}$, implying that choosing $\hat{\theta} < \overline{\theta}$ is never optimal.

First, suppose that $\delta < \underline{\delta}(\hat{\theta})$. Note that $\underline{\delta}(\hat{\theta})$ is decreasing in $\hat{\theta}$. When $\delta < \underline{\delta}(\overline{\theta})$, P's payoff is still 0 since $\delta < \underline{\delta}(\theta)$ for any θ. However, when $\delta \geq \underline{\delta}(\overline{\theta})$, it become positive for $\hat{\hat{\theta}}$ such that $\delta \geq \underline{\delta}(\hat{\hat{\theta}})$. Then choosing $\hat{\theta} < \overline{\theta}$ is weakly dominated by another $\hat{\hat{\theta}}$ such that $\delta \geq \underline{\delta}(\hat{\hat{\theta}})$.

Second, suppose that $\delta = \underline{\delta}(\hat{\theta})$. Then $q^*(\theta) = q^{FB}(\theta)$. Now consider $\hat{\hat{\theta}}$ which is slightly greater than $\hat{\theta}$. Since

$$\begin{aligned} \frac{d}{d\theta}\int_{\underline{\theta}}^{\theta} s(q^{FB}(z), \overline{e}, z)f(z)dz &= s(q^{FB}(\hat{\theta}), \overline{e}, \hat{\theta})f(\hat{\theta}) > 0 \\ \frac{d}{d\theta}\int_{\underline{\theta}}^{\theta} J(q^{FB}(z), \overline{e}, z)f(z)dz &= s(q^{FB}(\hat{\theta}), \overline{e}, \hat{\theta})f(\hat{\theta}) > 0 \end{aligned}$$

for all $\theta \in [\hat{\theta}, \hat{\hat{\theta}}]$,

$$\int_{\underline{\theta}}^{\hat{\hat{\theta}}} J(q^{FB}(z), \overline{e}, z)f(z)dz > \int_{\underline{\theta}}^{\hat{\theta}} J(q^{FB}(z), \overline{e}, z)f(z)dz$$

and

$$\frac{\delta}{1-\delta}\int_{\underline{\theta}}^{\hat{\hat{\theta}}} s(q^{FB}(z), \overline{e}, z)f(z)dz \geq \overline{e}$$

those imply that by changing $\hat{\theta}$ to $\hat{\hat{\theta}}$, P can achieve the payoff greater than the optimal value under $\hat{\theta}$.

Suppose next that $\delta \in (\underline{\delta}(\hat{\theta}), \overline{\delta}(\hat{\theta}))$. In this case, $q^*(\theta)$ satisfies the first order condition of the following Lagrangian;

$$\mathcal{L}(\hat{\theta}) = \int_{\underline{\theta}}^{\hat{\theta}} J(q^*(z), \overline{e}, z) f(z) dz + \lambda^*(\hat{\theta}) \left[\frac{\delta}{1-\delta} \int_{\underline{\theta}}^{\hat{\theta}} s(q^*(z), \overline{e}, z) f(z) dz - \overline{e} \right].$$

Note that $\int_{\underline{\theta}}^{\hat{\theta}} J(q^*(z), \overline{e}, z) f(z) dz = \mathcal{L}(\hat{\theta})$ in the neighbourhood of $\hat{\theta}$ and the envelope theorem implies that

$$\begin{aligned} & \frac{d}{d\hat{\theta}} \left[\int_{\underline{\theta}}^{\hat{\theta}} J(q^*(z), \overline{e}, z) f(z) dz \right] = \mathcal{L}'(\hat{\theta}) = \frac{\partial \mathcal{L}}{\partial \hat{\theta}}(\hat{\theta}) \\ = \; & J(q^*(\hat{\theta}), \overline{e}, \hat{\theta}) f(\hat{\theta}) + \lambda^*(\hat{\theta}) \frac{\delta}{1-\delta} s(q^*(\hat{\theta}), \overline{e}, \hat{\theta}) f(\hat{\theta}) > 0 \end{aligned}$$

meaning that slightly increasing $\hat{\theta}$ improves the objective function. Then such $\hat{\theta} < \overline{\theta}$ is never optimal.

Finally, suppose that $\delta \geq \overline{\delta}(\hat{\theta})$. Since $\overline{\delta}(\hat{\theta})$ is decreasing in $\hat{\theta}$, when $\hat{\theta}$ is increasing, the value of the objective function is still $\int_{\underline{\theta}}^{\hat{\theta}} J(q^{SB}(z), \overline{e}, z) f(z) dz$ and it is increasing in $\hat{\theta}$. Then slightly increasing $\hat{\theta}$ improves the objective functions. Then again such $\hat{\theta} < \overline{\theta}$ is never optimal.

So far we have shown that $\Theta^* = [\underline{\theta}, \overline{\theta}]$ or $[\underline{\theta}, \overline{\theta})$. In the case of $\Theta^* = [\underline{\theta}, \overline{\theta})$, the optimization problem achieves P's payoff as high as in the case of $\Theta^* = [\underline{\theta}, \overline{\theta}]$ demonstrated in Section 2.4. Thus both cases are indifferent for P and it is without loss of generality to focus on the case of $\Theta^* = [\underline{\theta}, \overline{\theta}]$.

2.9.3 Proof of Corollary 2.1

It is enough to demonstrate that the total surplus is decreasing and P's expected payoff is increasing in $\delta \in [\underline{\delta}, \overline{\delta})$. For $\delta \in [\underline{\delta}, \overline{\delta})$, $q^*(\theta)$ satisfies that

$$\int_{\underline{\theta}}^{\overline{\theta}} s(q^*(\theta), \overline{e}, \theta) f(\theta) d\theta = \frac{1-\delta}{\delta} \overline{e}.$$

The left hand side is the total surplus and the right hand side is decreasing in δ. On P's payoff, since it is characterized by the Lagrangian and the constraint is binding, the envelope theorem implies that

$$\frac{d}{d\delta} \int_{\underline{\theta}}^{\overline{\theta}} J(q^*(\theta), \overline{e}, \theta) f(\theta) d\theta = \frac{d\mathcal{L}}{d\delta} = \frac{\partial \mathcal{L}}{\partial \delta} = \frac{1}{(1-\delta)^2} \int_{\underline{\theta}}^{\overline{\theta}} s(q^*(\theta), \overline{e}, \theta) f(\theta) d\theta > 0.$$

2.9.4 Proof of Proposition 2.3

Note that $b^*(q^*(\theta), e)$ appears only in (2.2). Then by letting $b^*(q^*(\theta), e) = e$ for all θ and e, Lemma 2.9 implies that (2.2) is satisfied. Since $U^*(\theta)$ appears only in (2.5) and (2.3), (2.3) is obviously binding; $U^*(\overline{\theta}) = 0$. Then we obtain the problem that maximizes $\int J(q(\theta), e^*(\theta), \theta) d\theta$ subject to monotonicity of $q^*(\theta)$. By Assumption 2.1, the solution is $e^*(\theta) = \overline{e}$ and $q^*(\theta) = q^{SB}(\theta)$ for all θ.

2.9.5 Proof of Lemma 2.3

By substituting $U^C(\theta)$ and $b^C(q^C(\theta), e)$, (2.11) is written as

$$\int_{\underline{\theta}}^{\overline{\theta}} J(q^C(\theta), e^C, \theta) f(\theta) d\theta \geq \int_{\underline{\theta}}^{\overline{\theta}} [J(q^C(\theta), e^C, \theta) - b(q^C(\theta), e^C) + b(q^C(\theta), e')] f(\theta) d\theta.$$

Now $b^C(q^C(\theta), e)$ appears only in this inequality and (2.4). Then by substituting $b^C(q^C(\theta), e) = 0$ for any θ and e, these inequalities are satisfied. Since $U^C(\theta)$ appears only in (2.5) and (2.3), (2.3) is obviously binding; $U^C(\overline{\theta}) = 0$. Then we obtain the problem that maximizes $\int J(q(\theta), e^C, \theta) f(\theta) d\theta$ subject to monotonicity of $q^C(\theta)$. Assumption 2.1 implies that $e^C = \bar{e}$.

2.9.6 Proof of Lemma 2.4

By using the notation $U^{1S}(\theta) = p^S(\theta) - c(q^S(\theta), \theta)$ and $u^{2S} = b^S(e^S) - e^S$, the optimization problem for characterizing the OPPE under task separation is as follows;

$$\max \int_{\underline{\theta}}^{\overline{\theta}} \left[s(q^S(\theta), e^S, \theta) - U^{1S}(\theta) - u^{2S} \right] f(\theta) d\theta$$

subject to (2.1), (2.3), (2.12), (2.13), (2.14), and monotonicity of $q^S(\theta)$. If $e^S = 0$, the the objective is less than 0 meaning that P prefers to abstain from offering any mechanism. Then we focus on $e^S = \bar{e}$. Now (2.12) and (2.13) are reduced to $u^{2s} \geq \max\{b^S(0), 0\}$ and since $b^S(0)$ appears only in this and (2.14), it should be lower as long as it is possible. It implies that $b^S(0) = -\delta u^{2S}/(1-\delta)$ and $u^{2S} \geq 0$. Since u^{2S} should also be lower, it implies that $u^{2S} = 0$ (and then $b^S(0) = 0$). Substituting (2.1) shows that (2.3) is binding and the objective becomes

$$\int_{\underline{\theta}}^{\overline{\theta}} J(q^S(\theta), \bar{e}, \theta) f(\theta) d\theta$$

and since $b^S(\bar{e}) = u^{2S} + \bar{e} = \bar{e}$, (2.14) becomes (2.15).

$$\frac{\delta}{1-\delta} \int_{\underline{\theta}}^{\overline{\theta}} J(q^S(\theta), \bar{e}, \theta) f(\theta) d\theta \geq \bar{e}.$$

2.9.7 Proof of Proposition 2.5

Since $q^{SB}(\theta)$ is monotone and maximizes the objective, it is the solution if it is feasible, i.e.,

$$\frac{\delta}{1-\delta}\int_{\underline{\theta}}^{\overline{\theta}} J(q^{SB}(\theta),\bar{e},\theta)f(\theta)d\theta \geq \bar{e} \iff \delta \geq \underline{\delta}^{S} \equiv \frac{\bar{e}}{\bar{e}+\int_{\underline{\theta}}^{\overline{\theta}} J(q^{SB}(\theta),\bar{e},\theta)f(\theta)d\theta}.$$

If it is not feasible, then there is no feasible $q^{S}(\theta)$ and then P chooses not to offer any mechanism.

2.10 Appendix: The Optimal Contract without Assumption of Limited Commitment Ability

In this section, we characterize the optimal contract when both Assumption 2.3 and 2.4 are dropped.

Conditions for PPE: Without loss of generality, we can focus on the equilibrium in which both P and A honour informal agreements on the equilibrium path.[45] Let $(p^*(\cdot), b^*(\cdot,\cdot))$ be the equilibrium mechanism in period 0, $d^*(\theta) \equiv (q^*(\theta), e^*(\theta)) \in D$ be the agent's equilibrium decision in period 0 after he accepts $(p^*(\cdot), b^*(\cdot))$, and $(\pi^*_+(p,b,x), u^*_+(p,b,x))$ be the continuation payoff given mechanism (p,b) and decision $x \in \overline{D}$ and the informal agreement in period 0 honoured if $x \in D$. Denote $\Theta^* \equiv [\theta \in [\underline{\theta},\overline{\theta}] \mid d^*(\theta) \in D]$. Note that if either $p \neq p^*$, $b \neq b^*$, or $x \neq (q(\theta), e(\theta))$ for any θ, then those are observable deviations and without loss of generality $(\pi^*_+(p,b,x), u^*_+(p,b,x)) = (0,0)$ by Nash reversion due to Lemma 2.6.

Suppose that a strategy induces $(p^*(\cdot), b^*(\cdot,\cdot), \Theta^*, d^*(\cdot), \pi^*_+(\cdot,\cdot,\cdot), u^*_+(\cdot,\cdot,\cdot))$ and let $U^*(\theta)$ be A's average payoff when the type is θ in period 0. Then the one-shot deviation principle

[45] The proof of Lemma 2.7 is directly applicable.

implies that it is a PPE if and only if P's average payoff is greater or equal to 0,

$$
\begin{aligned}
{}^\forall\theta, \theta' \in \Theta^*, \quad & U^*(\theta) \geq U^*(\theta') + (1-\delta)[c(q^*(\theta'), \theta') - c(q^*(\theta'), \theta)], && (2.22)\\
{}^\forall\theta, \in \Theta^*, \theta' \notin \Theta^*, \quad & U^*(\theta) \geq \delta u^*_+(p^*, b^*, \omega) \\
& \geq U^*(\theta) + (1-\delta)[c(q^*(\theta), \theta) - c(q^*(\theta), \theta')], && (2.23)\\
{}^\forall\theta, \theta' \in \Theta^*, \quad & U^*(\theta) \geq U^*(\theta') + (1-\delta)\cdot \\
& [c(q^*(\theta'), \theta') - c(q^*(\theta'), \theta) + e^*(\theta') - e' - b^*(q^*(\theta'), e^*(\theta')) + b^*(q^*(\theta'), e')] \\
& -\delta u^*_+(p^*, b^*, d^*(\theta')), && (2.24)\\
& \frac{\delta}{1-\delta}\pi^*_+(p^*, b^*, q, e) \geq b^*(q, e) \geq -\frac{\delta}{1-\delta}u^*_+(p^*, b^*, q, e), && (2.25)
\end{aligned}
$$

and $(\pi^*_+(p, b, x), u^*_+(p, b, x))$ must be a PPE payoff vector for any (p, b, x).

We can show that Θ^* is an interval including $\underline{\theta}$ or an empty set by the similar proof of Lemma 2.2. Then from the envelope theorem (2.22) and (2.23) are equivalent to that $q^*(\theta)$ is non-increasing in $\theta \in \Theta^*$,

$$
\begin{aligned}
{}^\forall\theta \in \Theta^*, \quad & U^*(\theta) = U^*(\hat{\theta}) + (1-\delta)\int_\theta^{\hat{\theta}} c_\theta(q^*(z), z)dz, && (2.26)\\
\text{if } \Theta^* = [\underline{\theta}, \overline{\theta}], \quad & U^*(\overline{\theta}) \geq \delta u^*_+(p^*, b^*, \omega), && (2.27)\\
\text{if } \Theta^* \neq [\underline{\theta}, \overline{\theta}], \quad & {}^\forall\theta' \notin \Theta^*,\ U^*(\theta') = U^*(\hat{\theta}) = \delta u^*_+(p^*, b^*, \omega) && (2.28)
\end{aligned}
$$

where $\hat{\theta} \equiv \inf \Theta^*$.

Socially Optimal PPE: First, suppose that the PPE achieves the joint maximum payoff. Then it must maximize the following aggregate average payoff subject to the above constraints;

$$\overline{s}^{*} \equiv \int_{\underline{\theta}}^{\hat{\theta}} (1-\delta)s(d^{*}(\theta),\theta)f(\theta)d\theta + \delta\int_{\underline{\theta}}^{\hat{\theta}} s_{+}^{*}(p^{*},b^{*},d^{*}(\theta))f(\theta)d\theta + \delta(1-F(\hat{\theta}))s_{+}^{*}(p^{*},b^{*},\omega) \tag{2.29}$$

where $s_{+}^{*}(p^{*},b^{*},d^{*}(\theta)) \equiv \pi_{+}^{*}(p^{*},b^{*},d^{*}(\theta)) + u_{+}^{*}(p^{*},b^{*},d^{*}(\theta))$. Since $b^{*}(q,e)$ appears only in (2.24) and (2.25), without loss of generality

$$b^{*}(q^{*}(\theta),e) = \begin{cases} \dfrac{\delta}{1-\delta}\pi_{+}^{*}(p^{*},b^{*},q^{*}(\theta),e) & \text{if } e = e^{*}(\theta) \\ -\dfrac{\delta}{1-\delta}u_{+}^{*}(p^{*},b^{*},q^{*}(\theta),e) & \text{if } e \neq e^{*}(\theta). \end{cases} \tag{2.30}$$

Furthermore, with (2.26) and $u_{+}^{*}(p^{*},b^{*},q^{*}(\theta'),e') = 0$, (2.24) becomes

$$\frac{\delta}{1-\delta}s_{+}^{*}(p^{*},b^{*},q^{*}(\theta'),e^{*}(\theta')) + \int_{\theta}^{\theta'} \left[c_{\theta}(q^{*}(z),z) - c_{\theta}(q^{*}(\theta'),z)\right] dz \geq e^{*}(\theta') - e'$$

for all $\theta,\theta' \in \Theta^{*}$ and $e' \neq e^{*}(\theta)$. From Lemma 2.9, it is further simplified to

$$\frac{\delta}{1-\delta}s_{+}^{*}(p^{*},b^{*},q^{*}(\theta),e^{*}(\theta)) \geq e^{*}(\theta') - e' \tag{2.31}$$

for all $\theta \in \Theta^{*}$ and $e' \neq e^{*}(\theta)$. Increasing $s_{+}^{*}(p^{*},b^{*},d^{*}(\theta))$ improves the objective and relaxes the constraints. Then $s_{+}^{*}(p^{*},b^{*},d^{*}(\theta))$ should be the maximum joint PPE payoff for any $\theta \in \Theta^{*}$.

Suppose that

$$\frac{\delta}{1-\delta}E_{z}[s(q^{FB}(z),\bar{e},z)] < \bar{e}$$

or equivalently $\delta < \underline{\delta}$. Then for all $\theta \in \Theta^{*}$, $e^{*}(\theta) = 0$. It implies that if $\Theta^{*} \neq \emptyset$, then from

(2.29)

$$\overline{s}^* < \delta F(\hat{\theta})\overline{s}^* + \delta(1 - F(\hat{\theta}))s^*_+(p^*, b^*, \omega) \iff \overline{s}^* < \frac{\delta(1 - F(\hat{\theta}))}{1 - \delta F(\hat{\theta})} s^*_+(p^*, b^*, \omega) \le \frac{\delta(1 - F(\hat{\theta}))}{1 - \delta F(\hat{\theta})}\overline{s}^* \le \overline{s}^*,$$

which is a contradiction. Then $\Theta^* = \emptyset$ and from (2.29)

$$\overline{s}^* = \delta s^*_+(p^*, b^*, \omega) \le \delta \overline{s}^*,$$

which implies that $\overline{s}^* = 0$. Conversely, suppose that

$$\frac{\delta}{1 - \delta} E_z[s(q^{FB}(z), \overline{e}, z)] \ge \overline{e}$$

or equivalently $\delta \ge \underline{\delta}$. If a PPE implements $(q^{FB}(\theta), \overline{e})$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$ every period, then (2.31) implies that $\delta \ge \underline{\delta}$. Now we construct a stationary PPE to implement the first best every period and check the equilibrium condition for $\delta \ge \underline{\delta}$. Consider the following stationary mechanism; in each period P offers a mechanism such that

$$\begin{aligned} p(q^{FB}(\theta)) &= c(q^{FB}(\theta), \theta) + \int_{\theta}^{\overline{\theta}} c_\theta(q^{FB}(z), z)dz + \frac{\delta}{1 - \delta} \int_{\underline{\theta}}^{\overline{\theta}} c_\theta(q^{FB}(\theta), \theta)F(\theta)d\theta, \\ b^*(q^{FB}(\theta), e) &= e - \frac{\delta}{1 - \delta} \int_{\underline{\theta}}^{\overline{\theta}} c_\theta(q^{FB}(\theta), \theta)F(\theta)d\theta \end{aligned}$$

and if A rejects it, then the players start the Nash reversion from the next period (i.e., $u^*_+(p^*, b^*, \omega) = 0$). This mechanism generates that for all $\theta \in [\underline{\theta}, \overline{\theta}]$

$$U^*(\theta) = (1 - \delta) \int_{\theta}^{\overline{\theta}} c_\theta(q^{FB}(z), z)dz + \delta \int_{\underline{\theta}}^{\overline{\theta}} c_\theta(q^{FB}(z), z)F(z)dz,$$

$$u^*_+(p^*, b^*, q^{FB}(\theta), \overline{e}) = E_z[U^*(z)] = \int_{\underline{\theta}}^{\overline{\theta}} c_\theta(q^{FB}(z), z)F(z)dz,$$

$$\pi^*_+(p^*, b^*, q^{FB}(\theta), \bar{e}) = \int_{\underline{\theta}}^{\overline{\theta}} J(q^{FB}(z), \bar{e}, z) f(z) dz \geq 0$$

and it is easy to see that (2.24), (2.25), (2.26), and (2.27) are satisfied. Finally, since the mechanism is stationary, (2.24), (2.25), (2.26), and (2.27) assure that the players have no incentive to deviate at any history. Thus the stationary mechanism is a PPE.

Therefore the joint optimal PPE payoff is given as

$$\bar{s}^* = \begin{cases} E_\theta[s(q^{FB}(\theta), \bar{e}, \theta)] & \text{if } \delta \geq \underline{\delta} \\ 0 & \text{if } \delta < \underline{\delta}. \end{cases}$$

P's **Optimal PPE:** Now suppose that the PPE achieves *P*'s optimal average payoff. Then it must maximize the following function subject to the PPE constraints;

$$\int_{\underline{\theta}}^{\hat{\theta}} (1-\delta) \left[y(q^*(\theta), e^*(\theta)) - w^*(q^*(\theta), e^*(\theta)) \right] f(\theta) d\theta$$
$$+\delta \left[\int_{\underline{\theta}}^{\hat{\theta}} \pi^*_+(p^*, b^*, d^*(\theta)) f(\theta) d\theta + (1 - F(\hat{\theta})) \pi^*_+(p^*, b^*, \omega) \right]$$

where $w^*(q^*(\theta), e^*(\theta)) = p^*(q^*(\theta)) + b^*(q^*(\theta), e^*(\theta))$. By replacing w^* with U^*, *P*'s average payoff can be rewritten as

$$\begin{aligned} & \int_{\underline{\theta}}^{\hat{\theta}} (1-\delta) \left[y(q^*(\theta), e^*(\theta)) - c(q^*(\theta), \theta) - e^*(\theta) - \frac{1}{1-\delta} U^*(\theta) + \frac{\delta}{1-\delta} u^*_+(p^*, b^*, d^*(\theta)) \right] f(\theta) d\theta \\ & +\delta \int_{\underline{\theta}}^{\hat{\theta}} \pi^*_+(p^*, b^*, d^*(\theta)) f(\theta) d\theta + \delta(1 - F(\hat{\theta})) \pi^*_+(p^*, b^*, \omega) \\ = \ & (1-\delta) \int_{\underline{\theta}}^{\hat{\theta}} J(d^*(\theta), \theta) f(\theta) d\theta - F(\hat{\theta}) U^*(\hat{\theta}) \\ & +\delta \left[\int_{\underline{\theta}}^{\hat{\theta}} s^*_+(p^*, b^*, d^*(\theta)) f(\theta) d\theta + (1 - F(\hat{\theta})) \pi^*_+(p^*, b^*, \omega) \right]. \end{aligned} \tag{2.32}$$

Since $b^*(q,e)$ appears only in (2.24) and (2.25), without loss of generality

$$b^*(q^*(\theta),e) = \begin{cases} \dfrac{\delta}{1-\delta}\pi_+^*(p^*,b^*,q^*(\theta),e) & \text{if } e = e^*(\theta) \\ -\dfrac{\delta}{1-\delta}u_+^*(p^*,b^*,q^*(\theta),e) & \text{if } e \neq e^*(\theta). \end{cases} \tag{2.33}$$

Furthermore, with (2.26) and $u_+^*(p^*,b^*,q^*(\theta'),e') = 0$, (2.24) becomes

$$\frac{\delta}{1-\delta}s_+^*(p^*,b^*,q^*(\theta'),e^*(\theta')) + \int_\theta^{\theta'} \left[c_\theta(q^*(z),z) - c_\theta(q^*(\theta'),z)\right] dz \geq e^*(\theta') - e'$$

for all $\theta, \theta' \in \Theta^*$ and $e' \neq e^*(\theta)$. From Lemma 2.9, it is further simplified to (2.31) for all $\theta \in \Theta^*$ and $e' \neq e^*(\theta)$. Since increasing $s_+^*(p^*,b^*,d^*(\theta))$ improves the objective and relaxes the constraints, $s_+^*(p^*,b^*,d^*(\theta))$ should be the maximum joint PPE payoff for any $\theta \in \Theta^*$. Furthermore it is easy to see that regardless of Θ^*, (2.27) or (2.28) is binding. Those imply that (2.32) becomes

$$(1-\delta)\int_{\underline{\theta}}^{\hat{\theta}} J(d^*(\theta),\theta)f(\theta)d\theta + \delta\left[F(\hat{\theta})\left(\overline{s}^* - s_+^*(p^*,b^*,\omega)\right) + \pi_+^*(p^*,b^*,\omega)\right]$$

and the relevant constraints are monotonicity of $q^*(\theta)$ and (2.31) where the latter is rewritten by

$$\frac{\delta}{1-\delta}\overline{s}^* \geq e^*(\theta) - e'$$

for all $\theta \in \Theta^*$ and $e' \neq e^*(\theta)$.

Suppose that $\delta < \underline{\delta}$. Then $\overline{s}^* = 0$ and (2.31) implies that $e^*(\theta) = 0$ for all $\theta \in \Theta^*$. Since $\overline{s}^* \geq s_+^*(p^*,b^*,\omega)$, the objective is less than $\delta\pi_+^*(p^*,b^*,\omega)$. If $\pi_+^*(p^*,b^*,\omega) < 0$, then P can improve her average payoff by offering no mechanism after every history. Then

$\pi_+^*(p^*, b^*, \omega) \geq 0$. However, if $\pi_+^*(p^*, b^*, \omega) > 0$, then by starting the continuation strategy from period 0, she can obtain average payoff $\pi_+^*(p^*, b^*, \omega) > \delta\pi_+^*(p^*, b^*, \omega)$. Therefore $\pi_+^*(p^*, b^*, \omega) = 0$. Note that it can be achieved by the Nash reversion.

Suppose now that $\delta \geq \underline{\delta}$. Then since $q^*(\theta)$ and $e^*(\theta)$ affect only on $J(d^*(\theta), \theta)$ in the objective and $e^*(\theta) = \bar{e}$ is feasible, $d^*(\theta) = (q^{SB}(\theta), e^{SB}(\theta))$. Furthermore since $\overline{s}^* \geq s_+^*(p^*, b^*, \omega)$, the objective function is increasing in $\hat{\theta}$. Then $d^*(\theta) = (q^{SB}(\theta), e^{SB}(\theta))$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$.

Chapter 3

Relational Political Contribution under Common Agency

3.1 Introduction

In order to analyse the strategic use of monetary transfer between politicians and lobbyists of interest groups to influence political decisions. the common agency model is widely used as a standard framework. The seminal paper by Bernheim and Whinston (1986, hereafter BW), considers menu auction to be common agency such that the bidders offer a payment plan contingent on the auctioneer's decision on the allocation of the goods. This menu auction model has been applied in analyses of buying influence through political contribution from lobbyists to a politician to investigate many issues in political economy.

The implicit assumption behind analyses using menu auction models is a long term relationship between the politician and the lobbyists while the analysed model typically appears as a one shot game.[1] In BW, a payment plan offered by the principals is treated as a binding contract enforced by the court after the agent makes a decision. However to assume that lobbyists can offer such a binding contract contingent on political decisions to a politician is far from reality. Instead, in the context of politics, such a compensation contract must be just an implicit agreement and hence we need another justification for

[1] Grossman and Helpman (2001, p. 228) explicitly mention this assumption.

the lobbyists' ability to commit to the implicit agreement. Repeated relationship allows the lobbyists to commit the implicit agreements since reneging can be punished later by themselves.

Nevertheless repeated relationship cannot necessarily justify the full commitment assumption. Empirical evidence argues that although there are long-term interactions among politicians and leaders of interest groups, the amount of contributions is sensitive to the length of their relation.[2] It implies that it is important to ask under what conditions it is appropriate to assume that lobbyists can fully commit to their contribution schedule and what happens if the commitment assumption is not appropriate.

This chapter formally analyses the political contribution that takes a form of implicit agreements and then must be self-enforced. The literature on the theory of relational contracts has already studied self-enforced contracts in dynamic situations by applying the framework of the repeated game. Recently, tractable frameworks of relational contracts have been established (MacLeod and Malcomson (1989) and Levin (2003)) and these are the building blocks for our analysis. Our model departs from them in two features. First, while the existing model analyses the bilateral relations between a single principal and a single agent, our model allows multi-principal situations. Second, the players have no outside option, which is a more realistic assumption in the political process.

In labour contracts such as MacLeod and Malcomson (1989) and Levin (2003), it is reasonable to imagine that the employee's effort is meaningful to the firm only if their employment relation is established. In other words, in the model the players have a participation decision at the beginning of each period and if the players do not participate, then

[2]For example, McCarty and Rothenberg (1996) find from the data that in the U.S., credible commitment of campaign contribution by the Political Action Committee is significantly weak. Snyder (1992) provides empirical evidence that campaign contribution to young representatives in the U.S. House is larger than old representatives.

they obtain the (constant) reservation utility as the outside value. In political situations, by contrast, the lobbyists are always influenced by the political decision regardless of their access to the politician. For instance, in case of protection for sale, the industry group always cares about imports and/or exports from foreign countries influenced by the tariff policy. However the event to access the politician and to make an (implicit) agreement itself does not affect the trade quantity at all. This independence is similarly applicable to the politician. In this sense, the political players cannot escape from the agent's decision or equivalently they do not have an outside option.[3]

There are two interests in analysing our model. First, we investigate the optimal punishment strategy. The punishment on a deviating player supports the self-enforcing mechanism for the implicit agreements on the political contribution. Second, we fully characterize the stationary equilibrium payoff set where the agent chooses the same decision repeatedly. As an implication, we show that – contrary to the naive conjecture from the folk theorem in repeated game theory – an equilibrium in the standard menu auction model might not be supported on implicit agreements no matter how patient the players are.

The most severe punishment on a deviating principal is not as simple as in the standard model of the relational contract because of the lack of the outside option. Nevertheless we show that the optimal punishment on a principal is either of two types. The first one is an "Exclusion-type" in which the agent repeatedly chooses the same decision harmful to the deviating principal. It is credibly implemented if there is a decision that is undesirable for the deviating principal but desirable for the other principals. The second one is a "Sanction-type" in which the undesirable decision is chosen only in the first period. If

[3]It might be still suspected that there is a situation in which the lobbyist can escape to the outside option and avoid the effect of the decision made by the politician. Nevertheless our result is applicable for the model with the outside options. See footnote 11 section 3.2.2.

the decision which brings the deviating principal the harsh payoff is also undesirable for the other players, then the players are not willing to repeat it. Now think of the following statement; "the agent will take the undesirable decision as a sanction on the deviating principal. But once she pays some fine to the agent, we will stop the sanction." If she obeys this statement, she must incur the additional cost by paying the fine as well as the cost from the undesirable decision. However, once she pays the fine, the undesirable decision immediately stops and all the players, including her, become better off. Thus by setting an appropriate fine, the deviating principal is willing to pay the fine and it is enough for all the players to put up with the undesirable decision only in the first period. It is actually similar to the so-called "stick and carrot" punishment proposed by Abreu (1986) in the context of Cournot oligopoly.

Thanks to the analysis of the optimal punishment, we can relatively easily provide a full characterization of the set of the payoffs of the decision-stationary equilibrium in which the agent chooses the same decision repeatedly on the equilibrium path. It can then be compared with the equilibrium payoffs in the static menu auction model of BW. It easily provides us the necessary and sufficient condition under which the payoff of the equilibrium in the corresponding static menu auction cannot be supported by the equilibria in the corresponding relational political contribution model.[4] We see the existence of an upper bound for the credible amount of relational political contribution and show that a static equilibrium payoff is possible when the payment on the static equilibrium is below the upper bound. Not surprisingly, the upper bound is increasing in the discount factor implying that a static equilibrium payoff is more likely to appear on the relational equilibrium when the players are patient. However there could be a

[4]We later use a term "menu auction" to describe environments where binding contracts are allowed and "political contribution" to be those without binding contracts.

static equilibrium payoff which cannot be an equilibrium payoff in the relational political contribution for any discount factors.

Roughly speaking, this interesting case appears when there are more than two principals and each of them is faced with a threat of being isolated from all the other players. In particular, suppose that there exists a decision which is undesirable for one of the principals and desirable for the others. This is an isolating decision in the sense that except for the principal all the players are willing to implement. Now suppose that there are two principals being isolated in such a way but the isolating decision is not common between them. In the static menu auction, the agent can exploit the principal by using the isolating decision as a threat and then at least one of the two principals are forced to pay much amount of compensation for the agent to avoid the isolating decision. By contrast, in the relational political contribution, it is impossible for the principal to commit to such amount of payment since the punishment is invoked only after the deviation is observed and then the effect of the punishment must be delayed and discounted. This discounting makes it impossible for the agent to exploit the principal no matter how patient the players are. Therefore the exploitation from more than one principal is impossible by implicit agreements, whereas it is possible by binding contracts.

Our model is an extension of the static common agency model analysed by BW and Grossman and Helpman (1994) – the latter of which was the first application of the menu auction model to political lobbies – to a dynamic setting. Dynamic common agency has been studied by Bergemann and Välimäki (2003). However they assume that decision-contingent contracts within one period can be enforced by the third party, whereas these are not available in our model. Instead, they introduce a payoff-relevant state variable and characterize the Markov perfect equilibria.

The static menu auction with notion of the commitment problem is studied by Dixit et al. (1997) and Campante and Ferreira (2007). Dixit et al. (1997) generalizes BW's model and their model includes the case where the maximum feasible compensation is limited. Nevertheless their analysis has no economic implication from this assumption.[5] Our analysis suggests that it would be a quite realistic assumption and we discuss a new economic insight from this assumption, which is ignored in Dixit et al. (1997). Campante and Ferreira (2007) consider the menu auction model where the principal cannot fully commit to the compensation plan due to the wealth constraint and her own production technology. However they assume that the decision-contingent compensation contract is still feasible, which is different from our model.

Self-enforced contracts are recently analysed by a number of papers, especially in labour contracts (e.g., MacLeod and Malcomson (1989) and Levin (2003)) and the theory of the firm (e.g., Baker et al. (2002)). To the best of our knowledge, our analysis is the first step for introducing the self-enforcing contract into the study of political lobbies. Our contribution is to characterize the optimal punishment when there is no outside option. Furthermore, while there are several papers that study self-enforced contracts with multiple parties, such as Levin (2002) and Rayo (2007), no paper studies the situation in which multiple principals interact with a common agent *via* self-enforced contracts. Thus this chapter is also the first analysis of relational contracts in common agency.

The rest of this chapter is organized as follows. The next section describes the model and the approach to derive the equilibria. Section 3.3 characterizes the set of decision-stationary equilibrium payoffs. In section 3.4, we present the first main result: the characterization of the optimal punishment strategy. Section 3.5 compares a decision-stationary equilibrium payoff with that of the static menu auction and shows the second main result. Section 3.6 discusses some topics related to the main results and section 3.7 provides the

[5]Moreover, to the best of our knowledge, there is no paper explicitly taking this assumption into account.

conclusion. The proofs are in the appendix.

3.2 The Model

3.2.1 Environment

We start by describing the primitive of the model. There are N principals[6] (these are females, called 1, 2, ..., and N, respectively) and an agent (who is male and sometimes called 0), all of whom will live in periods $t = 0, 1, \dots$ until infinity. Denote the set of the principals by $\mathcal{N} := \{1, 2, \dots, N\}$.

In period t, the agent chooses a decision $a_t \in \mathcal{A}$ where $\mathcal{A}$ is a compact subset of $\mathbf{R}^L$. Each player $i \in \mathcal{N} \cup \{0\}$ gets the one-shot benefit $v^i(a_t) \in \mathbf{R}$ from the decision. Assume that $v^i(\cdot)$ is continuous for all $i \in \mathcal{N} \cup \{0\}$. These assumptions ensure the existence of the maximum and minimum of $v^i(a)$ and then for all $i \in \mathcal{N} \cup \{0\}$, let

$$\overline{A}^i := \arg\max_{a \in \mathcal{A}} v^i(a),\ \underline{A}^i := \arg\min_{a \in \mathcal{A}} v^i(a),\ \overline{v}^i := \max_{a \in \mathcal{A}} v^i(a),\ \underline{v}^i := \min_{a \in \mathcal{A}} v^i(a).$$

We sometimes use a notation $\overline{a}^i$ and $\underline{a}^i$ as a representative element of $\overline{A}^i$ and $\underline{A}^i$ respectively. $s(a) := \sum_{i=0}^N v^i(a)$ be the one shot total benefit and denote $A^* := \arg\max_{a \in \mathcal{A}} s(a)$ and $s^* := \max_{a \in \mathcal{A}} s(a)$.

Each principal has an opportunity to make a payment to the agent after he chose a decision. Given a decision a_t and a vector of payments $(b_t^1, \dots, b_t^N)$, principal j's one-shot net payoff is given by $v^j(a_t) - b_t^j$ and the agent's is $v^0(a_t) + B_t$ where $B_t := \sum_{k=1}^N b_t^k$.

In period t, the players play the following two stage game;

- stage 1: the agent chooses a decision a_t from the decision set $\mathcal{A}$,

[6] All the results in this chapter hold even when $N = 1$.

- stage 2: given a_t, each principal j simultaneously and non-cooperatively pays $b_t^j \in [0, M]$ to the agent where $M \geq \max_{j \in N} \delta(\overline{v}^j - \underline{v}^j)/(1-\delta)$.

The restriction on the action space of payment implies that the principals cannot make a negative transfer (i.e. receive money from the agent). The assumption on M is just for guaranteeing that the action space of payment is compact.

Assume that at the beginning of each period t, all past decisions and payments $\{a_\tau, \{b_\tau^j\}_{j=1}^N\}_{\tau=0}^{t-1}$ are observable to all the players. The purpose of each player in the entire game is to maximize the average discounted sum of her/his payoff with common discount factor $\delta \in [0, 1)$; that is, $(1-\delta)\sum_{\tau=0}^{\infty} \delta^\tau [v^j(a_\tau) - b_\tau^j]$ for principal j and $(1-\delta)\sum_{\tau=0}^{\infty} \delta^\tau [v^0(a_\tau) + B_\tau]$ for the agent.

The strategy of the repeated game is defined by a mapping from an observed history to a decision or payment in the current period. Formally let $\mathcal{H} := \cup_{t=1}^{\infty} (\mathcal{A} \times [0, M]^N)^t$. The strategy[7] is defined as $\sigma^0 : \{\phi\} \cup \mathcal{H} \to \mathcal{A}$ for the agent and $\sigma^j : (\{\phi\} \cup \mathcal{H}) \times \mathcal{A} \to P^j(\delta)$ for principal j where ϕ denotes the null history. Let $\sigma := (\sigma^0, \sigma^1, \ldots, \sigma^N)$ be its profile. The strategy profile generates the on-path outcome such as $((a_0, a_1, \ldots), (b_0^1, b_1^1, \ldots), \ldots, (b_0^N, b_1^N, \ldots))$ and then the discounted average payoffs can be computed from it. Denote player i's discounted average payoff by $u^i(\sigma)$. We adopt the subgame perfect equilibrium (SPE) as our equilibrium concept where, for any history $h_t \in \{\phi\} \cup \mathcal{H}$ or $(h_t, a_t) \in (\{\phi\} \cup \mathcal{H}) \times \mathcal{A}$, the strategy maximizes her/his own payoff given the others' strategy. Let Σ^* be the set of SPE.

3.2.2 Simple Strategy Representation

The equilibrium analysis is substantially simplified by recursive formulation and the simple strategy introduced by Abreu (1988).

[7] We focus on the pure strategy.

Let σ be an arbitrary SPE and $\underline{\sigma}(i) \in \Sigma^*$ be a SPE which yields the lowest equilibrium payoff for player i, i.e. $u^i(\underline{\sigma}(i)) \leq u^i(\sigma)$ for any $\sigma \in \Sigma^*$. Abreu (1988) calls $\underline{\sigma}(i)$ the "optimal penal code" (OPC).

Now consider the following "simple strategy" profile $\varsigma(\sigma, \underline{\sigma}(0), \underline{\sigma}(1), \ldots, \underline{\sigma}(N))$, meaning that

- to follow σ if no player has deviated,
- to pay nothing and follow $\underline{\sigma}(0)$ from next period if the agent deviated, and
- to change to $\underline{\sigma}(j)$ from next period if principal j deviated.

By construction of ς, if player i chooses strategy ς^i (provided that the other players follow ς^{-i}), she/he follows strategy σ^i if no player has deviated from the outcome path of σ and immediately moves to the strategy which yields the lowest equilibrium payoff for the deviating player otherwise.

Now the useful result by Abreu (1988, Proposition 2 and 5) can be directly applied.[8]

Lemma 3.1 1. *The optimal penal code exists for each* $i \in \mathcal{N} \cup \{0\}$.

2. *For each* $i \in \mathcal{N} \cup \{0\}$, *let* $\underline{\sigma}(i)$ *be the OPC. If* $\sigma \in \Sigma^*$, *then* $\varsigma(\sigma, \underline{\sigma}(0), \underline{\sigma}(1), \ldots, \underline{\sigma}(N)) \in \Sigma^*$.

The generating outcome of $\varsigma(\sigma, \underline{\sigma}(0), \underline{\sigma}(1), \ldots, \underline{\sigma}(N)) \in \Sigma^*$ is same as that of σ. Then Lemma 3.1 allows us to focus on the simple strategy with the OPC to analyse the equilibrium outcome. In what follows, we focus the simple strategy SPE with the OPC and denote this simply by $\hat{\sigma}$ unless noted explicitly.[9]

[8]Abreu (1988) shows it when the stage game is in normal form, while it is in extensive form here. Nevertheless it is enough to modify his proof a little for proving the result. The proof is omitted here.

[9]Notice that the OPC can also have a simple strategy representation such as $\underline{\sigma}(j) = \varsigma(\underline{\sigma}(j), \underline{\sigma}(0), \underline{\sigma}(1), \ldots, \underline{\sigma}(N))$.

The one-shot deviation principle implies that $\hat{\sigma}$ is a SPE if and only if (i) the players have no deviation incentive in period 0, and (ii) the continuation strategy profile is a SPE. We now formally describe these conditions.

From $\hat{\sigma}$ we can construct the equilibrium path such that $\hat{a}_0 = \hat{\sigma}^0(\phi)$, $\hat{b}_0^j = \hat{\sigma}^j(\hat{a}_0)$, and the continuation strategy profile $\hat{\sigma}_1(\cdot) \equiv \hat{\sigma}(\hat{a}_0, \hat{\boldsymbol{b}}_0, \cdot)$ where $\hat{\boldsymbol{b}}_0 := (\hat{b}_0^1, \dots, \hat{b}_0^N)$. First, look at the agent's deviation incentive in period 0. If the agent follows $\hat{\sigma}^0$ in period 0, the average payoff the agent obtains has the following recursive expression

$$u^0(\hat{\sigma}) = (1-\delta)\left[\hat{B}_0 + v^0(\hat{a}_0)\right] + \delta u^0(\hat{\sigma}_1)$$

where $\hat{B}_0 := \sum_{j \in N} \hat{b}_0^j$. If he deviates to another decision a', then he gains the benefit $v^0(a')$ in period 0 and, since the players start the punishment on the agent, he receives nothing and the continuation strategy profile would be $\underline{\sigma}(0)$. Then the utility is $(1-\delta)v^0(a') + \delta u^0(\underline{\sigma}(0))$. Notice that $(1-\delta)v^0(a') + \delta u^0(\underline{\sigma}(0)) \leq (1-\delta)\overline{v}^0 + \delta u^0(\underline{\sigma}(0))$ for all $a' \in \mathcal{A}$. Thus the agent does not deviate from $\hat{\sigma}$ in period 0 if and only if

$$\begin{aligned} & (1-\delta)\left[\hat{B}_0 + v^0(\hat{a}_0)\right] + \delta u^0(\hat{\sigma}_1) \geq (1-\delta)\overline{v}^0 + \delta u^0(\underline{\sigma}(0)) \\ \iff \quad & \hat{B}_0 \geq \overline{v}^0 - v^0(\hat{a}) - \frac{\delta}{1-\delta}\left[u^0(\hat{\sigma}_1) - u^0(\underline{\sigma}(0))\right]. \end{aligned} \tag{3.1}$$

Next, look at principal j's deviation incentive in period 0. However we only have to check it on the equilibrium path because, for the case where the agent has already deviated from $\hat{a}_0$, she plays the punishment strategy, paying nothing and following $\underline{\sigma}(0)$, in which she has no deviation incentive by the assumption that $\underline{\sigma}(0)$ is SPE.[10] On the equilibrium

[10]Strictly speaking, we have implicitly assumed that the players play $\underline{\sigma}(0)$ regardless of the payment in period 0 when the agent has already deviated. Expecting that, the principals would optimally choose to pay nothing.

path, if principal j follows $\hat{\sigma}^j$ in period 0, her utility can be expressed in a recursive way similar to the agent as

$$(1-\delta)[v^j(\hat{a}_0) - \hat{b}^j_0] + \delta u^j(\hat{\sigma}_1).$$

If she deviates to another payment $b' \geq 0$, then she pays only $b' \geq 0$ and since the players start the OPC on her from next period, her continuation payoff would be $u^j(\underline{\sigma}(j))$. Then the utility is $(1-\delta)(v^j(\hat{a}_0) - b') + \delta u^j(\underline{\sigma}(j))$. Notice that $(1-\delta)(v^j(\hat{a}_0) - b') + \delta u^j(\underline{\sigma}(j)) \leq (1-\delta)v^j(\hat{a}_0) + \delta u^j(\underline{\sigma}(j))$ for any $b' \geq 0$. Thus principal j does not deviate from $\hat{\sigma}$ in period 0 if and only if

$$(1-\delta)[v^j(\hat{a}_0) - \hat{b}^j_0] + \delta u^j(\hat{\sigma}_1) \geq (1-\delta)v^j(\hat{a}) + \delta u^j(\underline{\sigma}(j)) \iff \hat{b}^j_0 \leq \frac{\delta}{1-\delta}\left[u^j_1(\hat{\sigma}_1) - u^j(\underline{\sigma}(j))\right].$$

Note that since the payment must be non-negative, $\hat{b}^j_0 \geq 0$ for all $j \in \mathcal{N}$. By imposing it, we obtain that

$$0 \leq \hat{b}^j_0 \leq \frac{\delta}{1-\delta}\left[u^j_1(\hat{\sigma}_1) - u^j(\underline{\sigma}(j))\right]. \tag{3.2}$$

In order to check whether $\hat{\sigma}$ is a SPE or not, the following conditions are necessary and sufficient.[11]

Lemma 3.2 *$\hat{\sigma} \in \Sigma^*$ if and only if (i) (3.1) holds, (ii) (3.2) holds for all $j \in \mathcal{N}$, and (iii) $\hat{\sigma}_1 \in \Sigma^*$.*

We abuse Lemma 3.2 to characterize decision-stationary equilibria, where the agent chooses the same decision repeatedly and the OPC. In the next section, we first investigate

[11] It is worth noting that even if $\underline{\sigma}(i)$ is not the OPC, these conditions are still necessary and sufficient for that the simple strategy $\hat{\sigma}$ is a SPE as long as $\underline{\sigma}(i)$ is also a SPE for all $i \in \mathcal{N} \cup \{0\}$. For instance, these conditions can be used for the model where each player has an outside option by substituting the outside value into $u^i(\underline{\sigma}(i))$ for each $i \in \{0\} \cup \mathcal{N}$.

the decision-stationary equilibria given the OPC $\{\sigma(i)\}_{i=0}^N$.

3.3 Decision-Stationary Equilibria

This section focuses on the equilibria where the agent chooses the same decision $\hat{a}$ every period. Formally, the definition is as follows.

Definition 3.1 *$\hat{\sigma}$ is a decision-stationary strategy profile of $\hat{a} \in \mathcal{A}$ if it generates the outcome path where the decision $\hat{a}$ is repeatedly chosen every period.*

Let $\hat{\Sigma}^*(\hat{a})$ be the set of decision stationary equilibria of $\hat{a}$ and $\hat{U}^*(\hat{a}) := \{(u^0(\sigma), u^1(\sigma), \ldots, u^N(\sigma)) \mid \sigma \in \hat{\Sigma}^*(\hat{a})\}$ be its payoff set. Given the OPC $\{\underline{\sigma}(i)\}_{i=0}^N$, consider a strategy profile $\hat{\sigma}$ which is decision-stationary of $\hat{a}$. By applying Lemma 3.2, $\hat{\sigma}$ is a SPE if and only if

$$\hat{\psi}(\hat{a}) \geq \overline{v}^0 - v^0(\hat{a}) - \frac{\delta}{1-\delta}\left[u^0(\hat{\sigma}_1) - u^0(\underline{\sigma}(0))\right] \tag{3.3}$$

$$0 \leq \hat{\sigma}^j(\hat{a}) \leq \frac{\delta}{1-\delta}\left[u_1^j(\hat{\sigma}_1) - u^j(\underline{\sigma}(j))\right],\ j \in \mathcal{N} \tag{3.4}$$

$$\hat{\sigma}_1 \in \hat{\Sigma}^*(\hat{a})$$

where $\hat{\psi}(\hat{a}) := \sum_{k \in \mathcal{N}} \hat{\sigma}^k(\hat{a})$ and $\hat{\sigma}_1$ is the continuation strategy of $\hat{\sigma}$ on the equilibrium path. Note that $\hat{\sigma}_1$ must also be a decision-stationary SPE of $\hat{a}$.

While we have so far allowed that the equilibrium payment in some period to differ from that in another period, we will show that it is without loss of generality to restrict the payment schedule to being stationary for deriving the decision-stationary SPE payoff. The principal's strategy (or payment) is called stationary if the payment on the equilibrium path is determined only by the decision in the current period. The formal definition is as follows.

Definition 3.2 *Let $\hat{\sigma}$ be a strategy profile which generates the equilibrium path $(\hat{a}_t, \{\hat{b}_t^j\}_{j=1}^N)_{t=0}^\infty$ and denote $\hat{h}_0 := \phi$ and $\hat{h}_t := (\hat{a}_\tau, \{\hat{b}_\tau^j\}_{j=1}^N)_{\tau=0}^{t-1}$ for $t \geq 1$. $\hat{\sigma}$ is payment-stationary for principal j if $\hat{\sigma}^j$ can be described by*

$$\hat{\sigma}^j(\hat{h}_t, a) = \beta^j(a)$$

for any $a \in \mathcal{A}$ and $t \geq 0$. $\hat{\sigma}$ is stationary if it is decision-stationary and payment-stationary.

The following proposition implies that as long as the agent implements the same decision repeatedly, the stationary-payment strategy is without loss of generality no matter how many principals are there.[12]

Proposition 3.1 *Suppose that $\hat{\sigma} \in \hat{\Sigma}^*(\hat{a})$. Then there exists a stationary strategy profile in $\hat{\Sigma}^*(\hat{a})$ which generates the same payoff vector.*

When the strategy is stationary, the players repeat the same stationary strategy $\hat{\sigma}$ for the continuation game every period, which makes the continuation payoff identical in any period. Then in the stationary strategy of decision $\hat{a}$, (3.3) and (3.4) are equivalent to

$$\hat{\mathrm{B}}(\hat{a}) \geq \overline{v}^0 - v^0(\hat{a}) - \frac{\delta}{1-\delta}\left[u^0(\hat{\sigma}) - u^0(\underline{\sigma}(0))\right] \tag{3.5}$$

$$0 \leq \hat{\beta}^j(\hat{a}) \leq \frac{\delta}{1-\delta}(u^j(\hat{\sigma}) - u^j(\underline{\sigma}(j))) \tag{3.6}$$

where $\hat{\beta}^j(a)$ is the stationary payment of principal j and $\hat{\mathrm{B}}(\cdot) := \sum_{j=1}^N \hat{\beta}^j(\cdot)$. Since the players repeat the same stationary strategy every period, the one-shot net payoff is same across

[12] It should be emphasized that while the proof here is similar to Levin (2003), the limited liability might make his proof collapse. The key idea of his proof is that the variation of the continuation payoff, depending on the current decision can be completely transformed to the variation of the current payment. However under limited liability, this transformation might violate the limited liability constraint. Here we show that this limited liability does not matter when there is no asymmetric information.

all periods and identical to the average payoff in the entire game, i.e. $u^0(\hat{\sigma}) = v^0(\hat{a}) + \hat{\mathrm{B}}(\hat{a})$ and $u^j(\hat{\sigma}) = v^j(\hat{a}) - \hat{\beta}^j(\hat{a})$ for $j \in \mathcal{N}$. Thus (3.5) and (3.6) are further reduced to

$$u^0(\hat{\sigma}) \geq (1-\delta)\overline{v}^0 + \delta u^0(\underline{\sigma}(0)) \tag{3.7}$$

$$v^j(\hat{a}) \geq u^j(\hat{\sigma}) \geq (1-\delta)v^j(\hat{a}) + \delta u^j(\underline{\sigma}(j)). \tag{3.8}$$

With the restriction that the total net payoff among the players is always identical with $s(\hat{a})$, the set of payoff vectors of decision-stationary equilibria is characterized as follows;

$$\hat{U}^*(\hat{a}) = \left\{ \left(\begin{array}{c} u^0 \\ u^1 \\ \vdots \\ u^N \end{array} \right)' \left| \begin{array}{l} \sum_{j=0}^N u^j = s(\hat{a}), \\ u^0 \geq (1-\delta)\overline{v}^0 + \delta u^0(\underline{\sigma}(0)) \\ v^j(\hat{a}) \geq u^j \geq (1-\delta)v^j(\hat{a}) + \delta u^j(\underline{\sigma}(j)), {}^\forall j \in \mathcal{N} \end{array} \right. \right\}. \tag{3.9}$$

Now we investigate the condition for the existence of (decision-)stationary equilibria of $\hat{a}$. Combining (3.5) and (3.6) to eliminate $\hat{\beta}^j(\hat{a})$ yields the necessary condition for existence, that is,

$$\overline{v}^0 - v^0(\hat{a}) - \frac{\delta}{1-\delta}\left[u^0(\hat{\sigma}) - u^0(\sigma(0))\right] \leq \sum_{j=1}^N \frac{\delta}{1-\delta}\left[u^j(\hat{\sigma}) - u^j(\sigma(j)\right].$$

Since $\sum_{j=0}^N u^j(\hat{\sigma}) = s(\hat{a})$, it is

$$\frac{\delta}{1-\delta}\left[s(\hat{a}) - \sum_{j=0}^N u^j(\sigma(j))\right] \geq \overline{v}^0 - v^0(\hat{a}). \tag{3.10}$$

Conversely, given (3.10), if we can construct $\hat{\beta}^j(\hat{a})$ which satisfies (3.7) and (3.8) (where $u^j(\hat{\sigma}) = v^j(\hat{a}) - \hat{\beta}^j(\hat{a})$ for $j \in \mathcal{N}$ and $u^0(\hat{\sigma}) = v^0(\hat{a}) + \hat{\mathrm{B}}(\hat{a})$), the stationary strategy consisting of payment schedule $\{\hat{\beta}^j\}_{j=1}^N$ and decision $\hat{a}$ is a SPE, implying the sufficiency for the existence of a decision-stationary SPE. It is actually the case and the following proposition is obtained.

Proposition 3.2 *$\hat{\Sigma}^*(\hat{a}) \neq \emptyset$ if and only if (3.10) holds.*

Finally, we investigate the socially optimal equilibrium which maximizes the average discounted sum of the benefits among the players. The socially optimal equilibrium is interesting to us because not only could it be a focal point, but also the strategy which generates the maximum total benefit helps us to investigate the OPC later. The following proposition states that it is without loss of generality that the socially optimal equilibrium is decision-stationary.[13] Thus together with Proposition 3.2, (3.10) plays an important role for examining the OPC.

Proposition 3.3 *If $\sigma \in \Sigma^*$, then there exists a strategy profile $\hat{\sigma}^*(\hat{a}) \in \hat{\Sigma}^*(\hat{a})$ such that $\sum_{i=0}^N u^i(\hat{\sigma}^*(\hat{a})) \geq \sum_{i=0}^N u^i(\sigma)$.*

3.4 The Optimal Penal Code

In this section, we investigate the optimal penal code. We first briefly summarize this section. Our model is a special case of Wen (2002)'s repeated sequential game and Proposition 3 in his paper states that the equilibrium payoff must be larger than or equal to the effective minimax value. The effective minimax value in our model is given by $\overline{v}^0$ for the agent and $\underline{v}^i$ for principal i.[14] We will show that the OPC on the agent achieves $\overline{v}^0$ by the Nash reversion whereas it is not necessarily the case that either the OPC on the principal

[13]The optimal equilibrium for one of the players is not necessarily decision-stationary. See for instance Ray (2002).

[14]These are actually the same as the standard minimax value of the period game where all the players move simultaneously. In what follows we simply call it the minimax value.

is the Nash reversion or it can achieve $\underline{v}^j$ for principal j. On the OPC on the principal, we formulate and solve the minimization problem and demonstrate that, in general, the OPC would be the so-called two-phase scheme demonstrated by Abreu (1986) in the model of tacit collusion in repeated Cournot oligopoly.

3.4.1 The Optimal Penal Code on the Agent

If there exists an equilibrium strategy by which the agent's payoff becomes his minimax value, it would be the OPC. Consider the one-period game in our model. On the subgame perfect equilibrium of the one-period game, the principals pay nothing and the agent chooses $\overline{a}^0 \in \overline{A}^0$, which achieves the agent's payoff $\overline{v}^0$. Thus in the repeated game, the minimax payoff can be easily attained by repetition of it (after every history). Then $u^0(\sigma(0)) = \overline{v}^0$.

3.4.2 The Optimal Penal Code on the Principals

Characterization of the OPC payoff

Consider the OPC on principal k, $\underline{\sigma}(k)$. Let $(\underline{a}_0(k), (\underline{b}_0^1(k), \ldots, \underline{b}_0^N(k)))$ be the associated outcome path in period 0 generated by $\underline{\sigma}(k)$ and $\underline{\sigma}_1(k)$ be the continuation strategy from period 1 on the equilibrium path. Then, by Lemma 3.2, $\underline{\sigma}(k)$ is a SPE if and only if

$$(1-\delta)[v^0(\underline{a}_0(k)) + \underline{B}_0(k)] + \delta u^0(\underline{\sigma}_1(k)) \geq \overline{v}^0, \tag{3.11}$$

$$0 \leq \underline{b}_0^j(k) \leq \frac{\delta}{1-\delta}[u^j(\underline{\sigma}_1(k)) - u^j(\underline{\sigma}(j))],\ j \in \mathcal{N}, \tag{3.12}$$

and $\underline{\sigma}_1(k) \in \Sigma^*$.[15]

[15]In (3.11), it is taken into account that $u^0(\underline{\sigma}(0)) = \overline{v}^0$.

Since $\underline{\sigma}(k)$ leads to the lowest SPE payoff for principal k, the N+2-tuple $(\underline{a}_0(k), \{\underline{b}_0^j(k)\}_{j=1}^N, \underline{\sigma}_1(k))$ can be characterized by the solution of the minimization problem where the objective is $u^k(\underline{\sigma}(k)) = (1-\delta)[v^k(\underline{a}_0(k)) - \underline{b}_0^k(k)] + \delta u^k(\underline{\sigma}_1(k))$ subject to (3.11), (3.12), and $\underline{\sigma}_1(k) \in \Sigma^*$. Namely, it is the solution of the following problem;

$$\begin{aligned} \min_{(\underline{a}_0(k), \{\underline{b}_0^j(k)\}_{j=1}^N, \underline{\sigma}_1(k))} \quad & (1-\delta)[v^k(\underline{a}_0(k)) - \underline{b}_0^k(k)] + \delta u^k(\underline{\sigma}_1(k)) \\ \text{subject to} \quad & (3.11),\ (3.12),\ \underline{\sigma}_1(k) \in \Sigma^*. \end{aligned} \tag{3.13}$$

Suppose that the solution satisfies that $\underline{b}_0^k(k) < \delta(u^k(\underline{\sigma}_1(k)) - u^k(\underline{\sigma}(k)))/(1-\delta)$. By increasing $\underline{b}_0^k(k)$ slightly to keep to satisfy (3.12) for $j = k$, (3.13) is decreasing without any violations of the constraints. Thus it is not the solution, implying that $\underline{b}_0^k(k) = \delta(u^k(\underline{\sigma}_1(k)) - u^k(\underline{\sigma}(k)))/(1-\delta)$. Substituting it into (3.13) yields the following result.

Lemma 3.3 $u^k(\underline{\sigma}(k)) = v^k(\underline{a}_0(k))$ *for all* $k \in \mathcal{N}$.

Lemma 3.3 states that $u^k(\underline{\sigma}(k))$ must be identical to the benefit in the first period on $\underline{\sigma}(k)$. In other words, the average equilibrium payoff on the OPC can be represented by a decision in $\mathcal{A}$ and then in the minimization problem, it is enough to seek $\underline{a}_0(k)$ for minimizing principal k's benefit $v^k(\underline{a}_0(k))$ subject to the conditions for ensuring $\underline{\sigma}(k)$ a SPE.

We have already seen that (3.12) for $j = k$ is binding. Now look at (3.12) for principal $j \neq k$. When $\underline{b}_0^j(k)$ is not binding, increasing it slightly makes the degree of freedom in (3.11) greater without any effects on the other constraints and the objective function. Then without loss of generality (3.12) holds with equality at the upper bound for all $j \in \mathcal{N}$, i.e. $\underline{b}_0^j(k) = \delta(u^j(\underline{\sigma}_1(k)) - u^i(\underline{\sigma}(k)))/(1-\delta)$.[16] Substituting them to (3.11), we can reduce the

[16]Note that $u^j(\underline{\sigma}_1(k)) - u^j(\underline{\sigma}(j)) \geq 0$ by the definition of $\underline{\sigma}(j)$. Then it is obvious that $\underline{b}_0^j(k) \geq 0$ for all $j \in \mathcal{N}$.

constraints into $\underline{\sigma}_1(k) \in \Sigma^*$ and

$$(1-\delta)v^0(\underline{a}_0(k)) + \delta\left[\sum_{j=1}^{N}\left(u^j(\underline{\sigma}_1(k)) - u^j(\underline{\sigma}(j))\right)\right] + \delta u^0(\underline{\sigma}_1(k)) \geq \overline{v}^0.$$

Applying Lemma 3.3 for all $j \in \mathcal{N}$ and rearranging yields

$$\frac{\delta}{1-\delta}\left[\sum_{i=0}^{N} u^i(\underline{\sigma}_1(k)) - \left(\overline{v}^0 + \sum_{j=1}^{N} v^j(\underline{a}_0(j))\right)\right] \geq \overline{v}^0 - v^0(\underline{a}_0(k)).$$

Furthermore, due to Proposition 3.3, if $\underline{\sigma}_1(k)$ is a SPE, then there exists a decision-stationary SPE the total benefit of which is not less than $\sum_{i=0}^{N} u^i(\underline{\sigma}_1(k))$. Then without loss of generality we can restrict $\underline{\sigma}_1(k)$ on decision-stationary SPE. Since the total benefit is $s(\underline{a}_1(k))$ when $\underline{\sigma}_1(k)$ is a decision stationary strategy of $\underline{a}_1(k)$, this condition can be further rewritten as

$$\frac{\delta}{1-\delta}\left[s(\underline{a}_1(k)) - \left(\overline{v}^0 + \sum_{j=1}^{N} v^j(\underline{a}_0(j))\right)\right] \geq \overline{v}^0 - v^0(\underline{a}_0(k)). \tag{3.14}$$

Recall that thanks to Proposition 3.2, in order to check whether $\underline{\sigma}_1(k)$ is a decision stationary SPE of $\underline{a}_1(k)$ or not, condition (3.10) (for $\hat{a} = \underline{a}_1(k)$ and $u^j(\underline{\sigma}(j)) = v^j(\underline{a}_0(j))$) is necessary and sufficient, that is,

$$\frac{\delta}{1-\delta}\left[s(\underline{a}_1(k)) - \left(\overline{v}^0 + \sum_{j=1}^{N} v^j(\underline{a}_0(j))\right)\right] \geq \overline{v}^0 - v^0(\underline{a}_1(k)). \tag{3.15}$$

To summarize, the problem for finding the OPC for principal k is to minimize $v^k(\underline{a}_0(k))$ subject to (3.14) and (3.15);

Problem (k)

$$\min_{\underline{a}_0(k),\underline{a}_1(k)} \quad v^k(\underline{a}_0(k))$$

subject to $$\frac{\delta}{1-\delta}\left[s(\underline{a}_1(k)) - \left(\overline{v}^0 + \sum_{j=1}^{N} v^j(\underline{a}_0(j))\right)\right] \geq \overline{v}^0 - \min\left\{v^0(\underline{a}_0(k)), v^0(\underline{a}_1(k))\right\}.$$

Notice that the solution of Problem (k) depends on $\underline{a}_0(\ell)$ for $\ell \neq k$ which must be the solution of Problem (ℓ). It means that the $2n$-tuple $(\underline{a}_0(1), \underline{a}_1(1), \underline{a}_0(2), \underline{a}_1(2), \ldots, \underline{a}_0(N), \underline{a}_1(N))$ must be the solution of Problem (k) for all $k \in \mathcal{N}$ simultaneously. However this $2n$-tuple $(\underline{a}_0(1), \underline{a}_1(1), \underline{a}_0(2), \underline{a}_1(2), \ldots, \underline{a}_0(N), \underline{a}_1(N))$ can be reduced to the n+1-tuple $(\underline{a}_0(1), \underline{a}_0(2), \ldots, \underline{a}_0(N), \hat{a})$ thanks to the following lemma.

Lemma 3.4 *Suppose that the $2n$-tuple $(\underline{a}_0(1), \underline{a}_1(1), \underline{a}_0(2), \underline{a}_1(2), \ldots, \underline{a}_0(N), \underline{a}_1(N))$ is the solution of Problem (k) for all $k \in \mathcal{N}$ simultaneously. Then there exists $\hat{a} \in \mathcal{A}$ such that for all $k \in \mathcal{N}$ $(\underline{a}_0(k), \hat{a})$ is the solution of Problem (k).*

This lemma simplifies the problem into finding the $n+1$-tuple $(\underline{a}_0(1), \underline{a}_0(2), \ldots, \underline{a}_0(N), \hat{a})$ that minimizes $v^k(\underline{a}_0(k))$ for all $k \in \mathcal{N}$ simultaneously subject to only the following one constraint;

$$\frac{\delta}{1-\delta}\left[s(\hat{a}) - \left(\overline{v}^0 + \sum_{j=1}^{N} v^j(\underline{a}_0(j))\right)\right] \geq \overline{v}^0 - \min\left\{v^0(\hat{a}), \min_{k\in\mathcal{N}} v^0(\underline{a}_0(k))\right\}. \tag{3.16}$$

Denote $\boldsymbol{a}(\mathcal{N}) := (a(1), \ldots, a(N)) \in \mathcal{A}^N$[17] and define

$$\begin{aligned} A_1^{PC}(\boldsymbol{a}(\mathcal{N})) &:= \left\{\hat{a} \in \mathcal{A} \;\middle|\; \frac{\delta}{1-\delta}\left[s(\hat{a}) - \left(\overline{v}^0 + \sum_{j=1}^{N} v^j(a(j))\right)\right] \geq \overline{v}^0 - \min\left\{v^0(\hat{a}), \min_{k\in\mathcal{N}} v^0(a(k))\right\}\right\}, \\ A_0^{PC} &:= \left\{\boldsymbol{a}(\mathcal{N}) \in \mathcal{A}^N \mid A_1^{PC}(\boldsymbol{a}(\mathcal{N})) \neq \emptyset\right\}. \end{aligned}$$

Note that (3.16) is satisfied if and only if $\underline{\boldsymbol{a}}_0(\mathcal{N}) \in A_0^{PC}$. It means that A_0^{PC} is the feasible set of the problem and then the pair of the first period decisions of the OPC is chosen from A_0^{PC} in the following way.

[17]Similarly, denote $\underline{\boldsymbol{a}}_0(\mathcal{N}) := (\underline{a}_0(1), \ldots, \underline{a}_0(N))$ and $\boldsymbol{a}(\mathcal{N})' := (a(1)', \ldots, a_0(N)')$.

Proposition 3.4 $(u^1(\underline{\sigma}(1)), \ldots, u^N(\underline{\sigma}(N))) = (v^1(\underline{a}_0(1)), \ldots, v^N(\underline{a}_0(N)))$ *where* $\underline{\boldsymbol{a}}_0(\mathcal{N}) \in A_0^{PC}$ *and for all* $\boldsymbol{a}(\mathcal{N}) \in A_0^{PC}$, $v^k(\underline{a}_0(k)) \leq v^k(a(k))$ *for all* $k \in \mathcal{N}$.

Notice that Proposition 3.4 says that the pair of decisions $\underline{\boldsymbol{a}}_0(\mathcal{N})$ must weakly Pareto-dominate any pair of decisions in A_0^{PC} (in the negative sense) and then the existence of the pair satisfying Proposition 3.4 does not seem to be trivial. However it is assured in the following way.

Proposition 3.5 *The pair* $\underline{\boldsymbol{a}}_0(\mathcal{N})$ *satisfying the condition of Proposition 3.4 exists and if both* $\underline{\boldsymbol{a}}_0(\mathcal{N})$ *and* $\underline{\boldsymbol{a}}_0(\mathcal{N})'$ *satisfy the condition, then* $v^k(\underline{a}_0(k)) = v^k(\underline{a}_0(k)')$ *for all* $k \in \mathcal{N}$.

It is straightforward that if there exists $(\underline{a}^1, \ldots, \underline{a}^N) \in \prod_{j=1}^N \underline{A}^j$ such that $(\underline{a}^1, \ldots, \underline{a}^N) \in A_0^{PC}$, it is immediately the pair of first decisions in the OPCs. The condition for it can be simply expressed as follows.

Corollary 3.1 *The optimal penal code satisfies* $u^j(\underline{\sigma}(j)) = \underline{v}^j$ *for all* $j \in \mathcal{N}$ *if and only if there exists a vector* $(\hat{a}, \underline{a}^1, \ldots, \underline{a}^N) \in \mathcal{A} \times \prod_{j=1}^N \underline{A}^j$ *such that*

$$\frac{\delta}{1-\delta}\left[s(\hat{a}) - \left(\overline{v}^0 + \sum_{j=1}^N \underline{v}^j\right)\right] \geq \overline{v}^0 - \min\left\{v^0(\hat{a}), \min_{k \in \mathcal{N}} v^0(\underline{a}^k)\right\}. \tag{3.17}$$

Whether (3.17) is satisfied or not depends on the discount factor δ. For example, as an extreme case, when $\delta = 0$, (3.17) implies that $\min_k v^0(\underline{a}^k) \geq \overline{v}^0$ meaning that it is not satisfied unless $\underline{a}^k \in \overline{A}^0$ for all $k \in \mathcal{N}$. As δ becomes higher, the more severe punishment is available and, if δ is feasibly large enough, (3.17) holds meaning that the first best OPC is possible.

Proposition 3.6

1. *For all* $k \in \mathcal{N}$, $v^k(\underline{a}_0(k))$ *is non-increasing in* δ.
2. *There exists* $\overline{\delta}^{OPC*} \in [0, 1)$ *such that for* $\delta \in [\overline{\delta}^{OPC*}, 1)$, $u^k(\underline{a}_0(k)) = \underline{v}^k$ *for all* $k \in \mathcal{N}$.

The Path of the OPC on a Principal

Recall that, on strategy $\underline{\sigma}(k)$, the agent chooses decision $\underline{a}_0(k)$ in the first period and strategy $\underline{\sigma}_1(k)$ in the continuation periods, the latter of which is replaced with some decision-stationary strategy of $\hat{a}$. Thus the punishment scheme, in general, consists of two phases, $\underline{a}_0(k)$ and $\hat{a}$. From (3.13) for $i = k$,

$$u^k(\underline{\sigma}(k)) = (1-\delta)(v^k(\underline{a}_0(k)) - \underline{b}_0^k(k)) + \delta(v^k(\hat{a}) - \hat{b}^k)$$

where $\hat{b}^k$ is the stationary payment in the continuation strategy.[18] Lemma 3.3 implies that

$$v^k(\underline{a}_0(k)) = (1-\delta)(v^k(\underline{a}_0(k)) - \underline{b}_0^k(k)) + \delta(v^k(\hat{a}) - \hat{b}^k),$$

and if $\delta > 0$,[19] it is equivalently

$$v^k(\underline{a}_0(k)) = v^k(\hat{a}) - \frac{(1-\delta)}{\delta}\underline{b}_0^k(k) - \hat{b}^k.$$

Since the payment must be nonnegative, this equation implies that $v^k(\underline{a}_0(k)) \leq v^k(\hat{a})$.

First, suppose that $v^k(\underline{a}_0(k)) = v^k(\hat{a})$. It immediately implies that $\underline{b}_0^k(k) = \hat{b}^k = 0$ and principal k gains the same level of benefit and pays nothing over time. This is illustrated in Figure 3.1. Especially if $\hat{a} = \underline{a}_0(k)$,[20] the agent chooses the same decision and principal k has to be punished by the harsh decision $\underline{a}_0(k)$ over time. In this sense, she is excluded from the other players and we can call this punishment strategy an “Exclusion-type” of

[18]The stationarity of the payment is without loss of generality due to Proposition 3.1.

[19]If $\delta = 0$, the OPC is just the Nash reversion, i.e. the agent repeatedly chooses a decision from the set $\arg\max_{a \in \overline{A}^0} v^k(a)$ and the principals pay nothing.

[20]In general, $\underline{a}_0(k)$ could be different from $\hat{a}$. However notice that since $\underline{\boldsymbol{a}}_0(\mathcal{N}) \in A_0^{PC}$ and $v^k(\hat{a}) = v^k(\underline{a}_0(k))$, $(\underline{a}_0(0), \ldots, \underline{a}_0(k-1), \hat{a}, \underline{a}_0(k+1), \ldots, \underline{a}_0(N))$ is also in A_0^{PC}, which means that $\hat{a}$ can be also used as the first decision of $\underline{\sigma}(k)$. Thus it is without loss of generality.

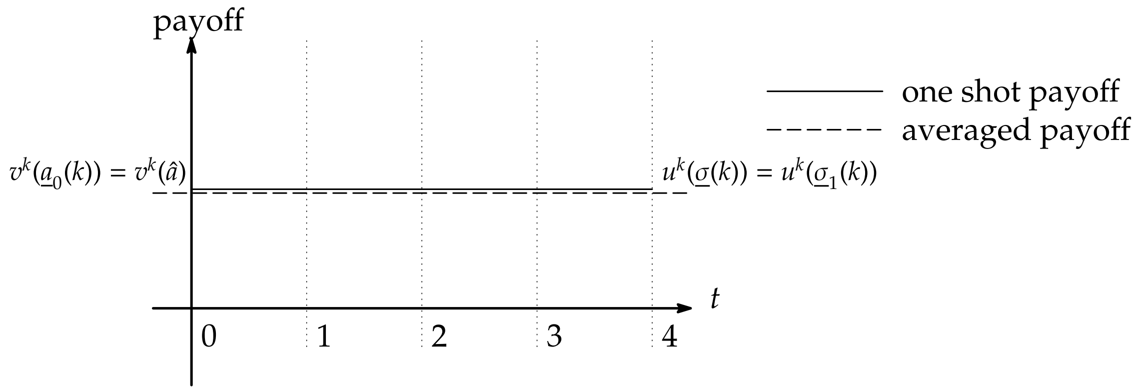

Figure 3.1: In Case of Exclusion-type

punishment.

Next, suppose that $v^k(\underline{a}_0(k)) < v^k(\hat{a})$. Note that

$$\begin{aligned} u^k(\underline{\sigma}_1(k)) = v^j(\hat{a}) - \hat{b}^k = v^k(\hat{a}) - \left(v^k(\hat{a}) - v^k(\underline{a}_0(k)) - \frac{(1-\delta)}{\delta}\underline{b}_0^k(k)\right) &= v^k(\underline{a}_0(k)) + \frac{(1-\delta)}{\delta}\underline{b}_0^k(k) \\ &\geq v^k(\underline{a}_0(k)) = u^k(\underline{\sigma}(k)) \end{aligned}$$

and since $\underline{b}_0^k(k) \geq 0$, we obtain $v^k(\underline{a}_0(k)) - \underline{b}_0^k(k) \leq v^k(\underline{a}_0(k)) = u^k(\underline{\sigma}(k))$. In words, the one-shot net payoff in the first period is less than the average payoff in the OPC and the average payoff in the continuation game must be more than that of the OPC. This process is illustrated in Figure 3.2. On the punishment path, principal k is severely punished first and rewarded later. Specifically, in the first period, after she incurs the cost from the undesirable decision, she must additionally make the positive amount of transfer. It can be interpreted as the "sanction fine" for deviation and once she pays it, the situation becomes "normal". In this sense, we call this punishment strategy a "Sanction-type" of punishment. This is qualitatively similar to the so-called "stick and carrot" strategy shown by Abreu (1986) where the first phase stands as the punishing "stick phase" and the remaining phase stands as the rewarding "carrot phase".

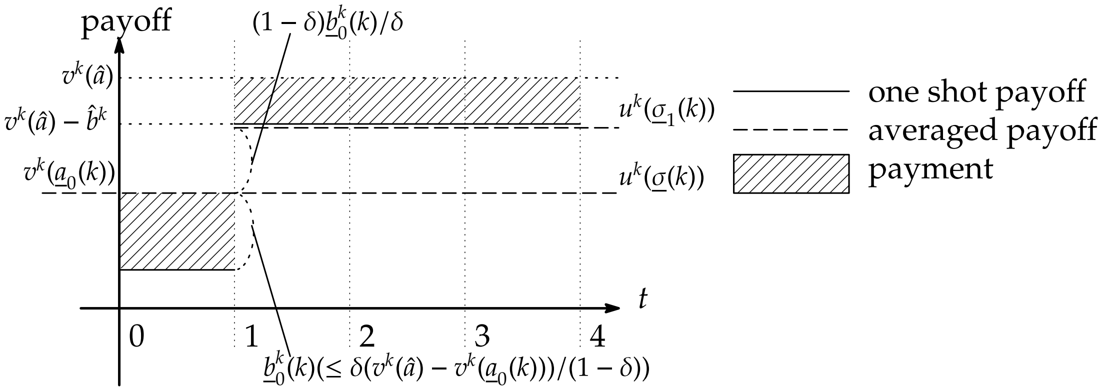

Figure 3.2: In Case of Sanction-type

Decision	$v^1(a)$	$v^0(a)$	$v^2(a)$
l	G^1	$-C$	$-D$
c	0	0	0
r	$-D$	$-C$	G^2

Table 3.1: Example 1

In the Cournot oligopoly in Abreu (1986), the optimal punishment on deviation is always the stick and carrot strategy. Then it seems that the OPC in our model is always a Sanction-type without loss of generality. However it is demonstrated by the following example that an Exclusion-type can be the unique OPC.

Example

The example used here is as follows. Suppose that $N = 2$ and $\mathcal{A} = \{l, c, r\}$ and the private benefit for each player is given in Table 3.1 where G^1, G^2, C, and D are all positive. We identify the threshold of the discount factor above which $u^j(\underline{\sigma}(j)) = -D$ for some $j = 1, 2$.[21]

First, assume that $G^j < C + D$ for $j = 1, 2$. Then c is socially optimal and can be implemented on a stationary equilibrium for any δ. Thus from corollary 3.1, $u^j(\underline{\sigma}(j)) = -D$

[21] If δ is less than the threshold derived later, it is easy to see that $u^j(\underline{\sigma}(j)) = 0$ for $j = 1, 2$.

for $j = 1, 2$ if and only if

$$\frac{\delta}{1-\delta}\left[0-(0-2D)\right] \;\geq\; 0-\min\{0,-C\} \iff \delta \geq \frac{C}{2D+C}.$$

The question is whether the players can achieve this punishment by an Exclusion-type strategy or not. If principal 2 can be punished by an Exclusion-type strategy, there must be a decision-stationary SPE of l. From (3.10), given $u^j(\underline{\sigma}(j)) = -D$ for $j = 1, 2$, it is equivalent to

$$\frac{\delta}{1-\delta}\left[(G^1-D-C)-(0-2D)\right] \;\geq\; 0-(-C) \iff \delta \geq \frac{C}{G^1+D}. \tag{3.18}$$

Notice that $C/(G^1+D)$ is greater than $C/(2D+C)$. Then, if $\delta \in [C/(2D+C), C/(G^1+D))$, the first best OPC can be achieved only through a Sanction-type strategy.

By contrast, we can obtain the opposite statement when the assumption on the parameters is changed. Now assume $G^1 > C + D \geq G^2$. Then l is the socially optimal decision. Given the punishment payoff being $-D$ for both principals, l can be implemented by a decision stationary SPE if and only if (3.18) holds. Further, since a decision-stationary SPE of l achieves a payoff no more than $-D$ for principal 2, the punishment payoff $-D$ for principal 2 can be achieved by an Exclusion-type strategy such that l is repeatedly chosen. The question is whether a Sanction-type strategy can achieve $-D$ for principal 2. For instance, if the punishment strategy is such that l is chosen first and c is repeatedly chosen after that, then (3.16) for $\hat{a} = c$ must be satisfied, that is,

$$\frac{\delta}{1-\delta}\left[0-(0-2D)\right] \;\geq\; 0-(-C) \iff \delta \geq \frac{C}{2D+C}.$$

Our hypothesis assures that $C/(2D+C)$ is greater than $C/(G^1+D)$. Then, if $\delta \in [C/(G^1 +$

$D), C/(2D + C))$, this Sanction-type punishment cannot attain the first best OPC. Similar analysis shows that with this discount factor, a Sanction-type punishment such that r is chosen from the second period cannot attain the first best OPC neither.[22]

The interesting implication from this example is that both Sanction-types and Exclusion-types could be the unique OPC. It is in sharp contrast to the optimal tacit collusion in Cournot oligopolistic markets. The characterization of the OPC here shares Abreu (1986)'s property in the sense that in general the OPC would be a two-phase scheme, the "punish" phase and the "reward" phase. In Cournot markets, in order to punish the deviating firm credibly, at the punish phase every firm must suffer from the punishment payoff led by the predatory behaviour and then every firm must be rewarded later by sharing the monopolistic profit in the industry. In this sense the phases are distinguishable from each other. By contrast, in our political contribution model, it is not necessary that all the players suffer in the punishment phase because there might be a decision which is quite terrible only for the punished principal and even if such a decision is not much preferable for the agent, monetary transfer can adjust the distribution of the total benefit so that the other principals can incentivize the agent to choose it. In the above example, this situation is clearly described when $G^1 > C + D$ because decision l is the worst for principal 2 but socially optimal and then principal 1 is willing to compensate the agent for choosing l (as long as the players are somewhat patient). Because the agent and principal 1 cannot benefit from deviating from this situation, exclusion of principal 2 proceeds.

[22]Precisely, this Sanction-type of punishment is impossible if and only if $\delta < C/(G^2 + D)$ and it is obvious that $C/(G^2 + D)$ is greater than $C/(2D + C)$.

3.5 Validity of Menu Auction for Political Contribution

This section discusses whether BW's menu auction is suitable for analyses of political contribution. As mentioned in the introduction, the menu auction assumes that the principals can fully commit to a compensation plan contingent on the agent's decision, which is not as realistic as our self-enforcing situation. If an equilibrium in the menu auction can be replicated by our equilibrium, the long term relationship between players can be a good justification for analysing political contribution using the menu auction model. However, if it cannot, we should keep in mind that the lack of enforcement power is a serious concern in studying political contribution. In order to address this question, we will compare the set of stationary equilibrium payoffs in our model with that of BW. More specifically, we derive the conditions under which the equilibrium payoff in BW is not included in our stationary equilibrium.

3.5.1 BW's Model

For reader's convenience, we first describe the corresponding static model of BW. Hereafter we call it "Static Menu Auction (SMA)" and an equilibrium in it a SMA-equilibrium. By contrast, we will call our repeated game environment "Relational Political Contribution (RPC)".

SMA is the following one-shot game. First, each principal simultaneously and non-cooperatively offers a contract $w^j : \mathcal{A} \to \mathbf{R}_+$, that is the payment schedule contingent on the agent's decision and then the agent accepts it.[23] Second, the agent makes a decision and the payment is enforced. Let $(\hat{a}, \{\{\hat{w}^j(a)\}_{a\in\mathcal{A}}\}_{j=1}^N)$ be the equilibrium decision and the payment contract contingent on the decisions in SMA and $(\hat{y}^0, \hat{y}^1, \ldots, \hat{y}^N) := (v^0(\hat{a}) +$

[23]Because we assume the nonnegative payment, the agent automatically accepts any offer.

$\hat{W}(\hat{a}), v^1(\hat{a}) - \hat{w}^1(\hat{a}), \ldots, v^N(\hat{a}) - \hat{w}^N(\hat{a}))$ be the equilibrium net payoff in SMA. While Lemma 2 in BW derives the necessary and sufficient condition for SMA-equilibria, the following lemma is sufficient for our analysis.

Lemma 3.5 *Let $(\hat{a}, \hat{w}^1(\cdot), \ldots, \hat{w}^N(\cdot))$ be a SMA-equilibrium and $(\hat{y}^0, \hat{y}^1, \ldots, \hat{y}^N)$ be its payoff vector. Then $\hat{y}^0 \geq \overline{v}^0$ and $\hat{y}^j \geq \underline{v}^j$ for each $j \in \mathcal{N}$.*

3.5.2 Comparison between SMA and RPC

Recall that, thanks to Proposition 3.1, all the decision-stationary RPC-equilibria can be expressed by stationary strategies consisting of the decision and the payment on the path, say $(\hat{a}, \{\hat{\beta}^j(\hat{a})\}_{j=1}^N)$. If there exists a stationary RPC-equilibrium such that $\hat{a}$ is chosen and $\hat{w}^j(\hat{a})$ is paid from principal j to the agent, the net payoff vector of the SMA-equilibrium $(\hat{a}, \hat{w}^1(\cdot), \ldots, \hat{w}^N(\cdot))$ can be supported by the RPC-equilibrium.

The set of decision-stationary RPC-equilibrium payoffs is characterized by (3.9). Then the question is whether a specific SMA-equilibrium payoff $(\hat{y}^0, \hat{y}^1, \ldots, \hat{y}^N)$ is included in it. Recall that due to Lemma 3.3, the OPC payoff can be described by $u^j(\underline{\sigma}(j)) = v^j(\underline{a}_0(j))$ for $j \in \mathcal{N}$ and $u^0(\underline{\sigma}(0)) = \overline{v}^0$. The following proposition says that the SMA equilibrium cannot be achieved by RPC-equilibria if and only if the amount of the equilibrium contribution exceeds some upper bound defined by $\hat{a}$ and $\underline{a}_0(j)$.

Proposition 3.7 *Let $(\hat{a}, \hat{w}^1(\cdot), \ldots, \hat{w}^N(\cdot))$ be a SMA-equilibrium and $(\hat{y}^0, \hat{y}^1, \ldots, \hat{y}^N)$ be its payoff vector. Then $(\hat{y}^0, \hat{y}^1, \ldots, \hat{y}^N) \notin \hat{U}^*(\hat{a})$ if and only if $\hat{w}^j(\hat{a}) > \delta[v^j(\hat{a}) - v^j(\underline{a}_0(j))]$ for some $j \in \mathcal{N}$.*

The term $\delta[v^j(\hat{a}) - v^j(\underline{a}_0(j))]$ is interpreted as the discounted benefit relative to deviation. If this value is not high enough to exceed the payment on the SMA-equilibrium $\hat{w}^j(\hat{a})$, she cannot credibly pay this amount in RPC-equilibria. Note that this upper bound is weakly increasing in δ due to two effects; the direct effect through caring more about the future benefit (i.e. δ is increasing)[24] and the indirect effect through more severe punishment (i.e. $v^j(\underline{a}_0(j))$ is weakly decreasing).

The next question is whether all SMA-equilibria can be supported by RPC-equilibria if the discount factor is sufficiently high. The answer is somewhat surprisingly negative. First, the following proposition provides the necessary and sufficient condition for this notion.

Proposition 3.8 *Let $(\hat{a}, \hat{w}^1(\cdot), \ldots, \hat{w}^N(\cdot))$ be a SMA-equilibrium. Then $(\hat{y}^0, \hat{y}^1, \ldots, \hat{y}^N) \notin \hat{U}^*(\hat{a})$ for any $\delta \in [0, 1)$ if and only if there exists $j \in \boldsymbol{N}$ such that*

$$\hat{w}^j(\hat{a}) = v^j(\hat{a}) - \underline{v}^j > 0. \tag{3.19}$$

It can be interpreted as a direct implication from Proposition 3.7. For principal j, the possible worst payoff is $\underline{v}^j$. Then the upper bound of the credible payment from principal j is at most $v^j(\hat{a}) - \underline{v}^j$ no matter how much δ is. Hence it is never a RPC-equilibrium to pay an amount no less than $v^j(\hat{a}) - \underline{v}^j$ to the agent for rewarding decision $\hat{a}$. Thus, if a SMA-equilibrium requires principal j to pay an amount no less than $v^j(\hat{a}) - \underline{v}^j$ on the equilibrium, it can never be implemented in RPC-equilibria. By contrast, if the equilibrium payment in SMA is less than $v^j(\hat{a}) - \underline{v}^j$, then by making the players sufficiently patient, punishment by the minimax payoff $\underline{v}^j$ becomes credible and hence it is possible to require principal j to pay it even in RPC.

So far we have referred to any given strategy $(\hat{a}, \hat{w}^1(\cdot), \ldots, \hat{w}^N(\cdot))$ satisfying Lemma 3.5 as a SMA-equilibrium. An important question is whether there exists such a SMA-

[24]If $v^j(\hat{a}) < v^j(\underline{a}_0(j))$, then it is easy to show that there is no stationary RPC-equilibrium of $\hat{a}$. Thus δ has the direct effect only if $v^j(\hat{a}) > v^j(\underline{a}_0(j))$.

equilibrium which satisfies (3.19). We now demonstrate that in some cases there actually exists such a SMA-equilibrium. Specifically, we illustrate it by truthful equilibria proposed by BW. The truthful equilibria are defined as follows.

Definition 3.3 *The payment schedule $\hat{w}^j(\cdot)$ is truthful relative to $\hat{a}$ if $\hat{w}^j(\cdot)$ satisfies*

$$\hat{w}^j(a) = \max\{\hat{w}^j(\hat{a}) + v^j(a) - v^j(\hat{a}), 0\}$$

for all $a \in \mathcal{A}$. $(\hat{a}, \hat{w}^1(\cdot), \ldots, \hat{w}^N(\cdot))$ is a truthful equilibrium in SMA (henceforth SMAT-equilibrium) if this is a SMA-equilibrium and $\hat{w}^j(\cdot)$ is truthful relative to $\hat{a}$ for all $j \in \mathcal{N}$.

BW show that SMAT-equilibria have some appealing properties; (i) it could be focal in the sense that there is always a truthful strategy which is the best response, (ii) SMAT-equilibria are (Pareto) efficient, and (iii) on SMAT-equilibria, the principals have no incentive to deviate jointly (i.e. coalition-proof). Because there typically exist many SMA-equilibria, many applied works adopt the SMAT-equilibrium as an equilibrium refinement. As we will see, however, this appealing equilibrium is sometimes vulnerable to the commitment problem explicitly introduced in our model.

(3.19) consists of two conditions, $\hat{w}^j(\hat{a}) = v^j(\hat{a}) - \underline{v}^j$ and $v^j(\hat{a}) > \underline{v}^j$. Note that from Lemma 3.5 if $\hat{w}^j(\cdot)$ is a SMA-equilibrium contract, then $\hat{w}(\hat{a})$ must be no more than $v^j(\hat{a}) - \underline{v}^j$. Thus the first condition can be interpreted as that where principal j pays the maximum payment on SMA-equilibria. On SMAT-equilibria, the sufficient condition for it is characterized as follows.

Lemma 3.6 *Let $(\hat{a}, \hat{w}^1(\cdot), \ldots, \hat{w}^N(\cdot))$ be a SMAT-equilibrium. Then $\hat{w}^j(\hat{a}) = v^j(\hat{a}) - \underline{v}^j$ if $\underline{A}^j \cap A^* \neq \emptyset$.*

The intuition is obtained from the sketch of the proof of the lemma. If there exists a decision $\underline{a}^j \in \underline{A}^j \cap A^*$, choosing $\underline{a}^j$ maximizes the total benefit. Moreover, since $\underline{a}^j$ minimizes principal j's benefit, the aggregate benefit in the coalition of all the players except for principal j (i.e. $\sum_{i \neq j} v^i(a)$) is also maximized. Then in SMA all the players except for principal j wish to implement $\underline{a}^j$ and are willing to reward for $\underline{a}^j$ rather than the other decisions. In order to avoid being "isolated" by $\underline{a}^j$ chosen, principal j has to pay the maximum credible amount.

In what follows, we call the set $\underline{A}^j \cap A^*$ and abusively its element "completely isolating decision of j". Now suppose that there exist two principals, say j and k, the completely isolating decisions of which exist, i.e. $\underline{A}^j \cap A^* \neq \emptyset$ and $\underline{A}^k \cap A^* \neq \emptyset$. If their preference is not congruent in the sense that they do not share the completely isolating decisions, i.e. $A^* \cap \underline{A}^j \cap \underline{A}^k = \emptyset$, then, for any $a^* \in A^*$, either $a^* \notin \underline{A}^j$ or $a^* \notin \underline{A}^k$ (or both). Since any SMAT-equilibrium decision must be socially efficient, i.e. $\hat{a} \in A^*$, either $v^j(\hat{a}) > \underline{v}^j$ or $v^k(\hat{a}) > \underline{v}^k$ must be satisfied. Then all possible SMAT-equilibria have principal j or k which satisfies (3.19).

Proposition 3.9 *Suppose that there exist $j, k \in \boldsymbol{N}$ (where $j \neq k$) such that $\underline{A}^j \cap A^* \neq \emptyset$, $\underline{A}^k \cap A^* \neq \emptyset$ and $A^* \cap \underline{A}^j \cap \underline{A}^k = \emptyset$. Then for any $\delta \in [0, 1)$ there are no RPC-equilibria which attain the SMAT-equilibrium payoff vector.*

To see the intuition further, suppose that $N = 2$ and there is a completely isolating decision of principal 2. Since it maximizes the joint benefit of the agent and principal 1, they prefer the completely isolating decision of principal 2 the most and since there is no enforcement cost in SMA, they would readily agree to the contract, which leads to the completely isolating decision of principal 2 unless principal 2 is there. This agreement works as a credible threat to exploit principal 2 by the agent. At the same time, if there is also a completely isolating decision of principal 1, the agent can make the same credible threat to principal 1. Thus the agent can exploit both principals.

If $\hat{a}$ is the SMAT-equilibrium decision, the exploitation has either of the following forms: i) the agent actually chooses the completely isolating decision (i.e. $\hat{a} \in \underline{A}^j \cap A^*$), or ii) the agent receives the maximum credible amount of transfer (i.e. $\hat{w}^j(\hat{a}) = v^j(\hat{a}) - \underline{v}^j > 0$). Now if $A^* \cap \underline{A}^1 \cap \underline{A}^2 = \emptyset$, the latter necessarily happens on the SMAT-equilibrium because the former is never satisfied for both $j = 1, 2$ simultaneously. However according to Proposition 3.8, the latter is impossible in the RPC.

Proposition 3.7 shows that the upper bound for principal j on RPC-equilibria of $\hat{a}$ is at most $\delta[v^j(\hat{a}) - \underline{v}^j]$ which never exceeds $v^j(\hat{a}) - \underline{v}^j$ for any $\delta \in [0, 1)$. The reason is that even if the players are patient enough to generate the strongest punishment payoff on principal j, $\underline{v}^j$, this effect is discounted by δ. In the RPC, reneging is punished endogenously by the players and the punishment is delayed by one period. Then the effect of the punishment is necessarily discounted and it does not allow principal j to pay $v^j(\hat{a}) - \underline{v}^j$ in the RPC.

It can be interpreted that under the condition in Proposition 3.9, principal j and k are in conflict *via* their completely isolating decisions. The idea that SMAT-equilibria are vulnerable due to conflict between the principals can be confirmed by redistributive politics specified as follows. Suppose that $s(a) = s^*$ for all $a \in \mathcal{A}$ so that the total benefit does not change at all and the decision determines only the distribution of the benefit among the players. Under this condition any non-trivial SMAT-equilibrium payoff – in that the agent's choice is indeed altered by the principals' payment schedule on the equilibrium – cannot be supported by RPC-equilibria.

Proposition 3.10 *Suppose that $s(a) = s^*$ for all $a \in \mathcal{A}$. Then any SMAT-equilibrium payoff where $\hat{a} \notin \overline{A}^0$ cannot be supported by RPC-equilibria of $\hat{a}$ for any $\delta \in [0, 1)$.*

Proposition 3.9 and 3.10 implies that the assumption of sufficient patience is sometimes

Decision	$v^1(a)$	$v^0(a)$	$v^2(a)$
l	$C+D$	$-C$	$-D$
c	0	0	0
r	$-D$	$-C$	$C+D$

Table 3.2: Example 2

not enough to justify the SMA model in political economy, especially when conflict among the players is severe.

3.5.3 Example

We revisit the example discussed in section 3.4 to demonstrate the results above. Now further assume that $G^1 = G^2 = C+D$ and then the benefit function is given in Table 3.2. Then a unique SMAT-equilibrium is[25][26] $(\hat{w}^1(l), \hat{w}^1(c), \hat{w}^1(r)) = (2D + C, D, 0)$, $(\hat{w}^2(l), \hat{w}^2(c), \hat{w}^2(r)) = (0, D, 2D + C)$, and the agent chooses c. The SMAT-equilibrium net payoff vector is then given by $(\hat{y}^0, \hat{y}^1, \hat{y}^2) = (2D, -D, -D)$.

This SMAT-equilibrium payoff cannot be achieved by any RPC-equilibria for any $\delta \in [0, 1)$. Notice that $s(c) = s(l) = s(r)$ and $\hat{w}^j(c) > 0$ for $j = 1, 2$, implying that the condition in Proposition 3.10 is satisfied.[27] Recall that from (3.9) the lower bound of the RPC-equilibrium payoff of c for principal j is $(1 - \delta)v^j(c) + \delta v^j(\underline{a}_0(j))$ which satisfies

$$(1 - \delta)v^j(c) + \delta v^j(\underline{a}_0(j)) = (1 - \delta) \cdot 0 + \delta v^j(\underline{a}_0(j)) = \delta v^j(\underline{a}_0(j)) \geq -\delta D > -D$$

for any $\delta \in [0, 1)$. Then on all the RPC-equilibria of c, principal j's net payoff must be

[25]Bergemann and Välimäki (2003) propose the marginal contribution equilibrium and they show that if the game has a marginal contribution equilibrium, then a SMAT-equilibrium payoff is unique and coincides with that of the marginal contribution equilibrium. By applying this property, it can be seen that a SMAT-equilibrium decision is uniquely given as c. See Bergemann and Välimäki (2003) for the detail of the marginal contribution equilibrium.

[26]It is still a SMAT-equilibrium as long as $G^j \leq C + D$ for $j = 1, 2$.

[27]It is easy to confirm that the conditions in Proposition 3.8 and 3.9 also hold.

strictly greater than the SMAT-equilibrium.

3.6 Discussion

3.6.1 Approximation

The above example has another interpretation that provides the opposite meaning. Given sufficiently high δ, according to (3.9), the RPC-equilibrium payoff that attains the lowest payoff for the principals is $(u^0, u^1, u^2) = (2\delta D, -\delta D, -\delta D)$. Then the principals' payoff is never less than $-D$ but the vector approaches $(2D, -D, -D)$ as δ goes to 1. In this sense, every SMA-equilibrium payoff can be approximately implemented in the RPC-equilibrium.

This approximation is in general true. Namely, even if the SMA-equilibrium payoff satisfies the condition in Proposition 3.9, it can be implemented in the RPC-equilibrium approximately.[28] If the approximation is appropriate, the assertion that the SMA-equilibrium could be vulnerable is no longer valid. Thus the validity of the approximation is an important factor for the economics implication.

The approximation is not realistic for the following reasons. First, to assume δ close to one seems to be unrealistic in many applications. The discount factor δ would be interpreted as the probability of continuing the same period game in the next period and the players obtain zero payoff on average when the game ends. Since many political positions stipulate a finite length for the term of office, the probability of continuing the game in the long run is obviously far from 1. Second, even if it is appropriate to assume that δ would be almost 1, the approximation can be achieved only if both the principals and

[28]It is directly from the folk theorem in Wen (2002). Since the payoff vector some of the components of which contain the exact minimax value is just a neighbourhood of the payoff which is strictly greater than the minimax payoff which can be supported for sufficiently high δ by his folk theorem.

the agents can coordinate their behaviour, especially for punishment. There are typically multiple SPE and then they need coordination for "choosing" the future equilibrium behaviour. Third, although we have not assumed any kinds of renegotiation-proofness on the equilibrium behaviour, it is often taken into consideration in the theory of relational contracts since it is natural to imagine that contracting parties in a long term relationship frequently meet each other. Imposing renegotiation-proofness on the equilibrium could makes the OPC weaker, which leads the impossibility of approximation.

In summary, the approximation is possible only when the players can overcome these difficulties; the players are fairly patient, rational in the sense that they can coordinate in a sophisticated way, and able to commit themselves not to renegotiate with each other. It would be unrealistic in many situations. Furthermore, Proposition 3.9 and 3.10 show that it could be the case that the SMA-equilibrium cannot be supported exactly even if the difficulties mentioned here can be ignored. From these factors, we would conclude that every SMA-equilibrium is not necessarily robust to the lack of commitment.

3.6.2 Conflict Makes the SMA Vulnerable

Proposition 3.9 and 3.10 clarify that conflict between principals makes it difficult to justify the SMAT-equilibria. This argument can be seen even if the conditions in Proposition 3.9 or 3.10 are not satisfied. In order to understand it, look at Example 3.1 with $0 < G^j - C < D$ for $j = 1, 2$. It is verified that a unique SMAT-equilibrium payoff is given by $(\hat{y}^0, \hat{y}^1, \hat{y}^2) = (G^1 + G^2 - 2C, -G^2 + C, -G^1 + C)$ and it is replicated by RPC-equilibria of c only if $\delta \geq \max\{(G^1 - C)/D, (G^2 - C)/D\}$.[29] Here $G^j - C$ is the aggregate benefit between the agent and principal j and as $G^j - C$ is larger, the total benefit of l and r is approaching that of c,

[29] Additionally, to assure the effective punishment, it is required that $\delta \geq C/(2D + C)$. See section 3.4.2.

meaning that the environment becomes more similar to the one described in Proposition 3.10. By observing the SMAT-equilibrium payoff vector, we see that as $G^j - C$ is larger, the agent can exploit principal $k(\neq j)$ more in the SMAT-equilibrium. At the same time, however, as $G^j - C$ becomes larger, δ must become higher to achieve the SMAT outcome in the RPC environment.

3.6.3 Other SMA Equilibria

While we have focused on the relation between SMAT-equilibria and RPC-equilibria in Proposition 3.9 and 3.10, we can also discuss the relation between SMA-equilibria (which are not necessarily SMAT-equilibria) and RPC-equilibria by Proposition 3.8. For instance, consider again Example 3.1 and suppose that $G^j - C - D < 0$ for $j = 1, 2$. Under this assumption, the following strategy profile is a SMA-equilibrium (which is not a SMAT-equilibrium); $(\hat{w}^1(l), \hat{w}^1(c), \hat{w}^1(r)) = (2D + C, D, 0)$, $(\hat{w}^2(l), \hat{w}^2(c), \hat{w}^2(r)) = (0, D, 2D + C)$, and the agent chooses c. It can be shown that this SMA-equilibrium satisfies the condition in Proposition 3.8 and then it cannot be supported in RPC-equilibria. It implies that there would be more environments where SMA-equilibria cannot be supported in RPC-equilibria than those conditioned by Proposition 3.9 and 3.10.

3.6.4 SMA with Caps on Transfer

The implication from the analysis of the RPC is that if the amount of compensation is restricted to be below some upper bound, SMA-equilibria could be implemented even if there is a lack of commitment *ex ante*. It means that one of the appropriate ways to analyse lobbying by SMA is to introduce an upper bound on the amount of compensation.

Consider Example 3.1 again with $G_1 = G_2 = G$. Now assume that there is a (common) upper bound $L \geq 0$ below which the principals' compensation must be, i.e. $w^j(a) \leq L$ for $a = l, c, r$ and $j = 1, 2$. This is a special case of Dixit et al. (1997) and the conditions for truthful equilibria are characterized by Proposition 3 in their paper. According to the conditions, it can be shown that when $L \geq 2G - C$, the SMAT-equilibrium satisfies $(\hat{w}^1(l), \hat{w}^1(c), \hat{w}^1(r)) = (2G - C, G - C, 0)$ and $(\hat{w}^2(l), \hat{w}^2(c), \hat{w}^2(r)) = (0, G - C, 2G - C)$ as before. By contrast, when $C < L < 2G - C$, the compensation schedule in the truthful equilibrium becomes $(\hat{w}^1(l), \hat{w}^1(c), \hat{w}^1(r)) = (L, (L - C)/2, 0)$ and $(\hat{w}^2(l), \hat{w}^2(c), \hat{w}^2(r)) = (0, (L - C)/2, L)$.

It means that the amount to be paid to the agent in the equilibrium is (weakly) decreasing as L is decreasing. However this is not necessarily due to the direct effect in that the principals' equilibrium payment is capped by L. In fact, if $G - C \leq L < 2G - C$, it is obviously possible for the principals to commit to pay $G - C$ when the agent chooses c.

Instead of the direct effect of caps on transfer, decreasing the amount of compensation in equilibrium is caused by the strategic effect. When $L < 2G - C$, principal 2 cannot commit to pay $2G - C$ when the agent chooses r and then the agent is less incentivized to choose r. Since r is the least preferred decision for principal 1, she can avoid r being implemented without a large amount of compensation for the other decisions. In other words, principal 2's inability to commit lowers principal 1's compensation.

This chapter argues the importance of commitment ability and it is captured by the maximum amount to be committed to the compensation. Thus the SMA-model with such caps on transfer should be adopted in the analysis of lobbying. It is actually a special case of the model of Dixit et al. (1997).[30] However, even though the framework has already been established by them, as discussed above, there have been no papers taking into account the upper bounds explicitly. Including the generalization of the above discussion, this

[30] Although their most important innovation seems to be to allow the payoff functions to be not quasi-linear in the transfer, it should be argued that to allow the institutional restriction on the compensation schedule is also important for analyses of lobbying.

should be a future research topic.

3.7 Conclusion

This chapter has formally analysed political contribution delivered only *via* self-enforcing agreements that is abstracted from in most of the literature. There are two main results. First, the optimal punishment is either an “Exclusion-type” in which the agent excludes the deviating principal forever or a “Sanction-type” in which a short-term punishing decision and fine are implemented. Second, when the contract must be self-enforcing, the outcome in the menu auction is not necessarily guaranteed even if the players are patient.

We conclude the chapter by proposing a direction for future research. The model presented in this chapter is a first step toward relational contracts under common agency. Incorporating asymmetric information would be an interesting and important extension. Moreover since our approach to characterize equilibria does not require patient players (i.e. our objective was not focused on proving the Folk theorem), it can be applied to many application works of special interest politics. In particular, our analysis on the OPC and comparison with SMA suggests that conflict in the principals’ preference could give us a new insight on political contributions which cannot be obtained from the menu auction model. Then embedding a political factor which causes conflict among lobbyists would stimulate the discussion on special interest politics.

3.8 Appendix: Proofs

3.8.1 Proof of Proposition 3.1

For an arbitrary decision-stationary strategy $\hat{\sigma}$ of $\hat{a}$, let principal j's stationary strategy (payment schedule) be

$$\hat{\beta}^j(\hat{a}) \quad := \quad \hat{\sigma}^j(\hat{a}) + \frac{\delta}{1-\delta}\left[u^j(\hat{\sigma}) - u^j(\hat{\sigma}_1)\right]$$

and $\hat{\beta}^j(a) = 0$ if $a \neq \hat{a}$, and consider the following stationary strategy profile;

- the agent always chooses $\hat{a}$ as long as no player has deviated,
- principal j pays $\hat{\beta}^j(\cdot)$ described above every period as long as no player has deviated,

and all players play the OPC once some player deviated.

Since the new strategy profile is also decision-stationary of $\hat{a}$, the total surplus does not change. Then, if each of the principals obtains the same payoff as $\hat{\sigma}$, the agent's payoff is also the same. Under this stationary strategy, principal j's average payoff is

$$\begin{aligned}
(1-\delta)\sum_{t=0}^{\infty}\delta^t[v^j(\hat{a}) - \hat{\beta}^j(\hat{a})] &= v^j(\hat{a}) - \hat{\beta}^j(\hat{a}) \\
&= v^j(\hat{a}) - \hat{\sigma}^j(\hat{a}) - \frac{\delta}{1-\delta}\left[u^j(\hat{\sigma}) - u^j(\hat{\sigma}_1)\right] \\
&= \frac{1}{1-\delta}\left[(1-\delta)(v^j(\hat{a}) - \hat{\sigma}^j(\hat{a})) + \delta u^j(\hat{\sigma}_1) - \delta u^j(\hat{\sigma})\right] \\
&= \frac{1}{1-\delta}\left[u^j(\hat{\sigma}) - \delta u^j(\hat{\sigma})\right] = u^j(\hat{\sigma}).
\end{aligned}$$

Thus the payoff vector is the same as $\hat{\sigma}$.

Now, we will show that the new stationary strategy is a SPE. Since the chosen decision and payment schedule are completely identical in each period, the on-path continuation strategy profile is the same among every periods. Thus according to Lemma 3.2, if the new stationary strategy satisfies (i) and (ii), then it also satisfies (iii) since there is no incentive to deviate after period 1 and then it is a SPE. Since, by construction, the continuation payoff is given by $u^i(\hat{\sigma})$ for $i \in \{0\} \cup \mathcal{N}$, the new stationary strategy satisfies (i) if and only if

$$\begin{aligned}
& \sum_{j\in\mathcal{N}} \hat{\beta}^j(\hat{a}) \geq \overline{v}^0 - v^0(\hat{a}) - \frac{\delta}{1-\delta}\left[u^0(\hat{\sigma}) - u^0(\underline{\sigma}(0))\right] \\
\Longleftrightarrow \quad & \hat{\psi}(\hat{a}) + \frac{\delta}{1-\delta}\sum_{j\in\mathcal{N}}\left[u^j(\hat{\sigma}) - u^j(\hat{\sigma}_1)\right] \geq \overline{v}^0 - v^0(\hat{a}) - \frac{\delta}{1-\delta}\left[u^0(\hat{\sigma}) - u^0(\underline{\sigma}(0))\right] \\
\Longleftrightarrow \quad & \hat{\psi}(\hat{a}) + \frac{\delta}{1-\delta}\left[\sum_{j=0}^{N} u^j(\hat{\sigma}) - \sum_{j=0}^{N} u^j(\hat{\sigma}_1)\right] \geq \overline{v}^0 - v^0(\hat{a}) - \frac{\delta}{1-\delta}\left[u^0(\hat{\sigma}_1) - u^0(\underline{\sigma}(0))\right]
\end{aligned}$$

in which $\sum_{j=0}^{N} u^j(\hat{\sigma}) = \sum_{j=0}^{N} u^j(\hat{\sigma}_1) = s(\hat{a})$ because both $\hat{\sigma}$ and $\hat{\sigma}_1$ are decision-stationary strategy. Thus (3.3) assures that it is true. The new stationary strategy satisfies (ii) for $j \in \mathcal{N}$ if and only if

$$\begin{aligned}
\hat{\beta}^j(\hat{a}) \leq \frac{\delta}{1-\delta}[u^j(\hat{\sigma}) - u^j(\underline{\sigma}(j))] \quad \Longleftrightarrow \quad & \hat{\sigma}^j(\hat{a}) + \frac{\delta}{1-\delta}\left[u^j(\hat{\sigma}) - u^j(\hat{\sigma}_1)\right] \leq \frac{\delta}{1-\delta}\left[u^j(\hat{\sigma}) - u^j(\underline{\sigma}(j))\right] \\
\Longleftrightarrow \quad & \hat{\sigma}^j(\hat{a}) \leq \frac{\delta}{1-\delta}\left[u^j(\hat{\sigma}_1) - u^j(\underline{\sigma}(j))\right],
\end{aligned}$$

which is implied by (3.4) and

$$\begin{aligned}
\hat{\beta}^j(\hat{a}) &= \hat{\sigma}^j(\hat{a}) + \frac{\delta}{1-\delta}\left[u^j(\hat{\sigma}) - u^j(\hat{\sigma}_1)\right] \\
&= \hat{\sigma}^j(\hat{a}) + \frac{\delta}{1-\delta}\left[(1-\delta)(v^j(\hat{a}) - \hat{\sigma}^j(\hat{a})) + \delta u^j(\hat{\sigma}_1) - u^j(\hat{\sigma}_1)\right] \\
&= \hat{\sigma}^j(\hat{a}) + \delta\left[v^j(\hat{a}) - \hat{\sigma}^j(\hat{a}) - u^j(\hat{\sigma}_1)\right] \\
&\geq \hat{\sigma}^j(\hat{a}) - \delta\hat{\sigma}^j(\hat{a}) = (1-\delta)\hat{\sigma}^j(\hat{a}) \geq 0
\end{aligned}$$

where the first inequality is due to decision stationarity of $\hat{\sigma}_1$ and the condition of non-negative payment implying that $v^j(\hat{a}) \geq u^j(\hat{\sigma}_1)$. Therefore the new stationary strategy is a SPE.

3.8.2 Proof of Proposition 3.2

It is enough to show the sufficiency. First of all, we will prove the following lemma.

Lemma 3.7 *If $\hat{a} \in \mathcal{A}$ satisfies (3.10), then $v^j(\hat{a}) \geq u^j(\underline{\sigma}(j))$ for all $j \in \mathcal{N}$.*

Proof (Lemma 3.7) *Let*

$$\underline{J} := \{j \in \mathcal{N} \mid v^j(\hat{a}) < u^j(\underline{\sigma}(j))\}$$

and $\overline{J} := \mathcal{N} \setminus \underline{J}$. Suppose that $\underline{J} \neq \emptyset$. Since $\hat{a}$ satisfies (3.10),

$$\begin{aligned}
&\frac{\delta}{1-\delta}\left[\sum_{j\in\overline{J}}[v^j(\hat{a}) - u^j(\underline{\sigma}(j))] + \sum_{j\in\underline{J}}[v^j(\hat{a}) - u^j(\underline{\sigma}(j))] + v^0(\hat{a}) - u^0(\underline{\sigma}(0))\right] \geq \overline{v}^0 - v^0(\hat{a}) \\
\iff\ &\delta\left[\sum_{j\in\overline{J}} v^j(\hat{a}) - \sum_{j\in\overline{J}} u^j(\underline{\sigma}(j))\right] \geq (1-\delta)\overline{v}^0 + \delta u^0(\underline{\sigma}(0)) - v^0(\hat{a}) - \delta\sum_{j\in\underline{J}}\left[v^j(\hat{a}) - u^j(\underline{\sigma}(j))\right],
\end{aligned}$$

which implies that, by the definition of $\underline{J}$,

$$\delta\left[\sum_{j\in\overline{J}} v^j(\hat{a}) - \sum_{j\in\overline{J}} u^j(\underline{\sigma}(j))\right] \geq (1-\delta)\overline{v}^0 + \delta u^0(\underline{\sigma}(0)) - v^0(\hat{a}). \tag{3.20}$$

Now we will construct a stationary strategy which is a SPE. Here, however, we do not use the simple strategy. Instead we use the following modified simple strategy where the players ignore *deviation by principal $j \in \underline{J}$ on the equilibrium path. Specifically, let*

$$\beta^j(\hat{a}) \;:=\; \begin{cases} \delta\left[v^j(\hat{a}) - u^j(\underline{\sigma}(j))\right] & \text{for} \quad j \in \overline{J} \\ 0 & \text{for} \quad j \in \underline{J}, \end{cases}$$

$\beta^j(a) = 0$ for all $a \neq \hat{a}$ and $j \in \mathcal{N}$. In addition, let $\mathrm{B}(\hat{a}) := \sum_{j \in \overline{J}} \beta^j(\hat{a})$. Notice that $\beta^j \geq 0$ for all $j \in \mathcal{N}$. The agent's strategy is

- *to choose $\hat{a}$ in the first period and keep it if no players have deviated,*
- *to change to $\underline{\sigma}^0(j)$ if principal $j \in \overline{J}$ deviated from paying $\beta^j(\hat{a})$, and*
- *to change to $\underline{\sigma}^0(0)$ if the agent deviated from $\hat{a}$.*
- *(Even if principal $j' \in \underline{J}$ deviated from the on-equilibrium path, he keeps to choose $\hat{a}$).*

Principal $j (\in \mathcal{N})$'s strategy is

- *to pay $\beta^j(\hat{a})$ in each period if no players have deviated,*
- *to change to $\underline{\sigma}^j(j')$, if principal $j' \in \overline{J}$ deviated, and*
- *to pay nothing and change to $\underline{\sigma}^j(0)$ if the agent deviated from $\hat{a}$.*
- *(Even if principal $j'' \in \underline{J}$ deviated from the on-equilibrium path, she keeps to pay $\beta^j(\hat{a})$).*

Because the strategy is stationary, each of the players gets the same net payoff in each period ($v^0(\hat{a}) + \mathrm{B}(\hat{a})$ for the agent and $v^j(\hat{a}) - \beta^j(\hat{a})$ for principal j) and their continuation payoff is also the same in each period. Furthermore, since $\underline{\sigma}(i)$ is a SPE for all $i \in \mathcal{N} \cup \{0\}$, the incentive on $\underline{\sigma}(i)$ does not need to be checked. Thus in order to check that it is a SPE, we only have to check (i) the agent does not deviate in period 0 and (ii) each of the principals does not deviate in period 0 given $a_0 = \hat{a}$.

The agent does not deviate in period 0 if and only if

$$
\begin{aligned}
& (1-\delta)\left(v^0(\hat{a}) + \mathrm{B}(\hat{a})\right) + \delta\left(v^0(\hat{a}) + \mathrm{B}(\hat{a})\right) \geq (1-\delta)\overline{v}^0 + \delta u^0(\underline{\sigma}(0)) \\
\iff \quad & \mathrm{B}(\hat{a}) \geq (1-\delta)\overline{v}^0 + \delta u^0(\underline{\sigma}(0)) - v^0(\hat{a}),
\end{aligned}
$$

which is equivalent to (3.20). Then (i) is satisfied.

For principal $j \in \overline{J}$, she does not deviate in period 0 if and only if

$$
\begin{aligned}
& (1-\delta)\left(v^j(\hat{a}) - \beta^j(\hat{a})\right) + \delta\left(v^j(\hat{a}) - \beta^j(\hat{a})\right) \geq (1-\delta)\left(v^j(\hat{a}) - 0\right) + \delta u^j(\underline{\sigma}(j)) \\
\iff \quad & \beta^j(\hat{a}) \leq \delta\left[v^j(\hat{a}) - u^j(\underline{\sigma}(j))\right],
\end{aligned}
$$

which is satisfied by the construction of $\beta^j(\hat{a})$. For principal $j \in \underline{J}$, when she follows $\beta^j(\hat{a}) = 0$, her payoff is $v^j(\hat{a})$. Moreover even if she deviates to $b^{j\prime} > 0$, her continuation payoff does not change from $v^j(\hat{a})$ since the players do not punish her. Then her payoff is

$$
(1-\delta)v^j(\hat{a}) - b^{j\prime} + \delta v^j(\hat{a}) = v^j(\hat{a}) - b^{j\prime},
$$

which is strictly less than the payoff without deviation, $v^j(\hat{a})$ for any $b^{j\prime} > 0$. Hence she prefers to follow this strategy. Therefore (ii) is also satisfied.

We have shown that the constructed stationary strategy is a SPE. Note that for $j \in \underline{J}$, the SPE payoff is $v^j(\hat{a})$ implying that, by the definition of $\underline{\sigma}(j)$, $v^j(\hat{a}) \geq u^j(\underline{\sigma}(j))$. It contradicts the hypothesis $v^j(\hat{a}) < u^j(\underline{\sigma}(j))$ we supposed first. □

Suppose that (3.10) holds. Now consider the following stationary strategy $\hat{\sigma}$;

- the agent always chooses $\hat{a}$ as long as no players have deviated,
- principal j pays $\hat{\beta}^j(\cdot)$ (described below) every period as long as no players have

deviated,

and all players play the OPC once some player deviated. $\hat{\beta}^j(\cdot)$ is defined as

$$\hat{\beta}^j(\hat{a}) := \delta\left[v^j(\hat{a}) - u^j(\underline{\sigma}(j))\right]$$

and $\hat{\beta}^j(a) = 0$ for all $a \neq \hat{a}$. Notice that due to Lemma 3.7, it is nonnegative.

Because the strategy is stationary, each of the players gets the same net payoff in each period and it is identical with the average payoff. Then $u^0(\hat{\sigma}) = v^0(\hat{a}) + \hat{\mathrm{B}}(\hat{a})$ where $\hat{\mathrm{B}}(\cdot) = \sum_{k=1} \hat{\beta}^k(\cdot)$ and $u^j(\hat{\sigma}) = v^j(\hat{a}) - \hat{\beta}^j(\hat{a})$ for $j \in \mathcal{N}$. Thus

$$\begin{aligned}
u^0(\hat{\sigma}) - (1-\delta)\overline{v}^0 - \delta u^0(\underline{\sigma}(0)) &= v^0(\hat{a}) + \hat{\mathrm{B}}(\hat{a}) - (1-\delta)\overline{v}^0 - \delta u^0(\underline{\sigma}(0)) \\
&= v^0(\hat{a}) + \sum_{j=1}^{N} \delta\left[v^j(\hat{a}) - u^j(\underline{\sigma}(j))\right] - (1-\delta)\overline{v}^0 - \delta u^0(\underline{\sigma}(0)) \\
&= (1-\delta)v^0(\hat{a}) + \delta\left[s(\hat{a}) - \sum_{j=0}^{N} u^j(\underline{\sigma}(j))\right] - (1-\delta)\overline{v}^0 \\
&= (1-\delta)\left[\frac{\delta}{1-\delta}\left(s(\hat{a}) - \sum_{j=0}^{N} u^j(\underline{\sigma}(j))\right) v^0(\hat{a}) - \overline{v}^0\right] \geq 0
\end{aligned}$$

where the last inequality is due to (3.10). It implies that (3.7) holds. For $j \in \mathcal{N}$,

$$\begin{aligned}
u^j(\hat{\sigma}) - (1-\delta)v^j(\hat{a}) - \delta u^j(\underline{\sigma}(j))) &= v^j(\hat{a}) - \hat{\beta}^j(\hat{a}) - (1-\delta)v^j(\hat{a}) - \delta u^j(\underline{\sigma}(j))) \\
&= v^j(\hat{a}) - \delta\left[v^j(\hat{a}) - u^j(\underline{\sigma}(j))\right] - (1-\delta)v^j(\hat{a}) - \delta u^j(\underline{\sigma}(j))) = 0
\end{aligned}$$

and $v^j(\hat{a}) - u^j(\hat{\sigma}) = \hat{\beta}^j(\hat{a}) \geq 0$. These imply that (3.8) holds.

Therefore, since the stationary strategy satisfies (3.7) and (3.8), it is a SPE implying the existence.

3.8.3 Proof of Proposition 3.3

Suppose that σ is a SPE and the associated equilibrium decision and payment path are given by $((a_0, a_1, \ldots), (b_0^1, \ldots, b_0^N), (b_1^1, \ldots, b_1^N), \ldots))$. Now define a sequence $\{\tilde{a}_t\}_{t=0}^{\infty}$ as follows; $\tilde{a}_0 := a_0$ and

$$\tilde{a}_t = \begin{cases} a_t & \text{if} \quad s(a_t) \geq s(\tilde{a}_{t-1}) \\ \tilde{a}_{t-1} & \quad \text{otherwise.} \end{cases}$$

Note that by construction $s(\tilde{a}_t) \geq s(\tilde{a}_{t-1})$ for all $t \geq 0$. Since $\mathcal{A}$ is compact, by Bolzano-Weierstrass theorem, we can choose a converging subsequence $\{\tilde{a}'_t\}_{t=0}^{\infty}$ of $\{\tilde{a}_t\}_{t=0}^{\infty}$. Let $\tilde{a}' := \lim_{t\to\infty} \tilde{a}'_t$. Now we will show that;

Claim 3.1 $s(\tilde{a}') \geq \sum_{j=0}^{N} u^j(\sigma_t)$ *for all* $t \geq 0$ *where* σ_t *is the on-path continuation strategy of* σ *in period* t.

Claim 3.2 $\hat{\Sigma}^*(\tilde{a}') \neq \emptyset$.

These claims imply that there exists a decision-stationary equilibrium of $\tilde{a}'$ which generates the total surplus at least $\sum_{j=0}^{N} u^j(\sigma)$.

Proof of Claim 3.1 Notice the following observations.

1. By construction of $\tilde{a}_t$, $s(\tilde{a}_t) \geq s(a_t)$ for all $t \geq 0$.

2. Since $\{\tilde{a}'_t\}_{t=0}^{\infty}$ is a subsequence of $\{\tilde{a}_t\}_{t=0}^{\infty}$, $s(\tilde{a}'_t) \geq s(\tilde{a}_t)$ and $s(\tilde{a}'_t) \geq s(\tilde{a}'_{t-1})$ for all $t \geq 0$.

These imply that $s(\tilde{a}'_t) \geq s(a_t)$ for all $t \geq 0$. Furthermore because $s(\cdot)$ is continuous and bounded and $\{s(\tilde{a}'_t)\}_{t=0}^{\infty}$ is a bounded and nondecreasing sequence, we obtain that for any

$t \geq 0$

$$s(\tilde{a}'_t) \leq \sup_{k \geq 0} s(\tilde{a}'_k) = \lim_{k \to \infty} s(\tilde{a}'_k) = s\left(\lim_{k \to \infty} \tilde{a}'_k\right) = s(\tilde{a}'), \tag{3.21}$$

implying that $s(a_t) \leq s(\tilde{a}')$ for any $t \geq 0$. It follows that

$$\begin{aligned}
\sum_{j=0}^{N} u^j(\sigma_t) &= \sum_{j=0}^{N} \left[(1-\delta) \sum_{k=t}^{\infty} \delta^{k-t} v^j(a_k)\right] \\
&= (1-\delta) \sum_{k=t}^{\infty} \sum_{j=0}^{N} \delta^{k-t} v^j(a_k) \\
&= (1-\delta) \sum_{k=t}^{\infty} \delta^{k-t} s(a_k) \\
&\leq (1-\delta) \sum_{k=t}^{\infty} \delta^{k-t} s(\tilde{a}') = s(\tilde{a}').
\end{aligned}$$

Proof of Claim 3.2 Since σ is a SPE, neither the agent nor the principals deviate in any period t, that is,

$$(1-\delta)(v^0(a_t) + B_t) + \delta u^0(\sigma_{t+1}) \geq (1-\delta)\left(\max_{a_t \in \mathcal{A}} v^0(a_t) + \delta u^0(\underline{\sigma}(0))\right) = (1-\delta)\overline{v}^0 + \delta u^0(\underline{\sigma}(0))$$

for the agent and

$$(1-\delta)(v^j(a_t) - b^j_t) + \delta u^j(\sigma_{t+1}) \geq (1-\delta)(v^j(a_t) - 0) + \delta u^j(\underline{\sigma}(j)).$$

for $j \in \mathcal{N}$. Combining them to eliminate b^j_t implies that

$$\frac{\delta}{1-\delta}\left[\sum_{j=0}^{N} u^j(\sigma_{t+1}) - \sum_{j=0}^{N} u^j(\underline{\sigma}(j))\right] \geq \overline{v}^0 - v^0(a_t)$$

for all t. Because $\sum_{j=0}^{N} u^j(\sigma_{t+1}) \leq s(\tilde{a}')$ by Claim 3.1 and $\tilde{a}'_t = a_s$ for some $s \geq 0$, we obtain that

$$\frac{\delta}{1-\delta}\left[s(\tilde{a}') - \sum_{j=0}^{N} u^j(\underline{\sigma}(j))\right] \geq \overline{v}^0 - v^0(\tilde{a}'_t)$$

for all $t \geq 0$. Since $\{\tilde{a}'_t\}_{t=0}^{\infty}$ is a converging sequence to $\tilde{a}'$ and $v^0(\cdot)$ is continuous, both sides are converging. Then the inequality still holds at the limit, which means that

$$\frac{\delta}{1-\delta}\left[s(\tilde{a}') - \sum_{j=0}^{N} u^j(\underline{\sigma}(j))\right] \;\; \geq \;\; \overline{v}^0 - \lim_{t\to\infty} v^0(\tilde{a}'_t) = \overline{v}^0 - v^0(\tilde{a}').$$

Hence by Proposition 3.2 we obtain Claim 3.2.

3.8.4 Proof of Lemma 3.4

Suppose that the $2n$-tuple $\boldsymbol{a} := (\underline{a}_0(1), \underline{a}_1(1), \underline{a}_0(2), \underline{a}_1(2), \ldots, \underline{a}_0(N), \underline{a}_1(N))$ is the solution of Problem (k) for all $k \in \mathcal{N}$ simultaneously. Let $\ell \in \arg\max_{j\in\mathcal{N}} s(\underline{a}_1(j))$. We show that for all $k \in \mathcal{N}$, $(\underline{a}_0(k), \underline{a}_1(\ell))$ is the solution of Problem (k).

Notice that for all $k \in \mathcal{N}$, $\underline{a}_1(k)$ is only in the constraint of Problem (k) and not in the objective function of Problem (k). Since $\underline{a}_0(k)$ achieves the minimum value in Problem (k), it is enough to check that $(\underline{a}_0(k), \underline{a}_1(\ell))$ satisfies the constraints of Problem (k), that is

$$\frac{\delta}{1-\delta}\left[s(\underline{a}_1(\ell)) - \left(\overline{v}^0 + \sum_{i=1}^{N} v^i(\underline{a}_0(i))\right)\right] \geq \overline{v}^0 - \min\left\{v^0(\underline{a}_0(k)), v^0(\underline{a}_1(\ell))\right\}. \tag{3.22}$$

Since $(\underline{a}_0(\ell), \underline{a}_1(\ell))$ satisfies the constraint of Problem (ℓ),

$$\begin{aligned}\frac{\delta}{1-\delta}\left[s(\underline{a}_1(\ell)) - \left(\overline{v}^0 + \sum_{i=1}^{N} v^i(\underline{a}_0(i))\right)\right] \;\; &\geq \;\; \overline{v}^0 - \min\left\{v^0(\underline{a}_0(\ell)), v^0(\underline{a}_1(\ell))\right\} \\ &\geq \;\; \overline{v}^0 - v^0(\underline{a}_1(\ell)) \qquad (3.23)\end{aligned}$$

and since $s(\underline{a}_1(\ell)) \geq s(\underline{a}_1(k))$ and $(\underline{a}_0(k), \underline{a}_1(k))$ satisfies the constraint of Problem (k),

$$\begin{aligned}
\frac{\delta}{1-\delta}\left[s(\underline{a}_1(\ell)) - \left(\overline{v}^0 + \sum_{i=1}^N v^i(\underline{a}_0(i))\right)\right] &\geq \frac{\delta}{1-\delta}\left[s(\underline{a}_1(k)) - \left(\overline{v}^0 + \sum_{i=1}^N v^i(\underline{a}_0(i))\right)\right] \\
&\geq \overline{v}^0 - \min\left\{v^0(\underline{a}_0(k)), v^0(\underline{a}_1(k))\right\} \\
&\geq \overline{v}^0 - v^0(\underline{a}_0(k)). \qquad (3.24)
\end{aligned}$$

(3.23) and (3.24) imply (3.22).

3.8.5 Proof of Proposition 3.5

For any $\boldsymbol{a}(\mathcal{N}), \boldsymbol{a}(\mathcal{N})' \in A_0^{PC}$, define $\boldsymbol{a}(\mathcal{N})'' := (a(1)'', \ldots, a(N)'')$ where $a(j)'' \in \arg\min_{a \in \{a(j), a(j)'\}} v^j(a)$ for all $j \in \mathcal{N}$. The following claim is useful to show the result.

Claim 3.3 *If $\boldsymbol{a}(\mathcal{N}), \boldsymbol{a}(\mathcal{N})' \in A_0^{PC}$, then $\boldsymbol{a}(\mathcal{N})'' \in A_0^{PC}$.*

Proof (Claim 3.3) *Suppose that without loss of generality* $\min_{k\in\mathcal{N}} v^0(a(k)) \leq \min_{k\in\mathcal{N}} v^0(a(k)')$. *Since $\boldsymbol{a}(\mathcal{N}) \in A_0^{PC}$, the following must hold; for any $\hat{a} \in A_1^{PC}(\boldsymbol{a}(\mathcal{N}))$,*

$$\frac{\delta}{1-\delta}\left[s(\hat{a}) - \left(\overline{v}^0 + \sum_{i=1}^N v^i(a(i))\right)\right] \geq \overline{v}^0 - \min_{k\in\mathcal{N}} v^0(a(k)) \qquad (3.25)$$

$$\frac{\delta}{1-\delta}\left[s(\hat{a}) - \left(\overline{v}^0 + \sum_{i=1}^N v^i(a(i))\right)\right] \geq \overline{v}^0 - v^0(\hat{a}). \qquad (3.26)$$

Since $\sum_{i=1}^N v^i(a(i)'') \leq \sum_{i=1}^N v^i(a(i))$ by the definition of $\boldsymbol{a}(\mathcal{N})''$, (3.25) and (3.26) imply that

$$\frac{\delta}{1-\delta}\left[s(\hat{a}) - \left(\overline{v}^0 + \sum_{i=1}^N v^i(a(i)'')\right)\right] \geq \overline{v}^0 - \min_{k\in\mathcal{N}} v^0(a(k)) \qquad (3.27)$$

$$\frac{\delta}{1-\delta}\left[s(\hat{a}) - \left(\overline{v}^0 + \sum_{i=1}^N v^i(a(i)'')\right)\right] \geq \overline{v}^0 - v^0(\hat{a}). \qquad (3.28)$$

If $\min_{k\in\mathcal{N}} v^0(a(k)'') \geq \min_{k\in\mathcal{N}} v^0(a(k))$, *then (3.27) and (3.28) imply that $\hat{a} \in A_1^{PC}(\boldsymbol{a}(\mathcal{N})'')$ and*

$\boldsymbol{a}(\mathcal{N})'' \in A_0^{PC}$ and the proof is completed.

Pick up $\ell \in \arg\min_{k\in\mathcal{N}} v^0(a(k)'')$. Notice that for each $j \in \mathcal{N}$, either $v^0(a(j)'') = v^0(a(j))$ or $v^0(a(j)'') = v^0(a(j)')$. If $v^0(a(\ell)'') = v^0(a(\ell))$, then

$$\min_{k\in\mathcal{N}} v^0(a(k)'') = v^0(a(\ell)'') = v^0(a(\ell)) \geq \min_{k\in\mathcal{N}} v^0(a(k))$$

and if $v^0(a(\ell)'') = v^0(a(\ell)')$, then

$$\min_{k\in\mathcal{N}} v^0(a(k)'') = v^0(a(\ell)'') = v^0(a(\ell)') \geq \min_{k\in\mathcal{N}} v^0(a(k)') \geq \min_{k\in\mathcal{N}} v^0(a(k)).$$

Therefore in both cases $\min_{k\in\mathcal{N}} v^0(a(k)'') \geq \min_{k\in\mathcal{N}} v^0(a(k))$. $\square$

Let

$$A_0^{OPC} := \left\{ \boldsymbol{a}(\mathcal{N}) \in A_0^{PC} \,\middle|\, \nexists \tilde{\boldsymbol{a}}(\mathcal{N})' \in A_0^{PC} \text{ such that } \begin{array}{ll} {}^\forall j \in \mathcal{N}, v^j(\tilde{a}(j)) \leq v^j(a(j)) & \text{and} \\ {}^\exists k \in \mathcal{N}, v^k(\tilde{a}(k)) < v^k(a(k)) & \end{array} \right\},$$

that is, the Pareto-frontier of A_0^{PC}. The proof is completed if $A_0^{OPC} \neq \emptyset$ and for any $\boldsymbol{a}(\mathcal{N}), \boldsymbol{a}(\mathcal{N})' \in A_0^{OPC}$, $v^j(a(j)) = v^j(a(j)')$ for all $j \in \mathcal{N}$. Since $\overline{\boldsymbol{a}} \in A_0^{PC}$ where $\overline{\boldsymbol{a}} := (\overline{a}^0, \ldots, \overline{a}^0) \in \overline{a}^0$, A_0^{PC} is nonempty. Moreover, it is obviously compact. These imply that A_0^{OPC} is also nonempty. Claim 3.3 implies that for any $\boldsymbol{a}(\mathcal{N}), \boldsymbol{a}(\mathcal{N})' \in A_0^{PC}$, $\boldsymbol{a}(\mathcal{N})'' \in A_0^{PC}$ and by the definition of $\boldsymbol{a}(\mathcal{N})'' \in A_0^{PC}$, $v^\ell(a(\ell)'') = \min\{v^\ell(a(\ell)), v^\ell(a(\ell)')\}$ for all $\ell \in \mathcal{N}$. Then, if, for some $\boldsymbol{a}(\mathcal{N}), \boldsymbol{a}(\mathcal{N})' \in A_0^{OPC}$, there exists $k \in \mathcal{N}$ such that (without loss of generality) $v^k(a(k)) < v^k(a(k)')$, then it can be seen that $v^j(a(k)'') = \min\{v^k(a(k)), v^k(a(k)')\} < v^k(a(k)')$ and $v^i(a(i)'') \leq v^i(a(i)')$ for all $i \in \mathcal{N}$. It contradicts the hypothesis that $\boldsymbol{a}(\mathcal{N})' \in A_0^{OPC}$. Hence $v^j(a(j)) = v^j(a(j)')$ for all $j \in \mathcal{N}$.

3.8.6 Proof of Corollary 3.1

It is a direct implication from Lemma 3.3 and Proposition 3.4.

3.8.7 Proof of Proposition 3.6

1. For treating δ explicitly, change to denote $A_1^{PC}(\boldsymbol{a}(\mathcal{N}))$, A_0^{PC}, and $\underline{\boldsymbol{a}}_0(\mathcal{N})$ by $A_1^{PC}(\boldsymbol{a}(\mathcal{N});\delta)$, $A_0^{PC}(\delta)$, and $\underline{\boldsymbol{a}}_0(\mathcal{N};\delta) \equiv (\underline{a}_0(1;\delta),\ldots,\underline{a}_0(N;\delta))$ respectively. We will show that if $1 > \delta' > \delta \geq 0$, then $v^j(\underline{a}_0(j;\delta')) \leq v^j(\underline{a}_0(j;\delta))$ for all $j \in \mathcal{N}$. It is straightforward that if $\hat{a} \in A_1^{PC}(\boldsymbol{a}(\mathcal{N});\delta)$, then $\hat{a} \in A_1^{PC}(\boldsymbol{a}(\mathcal{N});\delta')$, which means that $A_1^{PC}(\boldsymbol{a}(\mathcal{N});\delta) \subset A_1^{PC}(\boldsymbol{a}(\mathcal{N});\delta')$ for any given $\boldsymbol{a}(\mathcal{N}) \in \mathcal{A}^N$. It implies that if $A_1^{PC}(\boldsymbol{a}(\mathcal{N});\delta) \neq \emptyset$, then $A_1^{PC}(\boldsymbol{a}(\mathcal{N});\delta') \neq \emptyset$. Then we obtain $A_0^{PC}(\delta) \subset A_0^{PC}(\delta')$ and Proposition 3.4 implies that $\underline{\boldsymbol{a}}_0(\mathcal{N};\delta) \in A_0^{PC}(\delta')$. Proposition 3.4 also implies that for all $j \in \mathcal{N}$, $v^j(\underline{a}_0(j;\delta')) \leq v^j(a(j))$ for all $\boldsymbol{a}(\mathcal{N}) \in A_0^{PC}(\delta')$ and then $v^j(\underline{a}_0(j;\delta')) \leq v^j(\underline{a}_0(j;\delta))$.

2. If there exists $\hat{a} \in \mathcal{A}$ such that $s(\hat{a}) - \left(\overline{v}^0 + \sum_{i=1}^N \underline{v}^i\right) > 0$, the left-hand side of (3.17) is unbounded above in $\delta \in [0,1)$. Then we can pick $\overline{\delta}^{OPC*} \in [0,1)$ which satisfies (3.17).

Conversely suppose that $s(\hat{a}) - \left(\overline{v}^0 + \sum_{i=1}^N \underline{v}^i\right) \leq 0$ for all $\hat{a} \in \mathcal{A}$. However, no matter when we choose $\overline{a}^0 \in \overline{A}^0$, $s(\overline{a}^0) = \sum_{j=0}^N v^j(\overline{a}^0) = \overline{v}^0 + \sum_{j=1}^N v^j(\overline{a}^0) \geq \overline{v}^0 + \sum_{j=1}^N \underline{v}^j$, which implies that $s(\overline{a}^0) - \left(\overline{v}^0 + \sum_{j=1}^N \underline{v}^j\right) \geq 0$. Then there must exist a decision $\overline{a}^0$ which satisfies $s(\overline{a}^0) - \left(\overline{v}^0 + \sum_{j=1}^N \underline{v}^j\right) = 0$. It immediately implies that $\sum_{j=1}^N v^j(\overline{a}^0) = \sum_{j=1}^N \underline{v}^j$. Since $v^j(\overline{a}^0) \geq \underline{v}^j$ for all $j \in \mathcal{N}$, it can be seen that $v^j(\overline{a}^0) = \underline{v}^j$ for all $j \in \mathcal{N}$ implying that $\overline{a}^0 \in \underline{A}^j$ for all $j \in \mathcal{N}$. Then we see that a vector $(\overline{a}^0, \overline{a}^0, \ldots, \overline{a}^0) \in \mathcal{A} \times \prod_{j=1}^N \underline{A}^j$ satisfies inequality (3.17) (where both sides are 0).

3.8.8 Proof of Lemma 3.5

Suppose that $(\hat{a}, \hat{w}^1(\cdot), \ldots, \hat{w}^N(\cdot))$ is a SMA-equilibrium. If $\hat{y}^j < \underline{v}^j$ for some $j \in \mathcal{N}$, then principal j can improve her own payoff by offering another compensation plan $\tilde{w}^j(a) = 0$ for all $a \in \mathcal{A}$, a contradiction. If $\hat{y}^0 < \overline{v}^0$, then the agent can improve his own payoff by choosing $\overline{a}^0 \in \overline{\mathcal{A}}$, a contradiction.

3.8.9 Proof of Proposition 3.7

(Necessity) Due to Lemma 3.5 and the assumption that the compensation must be non-negative, if $(\hat{y}^0, \hat{y}^1, \ldots, \hat{y}^N)$ is a SMA-equilibrium net payoff vector, it satisfies

$$(\hat{y}^0, \hat{y}^1, \ldots, \hat{y}^N) \in Y(\hat{a}) := \left\{ (y^0, y^1, \ldots, y^N) \,\middle|\, \sum_{j=0}^{N} y^j = s(\hat{a}), y^0 \geq \overline{v}^0, v^j(\hat{a}) \geq y^j \geq \underline{v}^j, \forall j \in \mathcal{N} \right\}.$$

Now suppose that $\hat{w}^j(\hat{a}) \leq \delta[v^j(\hat{a}) - v^j(a_0(j))]$ for all $j \in \mathcal{N}$. Then it is straightforward to see that $Y(\hat{a})$ is a subset of $\hat{U}^*(\hat{a})$, which implies that $(\hat{y}^0, \hat{y}^1, \ldots, \hat{y}^N) \in \hat{U}^*(\hat{a})$.

(Sufficiency) Suppose that $\hat{w}^j(\hat{a}) > \delta[v^j(\hat{a}) - v^j(a_0(j))]$ for some $j \in \mathcal{N}$. Then the SMA-equilibrium payoff of principal j is $\hat{y}^j = v^j(\hat{a}) - \hat{w}^j(\hat{a}) < (1 - \delta)v^j(\hat{a}) + \delta v^j(a_0(j))$. It cannot be attained in any RPC-equilibria from (3.9).

3.8.10 Proof of Proposition 3.8

(Necessity) Recall that from Proposition 3.6 there exists $\overline{\delta}^{OPC*} \in [0, 1)$ such that for $\delta \in [\overline{\delta}^{OPC*}, 1)$, $u^j(a_0(j)) = \underline{v}^j$ for all $j \in \mathcal{N}$. Then for $\delta \in [\overline{\delta}^{OPC*}, 1)$,

$$\hat{U}^*(\hat{a}) = \left\{ (u^0, u^1, \ldots, u^N) \,\middle|\, \sum_{j=0}^{N} u^j = s(\hat{a}), u^0 \geq \overline{v}^0, v^j(\hat{a}) \geq u^j \geq (1 - \delta)v^j(\hat{a}) + \delta \underline{v}^j, \forall j \in \mathcal{N} \right\}.$$

Now suppose that every principal violates (3.19), i.e. for every $j \in \mathcal{N}$, $\hat{w}^j(\hat{a}) \neq v^j(\hat{a}) - \underline{v}^j$

if $v^j(\hat{a}) > \underline{v}^j$. Let $\tilde{J}(1) := \{j \in \mathcal{N} \mid v^j(\hat{a}) > \underline{v}^j\}$ and $\tilde{J}(2) := \mathcal{N} \setminus \tilde{J}(1)$. Since $v^j(\hat{a}) = \underline{v}^j$ for $j \in \tilde{J}(2)$, for $\delta \in [\overline{\delta}^{OPC*}, 1)$ $\hat{U}^*(\hat{a})$ can be written as

$$\hat{U}^*(\hat{a}) = \left\{ (u^0, u^1, \ldots, u^N) \,\middle|\, \sum_{j=0}^{N} u^j = s(\hat{a}), u^0 \geq \overline{v}^0, \begin{array}{ll} v^j(\hat{a}) \geq u^j \geq (1-\delta)v^j(\hat{a}) + \delta \underline{v}^j, & \forall j \in \tilde{J}(1) \\ u^j = \underline{v}^j, & \forall j \in \tilde{J}(2) \end{array} \right\}.$$

Note that the SMA-equilibrium net payoff satisfies that $\sum_{j=0}^{N} \hat{y}^j = s(\hat{a})$, $\hat{y}^j = v^j(\hat{a}) - \hat{w}^j(\hat{a}) \leq v^j(\hat{a})$ for $j \in \tilde{J}(1)$, $\hat{y}^j = \underline{v}^j$ for $j \in \tilde{J}(2)$, and $\hat{y}^0 \geq \overline{v}^0$. Then the SMA-equilibrium net payoff vector is in $\hat{U}^*(\hat{a})$ if $\hat{y}^j \geq (1-\delta)v^j(\hat{a}) + \delta \underline{v}^j$ for all $j \in \tilde{J}(1)$. Let $\tilde{\delta} := \max\{\overline{\delta}^{OPC*}, \max_{j \in \tilde{J}(1)} \hat{w}^j(\hat{a})/(v^j(\hat{a}) - \underline{v}^j)\}$. Since the SMA-equilibrium payoff is not less than the minimax value by Lemma 3.5 and $\underline{v}^j(\hat{a}) \neq v^j(\hat{a}) - \hat{w}^j(\hat{a})$ for each $j \in \tilde{J}(1)$, it is obtained that $v^j(\hat{a}) - \hat{w}^j(\hat{a}) > \underline{v}^j$ for $j \in \tilde{J}(1)$, which implies that $\tilde{\delta} < 1$. For $\delta \geq \tilde{\delta}$, it can be seen that $\hat{y}^j \geq (1-\delta)v^j(\hat{a}) + \delta \underline{v}^j$ for $j \in \tilde{J}(1)$.

(Sufficiency) Suppose that principal j satisfies (3.19). The SMA-equilibrium payoff of principal j is $v^j(\hat{a}) - \hat{w}^j(\hat{a}) = \underline{v}^j$. Notice that $v^j(a_0(j)) \geq \underline{v}^j$ for any $\delta \in [0, 1)$, which implies that $(1-\delta)v^j(\hat{a}) + \delta v^j(a_0(j)) \geq (1-\delta)v^j(\hat{a}) + \delta \underline{v}^j > \underline{v}^j$ for any $\delta \in [0, 1)$. Then from (3.9) the SMA-equilibrium payoff vector cannot be in $\hat{U}^*(\hat{a})$ for any $\delta \in [0, 1)$.

3.8.11 Proof of Lemma 3.6

For each $j \in \mathcal{N}$, let $M^j := s^* - \max_{a \in \mathcal{A}}[s(a) - v^j(a)]$ which is called the marginal contribution of principal j. Let $(\hat{a}, \hat{w}^1(\cdot), \ldots, \hat{w}^N(\cdot))$ be a SMAT-equilibrium and $(\hat{y}^0, \hat{y}^1, \ldots, \hat{y}^N)$ be the corresponding net payoff. Theorem 1 of Bergemann and Välimäki (2003) implies the following statement.

Claim 3.4 *$(\hat{y}^0, \hat{y}^1, \ldots, \hat{y}^N)$ is a SMAT-equilibrium payoff vector only if $\hat{y}^j \leq M^j$ for all $j \in \mathcal{N}$.*

Suppose that $\underline{A}^j \cap A^* \neq \emptyset$ and then there must exist a decision $\underline{a}^j \in \underline{A}^j \cap A^*$. Then this $\underline{a}^j$ satisfies $s(\underline{a}^j) = \max_{a\in\mathcal{A}} s(a) = s^*$ and $v^j(\underline{a}^j) = \min_{a\in\mathcal{A}} v^j(a) = \underline{v}^j$. Notice that

$$s(\underline{a}^j) - v^j(\underline{a}^j) \leq \max_{a\in\mathcal{A}}[s(a) - v^j(a)] \leq \max_{a\in\mathcal{A}} s(a) - \min_{a\in\mathcal{A}} v^j(a) = s^* - \underline{v}^j = s(\underline{a}^j) - v^j(\underline{a}^j),$$

which implies that $\max_{a\in\mathcal{A}}[s(a) - v^j(a)] = s(\underline{a}^j) - v^j(\underline{a}^j) = s^* - \underline{v}^j$. Then by applying Claim 3.4, it is obtained that $\hat{y}^j \leq s^* - \left[s^* - \underline{v}^j\right]$ or equivalently $\hat{y}^j \leq \underline{v}^j$. Then by Lemma 3.5 $\hat{y}^j = \underline{v}^j$. Since $\hat{y}^j = v^j(\hat{a}) - \hat{w}^j(\hat{a})$, $\hat{w}^j(\hat{a}) = v^j(\hat{a}) - \underline{v}^j$.

3.8.12 Proof of Proposition 3.10

Since $s(a) = s^*$ for all $a \in \mathcal{A}$, it is obvious that $A^* = \mathcal{A}$ and then $\underline{A}^j \cap A^* \neq \emptyset$ for all $j \in \mathcal{N}$. It implies that by Lemma 3.6, for any SMAT-equilibrium $(\hat{a}, \hat{w}^1(\cdot), \ldots, \hat{w}^N(\cdot))$, $\hat{w}^j(a) = v^j(a) - \underline{v}^i$ for all $a \in \mathcal{A}$ and $j \in \mathcal{N}$. Then thanks to Proposition 3.8, if the decision on the SMAT-equilibrium satisfies $v^j(\hat{a}) > \underline{v}^j$ for some $j \in \mathcal{N}$, the SMAT-equilibrium payoff cannot be that of RPC-equilibria. Conversely, suppose that $v^j(\hat{a}) = \underline{v}^j$ for all $j \in \mathcal{N}$. Since $\hat{a} \notin \overline{A}^0$, there exists $\overline{a}^0 \in \overline{A}^0$ such that $v^0(\overline{a}^0) > v^0(\hat{a})$. Since $s(\hat{a}) = s(\overline{a}^0)(= s^*)$, it implies that

$$\sum_{i\in\mathcal{N}} v^i(\overline{a}^0) < \sum_{i\in\mathcal{N}} v^i(\hat{a}) = \sum_{i\in\mathcal{N}} \underline{v}^i,$$

which is not true since $v^i(\overline{a}^0) \geq \underline{v}^i$ for all $i \in \mathcal{A}$.

Chapter 4

Strategic Candidacy *via* Endogenous Commitment

4.1 Introduction

In September 2007, Taro Aso, the former Prime Minister of Japan between 2008 and 2009, stood for election as the president of the Liberal Democratic Party (LDP) in Japan, standing opposed to sole rival Yasuo Fukuda. All of the media, politicians, and even Aso himself expected that Fukuda would almost certainly win.[1] Despite facing this adversity, Aso decided to stand for the election and lost as expected.

Aso's decision cannot be explained by the standard logic of strategic candidacy developed in the literature of political competition with policy-motivated candidates. Roughly speaking, it argues that strategic candidacy can emerge only if the candidates entry into the election can alter the winner (or probability distribution of the winner over the rivals in general). In that scenario, given that a candidate attempting strategic candidacy exists, there must be at least two other rivals in the election. By contrast, in the episode of Aso, Fukuda was the sole rival against him. As such, given that Aso had no chance to win the election, Fukuda would be the winner regardless of Aso's entry and Aso's decision was inconsistent with an intention to change the winner in the election.

[1] BBC News on 16 September 2007 (http://news.bbc.co.uk/1/hi/world/asia-pacific/6997190.stm).

To solve this puzzle, we develop a dynamic model of political competition with policy-motivated citizens. Specifically, our model is based on the canonical model of political competition with citizen candidates developed by Besley and Coate (1997, hereafter BC). We consider an infinitely repeated model of the political competition with citizen candidates and demonstrate the strategic candidacy Aso attempted, which is impossible to show using BC's model.

In BC's model, there are a finite number of citizens, each of whom decides whether to stand for the election with (small) positive cost or not and votes for one of the candidates, and then the elected citizen implements a policy. Since the political competition ends after implementing a policy, the representative would implement her most preferred policy on the equilibrium. In other words, each of the candidates cannot commit to implementing any policies different from her ideal policy.

In our model, while we focus on a simple trigger strategy on the equilibrium, it is enough to show that the repeated relationship could guarantee an ability to commit to an implemented policy *via* reputational concerns which makes a key difference from BC's political equilibrium. In the dynamic relationship, the citizens can alter the political process depending on implemented policies. Thus an elected representative may choose a policy different from her ideal policy for the fear of a worse political process after implementing the ideal policy. This can credibly expand the set of policies to be implemented.

More importantly, the implemented policy can be contingent on the set of the rival candidates. For instance, a candidate can attempt to implement a policy different from her ideal policy when and only when there are rival candidates in the election. Thus, even if the rival's candidacy cannot alter the identity of the winner in the election, it can alter the consequent policy through a change in the winner's implementing policy. It motivates him to stand for the election even when he has no chance to win and no possibility to change the winner.

To see it more intuitively, consider the Downsian spatial model in which there are two potential candidates, an extreme (female) leftist and a mild (male) rightist, and if each of them implements their ideal policy, then the mild rightist gains the strict majority of the votes. Now suppose that the leftist can credibly implement the median policy in addition to the left policy, but attempts to implement the median policy only when the mild rightist also stands for the election against her. Given her intention, the rightist has no chance to win since she would implement the policy closest to the median voter. Nevertheless the rightist still has an incentive to stand for the election. If he did not stand for the election, then the leftist would implement the left policy, which is less preferred to the median policy by him. In other words, he can induce policy compromises from the leftist simply by entering the race. Actually, this is consistent with Aso's intention, as he explains the reason for standing in the election for the president of the LDP as follows:

> In terms of history and international relations, my perspective is extremely opposed to Fukuda's even within the party. [...] Fukuda's ideology and beliefs are different from those that conservatism has been reviving in and dominating the LDP with recently. I thought of expressing a different position fairly. [...] I believed that only after fighting in the election for the president, the prime minister would get tough and could fight against Ozawa's Democratic Party of Japan (DPJ) in the decisive battle. (Aso, 2007, 110-111)[2]

At that time, Aso understood that DPJ, the growing party as the opposition to the LDP, would have been a tough rival in the next general election and that it was important to establish conservatism in the LDP. While Aso recognized that he was unlikely to win the

[2]Translated by the author.

election for the president of the LDP, he actually stood for election since he wanted to induce the stance of the conservatism from Fukuda, whose ideology was different from that of Aso.

This intuition is confirmed in the necessary condition for the existence of the political equilibrium, in which there are exactly two candidates and only one of them has a positive probability to win (Proposition 4.3). The necessary condition is not restricted to one-dimensional spatial competition and it states that if there exists such a dynamic political equilibrium, then (i) the loser's entry must induce political compromises, the benefit from which outweigh the cost of entry, (ii) the policy implemented by the winner must be supported by strict majority of the citizens, and (iii) the winner must be able to credibly implement a different policy from her ideal one.

In order to support our argument of strategic candidacy, we should also provide the sufficient conditions of its existence. Whereas the sufficient condition is stronger in general, the necessary condition is almost sufficient under one-dimensional spatial competition (Proposition 4.5).

As the necessary condition argues, it is essential for strategic candidacy in a two-candidate competition to induce political compromises from the winner. Nevertheless the converse is not necessarily true; strategic candidacy is not necessarily required for inducing political compromises. Actually, as long as the winner can credibly commit to multiple policies due to the fear of a worse political process after the deviation, the political compromise could be induced without any other candidates. With this in mind, we ask a question of whether a political compromise induced by strategic candidacy can be supported by the political process without the strategic candidacy.

We demonstrate that in one-dimensional spatial competition, if the compromised policy is not close enough to the median and either the median citizen or any citizen around the median cannot commit to policies far from their own ideal policy, then the compromised policy cannot be implemented without strategic candidacy whereas it is possible with strategic candidacy (Proposition 4.6). In this situation, the expected winner cannot deter the median citizen, or anyone around the median, from standing for the election opposed to her and then she is no longer the unique candidate. This result implies that strategic candidacy is not a trivial factor from the perspective of political compromises.

Possibilities of strategic candidacy have been already pointed out by Osborne and Slivinski (1996) and BC in the models of political competition with citizen candidates.[3] However, in their model, strategic candidacy emerges only when there are three or more citizens to stand for election. The contribution of this chapter is to show that in the dynamic political competition, strategic candidacy can emerge even when there are only two candidates.[4] Recently, Asako (2010) independently addresses the same question as ours in a modified model of political competition with citizen candidates. He considers a one-shot spatial model rather than a dynamic model and assumes "betrayal cost" which is increasing in the distance between the promises made before the election and the actual implementing policy.[5] While some of our results are shared with his paper, our model demonstrates it by studying the political reputation sustained endogenously in a dynamic political process instead of assuming the betrayal cost exogenously.[6] Furthermore our

[3]Recently, strategic candidacy is also studied in the theory of social choice. This literature takes axiomatic approaches and addresses under what rule of political competition implementation of social choice function is immune to the threat of strategic candidacy. See Dutta et al. (2001, 2002), Ehlers and Weymark (2003), Eraslana and McLennan (2004), Samejima (2007) among others.

[4]Besley and Coate (1998) investigate a finite repeated political competition with citizen candidates. They mainly study normative issues to ask whether efficient outcome can be implemented through the dynamic political competition and do not address strategic candidacy as we ask in this chapter.

[5]In addition, his model assumes sincere voting while our model adopts strategic voting.

[6]Our result could be interpreted as a microfoundation of the betrayal cost. Strictly speaking, however, it is somewhat difficult to justify his model by the repeated political competition since replicating the continuous betrayal cost function requires very sophisticated punishment strategy. We believe that if we construct a one-shot model with commitment ability which is justified by a repeated situation implicitly assumed, then the commitment ability should be defined as a set of the policies to be committed credibly rather than the continuous function increasing in the distance between the promise and the actual policy.

main results are not restricted to a one-dimensional spatial competition as he assumes.

Our model is also related to the repeated elections including Alesina (1988), Dixit et al. (2000), Duggan and Fey (2006), and Aragonès et al. (2007). These papers show that even when candidates are policy-motivated, they can implement a different policy from their most preferred one in the repeated relationship. However all of these papers consider a competition with two candidates (or two parties) and there is no endogenous decision to choose to stand for the election as there is in the citizen-candidate model. Our contribution to this literature is intended to uncover equilibrium characteristics of endogenous entry decisions in dynamic competition. In particular, we show that decisions to stand for election may be purely strategic in a two-candidate competition which cannot be explained in the models without endogenous entry decisions.

The rest of this chapter is organized as follows. Section 4.2 reviews a one-shot political competition model of BC. Section 4.3 extends the political competition to the repeated setting. Section 4.4 discusses a two-candidate equilibrium with strategic candidacy emerging in a simple example and identifies the conditions for the existence of such equilibria. Section 4.5 discusses the relationship between political compromises and strategic candidacy and shows that political compromises induced by strategic candidacy are not necessarily induced without the strategic candidacy. Section 4.6 concludes. The Appendix contains the proofs of the formal results and provides the supplementary results on voting decisions.

4.2 One-Shot Political Competition

4.2.1 Environment

We first describe one-shot political competition, denoted by G, which is essentially the same as BC. There are a set of citizens $\mathcal{N} = \{1, 2, \ldots, N\}$ with $N \geq 2$ and let $\mathfrak{N} := 2^{\mathcal{N}} \setminus \{\emptyset\}$, that is, the collection of non-empty subsets of $\mathcal{N}$. The policy space is defined by $\mathcal{A} \subset \mathbb{R}^K$ where K is a positive integer. The political process is described as the following three stage game.

- Stage 1 (Entry Stage): each citizen $i \in \mathcal{N}$ decides whether to stand in the election or not.
- Stage 2 (Voting Stage): given the set of candidates $C \in \mathfrak{N}$, each citizen votes for one of the candidates and the candidate who receives the most votes is elected as the representative.
- Stage 3 (Policy Stage): the elected representative I implements a policy $a^I \in \mathcal{A}$.[7]

There are several assumptions as follows. First, when a citizen stands for election, she incurs a utility cost for the candidacy $d > 0$.[8] Second, if the set of the candidates is empty after Stage 1, then the status quo policy $a^0 \in \mathcal{A}$ is immediately implemented and the game ends. Third, given $C \in \mathfrak{N}$, each of the citizens, including the candidates, can abstain from voting. Finally, given the votes, if there are two or more candidates who obtain the most votes, the representative is chosen with equal probability from the tying candidates.

[7] It implies that every citizen can choose any policy from $\mathcal{A}$ when she becomes representative. We can also assume that if citizen i becomes representative, she chooses a policy from $\mathcal{A}^i \subset \mathcal{A}$. This extension does not change our results qualitatively.

[8] We assume that the entry cost is common among all the citizens. None of the following results can be changed qualitatively even if the entry cost is heterogeneous among the citizens.

Given that policy $a \in \mathcal{A}$ is implemented, citizen i obtains political benefit $v^i(a)$.[9] The citizen's net payoff is the sum of the expected political benefit from the policy and the disutility from the decision to enter. More specifically, when policy a is chosen, the *ex post* payoff for citizen i is given by $v^i(a) - d$ if she stood for election and $v^i(a)$ if not.

Let $\overline{\mathcal{A}}^i \equiv \arg\max_{a \in \mathcal{A}} v^i(a)$ be the set of citizen i's ideal policy. Throughout the chapter, we assume that $\overline{\mathcal{A}}^i$ is singleton for each $i \in \mathcal{N}$ and let $\overline{\mathcal{A}}^i = \{\overline{a}^i\}$. We further assume that $\overline{a}^i \neq \overline{a}^j$ for any $i, j \in \mathcal{N}$ with $i \neq j$. The preference described above is assumed to be common knowledge for all the citizens. Further, we naturally assume that the set of candidates, the identity of the representative, and the implemented policy are all observable while voting choice is private information for each citizen.

4.2.2 Political Equilibria

A political equilibrium of G is defined as a subgame perfect equilibrium with some refinement, which is explained later. The backward induction implies that when candidate i becomes the representative, she will choose policy $\overline{a}^i$ regardless of whatever happened before. Then citizen i's pure strategy in game G is essentially defined by a duple $\chi^i \equiv (\gamma^i, \omega^i)$ where $\gamma^i \in \{0, 1\}$ is the probability to stand for the election and $\omega^i = (\omega^i_C)_{C \in \mathfrak{N}}$ with $\omega^i_C \in C \cup \{0\}$ is the voting decision meaning that citizen i votes for ω^i_C or abstains (when $\omega^i_C = 0$). A mixed strategy is defined by a lottery over the pure strategies. Specifically, with a slight abuse of notation, let $\gamma^i \in [0, 1]$ denote the probability of entry and $\Delta^i_{\omega_C}(j)$ denotes the probability to vote for $j \in C \cup \{0\}$ given a set of candidates C. In what follows, we use both ω^i_C and $\Delta^i_{\omega_C}$ for the notation of the decision of the vote and if it is expressed by ω^i_C, then citizen i's vote given C is implicitly supposed to be deterministic. Let $V^i(\chi)$ be the *ex ante*

[9] As BC, we can generalize the political benefit function to $v^i(a, I)$ where I is the representative.

payoff for citizen i where χ is a (possibly mixed) strategy profile.

In addition to subgame perfection, we impose weakly undominated voting decisions on our political equilibrium. Given $C \in \mathfrak{R}$, for any deterministic voting profile $\omega \equiv (\omega^1, \ldots, \omega^N) \in (C \cup \{0\})^N$, denote the set of the possible winning candidates by $W(\omega \mid C) \equiv \arg\max_{k \in C} \#\{i \in \mathcal{N} \mid \omega^i = k\}$.

Definition 4.1 *Given $C \in \mathfrak{R}$ and $a_C \equiv (a^j)_{j \in C} \in \mathcal{A}^{\#C}$, a (deterministic) voting decision $\omega^i_C \in C \cup \{0\}$ is weakly dominated with respect to (C, a_C) if there exists $j \in C \cup \{0\}$ such that*

$$\sum_{k \in W(j, \omega_C^{-j} \mid C)} \frac{v^i(a^k)}{\#W(j, \omega_C^{-i} \mid C)} \geq \sum_{k \in W(\omega_C^i, \omega_C^{-i} \mid C)} \frac{v^i(a^k)}{\#W(\omega_C^i, \omega_C^{-i} \mid C)}$$

for all $\omega_C^{-i} \in (C \cup \{0\})^{N-1}$ and it holds with strict inequality for some $\omega_C^{-i} \in (C \cup \{0\})^{N-1}$.

This definition is generalized from BC's requirement on their political equilibrium. We explicitly introduce the notion of a pair of policies to be implemented. This generalization is required in the dynamic model we investigate later since the implemented policy on the equilibrium does not necessarily coincide with the representative's ideal policy.

The following definition of our (static) political equilibrium is exactly the same as BC's definition.

Definition 4.2 *A strategy profile $\underline{\chi} \equiv (\underline{\gamma}, \underline{\Delta}_\omega)$ is a political equilibrium (hereafter PE) of G if $\underline{\chi}$ is a subgame perfect equilibrium of G and the (mixed) voting decision $\underline{\Delta}^i_\omega \equiv (\underline{\Delta}^i_{\omega_C})_{C \in \mathfrak{R}}$ satisfies that for each $C \in \mathfrak{R}$, if $\underline{\Delta}^i_{\omega_C}(j) > 0$, then to vote j (or to abstain if $j = 0$) is not weakly dominated with respect to $(C, \bar{a}_C)$ where $\bar{a}_C \equiv (\bar{a}^j)_{j \in C}$.*

In what follows, let $\Delta_W(k \mid \Delta_{\omega_C}, C)$ be the probability with which candidate k becomes the representative under a (mixed) voting decision profile Δ_{ω_C} given a set of candidates C.

Proposition 4.1 *1. There exists a PE of G.*

2. For any $C \in \mathfrak{N}$ and $a_C \equiv (a^i)_{i \in C} \in \mathcal{A}^{\#C}$, there exists a voting profile Δ_ω satisfying that for all $i \in \mathcal{N}$, $j \in \{j'' \mid \Delta^i_\omega(j'') > 0\}$ and for all $j' \in C$,

$$\sum_{n \in C} \Delta_W(n \mid (j, \hat{\Delta}^{-i*}_{\omega_C}), C) v^i(a^n) \geq \sum_{n \in C} \Delta_W(n \mid (j', \hat{\Delta}^{-i*}_{\omega_C}), C) v^i(a^n),$$

and to vote j is not weakly dominated with respect to (C, a_C).

Since the set of the strategy is finite, the standard fixed point argument implies not only the existence of PE but also that of weakly undominated Nash equilibria of subgames after the entry stage given (any) pair of policies to be implemented.[10] Let Ξ be the set of PE of G.

4.2.3 Impossibility of Strategic Candidacy with Two Candidates

In the one-shot political competition, there is never PE in which there are two candidates to stand for the election either of which has no chance to win.

Proposition 4.2 *Suppose that a strategy profile $\underline{\chi}$ satisfies that*

$$\underline{\gamma}^i = \begin{cases} 1 & \text{for } i = k, l \\ 0 & \text{otherwise,} \end{cases} \quad \text{and } \Delta_W(k \mid \underline{\Delta}_{\omega_{\{k,l\}}}, \{k, l\}) = 1.$$

Then $\underline{\chi}$ is not a subgame perfect equilibrium (and hence $\underline{\chi} \notin \Xi$).

The result is straightforward. If citizens k and l are the candidates and l has no chance to win, then regardless of l's entry, k becomes the representative and implements her

[10]The detail of the proof is omitted.

ideal policy $\bar{a}^k$ with probability 1. Hence l can save the entry cost d by withdrawing the candidacy. In other words, whenever there are exactly two candidates in the election, both of the candidates must have positive probability to become the representative in BC's setup and then there is no possibility of strategic candidacy.

The key assumption behind this impossibility result is that whatever happens in the political process, the representative implements her most preferred policy, which is due to the (implicit) assumption that candidates cannot make any binding agreements on implementing policy before the voting stage. Although this assumption is fairly reasonable under some situations, a reputational concern due to the repeated interaction can guarantee a credible commitment to the implementing policy which is different from the ideal policy. In the following sections, we analyze a repeated version of the political competition and show that when the implementing policy can be contingent on the set of the rival candidates due to political reputation, there could be a political equilibrium in which there are two candidates associated with strategic candidacy.

4.3 Repeated Competition

In the repeated model, all the citizens live in infinite periods $t = 0, 1, \ldots$, and they play the one-shot political competition G described above in each period. Each citizen maximizes the average discounted sum of the period net payoff with the common discount factor $\delta \in (0, 1)$.[11]

Denote the infinite repeated game by G^∞. Let $h_t^i \equiv (C_t, w^i, I_t, a_t)$ be the event in period t observable to citizen i, that is, the set of the candidates, i's vote,[12] the elected representative,

[11]None of the following results can be changed qualitatively if the discount factor is heterogeneous among the citizens.

[12]Recall that each citizen cannot observe the others' vote.

and the implemented policy, and $h^{it} \equiv (h^0, h^1, \ldots, h^{t-1})$ be the private history after period t. Let $\mathcal{H}^i$ be the set of citizen i's private histories. Citizen i's pure strategy in G^∞ is a pair of mappings $\sigma^i \equiv (\tilde{\gamma}^i, (\tilde{\omega}^i_C)_{C\in\mathfrak{N}}, (\tilde{\alpha}^i_C)_{C\in\{C'\in\mathfrak{N}|i\in C'\}})$ where $\tilde{\gamma}^i : \mathcal{H}^i \to \{0,1\}$ be the entry decision, $\tilde{\omega}^i_C : \mathcal{H}^i \to C \cup \{0\}$ be the voting decision, and $\tilde{\alpha}^i_C : \mathcal{H}^i \to \mathcal{A}$ be the implementing policy where $C \in \mathfrak{N}$ is the set of the candidates. Similar to the previous section, the mixed decision is defined as a lottery over the decisions.

In what follows, we focus on the following trigger strategy on the equilibrium. Citizen i's trigger strategy is denoted by $\sigma^{i*} \equiv (\hat{\chi}^{i*}, \underline{\chi}^*, F^*)$ where $\underline{\chi}^* \in \Xi$ is a PE of G and $F^* \subset \mathcal{N}$ is a subset of the citizens. It consists of two phases; Phase R ("reputational phase") and Phase D ("defection phase"). Under phase R, in each period citizen i plays according to $\hat{\chi}^{i*} := (\hat{\gamma}^{i*}, (\hat{\Delta}^{i*}_{\omega_C})_{C\in\mathfrak{N}}, (\hat{\alpha}^{i*}_C)_{C\in\{C'\in\mathfrak{N}|i\in C'\}})$, that is a pair of the decisions at the entry, voting, and policy stage. We assume that the policy decision must be deterministic while the entry and the voting decision can be mixed. In period 0, the phase is R and it is changed from R to D if and only if in the last period, the set of the candidates was C and the representative $j \in F^* \cap C$ implemented a policy different from $\hat{\alpha}^{j*}_C$.[13] Once the phase becomes D, it is never changed and the citizens play the game according to $\underline{\chi}^*$ forever.

We define a dynamic political equilibrium of G^∞ as follows.

Definition 4.3 *A trigger strategy σ^{i*} is a dynamic political equilibrium (hereafter DPE) of G^∞ if σ^{i*} is a subgame perfect equilibrium of G^∞ and the (mixed) voting decision $\hat{\Delta}^{i*}_\omega \equiv (\hat{\Delta}^{i*}_{\omega_C}(\cdot))_{C\in\mathfrak{N}}$ satisfies that for each $C \in \mathfrak{N}$, if $\hat{\Delta}^{i*}_{\omega_C}(j) > 0$, then to vote j (or to abstain if $j = 0$) is not weakly dominated with respect to $(C, \hat{\alpha}^*_C)$ where $\hat{\alpha}^*_C \equiv (\hat{\alpha}^{j*}_C)_{j\in C}$.*

The dynamic political equilibrium is a subgame perfect equilibrium with weakly undominated voting decisions. In the dynamic model, this definition is generalized from

[13]It means that when $j \notin F^* \cup C$, the phase is kept to be R even after deviation by the representative j.

BC's requirement of their political equilibrium. The implemented policy can be changed depending on the set of the rival candidates. The DPE requires that in addition to subgame perfection, given that the implemented policies at the period begin as expected, each of the citizens votes for a candidate which is not weakly undominated.

4.4 Strategic Candidacy in Two-Candidate Competition

4.4.1 Definition

Our main interest in this chapter is to explain the reason why strategic candidacy can emerge in two-candidate competition. This situation is formally described as follows.

Definition 4.4 *A trigger strategy σ^* causes (k, l) strategic candidacy (hereafter "(k, l)-SC") if*

$$\hat{\gamma}^{i*} = \begin{cases} 1 & \textit{for } i = k, l \\ 0 & \textit{otherwise,} \end{cases} \quad \textit{and } \Delta_W(k \mid \hat{\Delta}^*_{\hat{\omega}_{\{k,l\}}}, \{k, l\}) = 1.$$

On the equilibrium path of a strategy causing (k, l)-SC, two citizens, k and l, stand for the election and k wins in the election with probability 1. In one-shot political competition G, Proposition 4.2 states that this could not be the case. We now see that it is sustained by a trigger strategy DPE of G^∞ under some conditions. In order to understand the intuition behind the strategic candidacy, we first demonstrate it with a spatial example.

4.4.2 Example: A Spatial Model with Five Citizens

Suppose that there are five citizens ($N = 5$) and $\mathcal{A} = [-1, 1]$. Citizen i's political benefit is given by $v^i(a) = - \mid a - \overline{a}^i \mid$ where $\overline{a}^i = -1 + (i-1)/2$. The status quo is assumed to be $a^0 = 0$ and let $d \in (0, 1/2)$. For explanatory reason, we assume that citizen 2 is female and citizen

3 and 4 are male.

According to the propositions in BC, we can find a pure strategy PE of G satisfying the following;

- citizens 2 and 4 stand for the election,
- citizens 1 and 2 vote for 2, citizens 4 and 5 vote for 4, and citizen 3 abstains,
- 2 or 4 wins with probability 1/2 (and if $i \in \{2,4\}$ wins, then $\overline{a}^i$ is implemented).

On the equilibrium, there are exactly two candidates standing for the election and both of them have a positive probability to win. As shown in Proposition 4.2, if either candidate had zero probability to win, then she would strictly prefer not to stand for the election since it saves the entry cost d without altering the policy implemented by the rival. Thus, whenever there are exactly two candidates, strategic candidacy never arises.[14]

Denote this PE by $\underline{\chi} \equiv (\underline{\gamma}, \underline{\omega})$. Now consider the following trigger strategy σ^*;

$$\hat{\gamma}^{i*} = \underline{\gamma}^i,\ \hat{\omega}_C^{i*} = \begin{cases} 2 & \text{if } i = 3,\ C = \{2,4\}, \\ \underline{\omega}_C^i & \text{otherwise,} \end{cases} \qquad \hat{\alpha}_C^{i*} = \begin{cases} -\overline{a}^2 + q & \text{if } i = 2,\ C = \{2,4\}, \\ \overline{a}^i & \text{otherwise,} \end{cases}$$
$$\hat{\chi}^* = \underline{\chi},\ F^* = \{1,2,3\}$$

where $q \in [d, \delta/2]$. We assume that $\delta \geq 2d$ to guarantee the existence of such q.

This trigger strategy causes (2,4)-SC. Under Phase R, the set of the candidates is still $\{2,4\}$ which is the same as the PE of G. However, as figure 4.1 shows, the policy implemented by candidate 2 is more closed to the median by $q > 0$ from her most preferred policy $\overline{a}^2$. The voting decision under phase R implies that candidate 2 actually

[14]There is another pure strategy PE in which citizen 1 and 5 stand for the election. The same argument is exactly applied for this PE.

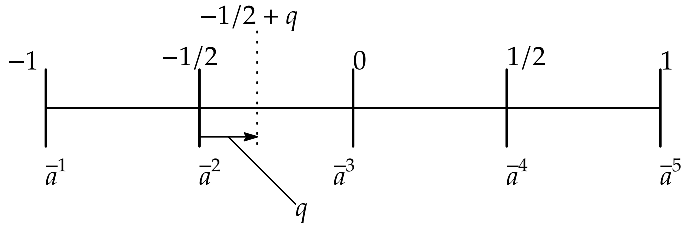

Figure 4.1: Consequent Policy

obtains the strict plurality in the election and then wins with probability 1. It obviously means that although citizen 4 actually stands for the election, he has no chance to win.

It can be shown that σ^* is a DPE. Here we check two relevant constraints to assure the equilibrium and discuss the implication.[15] The first constraint is that citizen 2 has an incentive to choose a policy different from $\overline{a}^2$. Specifically, when the phase is R and $C = \{2, 4\}$, candidate 2 chooses $\hat{\alpha}^{2*}_{\{2,4\}} = \overline{a}^2 + q \neq \overline{a}^2$. If she follows to implement $\overline{a}^2 + q$, then she obtains the current political benefit $-|q + \overline{a}^2 - \overline{a}^2|$ and the future discounted sum of the net payoff by $\hat{\chi}^*$, $(\delta/(1-\delta))(-|q + \overline{a}^2 - \overline{a}^2| - d)$. By contrast if she deviates $\overline{a}^2$ instead of $\overline{a}^2 + q$, she obtains the current political benefit by deviation $-|\overline{a}^2 - \overline{a}^2|$ and the future discounted sum of the net payoff by $\underline{\chi}^*$, $(\delta/(1-\delta))(-|\overline{a}^2 - \overline{a}^2|/2 - |\overline{a}^4 - \overline{a}^2|/2 - d)$. The comparison between these implies that the constraint is described as

$$-q + \frac{\delta}{1-\delta}(-q - d) \geq 0 + \frac{\delta}{1-\delta}\left(\frac{1}{2}0 + \frac{1}{2}(-1) - d\right) \iff \delta \geq 2q.$$

The second constraint is that under Phase R, citizen 4 stands for election. If he stands for election, then candidate 2 becomes the representative and implements $\overline{a}^2 + q$ with probability 1 and then citizen 4's current net payoff is $- \mid \overline{a}^2 + q - \overline{a}^4 \mid - d$. By contrast if he withdraws the candidacy, candidate 2 implements $\overline{a}^2$ for sure (rather than $\overline{a}^2 + q$) and

[15] The detail of the equilibrium conditions is stated later in Proposition 4.3 and 4.5.

then citizen 4 obtains the current net payoff $-\mid \bar{a}^2 - \bar{a}^4 \mid$. The comparison between these[16] implies that this constraint is described as

$$-\left|-\frac{1}{2}+q-\frac{1}{2}\right|-d \geq -\left|-\frac{1}{2}-\frac{1}{2}\right| \iff q \geq d.$$

Since $d \leq q \leq \delta/2$, both of these constraints are satisfied and it can be shown that the other constraints for σ^* being a DPE are also satisfied.

The intuition is as follows. When candidate 2 is opposed to candidate 4 in the election, citizen 2 wants to implement a policy close to the median in order to attract more votes to increase her probability of winning. The trigger strategy makes it possible; as long as the distance from her ideal policy is sufficiently small relative to the future benefit from the reputation, she can credibly commit to the policy closed to the median, the condition of which is given by $q \leq \delta/2$.

It is important to notice that citizen 2 implements the compromised policy only if citizen 4 stands against her. The trigger strategy σ^* stipulates that if citizen 4 does not enter, then citizen 2 would implement her ideal policy. In other words, if citizen 4 actually enters, he can induce a policy compromise from the winning rival. The condition $q \geq d$ expresses that the effect of the policy compromise from candidate 2 dominates the cost of entry for citizen 4. If this condition is satisfied, citizen 4 stands for election even she has no chance to win.

As we will see, to induce a political compromise from the winner is, in general, the necessary condition for causing strategic candidacy in two-candidate competition.[17] The key condition for inducing political compromises is that the winner can credibly undertake

[16]Since the entry decision does not change the phase, the continuation payoff from the next period is the same between these.

[17]It does not necessarily mean that strategic candidacy alters the rival's policy from her ideal policy. We explain the detail in footnote 18.

an opponent-dependent policy, which is supported by political reputation in our dynamic model. This kind of political compromise never emerges in political competition with citizen candidate without commitment as shown in Osborne and Slivinski (1996) and BC.

4.4.3 The Necessary and Sufficient Condition

We now investigate the necessary and sufficient conditions of the existence of DPE inducing strategic candidacy with two candidates in a more general setting. First, we provide the necessary conditions for the existence of DPE causing strategic candidacy in a two-candidate competition.

Proposition 4.3 *Suppose that there exists a trigger strategy DPE causing* (k, l)*-SC and on the equilibrium k chooses* a^k *on the equilibrium path. Then there exist* $(a^l, a^k_k, a^l_l) \in \mathcal{A}^3$ *and* $\underline{\chi} \in \Xi$ *such that*

1.
$$v^l(a^k) - d \geq v^l(a^k_k), v^k(a^k) - d \geq v^k(a^l_l),$$

2.
$$\#\left\{i \in \mathcal{N} \mid v^i(a^k) \geq v^i(a^l)\right\} > \#\left\{i \in \mathcal{N} \mid v^i(a^k) < v^i(a^l)\right\},$$

and

3. *for* $i = k, l$ *and* $a^{i\prime} = a^i, a^i_i$*, either* $a^{i\prime} = \overline{a}^i$ *or*
$$\frac{\delta}{1-\delta}\left[v^i(a^k) - d - V^i(\underline{\chi})\right] \geq v^i(\overline{a}^i) - v^i(a^{i\prime})$$

(or both).

Suppose that there exists a DPE $\sigma^* \equiv (\hat{\chi}^*, \underline{\chi}^*, F^*)$ causing (k,l)-SC. Then a^k, a^l, a^k_k, and a^l_l in the necessary condition above are corresponding to $\hat{\alpha}^{k*}_{\{k,l\}}$, $\hat{\alpha}^{l*}_{\{k,l\}}$, $\hat{\alpha}^{k*}_{\{k\}}$, and $\hat{\alpha}^{l*}_{\{l\}}$ respectively. In contrast to the one-shot political competition, $\hat{\alpha}^{i*}_C$ can be different from $\overline{a}^i$ in general. However in order to assure the implementation of a policy different from her ideal policy, the fear of the future loss from deviation is required. Part 3 in the proposition captures this condition and implies that depending on $\underline{\chi}^*$ and δ, the commitment ability is limited. Part 2 assures that given the pair of the implemented policies $(\alpha^{k*}_{\{k,l\}}, \alpha^{l*}_{\{k,l\}})$, candidate k can receive the strict majority and win the election with probability 1, implying that l's entry is purely strategic.

Part 1 in the proposition ensures that both k and l benefit from the entry. Importantly, although citizen l has zero probability of winning the election, he has an incentive to enter into the election since his entry alters the policy implemented by k from a^k_k to a^k. In particular, a^k must be strictly preferred to a^k_k by l.[18] It implies that the punishment on the deviation by citizen k is required; otherwise, citizen k cannot credibly commit to implementing a policy different from her ideal one. This is emphasized by the following corollary.

Corollary 4.1 *If a trigger strategy σ^* causing (k,l)-SC is a DPE, then $k \in F^*$ and $v^k(\hat{\alpha}^{k*}_{\{k,l\}}) - d > V^k(\underline{\chi}^*)$.*

In order to ensure the existence of DPE causing (k,l)-SC, the additional requirements must be satisfied; that is, the citizens other than k and l have no incentive to enter. One of the sufficient conditions for satisfying this requirement is as follows.

[18] In the example of Section 4.4.2, we have seen that $a^k_k = \overline{a}^k$, meaning that candidate k implements her ideal policy if there is no rival in the election. In general, however, it is not necessarily the case since the implemented policy by k does not necessarily coincide with $\overline{a}^k$ even when there is no rival in the election.

Proposition 4.4 *In addition to the conditions in Proposition 4.3, suppose that there exists* $(a_i^k, a_i^l, a_i^i)_{i\in\mathfrak{N}\setminus\{k,l\}} \in \mathcal{A}^{3(N-2)}$ *such that for each* $i \in \mathcal{N} \setminus \{k, l\}$*,*

1. *for each* $j = k, l, i$*, and* $a^{j\prime} = a_{j}^{i}$*, either* $a^{j\prime} = \overline{a}^j$ *or*

$$\frac{\delta}{1-\delta}\left[v^j(a^k) - \hat{\gamma}^{j*}d - V^j(\underline{\chi}^*)\right] \geq v^j(\overline{a}^j) - v^j(a^{j\prime})$$

   (or both), and

2. *there exists* $j \in \{k, l, i\}$ *such that* $v^i(a^k) \geq v^i(a_i^j) - d$ *and*

$$\#\left\{n \in \mathcal{N} \middle| v^n(a_i^j) = \max_{j'\in\{k,l,i\}} v^n(a_i^{j'})\right\} > \frac{N}{2} \text{ or} \quad (4.1)$$

$$N \geq 4 \text{ and } \#\left\{n \in \mathcal{N} \middle| v^n(a_i^j) \neq \min_{j'\in\{k,l,i\}} v^n(a_i^{j'}) \text{ or } v^n(a_i^k) = v^n(a_i^l) = v^n(a_i^i)\right\} > \frac{N}{2} + 1. \quad (4.2)$$

Then there exists a DPE causing (k, l)*-SC on which citizen k implements* a^k*.*

Part 1 assures that all the policies chosen under phase R are credibly feasible. Part 2 assures us to construct a voting decision under phase R so that the entry by citizen $i \neq k, l$ cannot improve on i's payoff given that the candidates are supposed to implement a_i^k, a_i^l, and a_i^i respectively. Proposition 4.4 allows the citizens to deter the entry of citizens other than k and l by a deterministic voting decision profile and it simplifies the proof of the existence of DPE.

Another approach for the sufficient condition is to assume one-dimensional spatial competition. When the policy space is one-dimensional and the political benefit only depends on the distance from the ideal policy, the necessary condition in Proposition 4.3 is almost sufficient. The one-dimensional spatial competition is a generalization of the example in Section 4.4.2 and satisfies the following assumption.

Assumption 4.1 *1. $\mathcal{A}$ is a closed interval of $\mathbb{R}$.*

2. For all $i \in N$, $v^i(a)$ is continuous in a and decreasing in $|a - \overline{a}^i|$.

3. $N \geq 5$ and N is an odd number.

4. (Without loss of generality) $\overline{a}^i < \overline{a}^{i+1}$ for all $i = 1, \ldots, N-1$ and denote $M \equiv (N+1)/2$.

Proposition 4.5 *Suppose that all of the conditions in Proposition 4.3 are satisfied under Assumption 4.1. Then, whenever*

1. $a^k = \overline{a}^M$ or

2. $a^k \neq a^l$ and

(a) $N = 5$ and $a^k, a^l \in [2\overline{a}^2 - \overline{a}^3, 2\overline{a}^4 - \overline{a}^3]$ or

(b) $N \geq 7$,

there exists a DPE causing (k,l)-SC in which citizen k implements a^k.

In a general setting, the construction of DPE to prevent citizens other than k and l from candidacy is not a trivial problem and then the additional sufficient condition is required as in Proposition 4.4. In one-dimensional spatial competition, we can construct a voting decision under phase R so that given that $j \neq k, l$ enters, at most one citizen votes for j. Under mild conditions as described in Proposition 4.5, this voting decision deters j's entry.

4.5 Is Candidacy Necessary for Political Compromises?

So far we have investigated DPE causing strategic candidacy in two-candidate competition and shown that a political compromise is necessary for the loser's strategic candidacy.

This result gives rise to the following question; whether the political compromise can be induced without strategic candidacy. Actually, even if there is only one candidate in the election, this candidate might implement policies different from her preferred one due to the fear of losing her office in the future after the deviation.

In this section, we demonstrate that under some conditions while a political compromise can be induced by strategic candidacy with two candidates, the same political compromise cannot be induced in a single-candidate political competition. Throughout this section, we maintain Assumption 4.1 and suppose $k < M$, meaning that citizen k is on the left side of the median citizen. The symmetric argument is applicable for $k > M$. Given that k is a left citizen, we define unopposed political compromise by k as follows.

Definition 4.5 *A trigger strategy σ^* induces unopposed political compromises by $k < M$ (hereafter "k-UPC") if*

$$\hat{\gamma}^{i*} = \begin{cases} 1 & \text{for } i = k \\ 0 & \text{otherwise} \end{cases} \quad \text{and } \hat{\alpha}^{k*}_{\{k\}} \in (\bar{a}^k, \bar{a}^M].$$

In a trigger strategy inducing k-UPC, citizen k is the unique candidate in the election (and then becomes the representative) and the implemented policy $\hat{\alpha}^{k*}_{\{k\}}$ is strictly closer to the median.

4.5.1 Example Revisited

We first revisit the example in Section 4.4.2 and show that there exists an interval of the discount factor for which a policy compromise can be achieved by strategic candidacy, but cannot without strategic candidacy. This example tells us that it is sometimes impossible to deter entry by the median citizen for constructing a DPE inducing unopposed political compromises.

Let $\underline{\chi}(2)$ be a PE which attains the worst net payoff for citizen 2 and suppose that $a^2 \in (\overline{a}^2 + d, -d)$ and

$$\frac{\delta}{1-\delta}\left[-a^2 + \overline{a}^2 - d - V^2(\underline{\chi}(2))\right] \geq a^2 - \overline{a}^2 \iff \delta \geq \underline{\delta}(a^2) \equiv \frac{a^2 - \overline{a}^2}{-(d + V^2(\underline{\chi}(2)))}, \tag{4.3}$$

$$a^2 + d < \hat{a}_3^1 \equiv -\frac{\delta}{1-\delta}\left[a^2 + 1\right] \iff \delta < \overline{\delta}(a^2) \equiv \frac{a^2 + d}{-(1/2 - d)}. \tag{4.4}$$

Note that since there exists a PE such that citizens 1 and 5 stand for election and the former wins with probability $d/2$, it is confirmed that $V^2(\underline{\chi}(2)) \leq -d/2(1/2) - (1 - d/2)(3/2) = -3/2 - d$ and for sufficiently small $d > 0$, there always exists a^2 such that $\underline{\delta}(a^2) < \overline{\delta}(a^2)$.

(4.3) assures that there exists a DPE causing $(2,4)$-SC in which candidate 2 implements a policy a^2. However, if (4.4) is satisfied, then there is no trigger strategy DPE inducing 2-UPC in which citizen 2 implements a^2.

To see the reason, suppose that a trigger strategy $\sigma^{**} \equiv (\hat{\chi}^{**}, \underline{\chi}^{**}, F^{**})$ inducing 2-UPC is a DPE in which $\hat{\alpha}^{2**}_{\{2\}} = a^2$. Then citizen 3, the median voter, has no incentive to stand for the election against citizen 2. Now consider $\hat{\alpha}^{3*}_{\{2,3\}}$, the policy implemented by candidate 3 when he is chosen as the representative from the set of candidates $\{2,3\}$. (4.4) means that $\hat{\alpha}^{3*}_{\{2,3\}}$ must not be far away from $\overline{a}^3$. This is illustrated in Figure 4.2. In the figure, given that candidate 2 implements a^2, the set of policies citizen 3 can credibly implement is within the interval from $\hat{a}^3_1$ to $\hat{a}^3_2$ where $\hat{a}^3_1$ is what we defined in (4.4) and $\hat{a}^3_2 = -\hat{a}^3_1$.[19] In particular, (4.4) implies that citizen 3 cannot credibly implement any policy which is less than $a^2 + d$ or in other words $\hat{a}^3_1 > a^2 + d$.

[19]Formally, $a \in [\hat{a}^3_1, \hat{a}^3_2]$ if and only if $\delta(a^2+1)/(1-\delta) \geq -\mid a \mid$. Note that in the one shot competition, citizen 3's worst political benefit is -1 which is achieved at $a = -1, 1$. Given the worst political benefit, by choosing not to stand for the election, he can assure the net payoff at least -1. Then there exists no PE χ such that $V^3(\chi) < -1$. Given this fact, citizen 3 must choose his policy from the interval $[\hat{a}^3_1, \hat{a}^3_2]$. See Lemma 4.2 in the appendix for the detail of the equilibrium conditions.

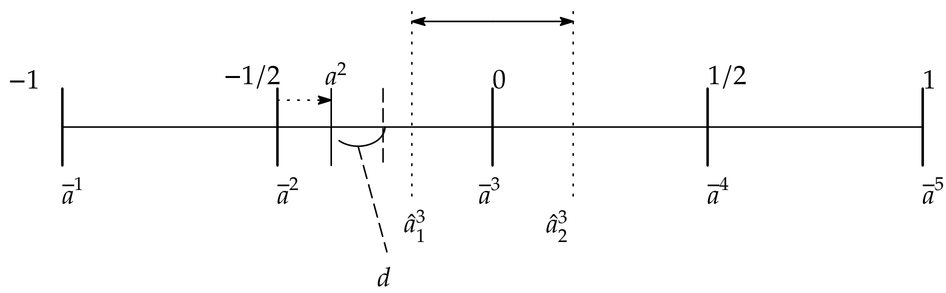

Figure 4.2: Policies by citizen 2 and 3

Now consider citizen 3's incentive to stand for election against citizen 2. Since there are two candidates in this case, the candidate whom the median voter prefers wins in the election. It means that citizen 2 wins with a positive probability only if $\hat{\alpha}^{2*}_{\{2,3\}} \in [\hat{a}^3_1, \hat{a}^3_2]$, that is, the policy citizen 2 would implement is closer to citizen 3's ideal policy. As a result, whoever wins in the election, the policy implemented by the representative is between $\hat{a}^3_1$ and $\hat{a}^3_2$ when citizen 3 stands for election against citizen 2. Recall that if citizen 3 does not stand for the election, then candidate 2 would implement a^2 and citizen 3's (one-shot) political benefit is a^2. Meanwhile, if citizen 3 does stand for election, then citizen 3's (one-shot) net payoff is $-|a^3| - d$ with $a^3 \in [\hat{a}^3_1, \hat{a}^3_2]$ which is strictly greater than a^2 since $- \mid a^3 \mid -d - a^2 \geq \hat{a}^3_1 - d - a^2 > 0$. It means that citizen 3 strictly prefers to stand for the election, which contradicts that σ^* inducing 2-UPC is a DPE.

4.5.2 Deterrence of the Median's Candidacy

The lesson from the above example is that for a strategy inducing k-UPC to be a DPE, the median voter must be able to implement the policies in a wide range. Otherwise the median citizen's entry cannot be prevented. This idea is more generally applicable. In particular, not only the exact median citizen, but also all the citizens closed to the median policy must be able to commit to policies far from their own ideal policy unless the implemented policy on the strategy inducing k-UPC is sufficiently close to the median.

Formally, it is stated as follows.

Proposition 4.6 *Suppose that under Assumption 4.1 a trigger strategy σ^* causing (k,l)-SC with $k < M$ and $\hat{\alpha}^{k*}_{\{k,l\}} \in (\overline{a}^k, \overline{a}^M]$ is a DPE and let*

$$\Xi^* \equiv \left\{ \chi \in \Xi \,\middle|\, \frac{\delta}{1-\delta}\left[v^k(\hat{\alpha}^{k*}_{\{k,l\}}) - d - V^k(\chi) \right] \geq v^k(\overline{a}^K) - v^k(\hat{\alpha}^{k*}_{\{k,l\}}) \right\}.$$

If there exists $k' \in \mathcal{N}$ and $\hat{a} \in (\hat{\alpha}^{k*}_{\{k,l\}}, \overline{a}^M]$ such that

$$v^{k'}(\hat{a}) - d > v^{k'}(\hat{\alpha}^{k*}_{\{k,l\}}), \tag{4.5}$$

and either of the following conditions is satisfied, then there exists no trigger strategy DPE inducing k-UPC where citizen k implements $\hat{\alpha}^{k*}_{\{k,l\}}$.

1. *$k' \in \{k' > k \mid \overline{a}^M > \overline{a}^{k'} \geq \hat{\alpha}^{k*}_{\{k,l\}}\}$ and for any $\underline{\chi} \in \Xi^*$.*

$$\frac{\delta}{1-\delta}[v^{k'}(\hat{\alpha}^{k*}_{\{k,l\}}) - V^{k'}(\underline{\chi})] < v^{k'}(\overline{a}^{k'}) - v^{k'}(\hat{a}), \text{ and} \tag{4.6}$$

$$\frac{\delta}{1-\delta}[v^{k}(\hat{\alpha}^{k*}_{\{k,l\}}) - V^{k}(\underline{\chi})] < v^{k}(\overline{a}^{k}) - v^{k}(\max\{\hat{a}, 2\overline{a}^{k'} - \hat{a}\}). \tag{4.7}$$

2. *$k' \in \{k' \geq M \mid \hat{a} \leq 2\overline{a}^M - \overline{a}^{k'}\}$ and for any $\underline{\chi} \in \Xi^*$.*

$$\frac{\delta}{1-\delta}[v^{k'}(\hat{\alpha}^{k*}_{\{k,l\}}) - V^{k'}(\underline{\chi})] < v^{k'}(\overline{a}^{k'}) - v^{k'}(2\overline{a}^M - \hat{a}). \tag{4.8}$$

If the conditions in Proposition 4.6 are satisfied, then we cannot construct a DPE inducing k-UPC since it cannot prevent citizen k' from entering. Let σ^{**} be a trigger strategy inducing k-UPC with $\hat{\alpha}^{k**}_{\{k\}} = \hat{\alpha}^{k*}_{\{k,l\}}$. The following figures are helpful to explain why the entry by citizen k' cannot be excluded.

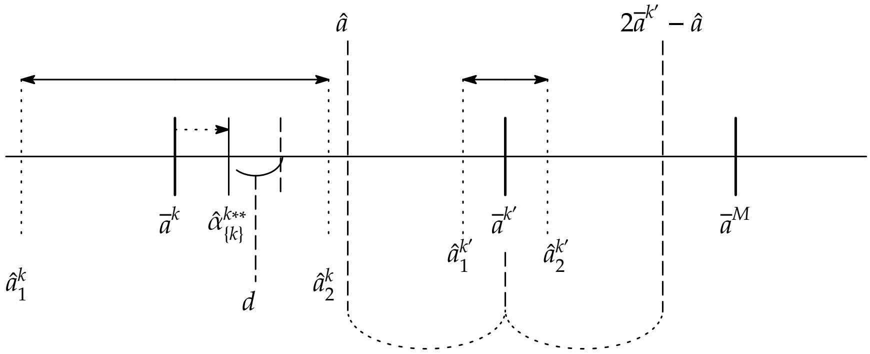

Figure 4.3: When $k' < M$

First, Figure 4.3 illustrates the case where $k' < M$.[20][21] (4.5) means that $\hat{a}$ is beneficial for citizen k' and since $k' > k$, $\hat{a}$ should be on the right side of $\overline{a}^k$. (4.6) implies that the policy implemented by candidate k', $\hat{\alpha}^{k'**}_{\{k,k'\}}$, is restricted into a "small" interval between $\hat{a}^{k'}_1$ to $\hat{a}^{k'}_2$. Then in order to prevent citizen k' from entering, the policy implemented by candidate k against k', $\hat{\alpha}^{k**}_{\{k,k'\}}$, must be outside of $[\hat{a}, 2\overline{a}^{k'} - \hat{a}]$; otherwise regardless of the election outcome the implemented policy is more preferred by citizen k' to $\hat{\alpha}^{k**}_{\{k\}}$. On the one hand, if $\hat{\alpha}^{k**}_{\{k,k'\}}$ is less than $\hat{a}$, it is not supported by the majority since $\hat{\alpha}^{k'**}_{\{k,k'\}}$ is closer to the median. On the other hand, (4.7) guarantees that candidate k cannot choose a policy greater than $2\overline{a}^{k'} - \hat{a}$; this constraint means that the greatest policy candidate k can implement, $\hat{a}^k_2$ in the figure, is less than $2\overline{a}^{k'} - \hat{a}$.

Figure 4.4 explains the reason why entry by citizen k' cannot be excluded when $k' \geq M$. Similar to (4.6) in the case of $k' < M$, (4.8) implies that $\hat{\alpha}^{k'**}_{\{k,k'\}}$, is restricted between $\hat{a}^{k'}_1$ to $\hat{a}^{k'}_2$. In addition, with condition $\overline{a}^{k'} < 2\overline{a}^M - \hat{a}$, it must be less than $2\overline{a}^M - \hat{a}$. Since $\hat{\alpha}^{k'**}_{\{k,k'\}}$ is preferable for citizen k', in order to prevent candidate k' from winning, the policy implemented by candidate k against k', $\hat{\alpha}^{k**}_{\{k,k'\}}$, must be close to the median policy, specifically, at least between $\hat{a}$ and $2\overline{a}^M - \hat{a}$. However by definition of $\hat{a}$, if $\hat{\alpha}^{k**}_{\{k,k'\}}$ is greater than $\hat{a}$, then it is

[20] Both Figure 4.3 and 4.4 implicitly assume that $v^i(a) = - \mid a - \overline{a}^i \mid$ for all $i \in \mathcal{A}$.
[21] Figure 4.3 assumes that $\hat{a} < \overline{a}^k$. The same result is established if $\hat{a} \geq \overline{a}^k$.

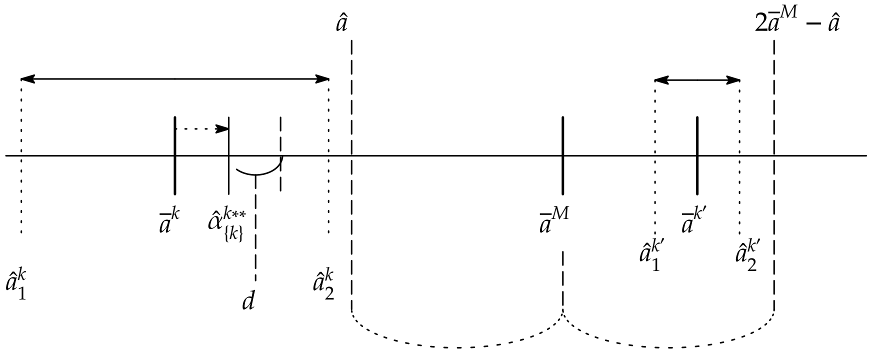

Figure 4.4: When $k' \geq M$

more preferred by citizen k' to $\hat{\alpha}^{k**}_{\{k\}}$. Thus, whoever the representative is, the implemented policy is more preferred by citizen k' to $\hat{\alpha}^{k**}_{\{k\}}$ and then citizen k' can enjoy the political benefit created by his entry.

In general, Proposition 4.6 means that if σ^{**} inducing k-UPC is a DPE, then all the citizens around the median citizen must implement a policy other than their own preferred one when each of them actually stands for election. It is obviously applicable for the exact median citizen M. Nevertheless it is somewhat strange in the sense that while the median citizen can implement his ideal policy which is the Condorcet winner, he must choose a different one. To take this notion into account, suppose that trigger strategy σ^{**} inducing k-UPC must satisfy that $\hat{\alpha}^{M**}_{\{k,M\}} = \overline{a}^M$. Then it can be a DPE only when the implemented policy on the equilibrium, $\hat{\alpha}^{k**}_{\{k\}}$, is sufficiently close to the median and the entry cost is sufficiently large.

Proposition 4.7 *Under Assumption 4.1, if a trigger strategy σ^{**} inducing k-UPC where $\hat{\alpha}^{M**}_{\{k,M\}} = \overline{a}^M$ is a DPE, then $v^M(\overline{a}^M) - d \leq v^M(\hat{\alpha}^{k**}_{\{k\}})$.*

In contrast with this result, given that the entry cost is small enough, there can be a DPE causing (k, l)-SC even if the equilibrium policy is not close to the median. For instance, in the example of Section 4.4.2, $\hat{\alpha}^{3*}_C = 0$ for any $C \in \mathfrak{N}$ such that $3 \in C$. The existence result is generally true if the conditions in Proposition 4.5 are satisfied and $l \neq M$.[22] Intuitively, the median citizen is less attractive to stand for the election in a strategy causing (k, l)-SC since he is the third entrant and cannot enjoy the advantage of the median voter theorem. Therefore, if we should require reasonable behaviour for the median citizen as mentioned above, political compromises may be more robust with strategic candidacy than without strategic candidacy.

In summary, a strategy inducing k-UPC can replicate the policy implemented by (k, l)-SC only when the implemented policy is close to median or all the citizens around the median can commit to implementing policies far from their preferred one. However, the latter is somewhat unreasonable since the citizens close to the median must commit to a policy far from the median to prevent an improvement of political benefit by entry. Thus, even if candidacy is costly, it is often essential for inducing the political compromise from the representative. In other words, strategic candidacy is often necessary for explaining political compromises.

4.6 Concluding Remarks

The outcome caused by a political process usually depends on the characteristics of the potential candidates who make a strategic decision to stand for election. As the example of Aso in the introduction points out, this strategic interaction may happen even when there are only two potential candidates. In this chapter, we have investigated a repeated model of political competition with citizen candidates and showed that strategic candidacy in two-candidate competition can emerge in an election when the candidate

[22] In the proof of Proposition 4.5, we construct a DPE strategy σ^* causing (k, l)-SC satisfying that $\hat{\alpha}_C^{j*} = \overline{a}^j$ for $j \neq k, l$. It implies this statement.

has a reputational concern and commits to implementing a policy different from her ideal policy. We have also demonstrated that it is often easier to induce political compromises from the representative with a strategic candidacy rather than an unopposed political process.

Throughout the chapter, we have mainly focused on situations in which there are exactly two candidates either one of which is sure to lose. For capturing the idea of strategic candidacy for inducing political compromises, it is enough to focus on two-candidate equilibria. Nevertheless to investigate equilibria with three or more candidates also seems to be important for future research as in equilibria with three or more candidates there might coexist two kinds of strategic candidacy, one for changing the winner, which the literature has pointed out, and one for inducing political compromise as studied in this chapter. The interaction of these motives could have interesting implications.

Beyond strategic candidacy, the repeated political competition with citizen candidates has other issues to be investigated for the future. One of the important topics is characterization of the set of policies implemented in repeated political competition which would suggest how an implicit agreement behind the society or some kind of atmosphere influences the implemented policy.[23] While the folk theorem in repeated game theory states that every individually rational payoff can be attained by a subgame perfect equilibrium for a sufficiently high discount factor, it does not necessarily argue that every action in the stage game can be implemented. In addition, a more complicated strategy than the trigger strategy we focus on is typically required for establishing the folk theorem. As a result, it is interesting to ask whether every policy can be implemented in a repeated political competition with citizen candidates and if so, whether it requires a complicated

[23]Besley and Coate (2001) and Felli and Merlo (2006, 2007) study the effect of lobbying on the implemented policy through the political competition with citizen candidates.

strategy or not.

4.7 Appendix: Proofs

4.7.1 Proof of Proposition 4.2

Suppose that $\underline{\chi}$ is a subgame perfect equilibrium. The backward induction tells us that if citizen l stands for election against k, then by assumption on $\underline{\Delta}_{\omega_{\{k,l\}}}$, candidate k implements $\overline{a}^k$ with probability 1 and then $V^l(\underline{\chi}) = v^l(\overline{a}^k) - d$. If citizen l withdraws the candidacy, then since k is the unique candidate, candidate k implements $\overline{a}^k$ with probability 1. Then citizen l's payoff is $v^l(\overline{a}^k)$ which is strictly greater than $V^l(\underline{\chi})$. Therefore $\underline{\chi}$ is not a subgame perfect equilibrium.

4.7.2 Proof of Proposition 4.3, 4.4, and 4.5 and Corollary 4.1

The following lemma is useful to prove the propositions.

Lemma 4.1 *Denote $\hat{\Gamma}^* \equiv \{k,l\}$ and $\hat{\mathfrak{N}}^* = \{\{k\},\{l\},\hat{\Gamma}^*,\hat{\Gamma}^* \cup \{1\},\dots,\hat{\Gamma}^* \cup \{N\}\}$. Then there exists a trigger strategy DPE causing (k,l)-SC if there exists a trigger strategy $\sigma^* \equiv (\hat{\chi}^*,\underline{\chi}^*,F^*)$ satisfying all of the following;*

$$v^l(\hat{\alpha}^{k*}_{\hat{\Gamma}^*}) - d \geq v^l(\hat{\alpha}^{k*}_{\{k\}}), \tag{4.9}$$

$$v^k(\hat{\alpha}^{k*}_{\hat{\Gamma}^*}) - d \geq v^k(\hat{\alpha}^{l*}_{\{l\}}), \tag{4.10}$$

$$\forall i \in \mathcal{N} \setminus \hat{\Gamma}^*, v^i(\hat{\alpha}^{k*}_{\{k,l\}}) \geq \sum_{j\in\hat{\Gamma}^*\cup\{i\}} \Delta_W(j \mid \hat{\Delta}^*_{\omega_{\hat{\Gamma}^*\cup\{i\}}}, \hat{\Gamma}^* \cup \{i\}) v^i(\hat{\alpha}^{j*}_{\hat{\Gamma}^*\cup\{i\}}) - d, \tag{4.11}$$

$$\#\left\{i \in \mathcal{N} \mid v^i(\hat{\alpha}^{k*}_{\hat{\Gamma}^*}) \geq v^i(\hat{\alpha}^{l*}_{\hat{\Gamma}^*})\right\} > \#\left\{i \in \mathcal{N} \mid v^i(\hat{\alpha}^{k*}_{\hat{\Gamma}^*}) < v^i(\hat{\alpha}^{l*}_{\hat{\Gamma}^*})\right\}, \tag{4.12}$$

$$\forall j \in \mathcal{N} \setminus \hat{\Gamma}^*,\ \forall i \in \mathcal{N},\ \forall m \in \left\{m' \mid \hat{\Delta}^i_{\omega_{\hat{\Gamma}^*\cup\{j\}}}(m') > 0\right\}, \forall m'' \in \hat{\Gamma}^* \cup \{j,0\},$$

$$\sum_{n\in\hat{\Gamma}^*\cup\{j\}} \Delta_W(n \mid (m, \hat{\Delta}^{-i*}_{\omega_{\hat{\Gamma}^*\cup\{j\}}}), \hat{\Gamma}^* \cup \{j\}) v^i(\hat{\alpha}^{n*}_{\hat{\Gamma}^*\cup\{j\}})$$

$$\geq \sum_{n\in\hat{\Gamma}^*\cup\{j\}} \Delta_W(n \mid (m'', \hat{\Delta}^{-i*}_{\omega_{\hat{\Gamma}^*\cup\{j\}}}), \hat{\Gamma}^* \cup \{j\}) v^i(\hat{\alpha}^{n*}_{\hat{\Gamma}^*\{j\}}), \tag{4.13}$$

and to vote m is not weakly dominated with respect to $(\hat{\Gamma}^* \cup \{j\}, \hat{\alpha}^*_{\hat{\Gamma}^*\cup\{j\}})$, (4.14)

$$^\forall i \in F^*, {}^\forall C \in \hat{\mathfrak{R}}^* \text{ such that } i \in C, \frac{\delta}{1-\delta}\left[v^i(\hat{\alpha}^{k*}_{\{k,l\}}) - \hat{\gamma}^{i*}d - V^i(\underline{\chi}^*)\right] \geq v^i(\overline{a}^i) - v^i(\hat{\alpha}^{i*}_C) \tag{4.15}$$

$$^\forall i \in \mathcal{N} \setminus F^*, {}^\forall C \in \hat{\mathfrak{R}}^* \text{ such that } i \in C, \hat{\alpha}^{i*}_C = \overline{a}^i. \tag{4.16}$$

Furthermore, if σ^* is a trigger strategy DPE causing (k,l)-SC, then there exists a trigger strategy DPE $(\hat{\chi}^*, \underline{\chi}^*, F^{*\prime})$ satisfying all of the conditions from (4.9) to (4.16) for $F^* = F^{*\prime}$.

Proof (Lemma 4.1) *(Necessity) Suppose that σ^* satisfies (4.9) to (4.16). Construct a trigger strategy $\sigma^{**} \equiv (\hat{\chi}^{**}, \underline{\chi}^{**}, F^{**})$ causing (k,l)-SC where*

$$^\forall i \in \mathcal{N}, \hat{\gamma}^{i**} = \hat{\gamma}^{i*},$$

$$^\forall i \in \mathcal{N}, {}^\forall C \in \mathfrak{R}^* \setminus \{\hat{\Gamma}^*\}, \hat{\Delta}^{i**}_{\omega_C} = \hat{\Delta}^{i*}_{\omega_C},\ \hat{\omega}^{i*}_{\hat{\Gamma}^*} = \begin{cases} k & \text{if } v^i(\hat{\alpha}^k_{\hat{\Gamma}^*}) \geq v^i(\hat{\alpha}^l_{\hat{\Gamma}^*}) \\ l & \text{if } v^i(\hat{\alpha}^k_{\hat{\Gamma}^*}) < v^i(\hat{\alpha}^l_{\hat{\Gamma}^*}), \end{cases},$$

$$^\forall C \in \mathfrak{R}^*, {}^\forall i \in C, \hat{\alpha}^{i**}_C = \hat{\alpha}^{i*}_C,\ {}^\forall C \notin \mathfrak{R}^*, {}^\forall i \in C, \hat{\alpha}^{i*}_C = \overline{a}^i$$

$$\underline{\chi}^{**} = \underline{\chi}^*, F^{**} = F^*,$$

and for each $C \notin \mathfrak{R}^*$ and $i \in \mathcal{N}$, $\hat{\Delta}^{i**}_{\omega_C}$ satisfies that for all $m \in \{m' \mid \hat{\Delta}^{i**}_{\omega_C}(m') > 0\}$ and $m'' \in C \cup \{0\}$,

$$\sum_{n\in C\cup\{0\}} \Delta_W(n \mid (m, \hat{\Delta}^{-i**}_{\omega_C}), C) v^i(\overline{a}^n) \geq \sum_{n\in C\cup\{0\}} \Delta_W(n \mid (m'', \hat{\Delta}^{-i**}_{\omega_C}), C) v^i(\overline{a}^n)$$

and to vote m is not weakly dominated with respect to $(C, \overline{a}_C)$ where $\overline{a}_C \equiv (\overline{a}^i)_{i\in C}$. Part 2 in Proposition 4.1 implies that for each $C \notin \mathfrak{R}^*$ and $i \in \mathcal{N}$, there exists $\hat{\Delta}^{i**}_{\omega_C}$ satisfying these constraints. Constraints (4.9) to (4.16) imply that on σ^{**} there is no citizen to have a deviation incentive after any history and by construction, the voting decision is not weakly dominated. Then the one-shot *deviation principle implies that σ^{**} is a DPE.*

(Sufficiency) Suppose that σ^* causing (k,l)-SC is a trigger strategy DPE. Then the one-shot deviation principle implies that (4.9) and (4.10) mean that citizens k and l prefer to stand for election, (4.11) means that the other citizens prefer to exit from the election. If (4.12) is not satisfied, then there is no voting profile which is weakly undominated with respect to $(\hat{\Gamma}^*, \hat{\alpha}_{\hat{\Gamma}^*})$ and candidate k wins with probability 1. *(4.13) and (4.14) mean that the voting profile is consistent with DPE for a set of the candidates $\hat{\Gamma}^* \cup \{j\}$. Finally, for $i \in F^*$, when i is chosen as the representative from C, she implements policy $\hat{\alpha}_C^{i*}$ if and only if*

$$\begin{aligned} & v^i(\hat{\alpha}_C^{i*}) + \frac{\delta}{1-\delta}[v^i(\hat{\alpha}_{\hat{\Gamma}^*}^{k*}) - \hat{\gamma}^{i*}d] \geq \sup_{a \in \mathcal{A} \setminus \{\hat{\alpha}_C^{i*}\}} v^i(a) + \frac{\delta}{1-\delta} V^i(\underline{\chi}^*) \\ \iff & \frac{\delta}{1-\delta}\left[v^i(\hat{\alpha}_{\hat{\Gamma}^*}^{k*}) - \hat{\gamma}^{i*}d - V^i(\underline{\chi}^*)\right] \geq \sup_{a \in \mathcal{A} \setminus \{\hat{\alpha}_C^{i*}\}} v^i(a) - v^i(\hat{\alpha}_C^{i*}). \end{aligned}$$

Note that it must be satisfied for all $C \in \mathfrak{N}$ such that $i \in C$. If $\hat{\alpha}_C^{i*} \neq \overline{a}^i$, then the right hand side is $v^i(\overline{a}^i) - v^i(\hat{\alpha}_C^{i*}) > 0$ while if $\hat{\alpha}_C^{i*} = \overline{a}^i$, then the right hand side is nonpositive. It implies that if there exists $C \in \{C' \in \mathfrak{N} \mid i \in C'\}$ such that $\hat{\alpha}_C^{i*} \neq \overline{a}^i$, then

$$\max_{C \in \{C' \in \mathfrak{N} | i \in C'\}} \left[\sup_{a \in \mathcal{A} \setminus \{\hat{\alpha}_C^{i*}\}} v^i(a) - v^i(\hat{\alpha}_C^{i*}) \right] = \max_{C \in \{C' \in \mathfrak{N} | i \in C'\}} \left[v^i(\overline{a}^i) - v^i(\hat{\alpha}_C^{i*}) \right]$$

which further implies that

$$\frac{\delta}{1-\delta}\left[v^i(\hat{\alpha}_{\hat{\Gamma}^*}^{k*}) - \hat{\gamma}^{i*}d - V^i(\underline{\chi}^*)\right] \geq \max_{C'' \in \{C' \in \mathfrak{N} | i \in C'\}} \left[v^i(\overline{a}^i) - v^i(\hat{\alpha}_{C''}^{i*})\right] \geq v^i(\overline{a}^i) - v^i(\hat{\alpha}_C^{i*})$$

for all $C \in \mathfrak{N}$. Thus (4.15) is satisfied. For $i \in \mathcal{N} \setminus F^*$, when i is chosen as the representative from

C, she implements policy $\hat{\alpha}^{i*}_C$ if and only if

$$v^i(\hat{\alpha}^{i*}_C) + \frac{\delta}{1-\delta}[v^i(\hat{\alpha}^{k*}_{\hat{\Gamma}^*}) - \hat{\gamma}^{i*}d] \geq \sup_{a \in \mathcal{A} \setminus \{\hat{\alpha}^{i*}_C\}} v^i(a) + \frac{\delta}{1-\delta}[v^i(\hat{\alpha}^{k*}_{\hat{\Gamma}^*}) - \hat{\gamma}^{i*}d]$$

which is satisfied only when $\hat{\alpha}^{i*}_C = \overline{a}^i$. Then (4.16) is satisfied.

Now consider the trigger strategy $(\hat{\chi}^*, \underline{\chi}^*, F^{*\prime})$

$$F^{*\prime} \equiv \left\{ i \in F^* \,\middle|\, ^{\exists}C \in \mathfrak{R}, i \in C, \hat{\alpha}^{i*}_C \neq \overline{a}^i \right\}.$$

The above discussion implies that the new trigger strategy satisfies all of the inequalities in the lemma. □

Proof (Proposition 4.3) Consider a trigger strategy DPE σ^* causing (k,l)-SC. Let $a^k = \hat{\alpha}^k_{\hat{\Gamma}^*}$, $a^l = \hat{\alpha}^l_{\hat{\Gamma}^*}$, $a^k_k = \hat{\alpha}^k_{\{k\}}$, $a^l_l = \hat{\alpha}^l_{\{l\}}$, and $\underline{\chi} = \underline{\chi}^*$. Then (4.9), (4.10), (4.12), (4.15), and (4.16) in Lemma 4.1 imply the results

Proof (Corollary 4.1) Suppose that $k \notin F^*$. Then (4.16) implies that $\hat{\alpha}^{k*}_{\hat{\Gamma}^*} = \hat{\alpha}^{k*}_{\{k\}} = \overline{a}^k$ which does not satisfy (4.9). Given $k \in F^*$, if $v^k(\hat{\alpha}^{k*}_{\{k,l\}}) - d \leq V^k(\underline{\chi}^*)$, then the left hand side of (4.15) for $i = k$ is nonpositive while the right hand side is positive for $C = \{k,l\}$ since $\hat{\alpha}^{k*}_{\{k,l\}} \neq \overline{a}^k$, a contradiction.

Proof (Proposition 4.4) For each $m \in \mathcal{N} \setminus \hat{\Gamma}^*$, let $\overline{j}(m)$ be j satisfying Part 2 of the proposition and $\overline{\mathcal{N}}(m) \equiv \{n \in \mathcal{N} | v^n(a^j_m) \neq \min_{j' \in \{k,l,m\}} v^n(a^{j'}_m) \text{ or } v^n(a^k_m) = v^n(a^l_m) = v^n(a^i_m)\}$.

Consider a trigger strategy σ^* satisfying the following;

$$^{\forall}i \in \mathcal{N}, \hat{\gamma}^{i*} = \begin{cases} 1 & \text{if } i = k, l \\ 0 & \text{otherwise,} \end{cases}$$

$$
{}^{\forall} i \in \mathcal{N}, {}^{\forall} j = k, l, \hat{\omega}^{i*}_{\{j\}} = j, \;\; \hat{\omega}^{i*}_{\hat{\Gamma}^*} = \begin{cases} k & \text{if } v^i(a^k) \geq v^i(a^l) \\ l & \text{if } v^i(a^k) < v^i(a^l), \end{cases}
$$

$$
{}^{\forall} m \in \mathcal{N} \setminus \hat{\Gamma}^*, \hat{\omega}^{i*}_{\hat{\Gamma}^* \cup \{m\}} \begin{cases} = \bar{j}(m) & \text{for } i \in \mathcal{N} \text{ if (4.1) is satisfied or} \\ & i \in \overline{\mathcal{N}}(m) \text{ if (4.2) is satisfied} \\ \in \arg\max_{j' \in \{k,l,m\}} v^i(a^{j'}_m) & \text{otherwise,} \end{cases}
$$

$$
{}^{\forall} j = k, l, \hat{\alpha}^{j*}_{\{j\}} = a^j_j, \hat{\alpha}^{j*}_{\hat{\Gamma}^*} = a^j, \;\; {}^{\forall} i \in \mathcal{N} \setminus \hat{\Gamma}^*, {}^{\forall} j = k, l, i, \hat{\alpha}^{j*}_{\hat{\Gamma}^* \cup \{i\}} = a^j_i, \;\; {}^{\forall} j \in \mathcal{N}, {}^{\forall} C \notin \mathfrak{R}^*, \hat{\alpha}^{j*}_C = \bar{a}^j,
$$

$$
\underline{\chi}^* = \underline{\chi}, \; F^* = \{i \in \mathcal{N} \mid v^i(a^k) - \hat{\gamma}^i d > V^i(\underline{\chi})\}.
$$

The conditions on a^j_i for $i \in \mathcal{N}$ and $j = i, k, l$ and $\underline{\chi}$ imply that the trigger strategy satisfies (4.9), (4.10), (4.11), (4.12), (4.15), and (4.16). Thus Lemma 4.1 implies that there exists a DPE causing (k, l)-SC if σ^* satisfies (4.13) and (4.14). Note that all the voting decisions are deterministic and according to Lemma 4.3 shown in Section 4.8, no voting decision is weakly dominated and then (4.14) is satisfied. When Part 4.1 in the proposition is satisfied, the voting profile implies that candidate $\bar{j}(m)$ obtains strictly more than $N/2$ votes and then wins in the election with probability 1. Changing citizen i's vote with $\omega^{i*}_{\hat{\Gamma}^* \cup \{m\}} \neq \bar{j}(m)$ cannot change the election outcome. When citizen i with $\omega^{i*}_{\hat{\Gamma}^* \cup \{m\}} = \bar{j}(m)$ changes his vote to other $n' \in \hat{\Gamma}^* \cup \{m\} \setminus \{\bar{j}(m)\}$, the probability the most preferable candidate for i, $\alpha^{\bar{j}(m)}_{\hat{\Gamma}^* \cup \{m\}}$, is winning is weakly decreased and that of less preferred policies is weakly increased. Then i has no incentive to change his vote. When Part 4.2 in the proposition is satisfied, the voting profile implies that candidate $\bar{j}(m)$ obtains strictly more than $N/2 + 1$ votes which means that no one can alter the winner by changing his vote. Then (4.13) is satisfied.

Proof (Proposition 4.5) Consider a trigger strategy σ^* satisfying the following;

$$
\begin{aligned}
&{}^\forall i \in \mathcal{N}, \hat{\gamma}^{i*} = \begin{cases} 1 & \text{if } i = k, l \\ 0 & \text{otherwise,} \end{cases} \\
&{}^\forall i \in \mathcal{N}, {}^\forall j = k, l, \hat{\omega}^{i*}_{\{j\}} = j, \;\; \hat{\omega}^{i*}_{\hat{\Gamma}^*} = \begin{cases} k & \text{if } v^i(a^k) \geq v^i(a^l) \\ l & \text{if } v^i(a^k) < v^i(a^l), \end{cases} \\
&{}^\forall j \in \mathcal{N} \setminus \hat{\Gamma}^*, \hat{\omega}^{i*}_{\hat{\Gamma}^* \cup \{j\}} = \begin{cases} j & \text{if } v^i(a^k) = v^i(a^l) < v^i(a^j) \\ \hat{\omega}^{i*}_{\hat{\Gamma}^*} & \text{otherwise,} \end{cases} \\
&\hat{\alpha}^{i*}_C = \begin{cases} a^i_i & \text{for } i = k, l, C = \{i\}, \\ a^i & \text{for } i = k, l, C \in \mathfrak{N}^* \setminus \{\{k\}, \{l\}\}, \\ \overline{a}^i & \text{otherwise,} \end{cases} \\
&\underline{\chi}^* = \underline{\chi}, \; F^* = \{i \in \mathcal{N} \mid v^i(a^k) - \hat{\gamma}^i d > V^i(\underline{\chi})\}.
\end{aligned}
$$

The conditions on a^k, a^l, a^k_k, a^l_l and $\underline{\chi}$ imply that the trigger strategy satisfies (4.9), (4.10), (4.12), (4.15), and (4.16). Thus Lemma 4.1 implies that there exists a DPE causing (k, l)-SC if σ^* satisfies (4.11), (4.13), and (4.14).

There are two cases to be considered.

When $a^k = a^l$: Note that by hypothesis, $a^k = a^l = \overline{a}^M$ and $\hat{\omega}^{i*}_{\hat{\Gamma}^*} = k$ for all $i \in \mathcal{N}$. Then, since $\hat{\omega}^{i*}_{\hat{\Gamma}^* \cup \{j\}}$ stipulates that each i votes for the candidate who would implement the most preferable policy, Lemma 4.3 implies that $\hat{\omega}^{i*}_{\hat{\Gamma}^* \cup \{j\}}$ is not weakly dominated for any $j \in \mathcal{N} \setminus \hat{\Gamma}^*$. Since $a^k = \overline{a}^M$, strictly more than half of the citizens vote for k meaning that candidate k wins with probability 1. Changing citizen i's vote with $\hat{\omega}^{i*}_{\hat{\Gamma}^* \cup \{j\}} \neq k$ cannot change the election outcome. When citizen i with $\hat{\omega}^{i*}_{\hat{\Gamma}^* \cup \{j\}} = k$ changes his vote to other $n' \in \hat{\Gamma}^* \cup \{j\} \setminus \{k\}$, the probability with which a^k is implemented is weakly decreased and that of a^l or $\overline{a}^j$ are weakly increased which implies that i has no incentive to change his vote since $v^i(a^k) = v^i(a^l) \geq v^i(\overline{a}^j)$. Thus (4.13) is satisfied. Finally, since $\Delta_W(k \mid \hat{\omega}^*_{\hat{\Gamma}^* \cup \{j\}}, \hat{\Gamma}^* \cup \{j\}) = 1$ for each $j \in \mathcal{N} \setminus \hat{\Gamma}^*$ and $\hat{\alpha}^{k*}_{\hat{\Gamma}^* \cup \{j\}} = \hat{\alpha}^{k*}_{\hat{\Gamma}^*}$, (4.11) is satisfied.

When $a^k \neq a^l$: We show the case where $a^k < a^l$. The symmetric argument can be applied for the case where $a^k > a^l$.

Since $N \geq 5$, Lemma 4.3 implies that for each $j \in \mathcal{N} \setminus \hat{\Gamma}^*$, $\omega^{i*}_{\hat{\Gamma}^* \cup \{j\}}$ is not weakly dominated for any $j \in \mathcal{N} \setminus \hat{\Gamma}^*$. Since $K = 1$, there is at most one citizen who is indifferent between a^k and a^l. Then for each $j \in \mathcal{N} \setminus \hat{\Gamma}^*$, there is at most one citizen who votes for j and then the probability for j to win is 0. Furthermore, Part 2 in Proposition 4.3 implies that the number of citizens voting for k is at least weakly greater than l.

We furthermore separate three cases to show that (4.13) is satisfied. First, suppose that all the citizens vote for either k or l. Changing the voting decision alters the implemented policy only when k obtains $(N+1)/2$ votes and l obtains $(N-1)/2$ votes. Even in this case, a citizen voting for l cannot alter the election outcome by changing her votes. Note that the voting profile implies that a citizen voting for k prefers a^k to a^l and when she changes her vote to either l , j, or abstaining, it weakly decreases the winning probability of k and weakly increases that of l. It means that she has no incentive to change her vote.

Second, suppose that there is a citizen voting for j. and candidate k obtains at least $(N+1)/2+1$ votes. It implies that k wins with probability 1. Note that k obtains at least 4 votes since $N \geq 5$. Thus a citizen voting for l cannot alter the election outcome by changing the vote. When a citizen voting for k changes her vote to either l, j, or abstaining, it weakly decreases the winning probability of k and weakly increases that of l. Since she weakly prefers a^k to a^l, she has no incentive to change her vote. Finally, citizen j has no incentive to change vote from j since $v^j(\overline{a}^j) > v^j(a^k) = v^j(a^l)$ and there is no possibility to make j win.

Third, suppose that there is a citizen that votes for j and both candidates k and l obtain exactly $(N+1)/2-1$ votes. When $N=5$, we see that $j=3$ and $a^k < \overline{a}^3 < a^l$. Since $a^k \geq 2\overline{a}^2 - \overline{a}^3$, citizens 1 and 2 prefer a^k to $\overline{a}^3$ and $\overline{a}^3$ to a^l and then vote for k. It implies that when citizen 1 or 2 changes her vote to l, 3 or abstaining, it weakly decreases the winning probability of k and weakly increases that of 3 or l. Then they have no incentive to change their vote. Similarly, citizen 4 and 5 have no incentive to change their vote. Finally, citizen 3 has no incentive to change vote since she is indifferent between a^k and a^l. When $N \geq 7$, citizens 1 to $M-1$ vote for k and $N+1$ to N votes for l and both k and l obtain at least 3 votes. When a citizen voting for k changes her vote to l, j or abstaining, it implies that l wins with probability 1. Since she weakly prefers a^k to a^l, she has no incentive to change her vote to l. Similarly, a citizen voting for l has no incentive to change her votes. Finally, a citizen M whose vote goes to j has no incentive to change vote since she is indifferent between a^k and a^l.

The voting profile discussed above implies that if there is no vote for j, then k wins with probability 1. By contrast, if there is one vote for j, then l wins with a positive probability only when k also wins with a positive probability and $v^j(a^k) = v^j(a^l)$. Since $\hat{\alpha}^{i*}_{\hat{\Gamma}^* \cup \{j\}} = \hat{\alpha}^{i*}_{\hat{\Gamma}^*}$ for $i = k, l$, (4.11) is satisfied for each $j \in \mathcal{N} \setminus \hat{\Gamma}^*$.

4.7.3 Proof of Proposition 4.6 and 4.7

The following lemma[24] is useful to prove the propositions.

Lemma 4.2 *There exists a trigger strategy DPE inducing k-UPC if there exists a trigger strategy*

[24]Lemma 4.2 holds without Assumption 4.1.

$\sigma^* \equiv (\hat{\chi}^*, \underline{\chi}^*, F^*)$ satisfying all of the following;

$$v^k(\hat{\alpha}^{k*}_{\{k\}}) - d \geq v^k(\overline{a}^0), \tag{4.17}$$

$$^{\forall} i \in \mathcal{N} \setminus \{k\}, v^i(\hat{\alpha}^{k*}_{\{k\}}) \geq \sum_{j \in \{k,i\}} \Delta_W(j \mid \hat{\Delta}^*_{\omega_{\{k,i\}}}, \{k,i\}) v^i(\hat{\alpha}^{j*}_{\{k,i\}}) - d, \tag{4.18}$$

$$^{\forall} n \in \mathcal{N},\ ^{\forall} i \in \mathcal{N} \setminus \{k\},\ ^{\forall} j \in \{k,i\} \text{ such that } \hat{\Delta}^n_{\omega_{\{k,i\}}}(j) > 0,\ v^n(\hat{\alpha}^{j*}_{\{k,i\}}) = \max_{m \in \{k,i\}} v^n(\hat{\alpha}^{m*}_{\{k,i\}}) \tag{4.19}$$

$$^{\forall} C \in \{\{k\}\} \cup \{\{k,j\} \mid j \in \mathcal{N} \setminus \{k\}\}, \frac{\delta}{1-\delta}\left[v^k(\hat{\alpha}^{k*}_{\{k\}}) - d - V^k(\underline{\chi}^*)\right] \geq v^k(\overline{a}^k) - v^k(\hat{\alpha}^{i*}_C), \tag{4.20}$$

$$^{\forall} i \in F^* \setminus \{k\},\ \frac{\delta}{1-\delta}\left[v^i(\hat{\alpha}^{k*}_{\{k\}}) - V^i(\underline{\chi}^*)\right] \geq v^i(\overline{a}^i) - v^i(\hat{\alpha}^{i*}_{\{k,i\}}), \tag{4.21}$$

$$^{\forall} i \in \mathcal{N} \setminus F^*, \hat{\alpha}^{i*}_{\{k,i\}} = \overline{a}^i. \tag{4.22}$$

Furthermore, if σ^* is a trigger strategy DPE inducing k-USC, then there exists a trigger strategy DPE $(\hat{\chi}^*, \underline{\chi}^*, F^{*\prime})$ satisfying all of the conditions from (4.17) to (4.22) are satisfied for $F^* = F^{*\prime}$.

Proof (Lemma 4.2) *(Necessity) Suppose that σ^* satisfies (4.17) to (4.22). Construct a trigger strategy σ^{**} inducing k-UPC where*

$$^{\forall} i \in \mathcal{N}, \hat{\gamma}^{i**} = \hat{\gamma}^{i*},$$

$$^{\forall} i \in \mathcal{N}, ^{\forall} C \in \{C' \in \mathfrak{R} \mid \#C' = 2, k \in C'\}, \hat{\Delta}^{i**}_{\omega_C} = \hat{\Delta}^{i*}_{\omega_C},$$

$$^{\forall} C \in \{C' \in \mathfrak{R} \mid \#C' \leq 2, k \in C'\}, ^{\forall} i \in C, \hat{\alpha}^{i**}_C = \hat{\alpha}^{i*}_C,$$

$$^{\forall} C \notin \{C' \in \mathfrak{R} \mid \#C' \leq 2, k \in C'\}, ^{\forall} i \in C, \hat{\alpha}^{i*}_C = \overline{a}^i$$

$$\underline{\chi}^{**} = \underline{\chi}^*,\ F^{**} = F^*$$

and for each $C \notin \{C' \in \mathfrak{R} \mid \#C' \leq 2, k \in C'\}$ and $i \in \mathcal{N}$, $\hat{\Delta}^{i**}_{\omega_C}$ satisfies that for all $m \in \{m' \mid \hat{\Delta}^{i**}_{\omega_C}(m') > 0\}$ and $m'' \in C \cup \{0\}$,

$$\sum_{n \in C \cup \{0\}} \Delta_W(n \mid (m, \hat{\Delta}^{-i**}_{\omega_C}), C) v^i(\overline{a}^n) \geq \sum_{n \in C \cup \{0\}} \Delta_W(n \mid (m'', \hat{\Delta}^{-i**}_{\omega_C}), C) v^i(\overline{a}^n)$$

and to vote m is not weakly dominated with respect to $(C, \bar{a}_C)$ where $\bar{a}_C \equiv (\bar{a}^i)_{i \in C}$. The fixed point argument implies that for each $C \notin \{C' \in \mathfrak{N} \mid \#C' \leq 2, k \in C'\}$ and $i \in \mathcal{N}$, there exists $\hat{\Delta}^{i**}_{\omega_C}$ satisfying these constraints. Conditions (4.17) to (4.22) imply that on σ^{**} there is no citizen to have a deviation incentive after any history and by construction, the voting decision is not weakly dominated. Then the one-shot deviation principle implies that σ^{**} is a DPE.

(Sufficiency) The one-shot deviation principle implies the above inequalities by the same procedure of the proof of the sufficiency part in Lemma 4.1. The detail is omitted. □

Proof (Proposition 4.6) Let σ^* be a trigger strategy DPE causing (k, l)-SC and suppose that there exists a trigger strategy DPE σ^{**} inducing k-UPC with $\hat{\alpha}^{k**}_{\{k\}} = \hat{\alpha}^{k*}_{\{k,l\}}$. We show that when the conditions in the proposition are satisfied, not all the constraints from (4.17) to (4.22) (where "$*$" is replaced with"$**$") are satisfied.

First, suppose that for some $k' = k+1, \ldots, M-1$ with $\hat{\alpha}^{k**}_{\{k\}} \leq \bar{a}^{k'}$, there exists $\hat{a}$ satisfying (4.5), (4.6), and (4.7). Let $w \equiv \mid \hat{a} - \bar{a}^{k'} \mid$. (4.6) implies that for any $a \in \mathcal{A} \setminus (\bar{a}^{k'} - w, \bar{a}^{k'} + w)$ and $\underline{\chi} \in \Xi^*$,

$$\frac{\delta}{1-\delta}[v^{k'}(\hat{\alpha}^{k**}_{\{k,l\}}) - V^{k'}(\underline{\chi})] < v^{k'}(\bar{a}^{k'}) - v^{k'}(a).$$

Since σ^{**} is a DPE inducing k-UPC and $\hat{\alpha}^{k**}_{\{k\}} = \hat{\alpha}^{k*}_{\{k,l\}}$, Lemma 4.2 implies that either $\hat{\alpha}^{k'**}_{\{k,k'\}} = \bar{a}^{k'}$ or

$$\frac{\delta}{1-\delta}\left[v^{k'}(\hat{\alpha}^{k**}_{\{k\}}) - V^{k'}(\underline{\chi}^{**})\right] \geq v^{k'}(\bar{a}^{k'}) - v^{k'}(\hat{\alpha}^{k'**}_{\{k,k'\}})$$

for some $\underline{\chi}^{**} \in \Xi^*$. These imply that $\hat{\alpha}^{k'**}_{\{k,k'\}} \in (\bar{a}^{k'} - w, \bar{a}^{k'} + w)$ and then since $\mid \hat{\alpha}^{k'**}_{\{k,k'\}} - \bar{a}^{k'} \mid < w = \mid \hat{a} - \bar{a}^{k'} \mid$, (4.5) further implies that $v^{k'}(\hat{\alpha}^{k'**}_{\{k,k'\}}) - d > v^{k'}(\hat{\alpha}^{k**}_{\{k\}})$. If $\hat{\alpha}^{k**}_{\{k,k'\}} \in (\bar{a}^{k'} - w, \bar{a}^{k'} + w)$, then the similar argument implies that $v^{k'}(\hat{\alpha}^{k**}_{\{k,k'\}}) - d > v^{k'}(\hat{\alpha}^{k**}_{\{k\}})$ and then (4.18) is not satisfied no matter what $\Delta_W(j \mid \hat{\Delta}^{**}_{\omega_{\{k,k'\}}}, \{k,k'\})$ is. If $\hat{\alpha}^{k**}_{\{k,k'\}} \leq \overline{a}^{k'} - w$, then since $\overline{a}^M > \overline{a}^{k'} - w$, $\mid \hat{\alpha}^{k**}_{\{k,k'\}} - \overline{a}^M \mid \geq \mid \overline{a}^{k'} - w - \overline{a}^M \mid > \mid \hat{\alpha}^{k'**}_{\{k,k'\}} - \overline{a}^M \mid$ which implies that with (4.19), $\Delta_W(k' \mid \hat{\Delta}^{**}_{\omega_{\{k,k'\}}}, \{k,k'\}) = 1$ and then (4.18) is not satisfied. Hence $\hat{\alpha}^{k**}_{\{k,k'\}} \geq \overline{a}^{k'} + w$. However note that since $\overline{a}^k < \overline{a}^{k'} + w$, $\mid \hat{\alpha}^{k**}_{\{k,k'\}} - \overline{a}^k \mid > \mid \overline{a}^{k'} + w - \overline{a}^k \mid \geq \mid \overline{a}^{k'} + \overline{a}^{k'} - \hat{a} - \overline{a}^k \mid = \mid 2\overline{a}^{k'} - \hat{a} - \overline{a}^k \mid$ and then $v^k(\hat{\alpha}^{k**}_{\{k,k'\}}) < v^k(2\overline{a}^{k'} - \hat{a})$. Furthermore, recall that by definition of $\hat{a}$, $\max\{2\overline{a}^{k\prime} - \hat{a}, \hat{a}\} > \overline{a}^k$. Then from (4.7)

$$\frac{\delta}{1-\delta}\left[v^k(\hat{\alpha}^{k**}_{\{k\}}) - V^k(\underline{\chi})\right] < v^k(\overline{a}^k) - v^k(\hat{\alpha}^{k**}_{\{k,k'\}})$$

for any $\underline{\chi} \in \Xi^*$ which contradicts (4.20). Therefore there is no $\hat{\alpha}^{k**}_{\{k,k'\}}$ satisfying the conditions.

Suppose now that for some $k' \geq M$ with $\hat{a} \leq 2\overline{a}^M - \overline{a}^{k'}$, there exists $\hat{a}$ satisfying (4.5) and (4.8). Let $y \equiv 2\overline{a}^M - \hat{a} - \overline{a}^{k'} \geq 0$. Then (4.8) implies that for any $a \in \mathcal{A} \setminus (\overline{a}^{k'} - y, \overline{a}^{k'} + y)$ and $\underline{\chi} \in \Xi^*$,

$$\frac{\delta}{1-\delta}[v^{k'}(\hat{\alpha}^{k**}_{\{k,l\}}) - V^{k'}(\underline{\chi})] < v^{k'}(\overline{a}^{k'}) - v^{k'}(a).$$

Note that since $\hat{a} \leq \overline{a}^M \leq \overline{a}^{k'} \leq 2\overline{a}^M - \hat{a}$, $\mid 2\overline{a}^M - \hat{a} - \overline{a}^{k'} \mid \leq \mid 2\overline{a}^M - \hat{a} - \overline{a}^M \mid = \mid \overline{a}^M - \hat{a} \mid \leq \mid \hat{a} - \overline{a}^{k'} \mid$. Then the above inequality is satisfied for $a = \hat{a}$. Meanwhile Lemma 4.2 implies that either $\hat{\alpha}^{k'**}_{\{k,k'\}} = \overline{a}^{k'}$ or

$$\frac{\delta}{1-\delta}\left[v^{k'}(\hat{\alpha}^{k**}_{\{k\}}) - V^{k'}(\underline{\chi}^{**})\right] \geq v^{k'}(\overline{a}^{k'}) - v^{k'}(\hat{\alpha}^{k'**}_{\{k,k'\}})$$

for some $\underline{\chi}^{**} \in \Xi^*$. These imply that $\hat{\alpha}^{k'**}_{\{k,k'\}} \in (\overline{a}^{k'} - y, \overline{a}^{k'} + y)$ and $y > 0$. Note that since $\mid \hat{\alpha}^{k'**}_{\{k,k'\}} - \overline{a}^{k'} \mid < y \leq \overline{a}^{k\prime} - \hat{a} < \overline{a}^{k\prime} - \hat{\alpha}^{k**}_{\{k\}} = \mid \hat{\alpha}^{k**}_{\{k\}} - \overline{a}^{k\prime} \mid$, $v^{k'}(\hat{\alpha}^{k'**}_{\{k,k'\}}) - d > v^{k'}(\hat{\alpha}^{k**}_{\{k\}})$. First, suppose that $\hat{\alpha}^{k**}_{\{k,k'\}} \geq \overline{a}^{k'} + y$. If $\overline{a}^M \leq \hat{\alpha}^{k'**}_{\{k,k'\}} < \overline{a}^{k'}$, then it is obvious that $\mid \hat{\alpha}^{k'**}_{\{k,k'\}} - \overline{a}^M \mid < \mid \hat{\alpha}^{k**}_{\{k,k'\}} - \overline{a}^M \mid$ and if $\overline{a}^M > \hat{\alpha}^{k'**}_{\{k,k'\}}$ then since $\hat{\alpha}^{k'**}_{\{k,k'\}} > \overline{a}^{k'} - y$ and $\overline{a}^{k'} \geq \overline{a}^M$, $\mid \hat{\alpha}^{k'**}_{\{k,k'\}} - \overline{a}^M \mid < \mid \overline{a}^{k'} - y - \overline{a}^M \mid \leq \mid -y \mid = 2\overline{a}^M - \hat{a} - \overline{a}^{k'} \leq \mid \hat{\alpha}^{k**}_{\{k,k'\}} - \overline{a}^{k'} \mid \leq \mid \hat{\alpha}^{k**}_{\{k,k'\}} - \overline{a}^M \mid$. Then in both cases (4.19) implies that

$\Delta_W(k' \mid \hat{\Delta}^{**}_{\omega_{\{k,k'\}}}, \{k,k'\}) = 1$ and then (4.18) is not satisfied. Next, suppose that $\hat{\alpha}^{k**}_{\{k,k'\}} \in [\hat{a}, \overline{a}^{k'} + y)$. If $\hat{\alpha}^{k**}_{\{k,k'\}} \in (\overline{a}^{k'} - y, \overline{a}^{k'} + y)$, then the argument similar to $\hat{\alpha}^{k'**}_{\{k,k'\}}$ implies that $v^{k'}(\hat{\alpha}^{k**}_{\{k,k'\}}) - d > v^{k'}(\hat{\alpha}^{k**}_{\{k\}})$ and if $\hat{\alpha}^{k**}_{\{k,k'\}} \in (\hat{a}, \overline{a}^{k'} - y]$, then since $\overline{a}^{k'} - y < \overline{a}^{k'}$, the definition of $\hat{a}$ implies that $v^{k'}(\hat{\alpha}^{k**}_{\{k,k'\}}) - d > v^{k'}(\hat{\alpha}^{k**}_{\{k\}})$. Then in both cases (4.18) is not satisfied no matter what $\Delta_W(j \mid \hat{\Delta}^{**}_{\omega_{\{k,k'\}}}, \{k,k'\})$ is. Finally suppose that $\hat{\alpha}^{k**}_{\{k,k'\}} \le \hat{a}$. Note that since $\hat{a} \le \overline{a}^{k'} - y < \hat{\alpha}^{k'**}_{\{k,k'\}}$. If $\overline{a}^M \ge \hat{\alpha}^{k'**}_{\{k,k'\}}$, then it is obvious that $\mid \hat{\alpha}^{k'**}_{\{k,k'\}} - \overline{a}^M \mid < \mid \hat{\alpha}^{k**}_{\{k,k'\}} - \overline{a}^M \mid$ and if $\overline{a}^M < \hat{\alpha}^{k'**}_{\{k,k'\}}$, then since $\hat{\alpha}^{k'**}_{\{k,k'\}} < \overline{a}^{k'} + y = 2\overline{a}^M - \hat{a}$, $\mid \hat{\alpha}^{k**}_{\{k,k'\}} - \overline{a}^M \mid \ge \mid 2\overline{a}^M - \hat{a} - \overline{a}^M \mid = \mid \hat{a} - \overline{a}^M \mid \ge \mid \hat{\alpha}^{k'**}_{\{k,k'\}} - \overline{a}^M \mid$. Then in both cases (4.19) implies that $\Delta_W(k' \mid \hat{\Delta}^{**}_{\omega_{\{k,k'\}}}, \{k,k'\}) = 1$ and then (4.18) is not satisfied. Therefore there is no $\hat{\alpha}^{k**}_{\{k,k'\}}$ satisfying the conditions.

Proof (Proposition 4.7) Suppose that a trigger strategy σ^{**} inducing k-UPC with $\hat{\alpha}^{M**}_{\{k,M\}} = \overline{a}^M$ is a DPE. Since σ^{**} is a DPE, all the constraints from (4.17) to (4.22) (where "$*$" is replaced with"$**$") must be satisfied. Suppose that $\hat{\alpha}^{k**}_{\{k,M\}} \neq \overline{a}^M$. Then under Assumption 4.1, more than half of the citizens strictly prefer $\overline{a}^M$ to $\hat{\alpha}^{k**}_{\{k,M\}}$ which implies that with (4.19), $\Delta_W(M \mid \hat{\Delta}^{**}_{\omega_{\{k,M\}}}, \{k,M\}) = 1$. Then from (4.18), $v^M(\hat{\alpha}^{k**}_{\{k\}}) \ge v^M(\overline{a}^M) - d$. Suppose now that $\hat{\alpha}^{k**}_{\{k,M\}} = \overline{a}^M$. Then, since $v^M(\overline{a}^M) = v^M(\hat{\alpha}^{k**}_{\{k,M\}})$, no matter what $\Delta_W(M \mid \hat{\Delta}^{**}_{\omega_{\{k,M\}}}, \{k,M\})$ is, (4.18) implies that $v^M(\hat{\alpha}^{k**}_{\{k\}}) \ge v^M(\overline{a}^M) - d$.

4.8 Appendix: Characterization of Weakly Dominated Voting

Lemma 4.3 *Given a set of candidates C and a profile of policies $(a_C) \equiv (a^i_C)_{i \in C}$, a weakly undominated voting decision for citizen i with respect to (C, a_C) is characterized as follows.*

- *If $v^i(a^x_C) = v^i(a^y_C)$ for all $x, y \in C$, any voting decision is not weakly dominated.*

- *If there exist $x, y \in C$ such that $v^i(a_C^x) > v^i(a_C^y)$,*
  - *to abstain or to vote for $l \in \arg\min_{j \in N} v^i(a_C^j)$ is weakly dominated,*
  - *to vote for $l \notin \arg\max_{j \in N} v^i(a_C^j)$ is weakly dominated if $N = 3$, $C = \{x, y, z\}$, and x, y, and z satisfy that*

$$v^i(a_C^x) > v^i(a_C^y) > v^i(a_C^z) \tag{4.23}$$

$$\frac{v^i(a_C^x) + v^i(a_C^z)}{2} \geq v^i(a_C^y), \tag{4.24}$$

    and
  - *otherwise the voting decision is not weakly dominated.*

Proof of Lemma 4.3

First, consider $i \in \mathcal{N}$ such that $v^i(a_C^x) = v^i(a_C^y)$ for all $x, y \in C$. Then her voting does not change the payoff $v^i(\cdot)$ at all and all the voting decisions including to abstain are indifferent for i regardless of the opponent's decision. Thus no voting decision is weakly dominated.

For the rest of the proof, consider voter $i \in \mathcal{N}$ such that there exist $x, y \in C$ such that $v^i(a_C^x) > v^i(a_C^y)$. The proof consists of several steps. Let $\underline{C} \equiv C \cup \{0\}$. Given voting profile $\omega \in \underline{C}$, denote

$$\begin{aligned}
\tilde{W}(\omega) &\equiv \arg\max_{j \in C} \#\{i \in \mathcal{N} \mid \omega^i(a_C) = j\}, \\
\tilde{L}(\omega) &\equiv \left\{ j \in C \,\middle|\, \#\{i \in \mathcal{N} \mid \omega^i(a_C) = j\} = \#\tilde{W}(\omega) - 1 \right\}, \\
\tilde{T}(\omega) &\equiv \left\{ j \in C \,\middle|\, \#\{i \in \mathcal{N} \mid \omega^i(a_C) = j\} = \#\tilde{W}(\omega) - 2 \right\}
\end{aligned}$$

those are respectively, the set of the winning candidates, that of the candidates obtaining exactly one vote less than the winners, and that of the candidates obtaining exactly two less votes than the winners. Fix $m \in \arg\max_{j \in C} v^i(a_C^j)$. If $\omega^i \neq m$, then we see that

$$\tilde{W}(m, \omega^{-i}) = \begin{cases} \{m\} & \text{if } m \in \tilde{W}(\omega) \text{ or } \omega^{-i} = (0, \dots, 0) \\ \tilde{W}(\omega) \cup \{m\} \setminus \{\omega^i\} & \text{if } m \in \tilde{L}(\omega) \\ \{\omega^i\} \cup \tilde{L}(\omega) \cup \{m\} & \text{if } m \in \tilde{T}(\omega) \text{ and } \tilde{W}(\omega) = \{\omega^i\} \\ \tilde{W}(\omega) \setminus \{\omega^i\} & \text{otherwise.} \end{cases}$$

Step 1: We show that when $\omega^i \in (\arg\min_{j \in C} v^i(a_C^j)) \cup \{0\}$, to vote for m makes voter i weakly better off for any $\omega^{-i} \in \underline{C}^{N-1}$ and strictly better off for some $\omega^{-i} \in \underline{C}^{N-1}$.

Fix $\omega^i \in (\arg\min_{j \in C} v^i(a_C^j)) \cup \{0\}$ and $\omega^{-i} \in \underline{C}^{N-1}$. Suppose that $m \in \tilde{W}(\omega)$ or $\omega^{-i} = (0, \dots, 0)$. Then the net payoff by changing to vote m from ω^i is

$$v^i(a_C^m) - \frac{\sum_{n \in \tilde{W}(\omega)} v^i(a_C^n)}{\#\tilde{W}(\omega)} = \frac{\sum_{n \in \tilde{W}(\omega)}}{\#\tilde{W}(\omega)} [v^i(a_C^m) - v^i(a_C^n)] \geq 0. \tag{4.25}$$

Suppose that $m \in \tilde{L}(\omega)$. When $\omega^i \in \tilde{W}(\omega)$ (implying that $\omega^i \neq 0$), $\#\tilde{W}(\omega) = \#\tilde{W}(m, \omega^{-i})$. Then the net payoff by changing to vote for m from ω^i is

$$\frac{\sum_{n \in \tilde{W}(\omega) \cup \{m\} \setminus \{\omega^i\}} v^i(a_C^n)}{\#\tilde{W}(\omega)} - \frac{\sum_{n \in \tilde{W}(\omega)} v^i(a_C^n)}{\#\tilde{W}(\omega)} = \frac{1}{\#\tilde{W}(\omega)} [v^i(a_C^m) - v^i(a_C^{\omega^i})] > 0. \tag{4.26}$$

When $\omega^i \notin \tilde{W}(\omega)$, $\tilde{W}(m, \omega^{-i}) = \tilde{W}(\omega) \cup \{m\}$ implying that $\#\tilde{W}(m, \omega^{-i}) = \#\tilde{W}(\omega) + 1$. Then the net payoff by changing to vote for m from ω^i is

$$\begin{aligned} \frac{\sum_{n \in \tilde{W}(\omega) \cup \{m\}} v^i(a_C^n)}{\#W(\omega) + 1} - \frac{\sum_{n \in \tilde{W}(\omega)} v^i(a_C^n)}{\#\tilde{W}(\omega)} &= \frac{1}{(\#\tilde{W}(\omega) + 1)\#\tilde{W}(\omega)} \sum_{n \in \tilde{W}(\omega)} \left[v^i(a_C^m) - v^i(a_C^n)\right] \\ &\geq 0 \end{aligned} \tag{4.27}$$

with strict inequality holds if there exists $l \in \tilde{W}(\omega)$ such that $l \notin \arg\max_n v^i(a_C^n)$.

Suppose that $m \in \tilde{T}(\omega)$. When $\tilde{W}(\omega) = \{\omega^i\}$, the net payoff by changing to vote for m from ω^i is

$$\frac{\sum_{n\in\tilde{L}(\omega)\cup\{m\}\cup\{\omega^i\}} v^i(a_C^n)}{\#\tilde{L}(\omega)+2} - v^i(a_C^{\omega^i}) = \frac{1}{\#\tilde{L}(\omega)+2} \sum_{n\in\tilde{L}(\omega)\cup\{m\}\cup\{\omega^i\}} [v^i(a_C^n) - v^i(a_C^{\omega^i})] > 0. \tag{4.28}$$

When $\tilde{W}(\omega) \neq \{\omega^i\}$ and $\omega^i \in \tilde{W}(\omega)$, $\#\tilde{W}(m, \omega^{-i}) = \#\tilde{W}(\omega) - 1$. Then the net payoff by changing to vote for m from ω^i is

$$\begin{aligned}
\frac{\sum_{n\in\tilde{W}(\omega)\setminus\{\omega^i\}} v^i(a_C^n)}{\#\tilde{W}(\omega)-1} - \frac{\sum_{n\in\tilde{W}(\omega)} v^i(a_C^n)}{\#\tilde{W}(\omega)} &= \frac{1}{(\#\tilde{W}(\omega)-1)\#\tilde{W}(\omega)} \sum_{n\in\tilde{W}(\omega)\setminus\{\omega^i\}} [v^i(a_C^n) - v^i(a_C^{\omega^i})] \\
&\geq 0. \qquad (4.29)
\end{aligned}$$

When $\tilde{W}(\omega) \neq \{\omega^i\}$ and $\omega^i \notin \tilde{W}(\omega)$, the set of the policies to possibly be implemented is not altered by changing the vote for m from ω^i implying that to vote m is indifferent from ω^i for i.

Finally, suppose that $m \notin \tilde{W}(\omega) \cup \tilde{L}(\omega) \cup \tilde{T}(\omega)$. When $\omega^i \in \tilde{W}(\omega)$ (implying that $\omega^i \neq 0$), $\tilde{W}(m, \omega^{-i}) = \tilde{W}(\omega) \setminus \{\omega^i\}$ and $\#\tilde{W}(m, \omega^{-i}) = \#\tilde{W}(\omega) - 1$. The net payoff by changing to vote for m from ω^i is

$$\begin{aligned}
\frac{\sum_{n\in\tilde{W}(\omega)\setminus\{\omega^i\}} v^i(a_C^n)}{\#\tilde{W}(\omega)-1} - \frac{\sum_{n\in\tilde{W}(\omega)} v^i(a_C^n)}{\#\tilde{W}(\omega)} &= \frac{1}{\#\tilde{W}(\omega)(\#\tilde{W}(\omega)-1)} \sum_{n\in\tilde{W}(\omega)} \left[v^i(a_C^n) - v^i(a_C^{\omega^i})\right] \\
&\geq 0. \qquad (4.30)
\end{aligned}$$

When $\omega^i \notin \tilde{W}(\omega)$, $\tilde{W}(m, \omega^{-i}) = \tilde{W}(\omega)$. Then the set of the alternatives possibly to be implemented is not altered by changing the vote for m from ω^i implying that to vote m is indifferent from ω^i for i.

Thus for any $\omega^{-i} \in \underline{C}^{N-1}$, to vote for m makes voter i weakly better off. To show that it makes her strictly better off for some ω^{-i}, consider

- For $\omega^i \in \arg\min_{n\in\mathcal{N}} v^i(n)$, $\omega^{-i} = (0, \ldots, 0)$
- For $\omega^i = 0$, $\omega^{-i} = (l, 0, \ldots, 0)$ where $l \in \arg\min_{n\in\mathcal{N}} v^i(a_C^n)$ (or $\omega^{-i} = l$ when $N = 2$)

In each case, it is easy to see that $m \in \tilde{L}(\omega)$ and (4.26) and (4.27) imply that to vote for m makes voter i strictly better off.

Step 2: We show that if $\#C = 2$, then to vote for $m \in \arg\max_{n\in\mathcal{N}} v^i(a_C^n)$ is not weakly dominated. Step 1 directly implies that if $\#C = 2$, then the unique voting decision which is not weakly dominated is to vote for m.

Step 3: We show that if $N = 3$, $C = \{m, y, z\}$, $v^i(a_C^m) > v^i(a_C^y) > v^i(a_C^z)$, and $(v^i(a_C^m)+v^i(a_C^z))/2 \geq v^i(a_C^y)$, then to vote for m makes voter i weakly better off than to vote $\omega^i = y$ for any ω^{-i} and strictly better off for some ω^{-i}.

By the same procedure as the proof of Step 1, we can show that if ω satisfies that $m \in \tilde{W}(\omega) \cup \tilde{L}(\omega)$ or $\omega^i \notin \tilde{W}(\omega)$, then to vote m makes voter i weakly better off for any of such ω^{-i} (and strictly better off for some of such ω^{-i}). Thus it is satisfied for any $\omega^{-i} \in \underline{C}^{I-1}$ if and only if

- for any ω^{-i} such that $m \in \tilde{T}(\omega)$ and $\tilde{W}(\omega) = \{\omega^i\}$, (4.28) is satisfied (with weak inequality),
- for any ω^{-i} such that $m \in \tilde{T}(\omega)$, $\tilde{W}(\omega) \neq \{\omega^i\}$, and $\omega^i \in \tilde{W}(\omega)$, (4.29) is satisfied, and
- for any ω^{-i} such that $m \notin \tilde{W}(\omega) \cup \tilde{L}(\omega) \cup \tilde{T}(\omega)$ and $\omega^i \in \tilde{W}(\omega)$, (4.30) is satisfied.

Note that whenever $m \in \tilde{T}(\omega)$ and $\omega^i = y \in \tilde{W}(\omega)$, it must be the case that $\#\{i \in \mathcal{N} \mid \omega^i(a_C) = m\} = 0$, $\#\{i \in \mathcal{N} \mid \omega^i(a_C) = y\} = 2$, and $\tilde{W}(\omega) = \{y\}$. Then the possible voting profile is $(\omega^i, \omega^{-i}) = (y, y, 0), (y, 0, y), (y, y, z), (y, z, y)$. For $(\omega^i, \omega^{-i}) = (y, y, 0), (y, 0, y)$, it is obvious that $\tilde{L}(\omega) = \emptyset$ implying that (4.28) is satisfied. For $(\omega^i, \omega^{-i}) = (y, y, z), (y, z, y)$, we see that $\tilde{L}(\omega) = \{z\}$ and then (4.28) is satisfied if and only if

$$\sum_{n=m,y,z} v^i(a_C^n) - 3v^i(a_C^y) \geq 0 \tag{4.31}$$

or equivalently $(v^i(a_C^m) + v^i(a_C^z))/2 \geq v^i(a_C^y)$.

When $m \notin \tilde{W}(\omega) \cup \tilde{L}(\omega) \cup \tilde{T}(\omega)$ and $\omega^i = y \in \tilde{W}(\omega)$, the unique possible voting profile is $(\omega^i, \omega^{-i}) = (y, y, y)$. Then $W(\omega) = \{y\}$ implying that (4.30) is satisfied with equality.

Step 4: We show that if $N = 3$, $C = \{m, y, z\}$, $v^i(a_C^m) > v^i(a_C^y) > v^i(a_C^z)$, and $(v^i(a_C^m)+v^i(a_C^z))/2 < v^i(a_C^y)$, then there is an opponents' voting file ω^{-i} such that to vote for m makes voter i strictly worse off than to vote $\omega^i = y$.

As (4.31) shows, given $\omega^{-i} = (y, z)$ or (z, y) to vote for m makes voter i strictly worse off than to vote $\omega^i = y$.

Step 5: We show that if $N > \#C = 3$ or $N \geq \#C \geq 4$, then for any $\omega^i \notin \arg\min_{n \in C} v^i(a_C^n) \cup \{0\}$, there is an opponents' voting profile $\omega^{-i} \in \underline{C}^{N-1}$ such that to vote for ω^i is the strict best response for ω^{-i}.

Note that the hypothesis implies that $N \geq 4$. Now for $\omega^i \notin \arg\min_{n \in N} v^i(a_C^n) \cup \{0\}$, consider the following opponent profile;

$$\omega^{-i} = (l, l, \omega^i, 0, \ldots, 0)$$

(or $\omega^{-i} = (l, l, \omega^i)$ when $N = 4$) where $l \in \arg\min_{n \in N} v^i(a_C^n)$. Then we see that $\tilde{W}(\omega) = \{\omega^i, l\}$ and $\tilde{W}(n, \omega^{-i}) = \{l\}$ for any $n \neq \omega^i$. The difference of the payoff between voting for $n \neq \omega^i$

and ω^i is

$$\frac{v^i(a_C^{\omega^i}) + v^i(a_C^l)}{2} - v^i(a_C^l) = \frac{v^i(a_C^{\omega^i}) - v^i(a_C^l)}{2} > 0$$

which means that to vote for ω^i is the strict best response for ω^{-i}.

Bibliography

Abreu, D. (1986): "Extremal Equilibria of Oligopolistic Supergame," *Journal of Economic Theory*, 39, 191–225.

——— (1988): "On the Theory of Infinitely Repeated Games with Discounting," *Econometrica*, 56, 383–396.

Alesina, A. (1988): "Credibility and Policy Convergence in a Two-Party System with Rational Voters," *American Economic Review*, 78, 796–805.

Alonso, R. and N. Matouschek (2007): "Relational Delegation," *RAND Journal of Economics*, 38, 1070–1089.

Aragonès, E., T. Palfrey, and A. Postlewaite (2007): "Political Reputations and Campaign Promises," *Journal of the European Economic Association*, 5, 846–884.

Asako, Y. (2010): "Why Do Some Candidates Run to Lose?: Partially Binding Platforms and Endogenous Candidates," *mimeo*, Bank of Japan.

Asanuma, B. (1989): "Manufacturer-Supplier Relationships in Japan and the Concept of Relation-Specific Skill," *Journal of the Japanese and International Economies*, 3, 1–30.

Aso, T. (2007): "Orega Atarashii Jimintoo Tsukuru (I Make the New LDP)," *Bungei Shunju*, 106–115, (in Japanese, Available at http://www.aso-taro.jp/lecture/2007_11bungei.pdf with the title "Jiminto Sosaisen (LDP President Election)").

Athey, S. and D. A. Miller (2007): "Efficiency in Repeated Trade with Hidden Valuations," *Theoretical Economics*, 299–354.

Baker, G., R. Gibbons, and K. J. Murphy (1994): "Subjective Performance Measures in Optimal Incentive Contracts," *Quarterly Journal of Economics*, 109, 1125–1156.

——— (2002): "Relational Contracts and the Theory of the Firm," *Quarterly Journal of Economics*, 117, 39–84.

Battaglini, M. (2005): "Long-Term Contracting with Markovian Consumers," *American Economic Review*, 95, 637–658.

Bergemann, D. and J. Välimäki (2003): "Dynamic Common Agency," *Journal of Economic Theory*, 111, 23–48.

Bernheim, B. D. and M. D. Whinston (1986): "Menu Auctions, Resource Allocation, and Economic Infuluence," *Quarterly Journal of Economics*, 101, 1–32.

——— (1998): "Incomplete Contracts and Strategic Ambiguity," *American Economic Review*, 88, 902–932.

Besley, T. and S. Coate (1997): "An Economic Model of Representative Democracy," *Quarterly Journal of Economics*, 112, 85–114.

——— (1998): "Sources of Inefficiency in a Representative Democracy: a Dynamic Analysis," *American Economic Review*, 88, 139–156.

——— (2001): "Lobbying and Welfare in a Representative Democracy," *Review of Economic Studies*, 68, 67–82.

Board, S. (2010): "Relational Contracts and the Value of Loyalty," *American Economics Review, forthcoming*.

Calzolari, G. and G. Spagnolo (2010): "Relational Contracts and Competitive Screening," Centre for Economic Policy Research Discussion Paper No. 7434.

Campante, F. R. and F. H. Ferreira (2007): "Inefficient Lobbying, Populism and Oligarchy," *Journal of Public Economics*, 91, 993–1021.

Daido, K. (2006): "Formal and Relational Incentives in a Multitask Model," *International Review of Law and Economics*, 26, 380–394.

Dewatripont, M., I. Jewitt, and J. Tirole (1999): "The Economics of Career Concerns, Part II: Application to Missions and Accountability of Government Agencies," *Review of Economic Studies*, 66, 199–217.

——— (2000): "Multitask Agency Problems: Focus and Task Clustering," *European Economic Review*, 44, 869–877.

Dixit, A., G. M. Grossman, and F. Gul (2000): "The Dynamics of Political Compromise," *Journal of Political Economy*, 108, 531–568.

Dixit, A., G. M. Grossman, and E. Helpman (1997): "Common Agency and Coordination: General Theory and Applicatoin to Government Policy Making," *Journal of Political Economy*, 105, 752–769.

Duggan, J. and M. Fey (2006): "Repeated Downsian Electoral Competition," *International Journal of Game Theory*, 35, 39–69.

Dutta, B., M. O. Jackson, and M. Le Breton (2001): "Strategic Candidacy and Voting Procedures," *Econometrica*, 69, 1013–1037.

——— (2002): "Voting by Successive Elimination and Strategic Candidacy," *Journal of Economic Theory*, 103, 190–218.

Ehlers, L. and J. A. Weymark (2003): "Candidate Stability and Nonbinary Social Choice," *Economic Theory*, 22, 233–243.

Eraslana, H. and A. McLennan (2004): "Strategic Candidacy for Multivalued Voting Procedures," *Journal of Economic Theory*, 117, 29–54.

Felli, L. and A. Merlo (2006): "Endogenous Lobbying," *Journal of the European Economic Association*, 4, 180–215.

——— (2007): "If You Cannot Get Your Friends Elected, Lobby Your Enemies," *Journal of the European Economic Association*, 5, 624–635.

Fong, Y.-F. and J. Li (2010): "Relational Contract, Efficiency Wages, and Employment Dynamics," *mimeo*, Northwestern Univesity.

Fudenberg, D., D. Levine, and E. Maskin (1994): "The Folk Theorem with Imperfect Public Information," *Econometrica*, 62, 997–1039.

Gibbons, R. (1998): "Incentives in Organizations," *Journal of Economic Perspectives*, 12, 115–132.

——— (2005a): "Four Formal(izable) Theories of the Firm?" *Journal of Economic Behavior and Organization*, 58, 200–245.

——— (2005b): "Incentives between Firms (and Within)," *Management Science*, 51, 2–17.

Grossman, G. M. and E. Helpman (1994): "Protection for Sale," *American Economic Review*, 84, 833–850.

——— (2001): *Special Interest Politics*, Cambridge, MA: MIT Press.

Halac, M. (2009): "Relational Contracts and the Value of Relationships," *mimeo*, Columbia University.

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

Iossa, E. and G. Spagnolo (2009): "Contracts as Threats: on a Rationale for Rewarding A While Hoping for B," *mimeo*, Brunel University.

Itoh, H. (1994): "Job Design, Delegation and Cooperation: A Principal-Agent Analysis," *European Economic Review*, 38, 691–700.

Itoh, H. and H. Morita (2010): "Formal Contracts, Relational Contracts, and the Threat-Point Effect," *mimeo*, Hitotsubashi University.

Jackson, M. O. and H. F. Sonnenschein (2007): "Overcoming Incentive Constraints by Linking Decisions," *Econometrica*, 75, 241–257.

Kovrijnykh, N. (2010): "Debt Contracts with Partial Commitment," *mimeo*, Arizona State University.

Kvaløy, O. and T. E. Olsen (2009): "Endogenous Verifiability and Relational Contracting," *American Economic Review*, 99, 2193–2208.

Levin, J. (2002): "Multilateral Contracting and the Employment Relationship," *Quarterly Journal of Economics*, 117, 1075–1104.

——— (2003): "Relational Incentive Contracts," *American Economic Review*, 93, 835–857.

MacLeod, W. B. (2007): "Reputations, Relationships, and Contract Enforcement," *Journal of Economic Literature*, 45, 595–628.

MacLeod, W. B. and J. M. Malcomson (1989): "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment," *Econometrica*, 57, 447–480.

——— (1998): "Motivation and Markets," *American Economic Review*, 88, 388–411.

Malcomson, J. M. (2010): "Relational Incentive Contracts," *mimeo*, Oxford Univesity.

Matsushima, H., K. Miyazaki, and N. Yagi (2010): "Role of Linking Mechanisms in Multitask Agency with Hidden Information," *Journal of Economic Theory*, 145, 2241–2259.

McCarty, N. and L. S. Rothenberg (1996): "Commitment and the Campaign Contribution Contract," *American Journal of Political Science*, 40, 872–904.

Meyer, M. A., T. E. Olsen, and G. Torsvik (1996): "Limited Intertemporal Commitment and Job Design," *Journal of Economic Behavior and Organization*, 31, 401–417.

Milgrom, P. and I. Segal (2002): "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, 70, 583–601.

Mookherjee, D. (2006): "Decentralization, Hierarchies, and Incentives: A Mechanism Design Perspective," *Journal of Economic Literature*, 44, 367–390.

Osborne, M. J. and A. Slivinski (1996): "An Model of Political Competition with Citizen Candidates," *Quarterly Journal of Economics*, 111, 65–96.

Pearce, D. G. and E. Stacchetti (1998): "The Interaction of Implicit and Explicit Contracts in Repeated Agency," *Games and Economic Behavior*, 23, 75–96.

Ray, D. (2002): "The Time Structure of Self-Enforcing Agreements," *Econometrica*, 70, 547–582.

Rayo, L. (2007): "Relational Incentives and Moral Hazard in Teams," *Review of Economic Studies*, 74, 937–963.

Samejima, Y. (2007): "Strategic Candidacy and Single-Peakedness," *Japanese Economic Review*, 58, 423–442.

Schmidt, K. M. and M. Schnitzer (1995): "The Interaction of Explicit and Implicit Contracts," *Economics Letters*, 48, 193–199.

Schöttner, A. (2008): "Relational Contracts, Multitasking, and Job Design," *Journal of Law, Economics and Organization*, 24, 138–162.

Shapiro, C. and J. Stiglitz (1984): "Equilibrium Unemployment as a Worker Discipline Device," *American Economic Review*, 74, 433–444.

Snyder, Jr., J. M. (1992): "Long-term Investing in Politicians; or, Give Early, Give Often," *Journal of Law and Economics*, 35, 15–43.

Strausz, R. (2006): "Deterministic versus Stochastic Mechanisms in Principal-Agent Models," *Journal of Economic Theory*, 128, 306–314.

Thomas, J. and T. Worrall (1994): "Foreign Direct Investment and the Risk of Expropriation," *Review of Economic Studies*, 61, 81–108.

Wen, Q. (2002): "A Folk Theorem for Repeated Sequential Games," *Review of Economic Studies*, 69, 493–512.

Wolitzky, A. (2010): "Dynamic Monopoly with Relational Incentives," *Theoretical Economics*, 5, 479–518.