Writing Legally Unenforceable Contracts to Facilitate Relationships[*]

Hideshi Itoh[†]

First Version: August 16, 2009

This Version: June 2011

[*]Preliminary and incomplete. Comments are welcome. I am grateful to Kohei Kawamura for helpful conversation at an earlier stage of this research, Junichiro Ishida, Shingo Ishiguro, Shinsuke Kambe, Hajime Kobayashi, Michael Riordan, and the seminar participants at Osaka University, University of Stavanger, Norwegian School of Economics and Business Administration (NHH), University of New South Wales, CTW, Hitotsubashi University, and Kobe University for helpful comments, and especially Andrew Daughety for detailed comments and encouragement.

[†]Graduate School of Commerce and Management, Hitotsubashi University, 2–1 Naka, Kunitachi, Tokyo, 186-8601, Japan. hideshi.itoh@r.hit-u.ac.jp

Abstract

Transacting parties sometimes write contracts that are unenforceable in courts. Why do they write such contracts despite “ink costs”? To answer this question, I analyze contractual relationships in context in the sense that there is a large population of principals and agents, a principal and an agent are randomly matched and engage in transaction, and at the end of each period, they can choose to continue or terminate the current partnership. I adopt an extreme assumption that written contracts are never legally enforced. I then show that writing a contract can help relational contracting between principals and agents more enforceable than relying on tacit understanding of their agreement for three reasons: (i) ink costs of writing a contract make a new match more costly and hence continuing the current match more valuable; (ii) the existence of a written document, with signatures of a principal and an agent, helps parties in the matching pool to identify (some of) those who reneged in the previous transaction; and (iii) the existence of a written document can raise motivation to engage in prosocial behavior (e.g., go to court to punish reneging parties), and hence increasing the probability that the matching pool learns about past deals.

JEL Classification Numbers: D86 (Economics of Contract: Theory), K12 (Contract Law), L14 (Transactional Relationships; Contracts and Reputation; Networks).

Keywords: relational contracting, community enforcement, random matching, legally unenforceable contract, prosocial behavior

There are contracts in societies that have no formal law enforcement machinery...Someone known not to perform his side of bargains will find it difficult to find people willing to make exchanges with him in the future.

(Posner, 2007, p.94)

1 Introduction

Why do trading parties write a contract? The main economic "rationale for contracting is to lock in a commitment *ex ante* that one or both parties would otherwise not wish to honor *ex post*... The use of a contract to establish such commitment is undermined,..., if the contract will not be enforced in the way the parties anticipate (Hermalin et al., 2007, p.99)."

However, as Djankov et al. (2003) argue and convincingly show, economists "have been generally most optimistic about courts as the institution securing property and enforcing contracts (p.454)." Standard economic theories of contracts assume that (i) contractual terms contingent on verifiable states and/or actions are perfectly enforced by courts; (ii) legal enforcement is all-or-nothing; and (iii) verifiability of actions or states are exogenously given. Although research based on these assumptions has contributed to our understanding of optimal contracts, they are also extreme. Judicial enforcement depends on contract law and courts' discretion (they fill gaps, interpret terms, supply default remedies, replace contractual terms with their terms, and so on), the parties' ex post costly action (e.g., submit evidence), and parties' ex ante costly contracting (e.g., costs of thinking of future contingencies and writing documents). Recent (law and) economics literature attempts to relax the standard extreme assumptions and incorporates some of these features into formal analysis.[1] However, they still take substantial (although imperfect) degree of legal enforcement for granted.

More attention has recently been paid to private/informal enforcement mechanisms alternative to courts, such as relational contracting under which contract enforcement is carried out within a bilat-

[1] For example, Ishiguro (2002), Krasa and Villamil (2000), and Bull and Watson (2004) study ex post costly verification, and Dye (1985), Anderlini and Felli (1994, 1999), Battigalli and Maggi (2002, 2008), Schwartz and Watson (2004), and Tirole (2009) concern ex ante costly contracting.

eral relationship, and community enforcement that third parties in the market play disciplinary roles. While these enforcement mechanisms are obviously important in developing/transition economies where legal protections are limited and unreliable (Dixit, 2004; Johnson et al., 2002; McMillan and Woodruff, 1999), the importance and prevalence of informal enforcement are also true even in economies with well-developed legal systems (Djankov et al., 2003; Macaulay, 1963). "The upshot is that private ordering is central to the performance of an economy whatever the conditions of lawfulness (Williamson, 2005, p.2)."

In this paper I also focus on informal enforcement and its interaction with formal contracts. In contrast to existing literature studying such interaction (Baker et al., 1994; Schmidt and Schnitzer, 1995; Pearce and Stacchetti, 1998; Battigalli and Maggi, 2008; Kvaløy and Olsen, 2009; Itoh and Morita, 2015), where a bilateral relationship is isolated from markets or communities, I analyze contractual relationships in context, by adopting the framework of matching games where a large population of buyers and sellers interact in bilateral relationships and they are neither tied permanently to one another, nor are they forced to dissolve their current partnerships at the end of every period.

There is relevant work in game theory that analyzes (mostly) prisoners' dilemma in a community setting (see Mailath and Samuelson (2006, Chapter 5) for an overview). Kandori (1992) and Ellison (1994) establish a folk theorem in a random matching game where the population is finite, each pair must break up exogenously at the end of each period, and each player can only observe the outcomes of the games he played previously.[2] Closer to my analysis are Ghosh and Ray (1996), Kranton (1996), Sobel (2006), Rob and Yang (2010), and Fujiwara-Greve and Okuno-Fujiwara (2009) where there is a continuum of players in the population and matched players can choose whether to continue or terminate the current partnership. However, their focus is on strategies in prisoner's dilemma, and hence no contract or transfer is considered. Exception is Greif (2006) who uses a matching game theoretic framework to solve for an optimal wage with community enforcement. In his model, however, contractual forms are exogenously restricted. I will discuss the differences of my model from his below.

[2] See Deb (2008) and Takahashi (2010) for recent extension.

By developing a "matching game theory meets relational contracting" framework, I examine roles of written contracts that *are not legally enforceable*. While the literature on relational contracting illuminates the logic of informal enforcement, it does not answer the following question: Do the parties need to *write* agreements in documents? It appears that tacit understanding of informal promises is enough and saves "ink costs." An answer to this question, formally analyzed by Kvaløy and Olsen (2009) and Sobel (2006), is that writing more costly contracts is more likely to be legally enforced, and can complement relational contracting under some conditions.

What if costly contracts are not legally enforceable? In this paper I adopt this extreme (and unrealistic) assumption in order to highlight roles of written contracts other than those related to legal enforcement. An influential article by a legal scholar Llewellyn (1931) in fact argues that "official aid on the contract side consists most commonly not in what we know as enforcement but rather in an official declaration—or merely official recognition...—that an obligation is owed and forfeit (p.711)." According to him, the main role of legal contract is to provide an adjustable framework that "almost never accurately indicates real working relations, but which affords a rough indication around which such relations vary, an occasional guide in cases of doubt, and a norm of ultimate appeal when the relations cease in fact to work (p.737)."

Writing a legally unenforceable contract is not entirely unrealistic. A classic paper Macaulay (1963), despite its emphasis on non-contractual business relationships, provides several examples of parties writing legally unenforceable contracts:

> ..., it is likely that businessmen are least concerned about planning their transactions so that they are legally enforceable contracts. For example, in Wisconsin requirements contracts—contracts to supply a firm's requirements of an item rather than a definite quantitb—probably are not legally enforceable. Seven people interviewed reported that their firms regularly used requirements contracts...None thought that the lack of legal sanction made any difference...Three of these people were house counsel who know the Wisconsin law before being interviewed. Another example of a lack of desire for legal sanctions is found in the relationship between automobile manufacturers and their

> suppliers of parts. The manufacturers draft a carefully planned agreement, but one which is so designed that the supplier will have only minimal, if any, legal rights against the manufacturers. The standard contract used by manufacturers of paper to sell to magazine publishers has a pricing clause which is probably sufficiently vague to make the contract legally unenforceable. The house counsel of one of the largest paper producers said that everyone in the industry is aware of this because of a leading New York case concerning the contract, but that no one cares (p.60).

Goldberg (2008) argues that the manufacturing agreement in the 1919 contract between Fisher Body and General Motors was legally unenforceable:

> Nothing precluded Fisher from selling some, or all, of its body production to Ford (p.1076). ... That is, if Fisher is free not to supply auto bodies if it so decides, then GM, despite the specific promises made in the contract, has no obligations; it is free to buy auto bodies from other suppliers (p.1078).

Goldberg (2008) further points out that their counsel should have known when drafting the agreement that it would not be enforceable. He conclude by suggesting that "the unenforceable agreements can be effective...If, as I suspect, such agreements are fairly common, any serious theory of the organization of economic activity will have to take this mechanism into account (p.1082)."

So how can costly written unenforceable contracts be useful in maintaining good relationships? An obvious answer is that writing down terms and obligations enable the parties to remember them (as well as communicate them with relevant members who belong to the same organization, as pointed out by Macaulay (1963)), and to minimize misunderstandings that might jeopardize repeated transactions. I do not pursue these roles of contracts, as well as the possibility that contracts serve as a signaling device (Bénabou and Tirole, 2003; Maskin and Tirole, 1990), and assume instead that there is no problem in communication between the trading parties.

I however show that when transactions are repeated, writing a particular form of contracts, although they are unenforceable, may help the market identify whose reputation must suffer should dissolution occur, and hence may facilitate self-enforcement of relational contracting. Ryall and Sampson (2009), who analyze a sample of joint technology development contracts in telecommunications and microelectronics industries, hint at this communication role of formal contracts:

> In other words, partners who deal with each other repeatedly may find it worthwhile to write a detailed agreement, including performance terms and related penalty clauses, not due to their usefulness in court, but instead, their usefulness in maintaining a smoothly functioning relational contract.
>
> ...
>
> Then, the public failure of such a relationship may lead the market to draw negative inferences about the firm's trustworthiness, thereby inhibiting its ability to conduct future business with others. By permitting firms to mitigate performance shortalls privately via prespecified side payments rather than seeking redress publically via the courts, contracts effectively regulate the conditions under which more substantive reputational costs can be imposed.
>
> ...
>
> This suggests a role for formal contracts that, to the best of our knowledge, has yet to receive theoretical attention: to complement relational enforcement by regulating the conditions under which performance disputes go public *and by publicly identifying the party at fault* (p.923, emphasis as in the original).

The empirical evidence found by Ryall and Sampson (2009) is that contracts become more detailed and more likely to include termination rights and penalty clauses, when at least one of the firms has prior deal experience. This finding is in contrast to a result by Tirole (2007), based on Macaulay (1963), that contracts are less complete in terms of smaller ex ante cognition costs under relational contracting.

The structure of the rest of the paper is as follows. In section 2, a matching game theoretic model of contractual relationships is introduced. The stage game is a simple principal (buyer)-agent (seller) model with symmetric information, where the agent chooses an action from a binary set and the principal offers a price. In contrast to Greif (2006) who assumes that a fixed wage is contractible, I assume no payment is contractible, and in the one-shot transaction the principal "holds up" the agent by expropriating all the benefit, and hence the agent does not choose the costly action. This incentive problem is not at all mitigated in repeated interaction, when there is a continuum of principals and agents and a principal and an agent are matched to play the game, given that no information about a new matched player is available in the matching pool: Each principal always chooses to pay nothing, terminate the current relationship, and return to the matching pool to start a new life. On the other hand, Greif (2006) assumes that with an exogenously given probability, each agent in the pool is not hired in the current period and hence suffers from costly delay.

Following existing literature on matching games, I introduce heterogeneity in the principal's population so as to create a cost of starting a new match:[3] Principals are either of "bad" type who never make positive payments or of "opportunistic" type, and look for equilibria where principals of the opportunistic type compensate their agents for choice of costly action. This creates a cost of starting a new match because the principal must guarantee higher pay to provide the agent, who does not know the principal's type, with incentives to choose an appropriate action. However, in section 3 where I assume no contract is written, I show that there exists no equilibrium in which opportunistic principals make positive payments: The cost of a new match is not large enough to prevent the principal from reneging every period.

In section 4, I assume a contract is written. Although writing a contract is costly, it is not at all enforced in courts. However, I show that writing a costly contract can help relational contracting between principals and agents more enforceable for three reasons. First, writing a contract makes a new match more costly via "ink costs." The role of ink costs is a familiar one, similar to that of "burning money" in literature on matching games.

Second, the existence of a written document, with signatures of a principal and an agent, serves as a communication device, by helping agents in the matching pool to distinguish old principals who did not pay, from new principals who do not have a bad record. I obtain conditions under which there

[3] There is no such heterogeneity in the model of Greif (2006).

exists an equilibrium where opportunistic principals make positive payments.

The written contract can serve as a communication device *only if agents in the matching pool can observe the document*. For example, the Securities and Exchange Commission (SEC) requires that publish firms submit contracts in some categories as part of their filings. However, for those transaction parties who are not subject to such requirements, voluntary disclosure is necessary. The needed disclosure is achieved if agents whose principals did not make the promised payments file lawsuits. However, while going to court is costly, no monetary benefit is expected since the agents win with probability zero.

Now the third reason why writing a contract could help comes in. Suppose that agents are heterogeneous in terms of their intrinsic preferences for prosocial activity, which is to file costly lawsuits in order to punish reneging principals. There is ample evidence showing that people engage in costly punishment.[4] Prosocial behavior may bring direct payoffs as well as reputational benefits from social or self signaling (Bénabou and Tirole, 2006). However, without a written document, intrinsic motivation may not be high enough for them to going to court. The existence of a written document can raise motivation to go to court in order to punish reneging principals, by making such behavior look "legitimate" personally and hence can increase the probability that the matching pool learns from the document.

In section 5, I discuss several extensions and conclude in section 6.

2 Model

Stage game There is a continuum of players called principals (buyers) and agents (sellers), each in the unit interval. In every period, a principal and an agent are matched and play the following simple contract game. The principal places an order for the delivery of a good to the agent, and the agent decides whether or not to accept it. If the agent rejects the offer, the stage game ends and the payoffs are zero for both players. After accepting the offer the agent chooses action $a \in \{0, 1\}$ with personal

[4] See, for example, Fehr and Gächter (2000) and Gintis (2009, Chapter 3) for an overview. According to the latter (p.51), "Recent neuroscientific evidence supports the notion that subjects punish those who are unfair to them simply because this gives them pleasure."

cost $d(a) = da$, where $d > 0$ is a constant representing the cost of action $a = 1$. The value of the good to the principal is $v(a) = va$, where $v > 0$ is a constant. The total surplus is written as $s(a) = sa$ where a constant $s = v - d > 0$ is the total surplus under $a = 1$. The principal, after observing a, chooses a payment $p \in [0, v]$ to the agent. The agent decides whether or not to accept the payment. If he accepts the payment, the payoffs in the period are $va - p$ for the principal and $p - da$ for the agent. If he rejects it, the payoffs are zero for both players.

All the relevant variables are observable but none of them is contractible, and hence there is no initial contract that can bind the principal and the agent ex post.[5] If the relationship is one-shot, it is optimal for the agent to accept any nonnegative price, and hence the principal expropriates all the benefit by offering $p = 0$. The agent, anticipating the principal's response, chooses $a = 0$. The principal could pay some amount in advance, before he chooses a. This is of course wasteful under spot transaction. I relegate most of the analysis of such advanced payments to a later section (subsection 5.3) and focus on ex post payments in the main sections.

Matching process At the beginning of each period, a principal and an agent are randomly matched and play the stage game. At the end of the period, after the outcome of the game realizes, the matched players simultaneously decide whether to continue or terminate the current relationship. If both choose to continue, they move to the next period to play the same stage game. If at least one of them chooses to terminate, they both go back to the matching pool where each of them is matched with a new player. Furthermore, at the end of each period (either before or after the separation decision), with exogenous probability $1 - \rho \in (0, 1)$ both players must leave the population and are replaced by new players who join the pool of unmatched players so as to keep the total population constant.[6] The players discount their payoffs with common discount factor $\delta_0 \in (0, 1)$, and the effective discount factor is denoted by $\delta = \delta_0 \rho$.

Note that there is no possibility that players in the pool are matched with someone they have previously met. I further assume that while players observe (and remember) the history of play in

[5] Note that in contrast to the standard relational contract literature such as Levin (2003), no fixed payment is contractible.

[6] Although the assumption that *both* players must leave the population is adopted to simplify calculation, the results do not essentially change if we instead allow the possibility that one player leaves while the other remains.

their own relationships, each player in the pool obtains no information about the new partner's past history of play with others, including the number of periods that the new partner has been in the population.

In this setting, it is immediate to observe that there is no stationary symmetric equilibrium where the agent chooses $a = 1$ every period.[7] Here stationarity means that each player chooses the same strategy every period. Symmetry means that the equilibrium strategy of all the principals is identical, and the same is true for all the agents. To see this claim, suppose instead that there is a symmetric stationary equilibrium in which every period each principal offers to pay $p \geq d$ if $a = 1$ and zero otherwise, and each agent chooses $a = 1$. It is optimal for the agent to choose $a = 1$ if he expects the principal to pay p. However, given the strategies of agents and other principals, the principal's optimal response is to renege by paying zero instead, and to terminate the relationship. When she goes back to the matching pool, she can make the same offer to a new agent who will choose $a = 1$ following the equilibrium strategy. Her per period payoff hence increases from $v - p$ in equilibrium to v. A contradiction.

The problem is that no one can punish the reneging principal: The matched agent cannot do so since the principal terminates the relationship, and new partners cannot, either, because they are unable to distinguish her from other principals. The principal hence incurs no cost for reneging and returning to the matching pool.[8]

Following Ghosh and Ray (1996), Kranton (1996), and Rob and Yang (2010), I introduce the following heterogeneity in the population as one source of a lower value of starting a new match

[7] Inducing the agent to choose $a = 1$ in later periods is possible in non-stationary equilibria. Suppose that in their first interaction a principal offers zero payment irrespective of action, a matched agent chooses $a = 0$, and both choose to continue the relationship. From the second period on, the principal offers to pay $p = d$ if $a = 1$, the agent chooses $a = 1$, and both continue. Any deviation leads to termination of the relationship. This strategy profile generates costs of starting a new match endogenously. See the analysis of repeated prisoners' dilemma in Mailath and Samuelson (2006, Chapter 5). One problem of this equilibrium is that each *pair* has an incentive to *jointly* deviate by choosing $a = 1$ at the beginning. This unwanted feature motivates Ghosh and Ray (1996) and Kranton (1996) to introduce heterogeneity of players, as I also do in this paper.

[8] Advanced payments do not help, either, because the agent can deviate by choosing $a = 0$ and terminate the relationship. It is impossible to distinguish the agent choosing $a = 1$ from reneging ones and to compensate him for a higher future utility. The principal-agent model in Greif (2006) is different in this respect. In his "individualistic strategy" combination where principals cannot take into account agents' past behavior when making hiring decisions, there is an exogenously given probability that each agent in the pool is not hired in the current period and hence suffers from costly delay. This feature of his model enables an appropriate fixed payment by principals to induce $a = 1$ every period.

than that of continuing the relationship. There are two types of principals in the population:[9] type B ("bad") in proportion $\beta \in [0, 1)$, and type O ("opportunistic") in proportion $\gamma = 1 - \beta$. The bad type behaves mechanically and never makes positive payments, for example, due to a large opportunity cost from paying for the partner, or being myopic with a discount factor close to zero. The focus of the analysis is hence the behavior of opportunistic principals whose payoffs are as specified above. Each principal's type is her private information. On the other hand, I assume all agents are opportunistic.[10]

Relationship-building equilibrium I restrict my attention to pure-strategy, symmetric equilibria with the following features.[11] First, all types of indistinguishable principals make an identical offer because I want to explore roles of contracting different from the well-studied signaling role (Bénabou and Tirole, 2003; Maskin and Tirole, 1990). Since I have not specified the preferences of type B principals, I can exclude "separating" equilibria by appropriately defining their payoff functions. Second, all the agents choose the efficient action $a = 1$ every period (unless they meet principals who are identified as type B).

Third, type O principals choose to pay a promised positive amount contingent on the agent's action choice $a = 1$. This implies that the agent, observing the principal's payment behavior in their first stage game, can know whether she is of type B or type O. Fourth, the equilibrium payment schemes depend only on whether or not the relationship between a principal and an agent is new. In their continuing match, in particular, payment schemes do not depend on how many times they play stage games. When the relationship is new, the agent does not know the principal's type initially, while he finds out her true type at the end of the period and terminates the relationship with type B. The agent does not learn any more information from the second game on in which he knows the principal is of type O. Following Ghosh and Ray (1996), I call a principal and an agent are in *stranger phase* (phase S) when they first interact and hence the agent does not know the principal's

[9]In subsection 5.1, I extend the model by introducing the third type, "good" principals. The main message of the paper is not affected by the existence of such a type.

[10]The analysis can be extended to the case in which there exist agents of bad type (who are myopic or never choose $a = 1$). The principal's reputation is at stake in the current model, and introducing heterogeneity of agents does not change the results substantially. See subsection 5.2 for more on this.

[11]Symmetry here means that the equilibrium strategy of all the principals *of the same type* is identical, and so is that of all the agents.

type, and they are in *friendly phase* (phase F) when they have already interacted and hence he knows the principal is of type O.

Fifth and finally, the equilibrium is in a steady state in the sense that the distribution of the principals' types in each of phases S and F does not change over time. From hereafter I call the equilibrium with all the features given above the "relationship-building" equilibrium.

3 Analysis: When No Contract Is Written

In this section I assume no contract is written, and examine the existence of the relationship-building equilibrium where type O principals make promised payments, and hence in phase S type B principals who do not pay are screened. The agent terminates the relationship if and only if the principal does not pay. If the principal makes payments (so that she is of type O), then they move to phase F (if they survive) by continuing the relationship. Type B principals go back to the matching pool and repeat phase S with probability ρ.

Steady state I first obtain the steady-state distribution of the principals' types under the equilibrium. Figures 1 and 2 summarize the transition of type B and type O principals, respectively. In the equilibrium, type B principals never move to phase F. Suppose proportion $x \in [0, \gamma]$ of type O principals is in phase S, of which ρx moves to phase F and $(1-\rho)x$ exits from the population. Proportion $\rho\beta$ of type B principals stays in and repeats phase S, and $(1-\rho)\beta$ dies. To keep the population constant, $1-\rho$ of newborn principals enter phase S who consist of $(1-\rho)\beta$ type B and $(1-\rho)\gamma$ type O principals. Note that for type B, the rate of inflow and that of outflow are equal to $(1-\rho)\beta$. For type O, the rate of inflow $(1-\rho)\gamma$ must be equal to that of outflow $(1-\rho)x + \rho x$, or $x = (1-\rho)\gamma$.

In phase F, there is proportion $\gamma - x$ of type O principals, of which $\rho(\gamma - x)$ repeats phase F and $(1-\rho)(\gamma - x)$ exits. Note that if $x = (1-\rho)\gamma$, then the rate of outflow $(1-\rho)(\gamma - x)$ is equal to that of inflow ρx. Given this steady-state distribution of type O principals in phase S, denote the probability

Figure 1: Transition of Type B Principals without Written Contracts

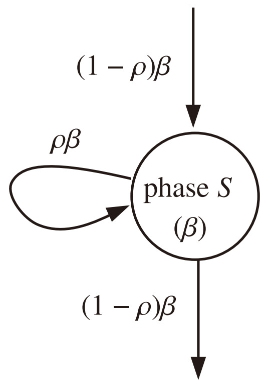
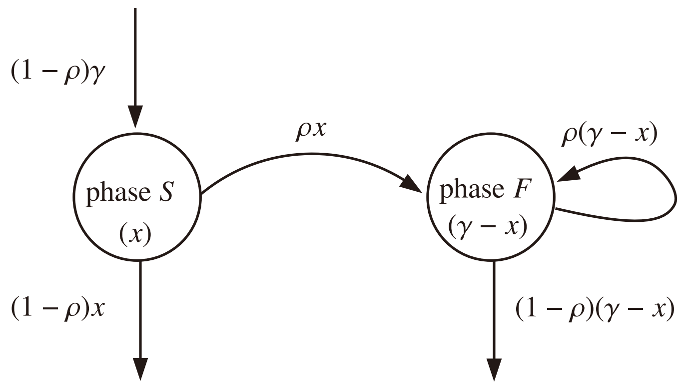


Figure 2: Transition of Type O Principals



of an agent's meeting a type B principal in phase S by

$$\phi = \phi(\beta) = \frac{\beta}{x+\beta} = \frac{\beta}{1-\rho+\rho\beta}. \tag{1}$$

Note that ϕ is increasing in β, $\phi = 0$ if $\beta = 0$, and $\phi \uparrow 1$ as $\beta \uparrow 1$.

Agents Consider the following informal agreements between a principal and an agent: In phase i, the agent chooses $a = 1$, and the principal pays p_i if $a = 1$, and pays nothing if $a = 0$. They terminate their relationship if and only if the principal fails to abide by the payment scheme. If the agent chooses $a = 0$, then the principal does not pay p_i and the relationship continues. In other words, the agent is, without loss of generality, provided with the incentive to choose $a = 1$ via the current payment only.

Let U_S and U_F be the agent's present values in phases S and F, respectively. They are obtained as follows:

$$U_S = (1 - \phi)(p_S - d + \delta U_F) + \phi(-d + \delta U_S)$$
$$U_F = p_F - d + \delta U_F$$

In phase S, each agent meets a type O principal with probability $1 - \phi$, and is paid p_S contingent on $a = 1$. Then they move to phase F. On the other hand, the agent meets a type B principal with probability ϕ, in which case he is paid nothing and hence terminates the relationship and moves back to phase S. In phase F, all the principals are of type O, and hence the agent is paid p_F and repeats phase F.

The agent's incentive compatibility constraints, implying he chooses $a = 1$ in both phases, are given as follows:

$$(1 - \phi)p_S - d \geq 0$$
$$p_F - d \geq 0$$

Since the principal makes a take-it-or-leave-it offer, these constraints bind. The optimal payments are thus $p_S^0 = p_S^0(\phi) = d/(1 - \phi)$ and $p_F^0 = d$, and the equilibrium values are $U_S^* = U_F^* = 0$.

Type O principals I now turn to the principal's payment decisions. Let V_S and V_F be the type O principal's present values in phases S and F, respectively, which are given as follows:

$$V_S = v - p_S^0 + \delta V_F$$
$$V_F = v - p_F^0 + \delta V_F$$

Note that the difference is

$$V_F - V_S = p_S^0 - p_F^0 \geq 0. \tag{2}$$

It is optimal for the principal to pay p_i^0 contingent on $a = 1$ in phase i if the following conditions hold:

$$V_S \geq v + \delta V_S$$
$$V_F \geq v + \delta V_S$$

In either phase, the principal can deviate by paying nothing and terminating the relationship, which yields the same right-hand side of the conditions. Since $V_F \geq V_S$, only the first constraint binds. Using (2) yields the following self-enforcing condition:

$$p_S^0 \leq \delta(V_F - V_S) = \delta(p_S^0 - p_F^0). \tag{3}$$

This condition, along with $p_i^0 > 0$ from the agent's incentive compatibility constraints, leads to the following result.

Proposition 1 The relationship-building equilibrium does not exist.

The proof is obvious from condition (3). It is in fact costly for the type O principal to go back to the matching pool because then she starts at phase S where she has to make a higher payment than in phase F. However, this difference in payment is not high enough to make the contract self-enforcing, even if the discount factor is close to one. The principal prefers enjoying the reneging temptation p_S^0 every period than paying as promised and moving to phase F, that only increases her future payoff by $\delta(p_S^0 - p_F^0)$.

Comparison with the bilateral repeated relationship setting It is instructive to contrast this result to the one in the standard setting where a principal and an agent engaged in repeated transaction infinitely, with zero reservation payoff for both players.[12] Each agent in phase S meets a type B principal with probability ϕ and a type O principal with probability $1-\phi$. The agent meeting type B earns zero payoff forever. The agent meeting type O moves to phase F and stays there forever. The agent's incentive compatibility constraints are the same as above, and hence $p_F^0 = d$ and $p_S^0 = d/(1-\phi)$.

In phase F, the type O principal's reneging temptation is $p_F^0 = d$ (not pay p_F^0) and the future loss is $[\delta/(1-\delta)]s$ (earns zero instead of $s = v - d$ every future period). Hence the principal's promise in phase F is self-enforcing if and only if $d \leq [\delta/(1-\delta)]s$ or

$$\delta \geq \frac{d}{v} \tag{4}$$

holds. Define $r(\delta) \equiv \delta v - d$. Condition (4) is rewritten as $r(\delta) \geq 0$.

Similarly, in phase S, the principal's reneging temptation is $p_S^0 = d/(1-\phi)$, while the future loss is $[\delta/(1-\delta)]s$. The self-enforcing condition is then given by $d/(1-\phi) \leq [\delta/(1-\delta)]s$ or

$$\phi \leq \phi_0(\delta) \equiv \frac{r(\delta)}{\delta s} = \frac{\delta v - d}{\delta(v-d)}. \tag{5}$$

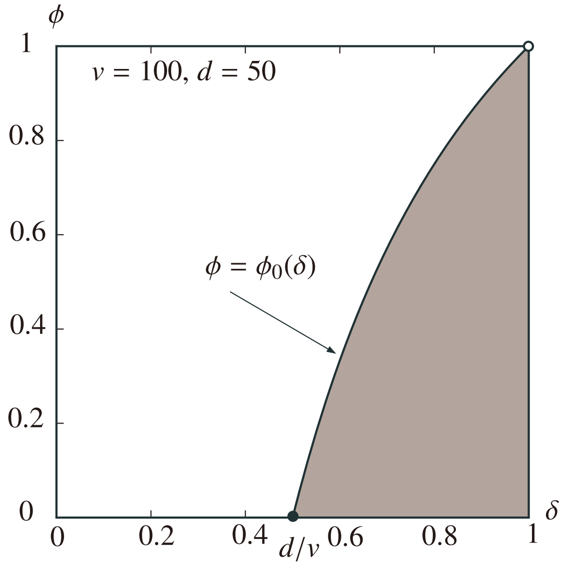
Note that $\phi_0(\delta)$ is increasing in δ and satisfies $\phi_0(d/v) = 0$ and $\lim_{\delta\to 1}\phi_0(\delta) = 1$. Since (5) is sufficient for (4), condition (5) is necessary and sufficient for $a = 1$ to be implemented every period.

In Figure 3, (δ, ϕ) pairs in the dark-filled region satisfy the condition. For a given discount factor, the probability of meeting a type B principal in phase S must be sufficiently low. If it is higher than $\phi_0(\delta)$, the incentive-compatible payment and hence the reneging temptation becomes too high to make it self-enforcing. The higher the discount factor is, the higher the future loss from reneging is, and the more likely the relationship-building equilibrium is to exist.

Denote by V_S^0 and V_F^0 the type O principal's equilibrium present values in phases S and F, re-

[12]See MacLeod and Malcomson (1989) and Levin (2003). Note that in contrast to their models where fixed transfers are contractible, the model here assumes that no contractible transfer is feasible.

Figure 3: Existence of Relationship-Building Equilibria in the Bilateral Repeated Relationship Setting



spectively, which are obtained as follows:

$$V_S^0 = \Pi(\delta) - \frac{\phi}{1-\phi}d \tag{6}$$

$$V_F^0 = \Pi(\delta) \tag{7}$$

where $\Pi(\delta) \equiv s/(1-\delta)$. Since $p_S^0 - p_F^0 = [\phi/(1-\phi)]d$, the present value is smaller in phase S than in phase F exactly because of the higher payment necessary for the agent's incentive to choose $a = 1$ in phase S.

4 Analysis: When Contracts Are Written

When a principal and an agent write a contract, two changes occur. First, writing a contract is costly. I assume that a principal, who writes a contract and makes an offer in the take-it-or-leave-it fashion, must incur "ink cost." The second change is that, as pointed out by Ryall and Sampson (2009) among others, a written contract serves as a communication device. I assume that the existence of written documents, each with signatures of a principal and an agent, enables each agent in the matching pool to identify with some probability whether or not his next partner has been in the population in previous periods, and to infer the partner's type from the content of the contract. For example, publish firms must submit M&A contracts, employment contracts, joint technology development contracts, and other material contracts as part of their filings, under the Securities and Exchange Commission (SEC)'s disclosure requirements. Even for those transaction parties who are not subject to such requirements, a public lawsuit between a principal and an agent can serve a similar role.

I thus assume that a written document in phase $i = S, F$ is revealed to the matching pool with probability λ_i, while with probability $1 - \lambda_i$, the matching pool remains ignorant. Let c_S be the ink cost in phase S, and c_F the ink cost in the *first* game in phase F. From the second period on in phase F, there is no need to modify the contract and hence the ink cost is assumed to be zero.

One interpretation is that when a principal wrote a contract with ink cost c_i but did not pay the specified amount p_i upon the agent's choice of $a = 1$, the agent goes to court and the contract becomes public with probability λ_i. By going to court he incurs the (expected) monetary cost of filing the lawsuit $\ell > 0$ while no monetary benefit is expected whether or not a contract is written, because it is not enforceable anyway and hence no remedy is paid. However, he enjoys private benefit from punishing the reneging principal, which may include direct intrinsic "joy" of engaging altruistic punishment, and/or reputational concern such as a desire to appear prosocial by others or by his later self (see the literature cited in Introduction). Let b be this private benefit from going to court, which is drawn from some probability distribution. Only those agents with $b \geq \ell$ file lawsuits when their principals renege on payments.

I assume that the very existence of a written document can increase the probability of going to court, by making such behavior look "legitimate" and hence increasing b. Suppose b is drawn from a probability distribution with density $\mu_0(b)$ under no contract, while the density is $\mu_i(b)$ if a contract is written in phase i, and the latter dominates the former in the sense of first-order stochastic dominance. Furthermore, suppose for simplicity that no agent goes to court under no contract. Then when a contract is written in phase $i = S, F$, the probability of going to court (and hence revealing the principal's type) λ_i is equal to $\lambda_i = \int_{b>\ell} \mu_i(b)db > 0$.[13]

Before analyzing the effects of these two changes mentioned above, I first examine in subsection 4.1, how the ink cost of writing a contract alone restores the existence of an equilibrium in which all the agents choose $a = 1$. There I assume $\lambda_i = 0$ for $i = S, F$: contracts do not serve as a communication device because no contract is disclosed or the contract disclosed is uninformative. The general case with communication is analyzed in subsection 4.2.

4.1 No Communication

Throughout this subsection I assume $\lambda_i = 0$ for $i = S, F$, and hence the transition of types B and O does not change from Figures 1 and 2, respectively. The only difference from the no contract case is that principals can incur ink costs.

Consider the following informal agreements between a principal and an agent: In phase i, each principal writes a contract which specifies ink cost c_i, payment p_S contingent on $a = 1$, and the agent' choice of $a = 1$. They terminate their relationship if the principal fails to offer the contract or to pay p_i contingent on $a = 1$. Contracts are said to *implement* the relationship-building equilibrium if there exist (c_S, p_S, c_F, p_F) under which a relationship-building equilibrium exists.

The present values and the incentive compatibility constraints of the agent do not change from the previous no contract case: The optimal payments are $p_S^0 = d/(1-\phi)$ and $p_F^0 = d$, and the present values are $U_S^* = U_F^* = 0$.

Let V_S be the type O principal's present value in phase S, and V_F be the present value at the first period of phase F. These are given as follows:

$$V_S = v - p_S^0 - c_S + \delta V_F = \Pi(\delta) - \frac{\phi}{1-\phi}d - c_S - \delta c_F \tag{8}$$

$$V_F = v - p_F^0 - c_F + \delta(V_F + c_F) = \Pi(\delta) - c_F \tag{9}$$

[13]On the other hand, I exclude the possibility that the agent uses the action to file a grievance to hold up the principal (such as "pay me for $a = 0$ or I file"), because such a threat is not credible: I assume that the agent does not enjoy b from going to court to the purpose of extracting rents from the principal.

where $V_F + c_F$ is the principal's present value from the second period on in phase F. Note that the difference is

$$V_F - V_S = p_S^0 - p_F^0 + c_S - (1-\delta)c_F = \frac{\phi}{1-\phi}d + c_S - (1-\delta)c_F \tag{10}$$

The principal makes promised payments if the following incentive compatibility constraints are satisfied:

$$\begin{aligned} V_S &\geq v - c_S + \delta V_S \quad &\Leftrightarrow \quad & p_S^0 \leq \delta(V_F - V_S) \\ V_F &\geq v - c_F + \delta V_S \quad &\Leftrightarrow \quad & p_F^0 \leq \delta(V_F - V_S + c_F) \end{aligned}$$

Since $p_S^0 \geq p_F^0$, the second condition is slack and can be ignored. Then the ink cost c_F in phase F only reduces the present values without relaxing the first constraint because $V_F - V_S$ is decreasing in c_F. Hence $c_F = 0$ must hold: no contract is written in phase F.

On the other hand, writing a contract in phase S relaxes the first incentive compatibility constraint:

$$p_S^0 = \frac{d}{1-\phi} \leq \delta(V_F - V_S) = \delta\left(\frac{\phi}{1-\phi}d + c_S\right). \tag{11}$$

Since this constraint does not hold for $c_S = 0$, writing a contract ($c_S > 0$) in phase S is necessary. Denote c_S satisfying (11) with equality by $\underline{c}^0$:

$$\underline{c}^0 = \underline{c}^0(\phi, \delta) = \frac{1-\delta\phi}{\delta(1-\phi)}d, \tag{12}$$

which is increasing in the probability of meeting type B principals (ϕ) and decreasing in the discount factor (δ).

If the principal deviated by not writing the contract in phase S, then both the principal and the agent would choose to terminate the relationship. Given this, the type O principal would optimally choose not to pay p_S^0, and hence the agent also would choose $a = 0$. The principal's payoff in the current period would be thus zero and the continuation payoff δV_S. The principal does not deviate from writing a contract if $V_S \geq \delta V_S$ or $V_S \geq 0$. The ink cost hence cannot be too high:

$$V_S = \Pi(\delta) - \frac{\phi}{1-\phi} d - c_S \geq 0. \tag{13}$$

Define c_S satisfying (13) with equality by $\overline{c}^0$:

$$\overline{c}^0 = \overline{c}^0(\phi, \delta) = \Pi(\delta) - \frac{\phi}{1-\phi} d, \tag{14}$$

which is decreasing in ϕ and increasing in δ.

Proposition 2 There exists a contract with ink cost $c_S \in [\underline{c}^0, \overline{c}^0]$ in phase S that, along with no contract in phase F, implements the relationship-building equilibrium, if and only if condition (5) $\phi \leq \phi_0(\delta)$ holds.

Proof There exists $c_S \in [\underline{c}^0, \overline{c}^0]$ if and only if $\underline{c}^0 \leq \overline{c}^0$, which is rewritten as

$$\Pi(\delta) = \frac{v-d}{1-\delta} \geq \frac{d}{\delta(1-\phi)}$$

This condition leads to (5):

$$\phi \leq \phi_0(\delta) = \frac{\delta v - d}{\delta(v-d)}. \tag{5}$$

Q.E.D.

Intuitively, as the proportion of type B principals is higher, the incentive compatible payment p_S^0 in phase S must be higher, which fact raises the principal's reneging temptation. Increasing the ink cost mitigates this incentive problem by making termination more costly. If the ink cost is very high ($c_S > v$), then the self-enforcing condition $c_S \geq \underline{c}^0$ is no longer an issue while the non-negative condition $c_S \leq \overline{c}^0$ must be satisfied.

By Proposition 2, the relationship-building equilibrium exists in the same dark-filled region as the bilateral repeated relationships sustain, as Figure 3 shows. Note however that different from the bilateral case, there is a welfare loss due to the ink cost in phase S, which must be positive and sufficiently large to generate costs of starting a new relationship.

As the relationship is more forward-looking (δ is larger), condition (5) is more likely to hold, and the relationship-building equilibrium is sustained for a broader range of ink cost c_S, since both $\underline{c}^0$ and $\overline{c}^0$ are increasing in δ. On the other hand, the more likely agents are to meet type B principals, the more difficult it is to satisfy the existence condition, and the narrower the range of admissible ink costs is.

4.2 Communication

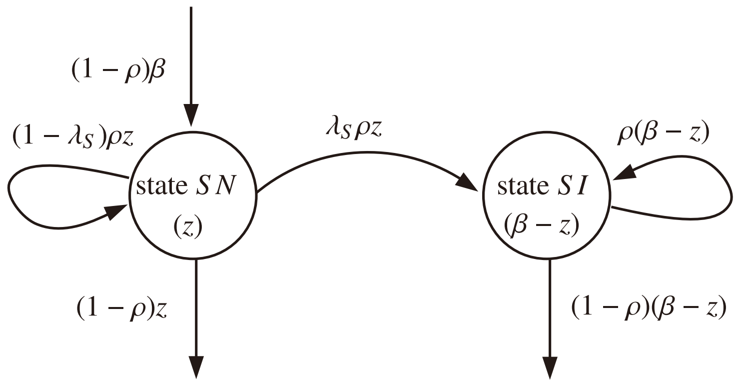
I now return to the general case in which written documents can serve as a communication device, by allowing $\lambda_i \geq 0$ for $i = S, F$. Consider an equilibrium where each type O principal writes an identical contract in phase i that looks like this: "The agent must choose $a = 1$. The principal pays p_i if and only if the agent chooses $a = 1$. If the principal fails to pay p_i contingent on $a = 1$, the relationship is terminated." Suppose that this contract, with signatures of the principal and the agent, is revealed to the matching pool. Suppose the equilibrium contracts are different between phase S and phase F. A new agent matched with a principal whose name is specified in a phase S contract can infer that the old principal did not make a specified payment previously, and hence she be of type B with probability 1.

In phase F, however, all the principals are of type O and hence are supposed to pay as promised in equilibrium. I assume that each agent in the pool who is matched with an old principal whose name is in the phase F contract (which never occurs in equilibrium) believes that she is of type O with probability 1. Then each agent matched to a principal in phase F cannot punish her by going to court, which is then only costly. Hence $\lambda_F = 0$: writing a contract in phase F does not serve as a communication device.

On the other hand, each principal matched with an agent whose name appears in the contract in either phase infers that the old agent is in the matching pool because his previous principal reneged, and hence he can start the new relationship without any disadvantage.

Steady state The steady-state distribution of type O principals is the same as in Figure 2: Proportion $x = (1-\rho)\gamma$ is in Phase S and $\rho\gamma$ is in Phase F. Type B principals are always in Phase F, as in the equilibrium without written contracts. However, now there are two kinds of type B principals, those who are old and not pay in phase S, and those indistinguishable from type O principals, that consist of new type B and old type B whose identities were not revealed in phase S. I say that the former type B principals are in state SI and the latter in state SN. The steady-state distribution of type B principals are summarized in Figure 4.

Figure 4: Transition of Type B Principals under Communication



Suppose proportion $z \in [0, \beta]$ of type B principals is in state SN, of which $(1-\lambda_S)\rho z$ stays in state SN, $\lambda_S\rho z$ moves to state SI, and $(1-\rho)z$ exits from the population. To keep the population constant, $(1-\rho)\beta$ of newborn type B principals enter state SN. From the conditions that the rate of inflow and that of outflow are equal in each of the states, the steady-state distribution of type B principals in state SN is obtained as

$$z = z(\lambda_S) = \frac{1-\rho}{1-\rho+\lambda_S\rho}\beta,$$

which is decreasing in λ_S. As λ_S goes to 1 (all the agents go to court), z approaches to $(1-\rho)\beta$, the proportion of new type B principals. That is, all the old type B principals go to and stays in state SI.

The probability distribution of principals in phase S is summarized in Table 1, where ϕ is the probability that an agent meets a type B principal in phase S, as previously defined in (1). And define

q, the proportion of those identified as type B to all type B principals, by

$$q = q(\lambda_S) = \frac{\lambda_S \rho}{1 - \rho + \lambda_S \rho}. \tag{15}$$

Then $z = (1 - q)\beta$. Note that $q(\lambda_S)$ is increasing in λ_S, with $q(0) = 0$ and $q(1) = \rho$.

Table 1: The Distribution of Principals in Phase S

information	type	probability
state SI	type B	$q\phi$
state SN	type B	$(1 - q)\phi$
	type O	$1 - \phi$

If an agent meets a principal in state SI, she believes that he be of type B who never compensates for $a = 1$. The agent hence chooses $a = 0$ even if he accepts a contract offered by the principal. The identified principal hence has no incentive to pay c_S to write a contract, and the payoffs are zero for both the principal and the agent for that period, and they repeat phase S again with probability ρ. If an agent meets a principal in state SN, she is of either type B or type O, with probabilities as summarized in Table 1.

Consider the following informal agreements between the principal and the agent: In phase S, each type O principal and type B principal in state SN writes a contract $(c_S, \lambda_S, p_S, a = 1)$. In phase F, each principal (who is of type O) writes a contract $(c_F, \lambda_F, p_F, a = 1)$. They terminate their relationship if the principal fails to offer the contract or to abide by the payment rule. When a principal in state SI and an agent meet, they earn zero payoff in the current period and terminate the relationship to go back to the matching pool.

Agents The agent's present value in phase S changes as follows:

$$U_S = (1 - \phi)(p_S - d + \delta U_F) + (1 - q)\phi(-d + \delta U_S) + q\phi\delta U_S$$

In phase S, each agent meets a type O principal with probability $1-\phi$, and is paid p_S upon his action choice $a = 1$. Then they move to phase F. The agent meets a type B indistinguishable principal with probability $(1-q)\phi$, in which case he is paid nothing though he chooses $a = 1$, and terminates the relationship and moves back to phase S. With probability $q\phi$, the agent meets an identified old principal who is of type B. He earns zero payoffs and repeats phase S.

The agent's incentive compatibility constraint in phase S is as follows:

$$(1-\phi)(p_S - d) + (1-q)\phi(-d) \geq 0,$$

which is binding. The optimal payment scheme in phase S is hence

$$p_S^*(\lambda_S, \phi) \equiv \left(1 + \frac{\phi}{1-\phi}(1 - q(\lambda_S))\right) d = \left(1 + \frac{\phi(1-\rho)}{(1-\phi)(1-\rho+\lambda_S\rho)}\right) d.$$

The optimal payment in phase F is, as before, $p_F^* = p_F^0 = d$. The agent's equilibrium values are $U_S^* = U_F^* = 0$. Define $m(\lambda, \phi)$ by

$$m(\lambda_S, \phi) = \frac{\phi(1-\rho)}{(1-\phi)(1-\rho+\lambda_S\rho)}$$

Then $p_S^*(\lambda_S, \phi) = (1 + m(\lambda_S, \phi))d$. Note that $m(\lambda_S, \phi)$ is decreasing in λ_S and increasing in ϕ, with $m(0, \phi) = \phi/(1-\phi)$ and $m(1, \phi) = \phi(1-\rho)/(1-\phi)$. Hence $p_S^*(\lambda_S, \phi)$ is lower than the optimal payment in phase S without written contracts, which is $p_S^0(\phi) = d/(1-\phi) = (1 + m(0, \phi))d$, because the agent need not be motivated to choose $a = 1$ when he meets an identified old principal. If old principals are more likely to be identified (λ_S higher), then the lower payment is needed to induce the agent to choose $a = 1$ in phase S. This payment, however, increases with the proportion of type B principals.

Type O principals The type O principal's present values in phase S and at the first period of phase F are similar to those in the no communication case:

$$V_S = v - p_S^* - c_S + \delta V_F = \Pi(\delta) - m(\lambda_S, \phi)d - c_S - \delta c_F$$
$$V_F = v - p_F^* - c_F + \delta(V_F + c_F) = \Pi(\delta) - c_F$$

Note

$$V_F - V_S = m(\lambda_S, \phi)d + c_S - (1 - \delta)c_F.$$

The type O principal makes promised payments if the following incentive compatibility constraints hold:

$$V_S \geq v - c_S + \delta(1 - \lambda_S)V_S$$
$$V_F \geq v - c_F + \delta V_S$$

To explain the right-hand sides, suppose the principal reneges on payments. In phase S, the agent goes to court with probability λ_S, and then the principal's present value becomes zero from the next period on. Otherwise, she goes back to the matching pool and restarts phase S. In phase F, no contract is revealed and hence the reneging principal starts phase S. These incentive compatibility constraints are rewritten as follows:

$$p_S^* \leq \delta(V_F - V_S + \lambda_S V_S) = \delta(m(\lambda_S, \phi)d + c_S - (1 - \delta)c_F + \lambda_S V_S) \tag{ICS}$$
$$p_F^* \leq \delta(V_F - V_S + c_F) = \delta(m(\lambda_S, \phi)d + c_S + \delta c_F) \tag{ICF}$$

The present values also have to be nonnegative:

$$V_S = \Pi(\delta) - m(\lambda_S, \phi)d - c_S - \delta c_F \geq 0 \tag{NS}$$
$$V_F = \Pi(\delta) - c_F \geq 0 \tag{NF}$$

I examine under what conditions all of these constraints hold and hence a relationship-building equilibrium exists.

Results

I first show that it is without loss of generality to confine attention to no contract in phase F ($c_F = 0$).

Lemma 1 If there is a relationship-building equilibrium with $c_F > 0$ and the principal's equilibrium present values (V_S, V_F), then there is another equilibrium such that $c_F^* = 0$ and the principal's present values (V_S^*, V_F^*) satisfy $V_S^* = V_S$ and $V_F^* > V_F$.

Proof Given a relationship-building equilibrium contract (c_S, λ_S, c_F), define a new contract $(c_S^*, \lambda_S, c_F^*)$ by $c_S^* = c_S + \delta c_F$ and $c_F^* = 0$. Then it is easy to check $V_S^* = V_S \geq 0$ and $V_F^* > V_F \geq 0$ hold. And the incentive compatibility constraints are also shown to hold as follows:

$$p_S^* \leq \delta(V_F - V_S + \lambda_S V_S) < \delta(V_F^* - V_S^* + \lambda_S V_S^*)$$

$$p_F^* \leq \delta(V_F - V_S + c_F) = \delta(V_F^* - V_S^*)$$

Q.E.D.

From now on I focus on phase S contracts (λ_S, c_S), assuming $c_F = 0$. The next proposition presents a necessary and sufficient condition for a relationship-building equilibrium to exist.

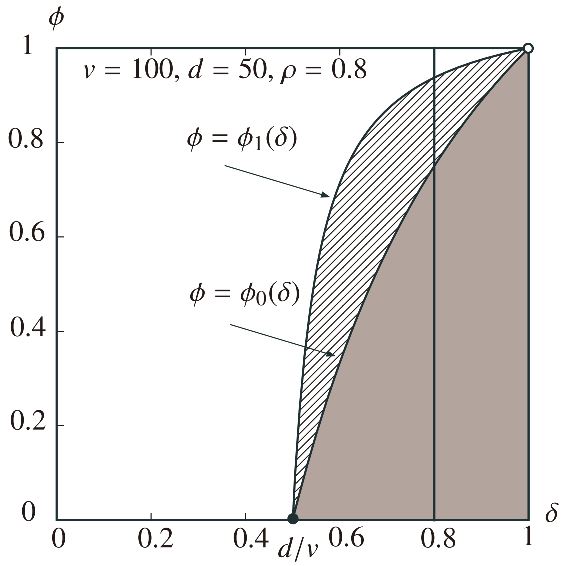
Proposition 3 If

$$\phi \leq \phi_1(\delta) \equiv \frac{r(\delta)}{r(\delta) + (1-\delta)(1-\rho)d} \tag{16}$$

holds, then there exists a contract (λ_S, c_S) in phase S implementing a relationship-building equilibrium. If (16) fails to hold, no written contract can implement the relationship-building equilibrium.

The proposition is formally proved in the appendix, where I show that a relationship-building equilibrium exists if and only if a contract with $(\lambda_S, c_S) = (1, v)$ satisfies all the conditions (ICS),

Figure 5: Existence of Relationship-Building Equilibria with or without Communication



(ICF), (NS), and (NF). Condition (16) is more likely to be satisfied as the parties are more forward-looking (higher ρ and δ_0).

The relationship-building equilibrium exists in both of the shaded region and the dark-filled region in Figure 5. Note that in the figure the curve $\phi = \phi_1(\delta)$ is drawn for $\rho = 0.8$, and hence δ can increase up to 0.8. The equilibrium is more likely to exist as the discount factor is higher and/or the probability of meeting a type B principal is lower. The figure also shows that the region expands from that under bilateral repeated interaction/written contracts without communication (ink cost only). The positive effect of communication is in particular strong in the north-west region where the discount factor is low and the proportion of type B principals is high. This is exactly the situation where the principal's self-enforcing conditions are hard to satisfy.

As I show in the appendix, condition (16) in Proposition 3 guarantees the existence of the most effective contract $(\lambda_S, c_S) = (1, v)$. I next ask what contracts (λ_S, c_S), along with $c_F = 0$, support the relationship-building equilibrium, given the existence condition. Three conditions (ICF), (ICS), and

(NS) are respectively rewritten as follows:

$$
\begin{aligned}
c_S &\geq f(\lambda_S) \equiv \left(\frac{1}{\delta} - m(\lambda_S, \phi)\right) d \\
c_S &\geq g(\lambda_S) \equiv v - \frac{1 - \delta(1 - \lambda_S)}{\delta(1 - \lambda_S)} \left(R(\delta) - m(\lambda_S, \phi) d\right) \\
c_S &\leq h(\lambda_S) \equiv \Pi(\delta) - m(\lambda_S, \phi) d,
\end{aligned}
$$

where $R(\delta) \equiv r(\delta)/(1-\delta) = (\delta v - d)/(1-\delta)$. It is easy to check that $f(\cdot)$ is increasing in λ_S and decreasing in δ and ϕ, $g(\cdot)$ is decreasing in λ_S and δ, and increasing in ϕ, and $h(\cdot)$ is increasing in λ_S and δ, and decreasing in ϕ.

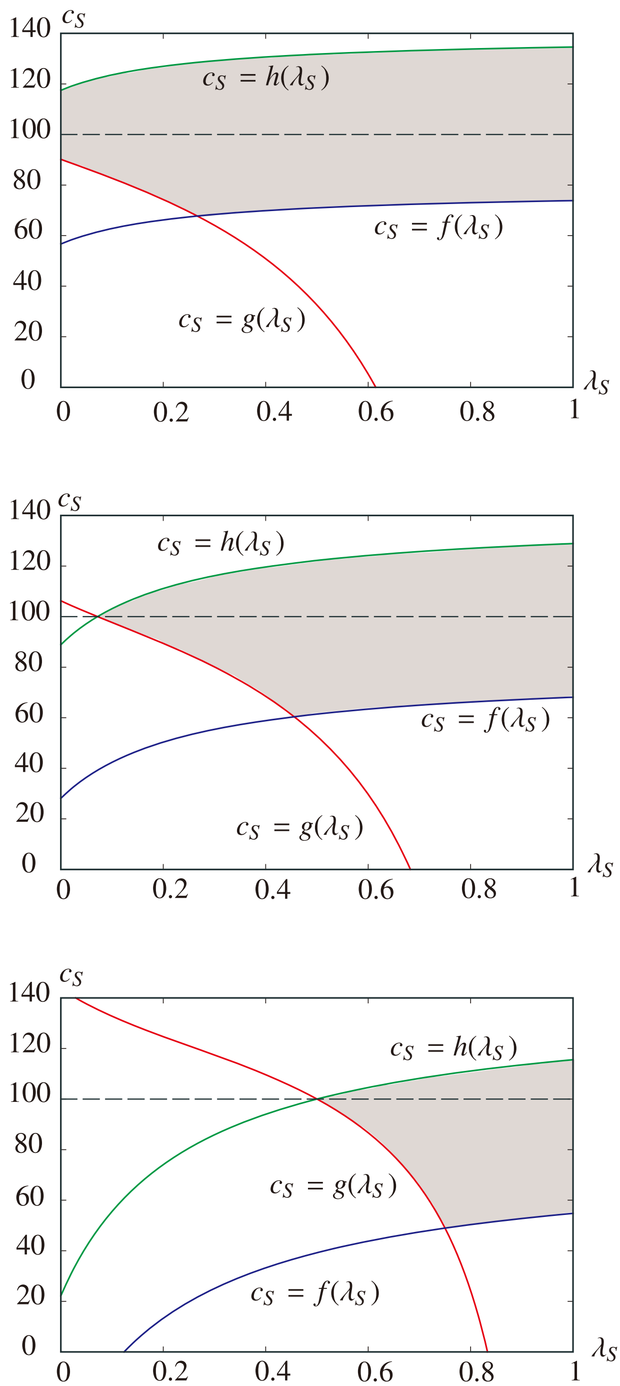
In Figure 6, I draw graphs of these three functions for each of three different values of ϕ. In each case, the equilibrium exists in the dark-filled region. Note that as I show in Proposition 3, the most effective contract $(\lambda_S, c_S) = (1, v)$ is in the dark-filled region for each value of ϕ. Note further that for $\phi = 0.5$ and $\phi = 0.7$, no contract can implement relationship-building equilibria when it does not serve as a communication device ($\lambda_S = 0$).

An interesting observation from the figure is that contracts located in the lower-right area (where ink costs are low and litigation probabilities are high) cannot implement relationship-building equilibria. If a contract whose cost is sufficiently small is very likely to be revealed, then (ICF) fails to hold: The type O principal in phase F has an incentive to deviate because two costs of starting a new relationship, the cost of writing a new contract and the payment difference between phase S and phase F, $m(\lambda_S, \phi)d$, are both small relative to the reneging temptation d.

Given the set of contracts implementing relationship-building equilibria, which one is "optimal" for principals? Let me call a contract optimal if it maximizes the type O principal's present value V_S in phase S subject to constraints (ICF), (ICS), and (NS). Note that by Lemma 1, $c_F = 0$, and constraint (NF) can be ignored. Since V_S depends on the contract only through the terms $m(\lambda_S, \phi)d$

Figure 6: Contracts Implementing Relationship-Building Equilibria

$v = 100,\ d = 50,\ \rho = 0.8,\ \delta = 0.64,\ \phi_0 = 0.4375,\ \phi_1 = 0.7954$



$\phi = 0.3$

$\phi = 0.5$

$\phi = 0.7$

and c_S, the problem can be written as follows:

$$\min_{(\lambda_S, c_S)} m(\lambda_S, \phi)d + c_S$$

subject to (ICF), (ICS), and (NS).

Since the first term of the objective function is decreasing in λ_S and the second term is increasing in c_S, the optimal contract should be located on $c_S = f(\lambda_S)$ in the dark-filled region. Rewriting this equation yields

$$m(\lambda_S, \phi)d + c_S = \frac{d}{\delta}, \tag{17}$$

that is, all the contracts on the segment of $c_S = f(\lambda_S)$ that belongs to the dark-filled region are equally optimal, and at optimum

$$V_S = \Pi(\delta) - \frac{d}{\delta} = \frac{r(\delta)}{\delta(1-\delta)}$$

holds.

Proposition 4 Suppose $\phi \leq \phi_1(\delta)$ holds so that a relationship-building equilibrium exists. Define $C(\lambda_S)$ by

$$C(\lambda_S) = f(\lambda_S) = \left(\frac{1}{\delta} - m(\lambda_S, \phi)\right)d. \tag{18}$$

Then there exists $\lambda^* \in (0, 1]$ such that for all $\lambda \in [\lambda^*, 1]$, $(\lambda_S, C(\lambda_S))$ is optimal.

Proof Define λ^* by

$$f(\lambda_S^*) = g(\lambda_S^*),$$

that is, it is the value of λ_S at the intersection of $c_S = f(\lambda_S)$ and $c_S = g(\lambda_S)$. Using the definitions of $f(\lambda_S)$ and $g(\lambda_S)$ yields

$$\lambda_S^* R(\delta) = m(\lambda_S^*, \phi)d. \tag{19}$$

Obviously $\lambda_S^* > 0$ must hold. And it is easy to check $\lambda_S^* \leq 1$ under condition $\phi \leq \phi_1(\delta)$. The rest follows from the arguments in the main text. **Q.E.D.**

Although I have assumed that c_S and λ_S can be independently chosen,[14] constraint (ICF) restricts principals' preferences for lower c_S and higher λ_S and generates a trade-off endogenously. For example, if λ_S is affected by factors beyond control, such as legal conditions, community size, and so on, then the more transparent the parties' previous deals are, the more detailed and costly contracts must be.

5 Extensions

5.1 "Good" Principals

The formulation of preferences in reputation models typically includes the third, "good" type. Suppose in addition to types B and O, there is a good (G) type in proportion α, where $\alpha + \beta + \gamma = 1$. The type G principals always pay what they promised, because, for example, they enjoy private benefit from making promised payments, they suffer from large negative disutility from reneging, their discount factor is close to one (or larger), and so on.[15]

The analysis of relationship-building equilibria where each type O principal makes a promised payment is not affected by the addition of type G, because type O principals behave exactly like type G principals. However, the existence of type G opens the possibility of another equilibrium under which type O principals do *not* pay in phase S. Only type G principals make promised payments and hence are screened to phase F. Types B and O principals go back to the matching pool and repeat phase S (with probability ρ).

In this "low-friendly-phase" equilibrium, type O principals behave exactly like type B principals,

[14] The analysis can be extended to a case where λ_S is determined by cost c_S via a given function $\lambda_S = \Lambda(c_S)$. I did not take this approach because it is difficult to make specific assumptions on function $\Lambda(\cdot)$: For example, it may not be monotonically increasing since more detailed contracts are sometimes more complex and thus less effective in communicating what happened in previous deals to the third parties. But, for now, suppose $\Lambda(\cdot)$ is increasing and the first-order condition solves for the cost minimizing $m(\Lambda(c_S), \phi)d + c_S$:

$$\frac{\phi\rho(1-\rho)}{(1-\phi)(1-\rho+\Lambda(c_S)\rho)^2}\Lambda'(c_S)d = 1.$$

Denote this cost by c_S^*. If $(\Lambda(c_S^*), c_S^*)$ is in the dark-filled region, then it is the optimal contract. If it is not, then there are two cases. If $\lambda_S = \Lambda(c_S)$ has an intersection with $c_S = f(\lambda_S)$ for $\lambda_S \geq \lambda_S^*$, then that intersection is optimal. Otherwise, the intersection with $c_S = g(\lambda_S)$ is optimal.

[15] However, I exclude the possibility that they make payment offers exceeding benefit v.

and hence I call them together "type NG" who stay in phase S. The steady-state distribution of types can be obtained in a way similar to the distribution under the previous relationship-building equilibrium. For type NG principals, the rate of inflow is $(1-\rho)(1-\alpha)$ which is equal to that of outflow. The remaining $\rho(1-\alpha)$ repeats phase S. For type G principals, the rate of inflow $(1-\rho)\alpha$ is equal to the sum of the exit rate $(1-\rho)y$ and the rate of moving to phase F equal to ρy, where $y \in [0, \alpha]$ is the proportion of type G in phase S. Hence $y = (1-\rho)\alpha$. In phase F, the rate of inflow $\rho y = \rho(1-\rho)\alpha$ is equal to that of outflow $(1-\rho)(\alpha - y) = \rho(1-\rho)\alpha$. Denote the probability of an agent's meeting a type G principal in phase S by

$$\tau = \frac{(1-\rho)\alpha}{(1-\rho)\alpha + 1 - \alpha} = \frac{(1-\rho)\alpha}{1-\rho\alpha}.$$

The agent's present values are then given as follows:

$$\begin{aligned} U_S &= \tau(p_S + \delta U_F) + (1-\tau)\delta U_S - d \\ U_F &= p_F + \delta U_F - d \end{aligned}$$

The incentive compatibility constraints are $\tau p_S \geq d$ and $p_F \geq d$, and hence the optimal payments are $\hat{p}_S = d/\tau$ and $\hat{p}_F = d$, and the agent's equilibrium values are $U_S^* = U_F^* = 0$.

The type O principal's present value in phase S is $V_S = v + \delta V_S = v/(1-\delta)$ since she does not compensate for $a = 1$. This is the maximum attainable value, and hence what positive level p_S is, she has no incentive to pay, and instead chooses not to pay and then terminate the relationship. The only condition to be satisfied for the existence of the low-friendly-phase equilibrium is that the payment in phase S, $\hat{p}_S = d/\tau$, cannot be higher than v, or

$$\frac{(1-\rho)\alpha}{1-\rho\alpha} \geq \frac{d}{v}. \tag{20}$$

Otherwise, even type G principals would not enter transactions.

Proposition 5 There exists $\overline{\alpha} \in (0, 1)$ such that if $\alpha > \overline{\alpha}$, a low-friendly-phase equilibrium exists.

Proof The left-hand side of (20) is increasing in α, goes to 0 as $\alpha \downarrow 0$, and goes to 1 as $\alpha \uparrow 1$. **Q.E.D.**

Note that the low-friendly-phase equilibrium is efficient (all the agents choose $a = 1$). There is hence no efficiency loss if the proportion of type G principals is sufficiently high. And no party has an incentive to write a costly contract in this equilibrium if it exists. My previous analysis thus can be interpreted as the case in which the proportion of type G principals is sufficiently low ($\alpha < \overline{\alpha}$ holds).

5.2 Heterogeneity in the Agents' Population

In the main model I have assume that all the agents are opportunistic. Introducing agents of myopic type with zero discount factor does not change the analysis because such a type can also be induced to choose $a = 1$ by compensation in the same period.

If the definition of type B agents is such that they never choose $a = 0$ (e.g., because of high opportunity costs), the main messages of the paper still go through with some modification. First, phase S becomes more costly for principals and hence an equilibrium in which type O principals make promised payments can exist even without written contracts. Let ψ be the stationary probability that a principal meets a type B agent in phase S, and consider the relationship-building equilibrium in which type O principals promise to pay p_S contingent on $a = 1$, and type O agents choose $a = 1$. Since type B agents choose $a = 0$ and type B principals choose not to pay p_S for $a = 1$, they repeat in phase S, and only type O players move to phase F.

Under no contract, the optimal payment scheme does not change: $p^0_S = d/(1 - \phi)$ and $p^0_F = d$. Type O principal's present values are given as follows:

$$V_S = (1 - \psi)(v - p^0_S + \delta V_F) + \psi\delta V_S$$
$$V_F = v - p^0_F + \delta V_F$$

She meets a type O agent with probability $1 - \psi$, in which case the agent chooses $a = 1$ and she pays p_S, and they move to phase F. With probability ψ she meets a type B agent who chooses $a = 0$ and hence she does not need to pay p_S, and then they go back to phase S. Her incentive compatibility constraint in phase S is given by

$$V_S \geq (1-\psi)(v+\delta V_F)+\psi\delta V_S$$

which is rewritten as

$$p_S^0 \leq \delta(V_F - V_S) = \frac{\delta\psi}{1-\delta\psi}s + \frac{\delta(1-\psi)}{1-\delta\psi}(p_S^0 - p_F^0).$$

This constraint never holds for sufficiently small ψ, in which case there is no relationship-building equilibrium without writing a contract.

Second, even type O principals with contracts may return to phase S if they meet type B agents. Then while the information value of written contracts is reduced, it does not disappear because type B principals *always* return to phase S.

5.3 Contractible Fixed Payments

Throughout the paper I have assumed that no payment is contractible, in contrast to literature in relational contracting such as Levin (2003) in which fixed transfers are contractible. The payment schedule in my analysis corresponds to a "bonus" contract where the principal compensates the agent by a bonus pay in the current period.

Suppose instead that fixed payments are contractible, or equivalently, principals can make advanced payments before their partner agents choose action. While principals need to give some rent in order to induce the agents to choose $a = 1$, the best response of the agents is to choose $a = 0$ and terminate their relationships, if no contract is written.

Similar to the previous analysis, writing a "termination" contract in which an agent is paid a fixed amount along with the termination clause after shirking can restore his incentive to choose $a = 1$, if it is the agent who must incur ink costs. However, there is no communication role in writing such a contract, because there is heterogeneity only in the principals' population, and hence it is the principals' reputation that is at stake. The parties should write a bonus contract if they want to facilitate informal enforcement via communication. What kind of contract to write thus matters.

This conclusion is reversed if only the agents' population is heterogeneous. In this case, the agents' reputation is at stake, and hence the parties should write a termination contract in which the principal has no chance to deviate while the agent's incentive to choose $a = 1$ is provided with future rent.

If both players' population is heterogeneous, what should be written in a contract is an issue. Writing either a bonus contract or a termination contract only serves as a device of communicating which side is at stake, while writing a contract where both sides obtain future rents reduces the information value of the contract. While this problem is to be explored in future research, it suggests that there is a reason for the parties to write a contract with unilateral penalties where only one party may exercise the penalty. Ryall and Sampson (2009) in fact find in their sample that of the 21 contracts with penalty clauses, 8 have bilateral penalties while 13 have unilateral penalties. And in the contracts with these unilateral penalty clauses, the party penalized is usually the party with fewer prior deals, or the one whose type is more uncertain in our formulation.

6 Concluding Remarks

To make a communication role of writing a contract suggested by empirical literature more precise, I have developed a simple but novel framework where there is a large population of principals and agents and they are matched and decide whether or not to continue their partnerships. And I have then shown that writing a costly contract, although it is unenforceable, contributes to its informal enforcement and facilitates relationships.

The current paper is just a start and does not offer an explanation of particular empirical evidence nor testable implications. Offering them is an obvious next step, and to this purpose it is important to extend the current model to heterogeneity in both principals' and agents' population. Another possible extension is to moral hazard, in which the agent's action is unobservable to the matched principal while her benefit is observable but stochastic.

References

Anderlini, L. and Felli, L. (1994), “Incomplete Written Contracts: Undescribable States of Nature,” *Quarterly Journal of Economics.* 109:1085–24.(URL).

Anderlini, L. and Felli, L. (1999), “Incomplete Contracts and Complexity Costs,” *Theory and Decision.* 46:23–50.(abstract).

Baker, G., Gibbons, R., and Murphy, K. J. (1994), “Subjective Performance Measures in Optimal Incentive Contracts,” *Quarterly Journal of Economics.* 109(4):1125–1156.

Battigalli, P. and Maggi, G. (2002), “Rigidity, Discretion, and the Costs of Writing Contracts,” *American Economic Review.* 92:798–817.(URL).

Battigalli, P. and Maggi, G. (2008), “Costly Contracting in a Long-Term Relationship,” *Rand Journal of Economics.* 39:352–377.

Bénabou, R. and Tirole, J. (2003), “Intrinsic and Extrinsic Motivation,” *Review of Economic Studies.* 70:489–520.

Bénabou, R. and Tirole, J. (2006), “Incentives and Prosocial Behavior,” *American Economic Review.* 96:1652–1678.

Bull, J. and Watson, J. (2004), “Evidence Disclosure and Verifiability,” *Journal of Economic Theory.* 118:1–31.

Deb, J. (2008), “Cooperation and Community Responsibility: A Folk Theorem for Repeated Random Matching Games,” mimeo.

Dixit, A. K. (2004), *Lawlessness and Economics: Alternative Modes of Governance.* Princeton NJ: Princeton University Press.

Djankov, S., La Porta, R., Lopez-De-Silanes, F., and Shleifer, A. (2003), “Courts,” *Quarterly Journal of Economics.* 118(3):453–517.

Dye, R. A. (1985), “Costly Contract Contingencies,” *International Economic Review.* 26:233–250.(URL).

Ellison, G. (1994), “Cooperation in the Prisoner’s Dilemma with Anonymous Random Matching,” *Review of Economic Studies.* 61:567–588.

Fehr, E. and Gächter, S. (2000), “Fairness and Retaliation: The Economics of Reciprocity,” *Journal of Economic Perspectives.* 14:159–181.

Fujiwara-Greve, T. and Okuno-Fujiwara, M. (2009), “Voluntarily Separable Repeated Prisoner’s Dilemma,” *Review of Economic Studies.* 76:993–1021.

Ghosh, P. and Ray, D. (1996), “Cooperation in Community Interaction Without Information Flows,” *Review of Economic Studies.* 63:491–519.

Gintis, H. (2009), *The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences.* Princeton University Press.

Goldberg, V. P. (2008), “Lawyers Asleep at the Wheel? The GM-Fisher Body Contract,” *Industrial and Corporate Change.* 17:1071–1084.

Greif, A. (2006), *Institutions and the Path to the Modern Economy: Lessons from Medieval Trade.* New York, NY: Cambridge University Press.

Hermalin, B. E., Katz, A. W., and Craswell, R. (2007), “Contract Law,” Vol. 1 . Amsterdam: North-Holland. .

Ishiguro, S. (2002), “Endogenous Verifiability and Optimality in Agency,” *Journal of Economic Theory.* 105:518–530.

Itoh, H. and Morita, H. (2015), “Formal Contracts, Relational Contracts, and the Threat-Point Effect,” *American Economic Journal: Microeconomics.* 7(3):318–346.

Johnson, S., McMillan, J., and Woodruff, C. (2002), "Courts and Relational Contracts," *Journal of Law, Economics, and Organization.* 18:221–277.

Kandori, M. (1992), "Social Norms and Community Enforcement," *Review of Economic Studies.* 59: 63–80.

Kranton, R. E. (1996), "The Formation of Cooperative Relationships," *Journal of Law, Economics, and Organization.* 12:214–233.

Krasa, S. and Villamil, A. P. (2000), "Optimal Contracts When Enforcement is a Decision Variable," *Econometrica.* 68:119–134.

Kvaløy, O. and Olsen, T. E. (2009), "Endogenous Verifiability and Relational Contracting," *American Economic Review.* 99(5):2193–2208.

Levin, J. (2003), "Relational Incentive Contracts," *American Economic Review.* 93(3):835–847.

Llewellyn, K. N. (1931), "What Price Contract?—An Essay in Perspective," *Yale Law Journal.* 40: 704–751.

Macaulay, S. (1963), "Non-Contractual Relations in Business: A Preliminary Study," *American Sociological Review.* 28(1):55–67.

MacLeod, W. B. and Malcomson, J. M. (1989), "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment," *Econometrica.* 57:447–480.

Mailath, G. J. and Samuelson, L. (2006), *Repeated Games and Reputations: Long-run Relationships.* Oxford University Press.

Maskin, E. and Tirole, J. (1990), "The Principal-Agent Relationship with an Informed Principal: The Case of Private Values," *Econometrica.* 58:379–409.

McMillan, J. and Woodruff, C. (1999), "Dispute Prevention Without Courts in Vietnam," *Journal of Law, Economics, and Organization.* 15(3):637–658.

Pearce, D. G. and Stacchetti, E. (1998), "The Interaction of Implicit and Explicit Contracts in Repeated Agency," *Games and Economic Behavior.* 23:75–96.

Posner, R. A. (2007), *Economic Analysis of Law.* New York: Aspen Publishers 7th. edition.

Rob, R. and Yang, H. (2010), "Long-Term Relationships as Safeguards," *Economic Theory.* 43:143–166.

Ryall, M. D. and Sampson, R. C. (2009), "Formal Contracts in the Presence of Relational Enforcement Mechanisms: Evidence from Technology Development Projects," *Management Science.* 55: 906–925.

Schmidt, K. M. and Schnitzer, M. (1995), "The Interaction of Explicit and Implicit Contracts," *Economics Letters.* 48:193–199.

Schwartz, A. and Watson, J. (2004), "The Law and Economics of Costly Contracting," *Journal of Law, Economics, and Organization.* 20:2–31.

Sobel, J. (2006), "For Better or Forever: Formal versus Informal Enforcement," *Journal of Labor Economics.* 24:271–297.

Takahashi, S. (2010), "Community Enforcement When Players Observe Partners' Past Play," *Journal of Economic Theory.* 145:42–62.

Tirole, J. (2007), "Bounded Rationality and Incomplete Contracts," mimeo.

Tirole, J. (2009), "Cognition and Incomplete Contracts," *American Economic Review.* 99:265–294.

Williamson, O. E. (2005), "The Economics of Governance," *American Economic Review Papers and Proceedings.* 95:1–18.

Appendix

In this appendix I prove Proposition 3 through the following three lemmas. By Lemma A1, I can confine attention to sufficiently high discount factors satisfying $r(\delta) = \delta v - d \geq 0$. In Lemma A2, I show that condition (16) on (δ, ϕ) in Proposition 3 is necessary and sufficient for the existence of a contract satisfying both (ICS) and (NS). In Lemma A3, I then show that a contract satisfying condition (16) satisfies (ICF) as well.

Lemma A1 There is no relationship-building equilibrium if $r(\delta) < 0$.

Proof (ICF) and (NS) are rewritten as follows:

$$c_S \geq \frac{d}{\delta} - m(\lambda_S, \phi)d$$
$$c_S \leq \Pi(\delta) - m(\lambda_S, \phi)d$$

And hence

$$\Pi(\delta) - \frac{d}{\delta} \geq 0$$

must hold for a relationship-building equilibrium to exist, which condition is equivalent to $r(\delta) \geq 0$. **Q.E.D.**

Hereafter assume $\delta \geq d/v$.

Lemma A2 Suppose no contract is written in phase F. There exists a contract (λ_S, c_S) in phase S satisfying both (ICS) and (NS) if and only if

$$\phi \leq \phi_1(\delta) \equiv \frac{r(\delta)}{r(\delta) + (1-\delta)(1-\rho)d} \tag{A1}$$

Proof (ICS) and (NS) are, respectively, calculated as follows:

$$m(\lambda_S, \phi)d \leq R(\delta) - \frac{\delta(1-\lambda_S)}{1-\delta(1-\lambda_S)}(v - c_S) \tag{A2}$$
$$m(\lambda_S, \phi)d \leq \Pi(\delta) - c_S \tag{A3}$$

where $R(\delta) \equiv r(\delta)/(1-\delta) = \Pi(\delta) - v$. It is easy to find that the right-hand sides are equal if $c_S = v$, and (A2) binds if $c_S < v$, while (A3) binds otherwise. Substituting $c_S = v$ yields

$$m(\lambda_S, \phi)d \leq R(\delta). \tag{A4}$$

Since the right-hand side of (A2) is increasing in c_S and that of (A3) is decreasing in c_S, condition (A4) is necessary and sufficient for a contract satisfying both (ICS) and (NS) to exist. Since the left-hand side is minimized at $\lambda_S = 1$, substituting $\lambda_S = 1$ into (A4) yields

$$\frac{\phi}{1-\phi}(1-\rho)d \leq \frac{r(\delta)}{(1-\delta)},$$

or

$$\phi \leq \phi_1(\delta) = \frac{r(\delta)}{r(\delta) + (1-\delta)(1-\rho)d}$$

which is condition (16) in Proposition 3. **Q.E.D.**

Lemma A3 Contracts with $c_S = y$ satisfy (ICF).

Proof Summarizing (ICF) yields

$$m(\lambda_S, \phi)d \geq \frac{d}{\delta} - c_S. \tag{A5}$$

When $c_S = v$, the right-hand side becomes

$$\frac{d}{\delta} - v = -\frac{r(\delta)}{\delta}$$

which is nonpositive by $r(\delta) \geq 0$. Condition (A5) hence holds for all contracts with $c_S = v$. **Q.E.D.**