Making and Breaking Promises When Their Costs Are Private Information*

Jin Li and Niko Matouschek

Abstract

We discuss how to maintain trust when promise-makers are privately informed about the costs of keeping their promises and efficient transfers are not feasible. To this end, we present a simplified version of the model in Li and Matouschek (2013) in which a principal and an agent are in an infinitely repeated relationship. The agent's effort and output are observable but not contractible and the principal is privately informed about the cost of paying the agent. We characterize the optimal relational contract, illustrate the methods used in solving games with one-sided asymmetric information and inefficient transfers, and discuss further applications.

Keywords: relational contracts, imperfect monitoring

JEL classifications: C73, D23, D82, J33

*Jin Li (corresponding author), The University of Hong Kong; Niko Matouschek: Northwestern University. We thank Hanzhe Li and Zhaoneng Yuan for excellent research assistance.

1 Introduction

Trust oils the wheels of organizations. But trust is no panacea for all organizational ills. It may allow superiors to improve on the formal contracts they write with their employees. But rarely does it allow them to complete those contracts by filling in all the open terms. The reason trust cannot solve all organizational problems is that it is based on self-interest, not blind faith. Employees do not trust their superiors to keep any promise. They only trust them to keep those promises that are even more costly to break than to keep.

Since the costs of breaking a promise typically lie in the future, while the benefits are immediate, the time preferences of superiors are a key determinant of how effective trust can be. The promise of a large performance bonus, for instance, does not inspire much trust if it is made by an impatient superior who discounts the future heavily. Her employees know that when it is time to pay the bonus, she will value the immediate benefit of saving the expense over the longer term cost of shattered trust. The classic papers on relational contracts explore precisely how time preferences impact trust and its ability to cure organizational ills (Bull (1987), MacLeod and Malcomson (1989), Baker, Gibbons, and Murphy (1994), and the first part of Levin (2003)).

The time preferences of superiors, though, are only one constraint on the effectiveness of trust. Another is their private information. Efficiency often requires promises to made contingent on factors that only superiors are informed about. In an efficiently designed incentive scheme, for instance, bonus payments may depend on conditions employees cannot observe themselves. Employees will then have a natural tendency to distrust their superiors whenever they claim that a bonus cannot be paid because those hard-to-observe conditions were just not met. This does not preclude superiors from making promises contingent on privately observed information. But it requires them to do so in a thoughtful manner that balances motivating employees with the need to keep themselves honest, and transparently so. This paper provides a brief introduction to the type of dynamics that arise when this balance is struck optimally.

Our focus is on situations in which capital markets are imperfect, in which case bonus payments need not be simple transfers and can impose efficiency costs that vary with the firms' market conditions. Bonus payments may, for instance, require firms to borrow the necessary funds at high rates from their banks, as Lincoln Electric did when it experienced large losses abroad, or preclude them from building capital and engaging in other investments, as executives at Credit Suisse First Boston (CSFB) explained when their bankers demanded higher bonuses (see Hastings (1999) on

Lincoln Electric and Stewart (1993) on CSFB). What makes such fluctuating opportunity costs problematic is that employees typically cannot observe them perfectly, if at all. No matter how adamant superiors are that the bonus pool is needed to cover losses in foreign markets, or build capital at home, employees may suspect that their superiors' true motive is not efficiency but greed. Such suspicions undermine trust and can lead to costly conflict, as they did at CSFB, and which Lincoln Electric was eager to prevent by borrowing funds to pay what had long been understood to be a *cash-sharing* bonus.

We explored such conflicts, and what can be done to keep their costs as small as possible, in Li and Matouschek (2013). A risk neutral principal is in an infinitely repeated relationship with a risk neutral agent whose funds are limited. The agent's effort and performance are not contractible but are observed by both parties. At the beginning of every period, the principal offers the agent a wage, to which she can commit contractually, and promises a bonus if the agent provides the desired effort. The key feature of the model is that the principal is privately informed about the cost of paying the agent. At the end of every period, she learns whether the per-dollar cost is one, or a fixed value greater than one, while the agent stays in the dark about the opportunity costs of transfer. Motivation then requires the agent to trust the principal to pay the promised bonus. But efficiency requires the promise to be contingent on the principal's privately observed opportunity costs of paying the bonus.

To keep the principal honest, the agent must punish her whenever he provides the desired effort and she fails to pay the promised bonus. In our setting, the optimal way to punish the principal is not to terminate the relationship, now or in the future. Instead, if the principal fails to pay a promised bonus this period, it is optimal for the agent to work a little less in the next. If the principal also fails to pay the bonus in the next period, the agent reduces his effort again in the following period, and so on. This continues until effort reaches a lower bound, at which point the agent punishes continued failures to pay the bonus by demanding higher and higher wage payments that have to be paid even if opportunity costs are high. The principal's optimal punishment is, therefore, gradual and evolves from the informal understanding that the agent is allowed to slack off to the formal requirement of higher, unconditional wage payments. This punishment continues until the principal's opportunity costs return to their low state, at which point she pays a large, once-off bonus to "settle up" and the relationship is re-set to where it was before the conflict. An optimally managed relationship, therefore, cycles indefinitely between peace and conflict, with each conflict getting gradually worse the longer it lasts, before ending abruptly and the relationship returning to its prior state.

In the next section, we explore a simplified version of this model in which effort is binary. Moreover, we assume that the principal's opportunity costs of paying the agent one dollar are either one or infinite, which precludes wage payments and focuses attention on bonuses. This simple model allows us to derive some of the key features of optimal relational contracts we just sketched, and to illustrate the, by now common, way in which such models can be solved more generally. It also serves as a springboard for a discussion of extensions and applications, which we provide in Section 5, before concluding in Section 6.

2 Model

A risk-neutral principal and a risk-neutral agent are in an infinitely repeated relationship. Time is discrete and denoted by $t \in \{1, 2, ...\infty\}$. We first describe the stage game and then move on to the repeated game.

The Stage Game At the beginning of each period t, the principal decides whether to offer a contract to the agent and the agent decides whether to accept the offer. Let $d_t^P \in \{0, 1\}$ denote the principal's decision, where $d_t^P = 0$ if she makes no offer and $d_t^P = 1$ otherwise. The contract consists solely of a noncontractible commitment to pay the agent a bonus $b_t \geq 0$. The opportunity cost of paying the bonus fluctuates and depends on the state of the world. Let $\theta_t \in \{L, H\}$ denote the state, where L and H stand for "low opportunity cost" and "high opportunity cost." In state L, the opportunity cost is one, so that it costs the principal one dollar to give the agent one dollar. In state H, instead, the opportunity cost is infinite and the principal simply cannot make any payment to the agent. The states are independently drawn from a binary distribution. In any period, the probability that the opportunity costs are high is given by $p \in (0, 1)$. The principal privately observes the state after output is realized.

If a contract is offered, the agent decides whether to accept it. Let $d_t^A \in \{0, 1\}$ denote the agent's decision, where $d_t^A = 0$ if he rejects the offer and $d_t^A = 1$ otherwise. If $d_t^P = 0$ or $d_t^A = 0$, both parties receive their outside options, which we normalize to zero. Otherwise, if a contract is offered and accepted, the agent chooses effort $e_t = \{0, 1\}$ at cost $c(e_t) = ce_t$, where $c > 0$. Output $Y(e_t)$ is given by $Y(1) = y$ if the agent works and $Y(0) = 0$ if he shirks. Effort is efficient, so that $y - c > 0$. Both effort and output are publicly observable but not contractible. In any period t, the principal can then expect to get

$$\hat{\pi}_t = \mathrm{E}[Y(e_t) - 1_{\{\theta_t = L\}} b_t]$$

and the agent can expect to get

$$\hat{u}_t = \mathrm{E}[1_{\{\theta_t = L\}} b_t] - c\,(e_t)\,.$$

Finally, after both parties have realized their payoffs, they observe the realization x of a public randomization device, after which time moves on to the next period.

The Repeated Game The principal and the agent share a common discount factor $\delta \in (0, 1)$. At the beginning of any period t, the per-period average of the principal's payoff is given by

$$\pi_t = (1-\delta) \sum_{\tau=t}^{\infty} \delta^{\tau-t} d_\tau^P d_\tau^A \hat{\pi}_\tau$$

and the agent's is given by

$$u_t = (1-\delta) \sum_{\tau=t}^{\infty} \delta^{\tau-t} d_\tau^P d_\tau^A \hat{u}_\tau.$$

In any period, the principal's and the agent's actions can depend on past histories, and may do so in complicated ways. It is, however, without loss to assume that strategies are public, that is, that they depend only on publicly available information and do not depend on the principal's past private information. Restricting attention to public strategies is without loss because the players' private information has no effect on their future payoffs (conditional on the publicly observable history; see, for instance, p. 330 in Mailath and Samuelson (2006) for a detailed discussion). When private information does affect future payoffs, the players can gain by using non-public strategies, and the recursive technique we discuss below cannot be applied directly (see, e.g., Fuchs (2007), Fong and Li (2017), Fahn and Klein (2019), and Li, Mukherjee, and Vasconcelos (2023)).

We follow the literature (see, e.g., Levin (2003)) and define a relational contract as a pure-strategy Perfect Public Equilibrium (henceforth PPE) in which the principal and the agent play public strategies and, following every history, strategies are a Nash Equilibrium of the continuation game. We define an optimal relational contract as a PPE that maximizes the principal's average per-period payoff (see, e.g., Levin, (2003)). In the following sections, we solve for the optimal relational contract. As will be clear below, our analysis characterizes, for each given payoff of the principal, the maximal equilibrium payoff of the agent. Our analysis, therefore, traces out all Pareto-efficient equilibria.

3 Preliminaries

Solving for equilibrium strategies in dynamic games, especially those with private information, may appear to be a daunting task. Strategies must describe actions after each history, the size of which grows exponentially over time. This makes it difficult to describe strategies and to check that they form an equilibrium. This task, however, is vastly simplified by Abreu, Pearce, and Stacchetti (1990), which shows that, instead of solving for the equilibrium strategies directly, it is sufficient to study the equilibrium payoffs sets, that is, the payoffs induced by the equilibrium strategies. Every equilibrium strategy, of course, generates equilibrium payoffs. The insight of Abreu, Pearce, and Stacchetti (1990) is that one can also use the set of equilibrium payoffs to generate equilibrium strategies.

Loosely speaking, Abreu, Pearce, and Stacchetti (1990) shows that, for a given equilibrium payoff, an equilibrium strategy can be constructed period-by-period using a recursive approach similar to that in dynamic programming. In particular, an action is an equilibrium action that *supports* the equilibrium payoff if, for each player, three conditions are satisfied (for each player). First, there is an continuation payoff—the discounted value of all future payoffs—so that the sum of the flow payoff from the action and the continuation payoff is equal to the equilibrium payoff. This is often referred to as the *"promise-keeping"* condition, which reflects the idea that current and future actions deliver the promised equilibrium payoff. Second, neither player wants to deviate from his or her action. This is referred to as the *"no-deviation"* constraint, which, in our context, includes the agent's incentive constraint to work and the principal's truth-telling constraint to be honest about the opportunity costs. Finally, the continuation payoff for each player again belongs to the equilibrium payoff set. This is referred to as the *"self-enforcing"* constraint because, once the continuation payoff is again an equilibrium payoff, we can repeat the procedure and look for the actions in the period.

Abreu, Pearce, and Stacchetti (1990) show that this procedure can be applied to games in which actions of the players are their private information and the outcomes are stochastic. In this case, the proper equilibrium concept is PPE as discussed above, so the equilibrium payoff set is referred to as the PPE payoff set. To make the analysis more tractable, it is typically assumed that players are allowed to use randomized actions and that there is a public randomization device that can be used to coordinate equilibrium actions. These assumptions ensure that the PPE payoff set is convex, implying that there is a single connected boundary that completely determines the PPE payoff set. We will also make these assumptions here.

We now apply the procedure of Abreu, Pearce, and Stacchetti (1990) to our model. Denote the PPE payoff set by $\mathcal{E}$ and define the PPE payoff frontier as

$$u(\pi) \equiv \max \left\{ u' : \left(\pi', u'\right) \in \mathcal{E} \right\}.$$

Let $\underline{\pi} = \min\{\pi : (\pi, u) \in \mathcal{E}\}$ be the principal's smallest PPE payoff and $\overline{\pi} = \max\{\pi : (\pi, u) \in \mathcal{E}\}$

her largest. In the appendix, we show that $u(\pi)$ is well defined, which follows because the actions of the players are essentially bounded and that the payoffs of the players are continuous in the actions. These two properties imply that the PPE payoff set is compact and has a well-defined frontier. Also note that $u(\pi)$ is a concave function because $\mathcal{E}$ is convex.

The most important property of $u(\pi)$ is that it is self-enforcing, that is, for any equilibrium payoff pair $(\pi, u(\pi))$, the continuation payoffs are also on the frontier. This property is important because it implies that the *information of the payoff frontier alone* is enough to trace out the entire optimal equilibrium strategy. In particular, suppose we know, for each payoff pair on the frontier, which actions support it and the values of the continuation payoffs. In this case, for each $(\pi, u(\pi))$, we know which actions will be chosen and the associated continuation payoffs. This property then implies that the associated continuation payoffs are again on the frontier, and we can again use our information to determine the actions to be chosen and the associated continuation payoffs. Proceeding accordingly, we can trace out the entire equilibrium action sequences as long as the continuation payoffs remain on the frontier. The self-enforcing property of the payoff frontier arises here because our game features one-sided private information: specifically, any deviation by the agent is publicly observable. As long as the agent does not deviate, there is no reason to punish him, as doing so does not relax any constraints, and the continuation payoff will remain on the payoff frontier.

We now use the self-enforcing property to characterize the payoff frontier. To do so, we need to determine, for each payoff pair on the frontier $(\pi, u(\pi))$, which action (or randomization of actions) is used to support it and what are the associated continuation payoffs. There are three types of actions to consider. First, the principal and the agent take their outside options (O). Second, both parties enter the relationship and the agent works (W). Finally, they enter the relationship, and the agent shirks (S). Define $u^j(\pi)$, for $j \in \{O, W, S\}$, as the agent's highest equilibrium payoff that is supported by action j when the principal's payoff is π, and define $u^j(\pi) = -\infty$ if there is no equilibrium that gives the principal a payoff of π. This then implies that

$$u(\pi) = \max_{\alpha_j \geq 0, \pi_j \in [\underline{\pi}, \overline{\pi}]} \sum_{j \in \{O,W,S\}} \alpha_j u^j(\pi_j)$$

such that

$$\sum_{j \in \{O,W,S\}} \alpha_j = 1$$

and

$$\sum_{j \in \{O,W,S\}} \alpha_j \pi_j = \pi.$$

The payoff frontier can, therefore, be characterized by answering two questions. First, for each equilibrium payoff of the principal π, which action, or the randomization of actions, supports the payoff frontier $(\pi, u(\pi))$? Second, if action $j \in \{O, W, S\}$ supports the payoff frontier, what are the associated continuation payoffs? Below, we first study the second question, and then move to the first.

LEMMA 1. *For any frontier payoff pair* $(\pi, u(\pi))$, *the following holds:*

(*i.*) *If the payoff pair is supported by the outside options, the principal's continuation payoff is given by* $\pi^O = \pi/\delta$.

(*ii.*) *If the payoff pair is supported by working, the principal's continuation payoffs in states* L *and* H *are given by*

$$\pi_L^W = \bar{\pi} \text{ and } \pi_H^W = \frac{\pi - (1-\delta)y}{\delta}$$

and the principal pays a bonus $b^W = y - (\pi - \delta\bar{\pi})/(1-\delta)$.

(*iii.*) *If the payoff pair is supported by shirking, the principal's continuation payoffs in states* L *and* H *are given by*

$$\pi_L^S = \bar{\pi} \text{ and } \pi_H^S = \frac{\pi}{\delta}$$

and the principal pays a bonus $b^S = -(\pi - \delta\bar{\pi})/(1-\delta)$.

The lemma describes the principal's continuation payoffs for all three types of actions. Part (i.) describes her continuation payoff when the players take their outside options. In this case, the continuation payoff is determined directly by the principal's promise-keeping constraint, that is, $\pi = \delta\pi^O$. This promise-keeping constraint arises because the principal's normalized payoff (π) is the weighted sum of her flow payoff and her continuation payoff (π^O). When the players take their outside options, the principal's flow payoff is zero.

Part (ii.) describes the principal's continuation payoffs when the agent works. To see how these continuation payoffs are determined, first note that when the agent works, the promise-keeping constraint for the principal is given by

$$\pi = (1-\delta)\left(y - (1-p)b^W\right) + \delta(p\pi_H^W + (1-p)\pi_L^W), \tag{1}$$

where π_H^W and π_L^W are the principal's continuation payoffs associated with states H and L. The promise-keeping condition takes the form above because, in state H (which occurs with probability p), the principal pays the agent zero and receives a continuation payoff of π_H^W, and in state L (which occurs with probability $1-p$), the principal pays the agent b^W and receives a continuation payoff of π_L^W.

Note that the state of the world is the principal's private information. When the state is L, the principal may be tempted to report that the state is H to avoid having to pay the bonus. To prevent the principal from doing so, the truth-telling constraint

$$-(1-\delta)b^W + \delta\pi_L^W \geq \delta\pi_H^W \tag{2}$$

must be satisfied. This constraint states that, when opportunity costs are low, the principal must prefer paying the bonus (and receiving a higher continuation payoff) to not paying the bonus (and receiving the continuation associated with state H).

Given these two constraints, we can pin down the continuation payoffs using the following two observations. First, the principal's truth-telling constraint is binding, that is, $\delta\left(\pi_L^W - \pi_H^W\right) = (1-\delta)b^W$. This observation follows from the concavity of the payoff frontier. If, to the contrary, $\delta\left(\pi_L^W - \pi_H^W\right) > (1-\delta)b^W$ one can find an alternative set of continuation payoffs on the frontier that satisfy all the relevant constraints with the property that the associated continuation payoffs of the principal $(\pi_L'^W, \pi_H'^W)$ are closer to each other $(\pi_L'^W - \pi_H'^W < \pi_L^W - \pi_H^W)$. The concavity of the payoff frontier will then imply that the expected continuation payoffs of the agent are higher under the alternative set of payoffs. This also implies that the agent's expected payoff is higher under the alternative continuation payoffs (since his current flow payoff is unchanged), violating the assumption that the agent's expected payoff is on the payoff frontier, i.e., it is at the highest level (given the principal's payoff). Therefore, when the payoff pair is on the frontier, the principal's truth-telling constraint is binding. Note that when the principal's truth-telling constraint is binding, her promise-keeping constraint immediately implies that

$$\pi = (1-\delta)\left(y - b^W\right) + \delta\pi_L^W = (1-\delta)\,y + \delta\pi_H^W.$$

This pins down π_H^W.

Second, the principal's continuation payoff should be made as large as possible. We use this observation to determine π_L^W. This observation follows because the principal's continuation payoff measures her stake in the relationship, and, therefore, reflects how much the agent can trust her. The relationship is more valuable when the agent trusts the principal more since this allows them to shelter more periods with high opportunity costs. This implies that the principal's continuation payoff π_L^W will go to $\bar{\pi}$, her highest feasible equilibrium payoff. The highest continuation payoff is obtained through paying the agent a corresponding level of bonus. In summary, after a state of L, the principal will pay the agent whatever bonus is necessary to settle past histories. The relationship essentially restarts in the next period, with the agent's trust of the principal staying at the highest level possible.

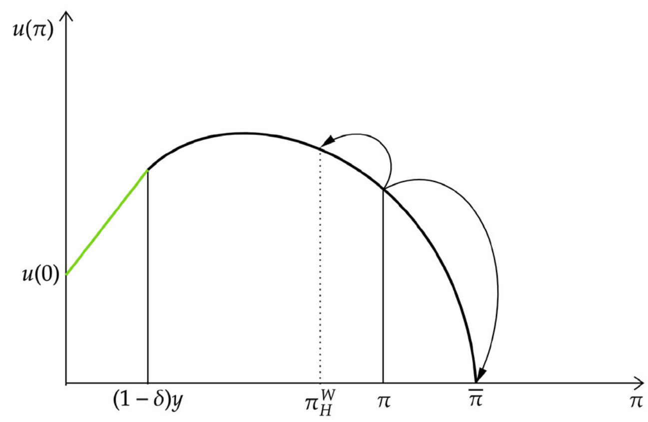

Figure 1: The PPE frontier.

Finally, Part (iii.) of the lemma describes the principal's continuation payoffs when the agent shirks. In this case, the continuation payoffs are again pinned down by the same two observations in Part (ii): that the principal's truth-telling constraint is binding and that her continuation payoff following a state of L is at the highest possible value.

Given the description of the continuation payoffs, Lemma 1 then implies that

$$\begin{aligned} u^O(\pi) &= \delta u\left(\frac{\pi}{\delta}\right), \\ u^W(\pi) &= (1-\delta)\left((1-p)y-c\right)-(1-p)(\pi-\delta\overline{\pi})+\delta\left(pu\left(\frac{\pi-(1-\delta)y}{\delta}\right)+(1-p)\,u(\overline{\pi})\right), \text{ and} \\ u^S(\pi) &= -(1-p)(\pi-\delta\overline{\pi})+\delta\left(pu\left(\frac{\pi}{\delta}\right)+(1-p)\,u(\overline{\pi})\right). \end{aligned}$$

Notice that we substituted out b^W and b^S using the observations above.

4 The Optimal Relational Contract

We now use these high equilibrium payoffs from different actions to characterize the payoff frontier and describe the optimal relational contract.

PROPOSITION. *There exists a cut-off level $\pi_W = (1-\delta)y$ such that the PPE frontier $u(\pi)$ can be divided into three regions:*

(i.) For $\pi = 0$, $u(\pi)$ is supported by shirking.

(ii.) For $\pi \in (0, \pi_W)$, $u(\pi)$ can be supported by randomization between working and shirking.

(iii.) For $\pi \in [\pi_W, \overline{\pi}]$, $u(\pi)$ is supported by working.

The PPE payoff frontier is illustrated in Figure **??**. The proposition shows that outside options are never used to support the payoff frontier. In addition, when the principal's equilibrium payoff π is zero, shirking is used to support the frontier; when π is above π_W, working is used; and when π is in the intermediate range, randomization is used so that working occurs with a certain probability. This pattern shows that the agent's effort decreases with his trust in the principal, that is, the principal's payoff π. The players, however, will not take their outside options even if the principal's payoff drops to zero.

To see why the proposition holds, consider first why the parties never choose their outside options. In particular, notice that when the principal's payoff drops to $\pi = 0$ and the outside option is used, the principal's continuation payoff will remain at zero. This implies that the players will take their outside options again next period, so that the relationship de facto terminates, giving both the principal and the agent a payoff of zero. However, if the players do not take their outside options but simply have the agent shirk, then there is some chance that the opportunity costs are low this period. In this case, the principal can pay a bonus to the agent, allowing the relationship to restart in the next period. This gives the agent a higher equilibrium payoff even if the principal's payoff is zero. In other words, $u^S(0) > u^O(0)$ so that, along the payoff frontier, the players will not take their outside options when the principal's payoff is zero, and a similar intuition can be used to show that outside options are never used to support the payoff frontier.

The result that the outside options are not used on the payoff frontier depends on a number of assumptions of our model. In particular, we assume that the principal is not liquidity-constrained so that she can always pay the bonus in the state L to bring her continuation payoff back to $\bar{\pi}$ next period. If she were liquidity-constrained, the outside options can be used to support the payoff frontier, resulting in the termination of the relationship (see Li and Matouschek (2013)). Another assumption is that, when the agent shirks, the joint payoff of the players is weakly higher than when they take their outside options. If the joint payoff of the players were lower than their outside options when the agent shirks, then it is possible that the players are better off taking their outside options (see Fong and Li (2017)).

Given that the outside options are not used on the payoff frontier, we now show why shirking is used when $\pi = 0$ (and with some probability when $\pi < \pi_W$). The reason for this lies at the heart of the logic for managing dynamic relationships with private information. Because the principal has private information about the state of the world, the agent does not know whether the principal is telling the truth when she reports that the opportunity cost of payment is high and she cannot pay the bonus. To make sure that the principal is telling the truth, the agent must punish the principal at some point through shirking (since, otherwise, the principal will never pay the bonus). Shirking, however, destroys the surplus of the relationship, and the best way to manage the relationship is to make it happen as infrequently as possible. The agent will, therefore, delay the punishment of the principal to the largest extent possible, so that shirking is happening only when it must. In particular, note that $\pi_W = (1 - \delta)y$ is the smallest payoff of the principal that allows the agent to work with probability one. For $\pi < \pi_W$, having the agent work will necessarily violate the promise-keeping condition of the principal (because she has the option of receiving the output and paying zero to the agent, and therefore pocketing π_W). The proposition, therefore, shows that shirking happens with positive probability only when $\pi < \pi_W$, and it happens with probability one when the agent's trust in the principal is completely lost ($\pi = 0$).

Given the proposition, we can now describe the dynamics of the optimal relational contract.

COROLLARY. *The dynamics of the optimal relational contract involve a working phase (in which the agent works) and a shirking phase (in which he shirks). The relationship starts in the working phase, where the principal's payoff is given by* $\bar{\pi}$.

(*i.*) *Working phase: if the state is* H*, the principal's continuation payoff is given by* $\pi_H^W(\pi) < \pi$*. But if the state is* L*, her continuation payoff jumps to* $\bar{\pi}$*. The relationship stays in the working phase as long as* $\pi \geq \pi_W$*. If* $\pi < \pi_W$*, the relationship transitions to the shirking phase with positive probability.*

(*ii.*) *Shirking phase: if the state is* H*, the principal's continuation payoff goes to zero. But if the state is* L*, her continuation payoff jumps to* $\bar{\pi}$*, and the relationship returns to the working phase.*

The corollary shows that when the relationship is managed optimally, it cycles between two phases indefinitely. The cycle shows features akin to those in Li and Matouschek (2013): the relationship declines gradually when the principal does not pay the bonus, and it recovers instantaneously when she does so. Each time the principal fails to pay the bonus, her continuation payoff drops. When the continuation payoff drops below π_W, the agent shirks with positive probability. The lower is the principal's continuation payoff, the less likely the agent works. And when the principal's payoff goes to zero, the agent does not work and shirks with probability one. In any stage of the relationship, however, the relationship restarts (at payoff $(\bar{\pi}, u(\bar{\pi}))$) in the next period whenever the principal pays the bonus.

The reason why declines are gradual and recoveries instantaneous is that joint surplus is increasing in expected profits. The higher expected profits are, the less tempted the principal is to renege on her promises, and, thus, the more willing the agent is to work hard. As a result, it is efficient to reduce expected profits only gradually during periods of high opportunity costs and to let them recover to their upper bound as quickly as possible when opportunity costs are low again. This, of course, requires the principal to have the liquidity to pay a sufficiently large bonus after, potentially many, periods of high opportunity costs. In Li and Matouschek (2013) we consider an extension in which the principal may not be able to do so because the bonus she is able to pay in any period is limited by the output generated in that period. We show that such liquidity constraints can slow down recoveries by forcing the principal to settle up over multiple periods.

5 Discussion

The simple model we just explored, and the more general model in Li and Matouschek (2013), make a number of predictions that are, at least in principle, testable: (i.) superiors' private information about the opportunity costs of paying employees impacts employees' effort choices and, thus, their productivity, (ii.) in the absence of liquidity constraints, effort and productivity cycle and the cycles are characterized by gradual declines and instantaneous recoveries, and (iii.) liquidity constraints slow down recoveries but not declines. Related models share some, but not all of these predictions, which makes it possible to distinguish between them empirically. Englmaier and Segal (2013), for instance, explore a model and mechanism that is similar to ours but focus on Green-Porter type equilibria in which the employment relationship alternates between high-effort cooperation phases and low effort punishment phases. As such, their model generates prediction (i.) but not predictions (ii.) and (iii.). In contrast, Halac and Prat (2016) explore a model in which the superior has to invest in a monitoring technology in order to recognize and reward the employee's efforts. Their model predicts effort and productivity dynamics that are similar to (ii.). The mechanism, though is very different from those described above and, as such, the model does not generate predictions (i.) and (iii.).

One feature of the simple model we just explored is that superiors do not have many tools at their disposal to alleviate their employees' distrust. All they control are the bonus payments which, incidentally, is what is sowing distrust in the first place. The model shows that when handled just right, though, this tool is enough to maintain trust, at least periodically. The key is for superiors to condition their current payments, not just on current effort and opportunity costs, but also on the employees' past disappointments, and to compensate them for those disappointments, even if they responded to them by slacking off.

In practice, there are other tools in the superiors' toolkit. One is establishing the right kind of transparency. The notion that transparency facilitates the organization of agency relationships has a long history in economics. Holmström (1979), in particular, observed that in moral hazard problems the principal is better off if she knows more about the agent (see Prat (2005) and Fong and Li (2017) for exceptions). In line with this "*Informativeness Principle*," firms employ myriad means to gather information about the actions of their employees and the circumstances they operate in. Our model highlights the value of the opposite kind of transparency that allows employees to learn more information about the circumstances their employers face. If bonus payments depend on the principal's private information, the principal is better off if the agent knows more about her.

This reasoning is consistent with the experience at Lincoln Electric. As we mentioned earlier, the firm responded to the vocal discontent of its employees by borrowing the money needed to pay the "cash-sharing" bonus they expected. Lincoln's CEO Donald Hastings, though, did not stop there. Instead, he also tried to prevent future conflicts by improving the employees' understanding of the firm's financial situation and instituting "*...a financial education program so that employees would understand that no money was being hidden from them...*" (Hastings 1999, p.172). The notion that superiors can benefit when agents know more about them is also consistent with the practice of co-determination in Germany that assigns board seats to worker representatives. Apart from giving these representatives a voice and vote in corporate decision making, it provides them with better information about the firm's circumstances (see, for instance, the interviews with worker representatives in Addison, Teixeira, and Zwick (2010); see also Jäger, Schoefer, and Heining (2020)). Such information sharing can help convince employees that unpopular corporate decisions are based on efficiency not greed, which reduces the costs of conflicts and allows firms to adapt to adverse economic shocks more effectively (for a discussion and exploration of these issues see Englmaier and Segal (2013)).

Another tool in the superiors' toolkit is corporate cash management. A simplifying assumption we made above is that when opportunity costs are low, the superior does not face any constraints in how much she can pay the employee. In Li and Matouschek (2013), we explore an extension in which the amount the principal can pay in any given period is constrained by the output the agent generates in that period. Even in that extension, though, we do not allow the principal to save cash in good times to maintain trust during bad ones. As a practical matter this is clearly not right. Firms do routinely keep significant cash reserves, both to reduce their tax bill and as a form of precautionary savings (see, for instance, Faulkender, Hankins, and Petersen (2019)). The model suggests that in addition to the standard motive for precautionary corporate savings—enabling firms to invest without delay—such savings can serve as a human resources tool that allows firms to manage trust more effectively. To our knowledge, this trust-based motive for precautionary corporate savings has not yet been explored formally (see Barron, Li, and Zator (2021) for an exploration of the related notion that debt can harm employee morale).

Another popular management practice that requires promises to be made contingent on superiors' private information is subjective performance evaluation. Such evaluations are based on the personal impressions, opinions, and experiences of superiors that are not only difficult to specify in formal contracts but cannot even be observed by the employees. The evaluation of an employee's leadership potential, for instance, is inherently subjective. Subjective performance evaluations are helpful when superiors cannot observe their employees' effort choices, and cannot infer them from other, objective performance measures. Superiors will then want to promise employees a higher bonus if their subjective evaluations are better. But they will also need to keep themselves honest by accepting that poor evaluations impose costs, not only on the employees, but also on themselves.

One way to impose costs on superiors is to tie poor evaluations to higher probabilities that the employment relationship is terminated, and the superiors' rents are destroyed. Levin (2003) shows that a simple form of such termination contracts is optimal among all contracts in which superiors give full performance evaluations after every period, what he terms contracts with a "*full review property*." His work, and the contemporaneous work by MacLeod (2003), has spurred a literature on subjective performance evaluation, of which Fuchs (2007) is another prominent example. A key feature of the settings these papers explore is that asymmetric information is two-sided, with superiors being privately informed about their subjective evaluations and employees about their effort choices. This feature has so far precluded the literature from fully characterizing optimal general contracts.

Superiors, of course, do not only make promises about bonus payments. In their efforts to attract, retain, and motivate employees they also often promise other, non-monetary benefits, such as favorable work schedules, attractive assignments, recognition, and the like. The costs of keeping such promises often vary over time in ways that are difficult for employees to observe, just as those of bonus payments do. By itself this does not have to undermine trust. If superiors can make transfers to employees, and the costs of those transfers are constant, they can always compensate employees for not delivering on a promised non-monetary benefit by paying them its value in cash. In practice, however, superiors are often not authorized to make such transfers. A department chair cannot pay a faculty member cash for serving on a committee that she had promised to keep him from. Even superiors who do have the authority to make such transfers—the Dean—often find it optimal to commit themselves, and the members of their organizations, not to do so (see, for instance, Prendergast and Stole (1999)). The combination of superiors' private information about the cost of keeping non-monetary promises, and their inability to make transfers to their employees, then gives rise to similar challenges as the ones we explored above. How to best overcome these challenges, though, depends on the characteristics of the non-monetary benefit, which may differ in important ways from those of bonus payments.

Take, for instance, the promise to "empower" employees by delegating decision rights to them, which is a common management practice. Legally the rights to make decisions always reside at the top of corporate hierarchies and cannot be relegated to lower ranks (see, for instance, Bolton and Dewatripont (2013)). As such, the delegation of decision rights is always informal. It is based on superiors' promises to rubber-stamp reasonable recommendations and not on any contractual commitments to do so. Over the last few years, a number of papers have explored such informal, trust-based delegation schemes (see, for instance, Guo and Hörner (2015), Li, Matouschek, and Powell (2017), and Lipnowksi and Ramos (2020)). These papers follow the broader literature on delegation that goes back to Holmström (1977) in assuming that transfers are not feasible and by focusing on settings in which private information resides with the employees (for an overview of the literature see, for instance, Gibbons, Matouschek, and Roberts (2013)). Employees, for instance, often know more about the availability and payoffs of different projects because they are closer to customers and suppliers. Superiors, however, also often have private information about corporate decision making. They may, for instance, know more about the impact of the employees' recommended actions on other parts of the business or its implications for future investments that have not yet been made public. Such information may lead a superior to override an employee's recommendation and keep the employee in the dark about her true motive. To keep the superior honest, her veto has to come at a cost to her, either in the form of the agent slacking off, as in our model, or, possibly, in the form of more discretion for the agent. In contrast to our model, though, it cannot take the form of large "settling-up" payments at the end of the conflict, since such transfers are never feasible. How to optimally punish the superior, and thus best structure relational delegation in such a setting, is an open question.

More generally, trust matters when complete contracts are not feasible or at least costly to write. This is naturally the case within firms and organizations, which arise precisely when markets fail. The absence of complete contracts, though, is also a feature of many real-world markets, which work well enough to not be replaced by firms but fall short of the Arrow-Debreu ideal. To the extent that market participants interact repeatedly, trust can improve the functioning of such markets, just as it improves the functioning of firms and organizations (see, for instance, Macchiavello and Morjaria (2015, 2021) and Gil, Kim, and Zanarone (2022)). Even between firms, though, there can be frictions in transfers, which undermine trust when promise-makers have private information about the costs of keeping their promises. Take, for instance, a liquidity-constrained entrepreneur and her investor who are in a long term relationship that involves multiple rounds of investments. To structure their relationship, the parties could write a debt-like contract in which the entrepreneur commits to a specific repayment scheme that gives her limited flexibility to adjust payments to changes in the business environment. Alternatively, they could write an equity-like contract in which the entrepreneur promises to pay out free cash-flow as it becomes available but makes limited formal commitments. This latter arrangement is more efficient in principle, but requires the investor to trust the entrepreneur to be honest about the firm's finances. The interplay of formal and informal commitments to pay investors is related to that of wage and bonus payments in Li and Matouschek (2013). How it shapes optimal financial contracts in a market setting, though, remains to be seen.

Finally, beyond organizations and markets, promises are central for the everyday activity of asking for, and granting, favors. By definition, people do not pay for favors. They ask others to do something good for them because they cannot afford to pay them to do so, or because cultural and legal norms keep them from doing so. At the same time, most favors are not free. Their price is the, often unspoken, promise to return it in the future. People, in other words, trade favors, as in Möbius (2001) and Hauser and Hopenhayn (2008). A feature of these models is that they assume that people are privately informed about the opportunity to do someone a favor and then explore simple mechanisms to trade favors. A natural, alternative specification is one in which what is private information is not whether I can do you a favor. Rather, it is my cost of doing you a favor we are both aware of. I may then want to refuse to return a favor because the cost of the one you are asking for today is just too high. At the same time, I worry that you would interpret such a refusal as a betrayal of trust, which would undermine our ability to trade favors in the future. Favor trading then has to overcome the same kind of challenges as in our simple model, but in a richer setting with two-sided asymmetric information and the absence of any transfers. Characterizing optimal favor trading in such a setting is difficult. Progress, though, can be made by focusing on natural forms of favor trading, as in the papers we mentioned above, and as in Abdulkadiroglu and Bagwell (2013), who explore an even richer setting in which agents are privately informed about their ability to both, do favors and return them.

6 Conclusions

Trust is delicate and needs to be managed carefully. In this paper we discussed how to do so when promise-makers are privately informed about the costs of keeping their promises and doubts cannot be dispelled through efficient transfers. The combination of such private information and the absence of efficient transfers does not preclude parties from establishing trust, but it does complicate it enough to necessitate an efficiency cost. The central question then is how the parties should act to keep this cost as small as possible. The goal of this paper was to illustrate the answer to this question in the context of a simple model of repeated moral hazard. Different versions of the same problem arise in many economically important contexts, some of which we sketched above and hope will be explored in future research.

References

[1] Abdulkadiroğlu, Atila, and Kyle Bagwell (2013), "Trust, Reciprocity, and Favors in Cooperative Relationships," *American Economic Journal: Microeconomics*, 5(2), 213–59.

[2] Abreu, Dilip, David Pearce, and Ennio Stacchetti (1990), "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring," *Econometrica*, 58(5), 1041–1063.

[3] Addison, John, Paulino Teixeira, and Thomas Zwick (2010), "German Works Councils and the Anatomy of Wages," *Industrial and Labor Relations Review Review*, 63(2), 247–270.

[4] Baker, George, Robert Gibbons, and Kevin Murphy (1994), "Subjective Performance Measures in Optimal Incentive Contract," *Quarterly Journal of Economics*, 109(4), 1125–1156.

[5] Barron, Daniel, Jin Li, and Michał Zator (2021), "Morale and Debt Dynamics," *Management Science*, 68(6), 4496–4516.

[6] Bolton, Patrick and Mathias Dewatripont (2013), "Authority in Organizations," in: Robert Gibbons and John Roberts (eds.), *Handbook of Organizational Economics*, Princeton University Press, Princeton (NJ), 343–372.

[7] Bull, Clive (1987), "*The Existence of Self-Enforcing Implicit Contracts*," *Quarterly Journal of Economics*, 102(1), 147–159.

[8] Englmaier, Florian and Carmit Segal (2013), "Unions, Communication, and Cooperation in Organizations," Working Paper.

[9] Faulkender, Michael, Kristine Hankins, and Mitchell Petersen. 2019. Understanding the Rise in Corporate Cash: Precautionary Savings or Foreign Taxes. *The Review of Financial Studies*, 32 (9): 3299–34.

[10] Fahn, Matthias, and Nicolas Klein (2019), "Relational Contracts with Private Information on the Future Value of the Relationship: The Upside of Implicit Downsizing Costs," *American Economic Journal: Microeconomics*, 11(4), 33–58.

[11] Fong, Yuk-fai, and Jin Li (2017), "Relational Contracts, Limited Liability, and Employment Dynamics," *Journal of Economic Theory*, 169, 270–293.

[12] Fuchs, William (2007), "Contracting with Repeated Moral Hazard and Private Evaluations," *American Economic Review*, 97(4), 1432–1448.

[13] Gibbons, Robert, Niko Matouschek, and John Roberts. 2013. “Decisions in Organizations,” in: Robert Gibbons and John Roberts (eds.), *Handbook of Organizational Economics*, Princeton University Press, Princeton (NJ), 373–431.

[14] Gil, Ricard, Myongjin Kim, and Giorgio Zanarone (2022), “Relationships Under Stress: Relational Outsourcing in the US Airline Industry After the 2008 Financial Crisis,” *Management Science,* 68(2), 1256–1277.

[15] Guo, Yingni, and Johannes Hörner (2015), “Dynamic Mechanisms without Money,” Working Paper.

[16] Halac, Marina, and Andrea Prat (2016), “Managerial Attention and Worker Performance,” *American Economic Review*, 106(10), 3104–3132.

[17] Hastings, Donald (1999), “Lincoln Electric’s Harsh Lessons from International Expansion,” *Harvard Business Review*, 77(3), 162–178.

[18] Holmström, Bengt (1979), “Moral Hazard and Observability,” *The Bell Journal of Economics*, 74–91.

[19] Holmström, Bengt (1977), “On Incentives and Control in Organizations,” Ph.D. thesis, Stanford University, Palo Alto (CA).

[20] Hauser, Christine, and Hugo Hopenhayn (2008), “Trading Favors: Optimal Exchange and Forgiveness,” *Carlo Alberto Notebooks*, 88, 864.

[21] Jäger, Simon, Benjamin Schoefer, and Jörg Heining (2020), “Labor in the Boardroom,” NBER Working Paper No. 26519, National Bureau of Economic Research, Boston (MA).

[22] Levin, Jonathan (2003), “Relational Incentive Contracts,” *American Economic Review*, 93(3), 835–857.

[23] Li, Jin and Niko Matouschek (2013), “Managing Conflicts in Relational Contracts,” *American Economic Review*, 103(6), 2328–2351.

[24] Li, Jin, Niko Matouschek and Michael Powell (2017), “Power Dynamics in Organizations,” *American Economic Journal: Microeconomics*, 9(1), 217–241.

[25] Li, Jin, Arijit Mukherjee and Luis Vasconcelos (2023), “What Makes Agility Fragile? A Dynamic Theory of Organizational Rigidity,” *Management Science*, 69(6), 3578–3601.

[26] Lipnowski, Elliot, and Joao Ramos (2020), "Repeated Delegation," *Journal of Economic Theory*, 188, 105040.

[27] Macchiavello, Rocco, and Ameet Morjaria (2015), "The Value of Relationships: Evidence from a Supply Shock to Kenyan Rose Exports," *American Economic Review*, 105(9), 2911–2945.

[28] Macchiavello, Rocco, and Ameet Morjaria (2021), "Competition and Relational Contracts in the Rwanda Coffee Chain," *The Quarterly Journal of Economics*, 136(2), 1089–43.

[29] MacLeod, Bentley and James Malcomson (1989), "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment," *Econometrica*, 57(2), 447–480.

[30] MacLeod, Bently (2003), "Optimal Contracting with Subjective Evaluation," *American Economic Review*, 93(1), 216–240.

[31] Mailath, George J., and Larry Samuelson (2006), *Repeated Games and Reputations: Long-Run Relationships*, Oxford University Press, Oxford.

[32] Möbius, Markus (2001), "Trading Favors," Working Paper.

[33] Prat, Andrea (2005), "The Wrong Kind of Transparency," *American Economic Review*, 95(3), 862–877.

[34] Prendergast, Canice, and Lars Stole (1999), "Restricting the Means of Exchange within Organizations," *European Economic Review*, 43(4–6), 1007–1019.

[35] Stewart, James (1993), "Taking the Dare," *The New Yorker*, July 26, 34–39.

Appendix

We first prove some properties of the PPE payoff frontier.

LEMMA 2. *The PPE payoff set $\mathcal{E}$ has the following properties:* (i) *it is compact,* (ii) $u(\pi)$ *is concave,* (iii) for any $(\pi, u(\pi))$, the continuation payoffs are also on the frontier, and (iv) $\underline{\pi} = 0$.

Proof. Part (i): As there are only a finite number of actions, and the bonus paid is bounded by the value of relationship, standard arguments implies that the PPE payoff set $\mathcal{E}$ is compact

Part (ii) follows directly from the availability of a public randomization device since any payoff on the line segment between $(\bar{\pi}, u(\bar{\pi}))$ and $(\underline{\pi}, u(\underline{\pi}))$ can be supported as a PPE payoff.

Part (iii): Suppose there is a payoff pair on the PPE frontier supported by working $(\pi, u) = (\pi, u(\pi))$, and gives the continuation payoffs (π_H^W, u_H^W) and (π_L^W, u_L^W). Suppose that $u_L^W < u(\pi_L^W)$. Now consider an alternative strategy that specifies working and offers continuation payoffs (π_H^W, u_H^W) and $(\pi_L^W, u_L^{W\prime})$ where $u_L^{W\prime} = u_L^W + \epsilon < u(\pi_L^W), \epsilon > 0$. We can check that the alternative strategy satisfies all constraints and is therefore a PPE strategy. The promise-keeping constraints implies that this alternative strategy also gives the principal π but increases the agent payoff by $(1-p)\delta\epsilon$. This contradicts the fact that u is the highest PPE payoff to agent given the principal's payoff is π. Therefore, we must have $u_L^W = u(\pi_L^W)$. When the agent's effort can be observed, it is unnecessary to punish him below the frontier. With the same reasoning, we can show that all continuation payoffs after all actions remain on the frontier.

Part (iv) follows from first, 0 is the principal's maxmin payoff, and second, payoff pairs $(0, 0)$ is a PPE payoff sustained by taking outside option forever. □

Proof of LEMMA 1 We have shown in the main text that the continuation payoffs of state H are calculated by combining the principal truthtelling constraint and promise-keeping constraint. We now show that the continuation payoffs of state L equal $\bar{\pi}$. Mathematically, this is reflected by the result that $u'(\pi) > -1$, so $u(\pi) + \pi$ is increasing in π.

LEMMA 3. $u'(\pi) > -1$ *for all* $\pi \in (0, \bar{\pi})$.

Proof. Define $\Pi := \{\pi : u'(\pi) > -1\}$. The set is nonempty because $0 \in \Pi$, and is an interval because $u(\cdot)$ is concave. Suppose that $(0, \bar{\pi}) \nsubseteq \Pi$, and let $\pi^* = \sup \Pi$. We know from Darboux's theorem that $u'(\pi^*) = -1$.

Suppose $(\pi^*, u(\pi^*))$ is supported by working, and the bonus and continuation payoffs are given by b^W, u_L^W, u_H^W. The promise-keeping constraint (1) and the binding truth-telling constraint (2) together imply

$$\pi^* = (1-\delta)(y - b^W) + \delta\pi_L^W = (1-\delta)y + \delta\pi_H^W.$$

Then, we can write $u(\pi^*)$ as

$$u(\pi^*) = u^W(\pi^*) = (1-\delta)((1-p)y-c)-(1-p)(\pi^*-\delta\pi_L^W)+\delta\left(pu\left(\frac{\pi^*-(1-\delta)y}{\delta}\right)+(1-p)u(\pi_L^W)\right).$$

By the definition of the PPE frontier, $\pi_L^W \in \arg\max_{\pi' \in [0,\bar{\pi}]} u(\pi') + \pi'$, which does not depend on the value of π. In particular, π_L^W can be taken as π^* for any $\pi \in (0, \bar{\pi})$. Taking the derivative, we have

$$u'(\pi^*) = pu'\left(\frac{\pi^*-(1-\delta)y}{\delta}\right) - (1-p),$$

which then implies $u'(\pi_H^W) = -1$. Hence, $u(\cdot)$ is affine on $[\pi_H^W, \pi^*]$ with the slope of -1. The above observations allow us to show $u(\pi^*) = y - c - \pi^*$, for which the total surplus attains the first-best. This contradicts with the possibility that the state is H. Similarly, if $u(\pi^*) = -1$ for $\pi^* \in (0, \bar{\pi})$, contradictions arise when $(\pi^*, u(\pi^*))$ is supported by shirking or the outside option. Therefore, $u'(\pi) > -1$ for all $\pi \in (0, \bar{\pi})$. $\square$

Now we are able to prove that the continuation payoffs when opportunity costs are low must equal $\bar{\pi}$. Suppose there exists frontier payoff pair $(\pi, u(\pi))$ that is supported by working, but instead gives $\pi_L^W < \bar{\pi}$. Consider an alternative strategy that increases π_L^W by $(1-\delta)\epsilon$ and also increases b^W by $\delta\epsilon$. The alternative strategy satisfies all the constraints and therefore constitutes an PPE strategy. Under the alternative strategy, the principal's payoff remains unchanged while the agent's payoff changes by

$$(1-\delta)(1-p)\delta\epsilon + \delta(1-p)(u(\pi_L^W + \delta\epsilon) - u(\pi_L^W)) > 0$$

where the inequality follows from $u'(\pi) > -1$. This implies the alternative strategy generates higher payoff to the agent. This contradicts with the definition of the frontier. Similarly, we can prove for the case of shirking that $\pi_L^S = \bar{\pi}$.

Proof of PROPOSITION We proceed the proof with the following steps.

Step 1: Outside option is not used on the frontier.
We prove this by showing that $u^S(\pi) > u^O(\pi)$ for all π. Notice that

$$\begin{aligned} u^S(\pi) - u^O(\pi) &= (\pi + u^S(\pi)) - (\pi + u^O(\pi)) \\ &= \delta[p(\pi_H^S + u(\pi_H^S)) + (1-p)(\pi_L^S + u(\pi_L^S)] - \delta(\pi^O + u(\pi^O)) \\ &= \delta(1-p)[(\bar{\pi} + u(\bar{\pi})) - (\pi^O + u(\pi^O))], \end{aligned}$$

where recall that $\pi_H^S = \pi/\delta = \pi^O, \pi_L^S = \bar{\pi}$. As $\pi + u(\pi)$ is increasing in π and $\pi^O < \bar{\pi}$, we have that $u^S(\pi) > u^O(\pi)$. For any principal's payoff π, shirking always generates higher payoff to the agent than taking outside option. This is because when they stay in the relationship, there is positive probability the opportunity costs are low this period and the relationship can restart in the next period. When they restart the relationship, the total value from the relationship $(\pi + u(\pi))$ is the highest.

Step 2: Working is more likely to be used when π is higher.

LEMMA 4. *If $u^W(\pi') \geq u^S(\pi')$ for some π', then $u^W(\pi) \geq u^S(\pi)$ for all $\pi \geq \pi'$.*

Proof. This result, at some level, arises directly from the comparison of $u^W(\pi)$ and $u^S(\pi)$. Notice that

$$u^W(\pi) - u^S(\pi) = (\pi + u^W(\pi)) - (\pi + u^S(\pi)) = \delta p(u(\pi_H^W) - u(\pi_H^S)) + (1-\delta)(y-c),$$

where recall that $\pi_H^W = (\pi - (1-\delta)y)/\delta < \pi_H^S = \pi/\delta$.

This expression describes the benefit and cost of having the agent work. On the one hand, working generate surplus in the current period $(y - c > 0)$. On the other hand, if the opportunity costs are high, the principal's continuation payoff is lower when the agent works, i.e., requiring working results in a larger loss in trust. Notice that the benefit of working is independent of the agent's trust of the principal (π). The cost, however, is decreasing in π so that the cost from loss of trust is higher when the agent has lower trust in the principal. This is because when π is smaller, it becomes more likely that the principal's continuation payoff will drop below a threshold so that the agent will shirk. In other words, the lower is π, the smaller is the cushion in the relationship (before surplus destruction occurs). In this case, the principal is more likely to ask the agent to work when the trust (π) is high.

Formally, we have

$$u_+^{W\prime}(\pi) - u_+^{S\prime}(\pi) = p(u_+'(\pi_H^W) - u_+'(\pi_H^S)) \geq 0$$

as $\pi_H^W < \pi_H^S$ and the frontier $u(\pi)$ is concave. □

Step 2: LEMMA 5. *$(0, u(0))$ is supported by shirking, and $(\bar{\pi}, u(\bar{\pi}))$ is supported by working. Furthermore, $u(\pi) = u^S(\pi)$ if and only if $\pi = 0$, and $u(\pi) = u^W(\pi)$ if and only if $\pi \geq \pi_W = (1-\delta)y$.*

Proof. Suppose $(0, u(0))$ is instead supported by working. Then the continuation payoff when the opportunity costs are high must be negative

$$\pi_H^W = \frac{0 - (1-\delta)y}{\delta} < 0.$$

But then this contradicts with $\underline{\pi} = 0$. When the agent has no trust in the principal, the principal cannot motivate him to work. Therefore, $(0, u(0))$ must be supported by shirking.

Similarly, we can prove that $u(\bar{\pi}) = u^W(\bar{\pi})$ by showing that if $(\bar{\pi}, u(\bar{\pi}))$ is supported by shirking, the continuation payoff when the opportunity costs are high $\pi_H^W = \frac{\bar{\pi}}{\delta}$ must exceed $\bar{\pi}$.

Suppose there exists $\tilde{\pi} > 0$ such that the PPE payoff frontier is supported by shirking, $u(\tilde{\pi}) = u^S(\tilde{\pi})$. From promise-keeping constraint we have

$$u^S(\tilde{\pi}) = -(1-p)(\tilde{\pi} - \delta\bar{\pi})) + \delta(pu(\frac{\tilde{\pi}}{\delta})) + (1-p)u(\bar{\pi}))$$

As $u^S(\pi)$ is well defined

$$u_-^{S\prime}(\pi) = -(1-p) + \delta u_-'(\frac{\pi}{\delta}) \leq -(1-p) + pu_-'(\pi) < u_-'(\pi)$$

for all $\pi \in (0, \tilde{\pi}]$, where the first inequality follows from the concavity of PPE frontier and the second follows from $u'(\pi) > -1$. But as $u(0) = u^S(0)$, the above inequality implies that $u(\tilde{\pi}) > u^S(\tilde{\pi})$. This is a contradiction. Therefore, the frontier is supported by shirking only at $\pi = 0$.

Finally, we show that $\pi^W = (1-\delta)y$ is the smallest π supported by working. Suppose there exists $\pi < \pi_W$ supported by working, then the continuation payoff when the opportunity costs are high must be negative

$$\pi_H^W = \frac{\pi - (1-\delta)y}{\delta} < 0$$

This is a contradiction.

□

Step 4 : $u(\bar{\pi}) = 0$.

Suppose $u(\bar{\pi}) > 0$ so that the agent's IC is slack. As $(\bar{\pi}, u(\bar{\pi}))$ is supported by working, the continuation payoffs are given by $\pi_H^W < \bar{\pi}$ and $\pi_L^W = \bar{\pi}$, and the payment by $b^W = y - \bar{\pi} > y - (y - c) > 0$. Now consider an alternative strategy that specifies working, offers the principal continuation payoffs $\pi_H^{W\prime} = \pi_H^W + \epsilon < \bar{\pi}, \epsilon > 0$ and π_L^W, and lowers the bonus payment $b'^W = b^W - \delta\epsilon/(1-\delta) > 0$. We can check that this strategy satisfies all the constraints and therefore is a PPE strategy. But from the promise-keeping constraint, the principal's payoff under this strategy is given by

$$\pi = \bar{\pi} + \delta\epsilon > \bar{\pi}.$$

This is a contradiction. When the agent has most trust in the principal, the principal does not leave him with any rent and uses the trust alone to motivate the agent to work.

LEMMA 4. *The following holds:*

(*i.*) $(0, u(0))$ *is supported by shirking.*

(*ii.*) $(\bar{\pi}, u(\bar{\pi}))$ *is supported by working. Moreover,* $u(\bar{\pi}) = \underline{u} = 0$.

(*iii.*) *If* $u^W(\pi') \geq u^S(\pi')$ *for some* π', *then* $u^W(\pi) \geq u^S(\pi)$ *for all* $\pi \geq \pi'$.

For part (i.), it can be checked that $(0, u(0))$ cannot be supported by working as the continuation payoff $\pi_s^S(0) < 0$ according to Lemma 3. [#I think more context would help here. Perhaps express in words what each part says before providing an intuition?#] When it is supported by outside option, $u(0) = 0$. When it is supported by shirking, from the promise-keeping constraint and notice that $b^S(0) = \delta\bar{\pi}/(1-\delta) > 0$, we have $u(0) > 0$ Therefore, $(0, u(0))$ is supported by shirking [#I think this is either too much or too little. As it stands, I cannot easily follow this?#] . Similarly, for part (ii.), $(\bar{\pi}, u(\bar{\pi}))$ cannot be sustained by shirking or outside option as the continuation payoffs may exceed $\bar{\pi}$. And because effort is observable, at the maximal payoff for the principal, the agent does not earn any rents. Finally, part (iii) states that the frontier supported by shirking and that supported by working crosses only once. This follows from direct calculation of adding up the promise-keeping constraints and taking the difference:

Jin Li, Faculty of Business and Economics, The University of Hong Kong, Pok Fu Lam, Hong Kong.

Niko Matouschek, Kellogg School of Management, Northwestern University, Evanston, IL, USA.