A Model of Loyalty and Competence

Alexander F. Wagner

First version: July 2006
Current version: July 2006

This research has been carried out within the NCCR FINRISK project on “Corporate Finance”
A model of loyalty and competence

Alexander F. Wagner*
University of Zurich
July 3, 2006

Abstract

An agent shows loyalty to his manager by bearing personal costs to the superior’s benefit. In return, the manager may offer various forms of rewards. If this exchange is not contractible, typically repeated interaction will be required to sustain it. Beyond loyalty, the manager is interested in the employee’s competence. This analysis looks at the inevitable tradeoff between these two valued qualities when, as is to be expected, more competent agents have more attractive outside opportunities. It finds that: (1) the manager may find it optimal to hire a less competent employee so as to induce loyalty (as is frequently observed in practice), (2) even when loyalty is both feasible and net productive, it may not be optimal (i.e., the manager prefers spot contracting), and that (3) loyalty will be a signal of low competence. JEL-Code: D23, D82, M51.

*Assistant Professor of Finance and Financial markets at the University of Zurich and Research Associate at the Kennedy School of Government, Harvard University. Office address: Plattenstrasse 14, CH-8032 Zurich, Switzerland. Phone: 0041-44-634-3963, Email: wagner@isb.unizh.ch. I thank Philippe Aghion, Alberto Alesina, Nolan Miller and in particular Richard Zeckhauser for encouragement and comments. Randall Akee, George Baker, Charles Cohen, Robert Gibbons, Oliver Hart, Sonal Pandya, Catherine Thomas, Robert Urbatsch, Lucy White, and seminar participants at Analysis Group, Boston University, George Mason University, Harvard, MIT, the University of Chicago, and the University of Zurich provided helpful comments. I would also like to thank the Austrian Academy of Sciences for support through the APART fellowship. This research was supported by the NCCR FINRISK and by the University Research Priority Program / URPP "Finance and Financial Markets" a research instrument of the University of Zurich.
1 Introduction

I’ll take fifty percent efficiency, 
if I can get 100% loyalty.
Samuel Goldwyn

Organizations need competent members to fulfill their missions and to survive. Consequently, the overwhelming majority of economic theories of organizations focus on productivity – or competence, as it will be broadly referred to in this paper – as the key desired characteristic in job candidates. But the real life of many organizations poses puzzles for such theories. Anecdotes abound of managers of for-profit companies hiring less competent subordinates than possible (after taking into account different wages of more and less competent workers). A specific example is that family firms frequently employ dynastic management, i.e., they use family members to manage their business, rather than more competent professional managers.

When confronted with these facts, practitioners (and management books) frequently point out that organizations also need to ensure a sufficient degree of "loyalty." But what is loyalty? And why would this help to explain the puzzles that have to do with apparently inefficiently low levels of competence in some organizations? Motivated by these questions, this paper begins by proposing a simple economic model to capture part of the elusive notion of "loyalty." It broadly defines loyalty as the willingness of an individual to bear personal costs, perhaps opportunity costs, for another economic agent. Using this definition, it explores the idea that there is an important relationship between competence and loyalty, which often leads to a tradeoff between the two, i.e., more loyalty means less competence and vice versa.

This paper analyses loyalty as an optimally chosen behavior, not as a given trait.\footnote{In Hirschman (1971), loyalty refers to an individual staying in a declining organization, thus facilitating the voice option rather than the exit option. Hirschman’s seminal work therefore does not address the primary concern of this paper, namely, the interplay between competence and loyalty, and its consequences for the level and allocation of competence in organizations. Moreover, this paper does not only concern organizations in decline. In the organizational behavior literature, research on “organizational citizenship}
cusing instead on loyalty as an individual characteristic, the literature on cooperation has suggested that selection of partners and thus also of employees who will cooperate nicely is crucial, but it does not consider a tradeoff involved in that choice. In particular, that literature suggests that a manager should pick the person with whom a relational contract can be sustained at minimal costs, but does not take into account whether that person is also sufficiently productive. This assumption seems as restrictive as the approach of the literature that focuses solely on productivity. This paper remedies the respective problems in both literatures by bringing out the common price of loyalty, namely, reduced competence.

If there is a tradeoff between competence and loyalty, this would explain the above puzzles. Indeed, economic theories sometimes assume — consistent with Samuel Goldwyn’s dictum, but without a fundamental explanation — that high productivity types find it more expensive to act according to the needs of others. It is also a common theme in the popular management literature and in practice that managers hire less competent workers in order to have more loyal subordinates. Consistent with this view, Larry Ellison’s hiring practices for his senior executives at Oracle have been interpreted as favoring loyalty over competence, and of hiring more dependent workers as a means to continually ensure loyalty. But under

behavior” (OCB) has mushroomed since the contributions of Bateman and Organ (1983) and Smith, Organ and Near (1983). My definition of loyalty is closest to the notion of “organizational loyalty” as used as a subset of this concept, namely “allegiance to organizational leaders and the organization as a whole, transcending the parochial interests of individuals” (Graham 1991). For a review of the many different concepts related to OCB, see Podsakoff, MacKenzie, Paine and Bachrach (2000).

For example, Fehr and Falk (2002) define loyalty as preferences for taking into consideration the preferences of the partner, e.g., the employer. They thus look at loyalty as a matter of character, not as an endogenous behavior. The same is true for the “congruence parameter” in Aghion and Tirole (1997).

This is not to say that personality plays no role for loyalty; of course, it does. Moreover, even if competence and loyalty were independently distributed immutable personal characteristics, a manager who focuses on loyalty will, on average, choose less competent workers than a manager who focuses on competence. The analysis here demonstrates that there is an additional, endogenous cost of strategically optimal loyalty in terms of competence.

This assumption is made, e.g., in Fryer (2003) and Austen-Smith and Fryer (2005) where more productive individuals have higher costs of contributing to their peer groups.

See, for example, Heckscher (1995), Williams (1996), Joni (2004), and the discussion in section 3.

Manager-Magazin, January 13, 2004. The same argument also is often brought up where there is no
which conditions is this intuition correct?

To analyze these questions, in section 2, I consider an organization, for example, a consultancy, where a (female) manager (M) and a (male) worker (W) produce output. The more competent W is, the more output he produces; he writes better reports for the clients of the consultancy. W acts *loyally* when he is willing to bear personal costs to benefit M. I first concentrate on the case where the value of loyalty to the person receiving it is not related to the competence of the person giving it. For example, loyalty could mean casting a vote for M by lauding M’s abilities in a client meeting. W acts *independently* when he instead chooses to obtain private benefits, perhaps by using the meeting with the client to establish some personal business relationships.\(^7\) More competent W’s have more such opportunities for private benefits. Of course, W is willing to behave loyally only if he receives appropriate rewards.

Rarely will M and W be able to write a binding contract about allegiance and its rewards. Rather, loyalty and its results are usually non-contractible.\(^8\) It should be pointed out that it is possible to think of "loyalty" as "non-contractible effort." However, as described momentarily the setup here leads to quite different results than efficiency-wage theories. The substance of the analysis is more important than the terminology; the term "loyalty" primarily helps

\(^7\) The term “independence” suggests no actively hostile act. In contrast, “disloyalty” would have that more negative connotation. Loyal employees are “yes-men” in the sense that they help when they are needed. However, in contrast to Prendergast (1993), my model does not require an informational story for the core tradeoff.

\(^8\) I focus on the case where loyalty is observable. Even in this simple case interesting results arise which are expected to continue to hold in the more complicated case where loyalty cannot be easily monitored.
to provide a shortcut for the interpretation of anecdotes in organizational studies.

The non-contractibility of W’s actions suggests that the essence of loyalty lies in relations, thus making a repeated game the natural framework of analysis. I extend the theory of relational contracts by explicitly assuming that the inside and outside productivity of agents are correlated.9 Though small, this plausible assumption turns out to go a long way towards answering all of the above questions, and it provides several additional interesting predictions.

First, loyalty may be feasible only for less than fully competent individuals. This result arises because highly competent workers require enormous rewards for loyalty (due to their better outside options on the reneging path), but these rewards may not be credibly promised by the manager. This finding is an especially important case of the notion that loyalty can only be obtained where M can effectively and credibly punish W for disloyalty. Moreover, the less M reaps the benefits of higher productivity of W, the more is willing to sacrifice competence to induce loyalty. Recasting loyalty as effort, the difference to efficiency-wage and other effort-cost models becomes clear: Here, more competent W’s have higher costs of effort.

Second, even where loyalty is feasible and has net positive value for the manager, the manager may prefer a more competent worker with whom she keeps repeating the statically optimal equilibrium of the loyalty game – termed “spot interaction” in what follows.10 Sec-

9 The seminal contributions on relational contracts of Bull (1987), MacLeod and Malcomson (1989) and Levin (2003) do of course recognize that higher outside options make relations more difficult. But because the existing literature on self-enforcement does not allow for correlations between inside and outside productivity, taken literally, these models predict that it is optimal to pick the counterparty with the lowest possible reservation utility as the partner for interaction. For static contracts, see Lewis and Sappington (1989), Maggi and Rodriguez-Clare (1995), and Jullien (2000).

10 The result that M may choose to forego a feasible loyalty contract is related to the finding in Baker, Gibbons and Murphy (1994) that, in some cases, the presence of explicit contracts can make implicit contracts infeasible. In their model, the fallback option is determined by the quality of explicit contracts. Here, instead, the focus is on the competence of the worker, which in turn determines the attractiveness of the explicitly contracted part of the relationship, namely, expected output. More importantly, because I allow M to choose the worker’s competence, my model yields empirical predictions for which contracts as well as which patterns of task assignments and hierarchies will be observed in equilibrium.
tion 3 uses these two theoretical predictions to interpret evidence on the level of competence in family firms and other organizations.

Third, somewhat surprisingly, conditional on the optimality of a loyalty regime, when M values loyalty more, she will increase competence. This is so because with a higher valuation of loyalty, M can more credibly promise rewards.

Fourth, loyal behavior can be a signal of low competence. In other words, the endogenous tradeoff between loyalty and competence serves a screening condition. In particular, independence and loyalty can be signals of competence (to an uninformed outside market) made effective by the different marginal costs of loyalty that workers of differing competence levels face. Section 4 demonstrates, however, that whether less competent workers want to mimic more competent ones or whether the opposite is true depends on the extent to which the market values competence or loyalty. This prediction distinguishes this research from existing theories about adverse selection and job market signaling following Spence (1973), and it matches with some otherwise puzzling pieces of empirical evidence.

The analysis so far has assumed that the value of loyalty does not vary with competence. Naturally, one might expect things to be different where competence enters the production function of M’s loyalty gains. The result that emerges in section 5 is that the tradeoff continues to exist even in this general case as long as the range of ways in which the worker’s competence might be useful for loyalty within the firm is limited compared to what a worker might do with his competence in the outside world. As discussed in that section, this is true in many real world cases for within-firm relationships, but less likely to be true between firms.

Section 6 mentions a few other applications and offers concluding remarks.

2 A model of loyalty and competence

2.1 Setup

The model considers the interaction between a manager (M) and a worker (W) and posits a simple environment as follows. Both the manager and the worker are risk-neutral and
indefinitely-lived with common discount factor $\delta$, which will be assumed to be strictly less than unity throughout. The worker’s competence level is $\theta \in [0, 1]$.

2.1.1 Basic payoffs

M is the residual claimant of output. In each period, W produces output $\bar{y}$ with probability $\theta$ and $y$ with probability $(1 - \theta)$, where $\Delta y = \bar{y} - y > 0$. Output flows solely from competence; effort plays no role in the model. M pays a wage $w$ to W. This wage is market-determined, as discussed below. Thus, M’s expected payoff before calculating benefits and costs of loyalty is $Ey - w = \bar{y} + \theta \Delta y - w$. Similarly, W’s payoff in the absence of loyalty is $w$.

2.1.2 W’s side opportunities

Besides producing more output for M, more competent individuals have better and more side opportunities. I capture this notion by assuming that in each period, a side opportunity of value $b > 0$ arises for W with probability $p(\theta)$. For simplicity, let us assume that $p(\theta) = \theta$.\textsuperscript{11} If W chooses to pursue this opportunity, he acts independently.

The side opportunities can take many forms. For example, a highly competent employee of a software company may engage in consulting activities. An employee of a regulatory agency may be able to give a presentation on his field of expertise at a conference of professionals.

The wage $w$ and expected private benefits $\theta b$ are additively separable in the worker’s utility function. It does not directly harm M when W uses his side opportunity.

2.1.3 M’s gains from loyalty

The manager can, in each period, realize a potential gain (or avoided loss) $v > 0$, where $v$ is commonly known and commonly observed when it accrues to M. However, allowing M to obtain $v$ requires W to bear some personal costs. These costs of cooperation are here

\textsuperscript{11}For now, it is sufficient to say that we should think of W as an employee with a sufficiently broad portfolio of activities, such as ability to travel on the job or to have some other outside contacts.
modeled as foregoing the side opportunities \( b \).\(^{12}\) This is what it means for a W to be *loyal*. In other words, due to the worker’s limited time, he must make an either-or choice whenever a side opportunity arises: Either he acts loyally or he acts independently.\(^{13}\) The fact that W has to bear costs in order for M to realize the gains from loyalty is a defining characteristic of loyalty. Throughout, I assume that everybody can observe whether the worker is loyal. Let us say that loyalty has *positive social (joint) value* when \( v > b \).

Two remarks are in order: First, as discussed in the introduction, one might think of W’s decision as his effort choice. Previewing the analysis that follows, however, note that the equilibrium costs of effort for W will be determined by the strategic interaction with M. This will ultimately yield exactly the opposite prediction for how the cost of effort depends on competence than is implicit in a standard efficiency-wage model. Second, I begin to conduct the analysis under the assumption that \( v \) does not vary with competence. One example, relevant in many circumstances in the real world, is a situation of “one man, one vote.” In other words, an act of loyalty may literally involve casting the right vote in a democratic decision process, e.g., a committee or a board in an organization. It may also simply be taking a decision, like approving a project. (That the projects happen to be the same size every period is immaterial.) We will see later the precise way in which introducing a \( v \) that varies with competence alters the results.

### 2.1.4 Rewards for loyalty

The manager needs to offer the worker something in exchange for bearing personal costs. The most direct way to accomplish this is for the manager to pay the worker for loyalty, and this is the case I focus on. Risk aversion is not necessary to create a potentially beneficial role for the relation, and so I exclude it. The monetized value of this reward – say that of a company car, or invitations to basketball games – is given by \( x \geq 0 \), and is a choice variable.

\(^{12}\) The real costs of loyalty to one person may also consist in not being loyal to another person. As long as there is no strategic interaction between the two “liege lords” this case is covered here, too.

\(^{13}\) This is obviously an extreme assumption, but is allows us to bring out the basic insight in the sharpest possible way. I comment on less-than-full loyalty below.
for M. It is assumed to be stationary for reasons discussed later. It is also commonly observed.

| Stage 1 | 1. M offers wages $w$ such that W (of competence $\theta$) works for him.  
|         | 2. M receives output $y$ (with probability 1-0) or $\bar{y}$ (with probability 0) |
|         | Stage 2 | Probability 0: \( \theta \): Opportunity for W arises  
|         |         | Probability 1-0: No opportunity for W arises |
|         |         | Stage 3 | loyalty game (Table 1) |
|         |         |         | (Loyal, reward)  
|         |         |         | (Independent, reward) |
|         |         |         | (Loyal, not reward)  
|         |         |         | (Independent, not reward) |
|         |         |         |         | Next period: Restart at stage 1 |

**Figure 1: Sequence of events in each period**

<table>
<thead>
<tr>
<th>Table 1: Payoff matrix in the loyalty game</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>M rewards</td>
</tr>
<tr>
<td>M does not reward</td>
</tr>
</tbody>
</table>

Some real-world scenarios are more characterized by delayed, but bigger rewards, but for simplicity the analysis concentrates on pay-as-you-go rewards.
2.1.5 Summary, parameter restrictions, and timing

The economic environment in which M and W live is summarized as the game \( \Gamma (b, v, \delta, Y, \theta) \in \mathcal{G} \), where \( \mathcal{G} \) denotes the universe of possible games and \( Y \) summarizes information regarding the production technology \((y, \tilde{y})\) and the labor market (which is discussed momentarily). To ensure that M does not want to hire a W of negative competence, I assume that \( \delta^2 \geq \frac{1}{1-\delta} \cdot \frac{b}{v} \). This is guaranteed to hold as long as M and W are patient enough. Figure 1 shows the timing within each period. I assume throughout that M has all the bargaining power and gets all the surplus because she proposes the contracts. Each period, whenever an outside opportunity arises, M and W face a game characterized by the stage game payoffs given in Table 1, where payoffs are in the format \( U_M, U_W \). When no opportunity for W arises, it is assumed that W is loyal when a (explicit or self-enforcing) contract of loyalty is in place.

2.1.6 Labor market

For participation of W, we need to set a level of utility M has to guarantee him. I use a framework which will later allow predictions about how W’s productivity at M relative to other employment opportunities affects M’s preferences for loyalty.

As a baseline, assume that the only outside employment possibility for W is self-employment. Assume that M and W are uncertain about the possibilities that W will have to leverage his competence in his own business. I capture this notion by assuming that W’s self-employment opportunity i’s output takes the form: \( \theta \tilde{y} + (1 - \theta) y + \theta \tilde{\alpha}_i \). Neither loyalty nor side opportunities are relevant when W is self-employed. Let us assume that \( \tilde{\alpha}_i \) is distributed with some non-degenerate distribution with mean \( \bar{\alpha} \). Suppose that when W leaves M, he is randomly matched to one of these technologies and is self-employed from then on. We can interpret \( \bar{\alpha} \) as the additional marginal product of competence in self-employment. Comparing output in self-employment and output at M, we can also see that an appropriate interpretation of \( \bar{\alpha} \) is as an indicator of the outside employment opportunities’ relative productivity. A low value of \( \bar{\alpha} \) indicates that M has a relative advantage at using W’s competence in production. For brevity, and slightly abusing terminology, I will say that \( \bar{\alpha} < 0 \) means that M is relatively productive and that \( 0 < \bar{\alpha} < b \) means that M is relatively unproductive. I will throughout
focus on the case where $\bar{\alpha} < b$.\textsuperscript{15}

M and W expect a self-employed W to earn

$$\bar{E}y = \theta \bar{y} + (1 - \theta) y + \theta \bar{\alpha}$$

where $\bar{E}$ denotes the expectations operator taken over all the available technologies, and conditional on the expectation of productivity being $\theta$. In this perfect labor market, equation (1.1) therefore gives the reservation utility, denoted $\bar{u}(\theta) = \theta \bar{y} + (1 - \theta) y + \theta \bar{\alpha}$ that M has to guarantee W. M’s reservation utility is assumed to be sufficiently low to make the match between M and W attractive for M.

2.2 Analysis

I consider the case where the manager chooses the worker’s competence $\theta \in \Theta$ (where $\Theta$ are all numbers between 0 and 1), and, therefore, chooses which game $\Gamma(b, v, \delta, Y, \theta)$ to play. A natural interpretation of this assumption is that of M hiring exactly one W. (Alternatively, we can think of M as choosing W’s training, which is assumed to have a fixed cost.) W can take exactly one job. In all games, competence and loyalty are fully observable (monitorable) by everybody. I begin by assuming that loyalty and its rewards can be verified by a third party and are thus court-enforceable, or contractible. In contrast, the main analysis posits that loyalty and its rewards cannot be proven in a courtroom. In this case, loyalty is non-verifiable and thus non-contractible.

2.2.1 Statically optimal behavior (spot interaction)

For future reference, I define spot interaction as the equilibrium that arises out of statically optimal behavior by the players. When M and W cannot contract on loyalty, this would be the only equilibrium that arises if indeed M and W were meeting for only one round, and it remains a subgame-perfect equilibrium outcome also when the game is repeated.\textsuperscript{16}

\textsuperscript{15}In fact, $\bar{\alpha} > b$ would mean that M is not as productive as W’s self-employment opportunity even if M could completely internalize all of W’s outside opportunities into M’s own production.

\textsuperscript{16}The reference to this equilibrium as having “spot” characteristics is meant to distinguish it from a relational equilibrium which requires long-term interactions, as discussed below. The literature sometimes
In this equilibrium, M does not pay for loyalty \((x = 0)\), and it is thus optimal for W to act independently, which in turn rationalizes M’s strategy. Anticipating this equilibrium, M chooses the optimal competence level by solving

\[
\max_{\theta} U_M = y + \theta \Delta y - w
\]

\[
s.t. U_W = w + \theta b \geq \bar{u}(
\]

The constraint holds that inside utility, given by wages plus the expected value of side opportunities, must, on average, be no less than reservation utility. To minimize wage costs, M will offer \(w = \bar{E}y - \theta b\), i.e., she will make W’s participation constraint bind.\(^{17}\) Simplifying (1.2) by using the definition of \(\bar{u}(\theta) = \theta \bar{y} + (1 - \theta) y + \theta \bar{\alpha}\), M’s problem (1.2) is equivalent to

\[
\max_{\theta} \theta [b - \bar{\alpha}].
\]

If \(\bar{\alpha} < 0\), i.e., if M is relatively productive (or competence is relatively important) \(\theta = 1\) is optimal, leading to a payoff of \(b - \bar{\alpha} > b\). If \(0 < \bar{\alpha} < b\), i.e., if M is relatively unproductive (or competence is relatively unimportant), maximal competence is also optimal. The reason is that with a competent worker, M can economize on wage costs because the savings due to W’s outside opportunities outweigh his higher reservation utility. In this case, M’s payoff is \(b - \bar{\alpha} < b\).

2.2.2 Contractible loyalty

As a benchmark for the following analysis, consider now a scenario where loyalty and the rewards for loyalty are not only observable, but verifiable as well. In this case, because the manager proposes the contract, she can simply ask for loyalty and offer to pay the worker an amount \(x = b\) whenever an outside opportunity arises.\(^{18}\) The expected loyalty costs in uses the terminology of “no trade” (MacLeod and Malcomson 1998), but the notion of statically optimal behavior as used, for example, by Levin (2003) seems both more general and more descriptive here.

\(^{17}\)The assumption that \(\bar{u}(\theta) = \bar{E}y\) rather than \(\bar{u}(\theta) = \bar{E}y + \theta b\) is a way of saying that there must be some benefits for M to get a high-competence worker if he is affordable.

\(^{18}\)This is exactly equivalent to paying the expected present value of outside opportunities to W, up front. As is obvious from M’s maximization problem below, wages and loyalty rewards are perfect substitutes
each period, therefore, are \( \theta b \), which add to the wage bill. Thus, assuming for the moment that M indeed wants to induce loyalty, in order to identify the optimal competence level, the manager maximizes

\[
\max_{\theta} U_M = y + \theta \Delta y + v - (w + \theta b).
\] (3)

s.t. \( U_W = w + \theta b \geq \bar{u}(\theta) \).

The constraint now indicates that inside utility, given by wages plus rewards for loyalty, must, on average, be no less than reservation utility. To minimize wage costs, M will again offer \( w = \bar{E}y - \theta b \). Thus, we see that the manager’s payoff is maximized by

\[
\max_{\theta} -\theta \bar{\alpha} + v.
\]

Thus, if M is relatively productive, i.e., \( \bar{\alpha} < 0 \), maximal competence \( \theta = 1 \) is optimal, leading to a payoff of \( v - \bar{\alpha} \). When M is relatively unproductive, \( \theta = 0 \) is optimal, leading to a payoff of \( v \). Comparing the payoffs under contractible loyalty and spot interaction, we have

**Proposition 1** When loyalty is contractible and has positive social (joint) value, it is preferred to spot interaction.

**Proof.** See the Appendix. ■

This result is intuitive. The Proposition tells us that when loyalty has positive social value \( (v > b) \), the alternative – spot interaction – is never relevant. In other words, a contractible loyalty (static multi-tasking, contractible effort) model with positive correlation between inside and outside productivity therefore holds no predictions for the optimal level of competence except those stemming from comparative technology advantages (which translate into wage savings in the presence of a labor market). This situation changes drastically when we relax the assumption that loyalty is contractible.

under contractibility of rewards. The point is that M must somehow match W’s expected reservation utility.
2.2.3 Noncontractible loyalty

In many cases, neither allegiance nor rewards for loyalty is contractible. In this case, Table 1 shows that loyalty cannot be sustained in the one-shot game. Because $\theta \geq 0, Ey \geq 0, w \geq 0, x \geq 0, b \geq 0$, the unique Nash equilibrium of the stage game is spot interaction. Indeed, note that the game is similar to that of an asymmetric Prisoner’s Dilemma. Repeated interaction is required to open up the possibility for M and W to obtain the loyalty outcome. The crucial difference of this analysis from that of a standard repeated Prisoner’s Dilemma is that M can choose which amount $x$ to offer and with whom to play the game.\footnote{Some recent literature has analyzed mutual partner selection where purposeful matching replaces random or tournament matching. See, e.g., Orbell and Dawes (1991), Orbell and Dawes (1993), Morikawa, Orbell and Runde (1995), and Hauk (2001), among others. However, in these papers, being cooperative is assumed to be a trait (sometimes limited to a given time-period as in Hauk’s analysis), and the problem is to learn about others’ characters. Here, by contrast, “cooperation” (loyalty) is endogenous, and before selecting a worker, the manager knows exactly what the outcome of the game with the worker will be after everybody has optimized their behavior.}

In this setting, I consider self-enforcing (relational) loyalty with rewards contracts. I concentrate on stationary contracts of the following form: W promises to be loyal to the manager (that is, to not be independent even if the opportunity to do so appears but instead always help the manager to obtain $v$). The manager promises to pay $x$ in each period.\footnote{Levin (2003) provides a general theorem that shows we can limit our attention to stationary contracts in settings like the one considered here. There would be state-varying payments if the outside wage were also state-varying (Thomas and Worrall 1988). That rewards are paid every period and not merely whenever loyalty is costly for W is a convenient assumption, but it can be shown that the result is identical when M only pays $x$ when there indeed is a side opportunity and only M is tempted. One interesting difference arises, when W is also tempted in that setting (i.e., when he can take $b$ and $x$ at the same time). Then it becomes difficult for M to promise sufficient loyalty rewards for low-competence workers because they realize that taking a side opportunity now has huge value as compared to waiting for opportunities to show up in the future and then receiving payments.}

The timing of events becomes relevant. In particular, we need to ask whether M can take $v$, but not pay $x$, and whether W can take $x$, but still do $b$. That is, can the parties deviate from the loyalty contract and still obtain the other side’s cooperative contribution? In the
standard Prisoner’s Dilemma, the answer to this question is yes. Even when M and W do not literally move simultaneously, this is the correct assumption to make when they do not learn the other party’s move until later in the period. I will therefore proceed under this assumption.\textsuperscript{21}

\subsection*{2.2.4 Non-reneging constraints}

I now present the conditions for loyalty with rewards to be an equilibrium supported by trigger strategies, i.e., strategies in which M and W promise each other allegiance and rewards, and any one-time deviation results in both players exerting the statically optimal behavior in all future periods.\textsuperscript{22}

For the worker, honoring this promise of loyalty is preferred if and only if deviating today (and thus obtaining extra utility $b$) plus spot interaction going forward (and thus obtaining expected utility $w + \theta b$ at all future dates) is dominated by foregoing $b$, but obtaining $x$ in addition to the basic wage $w$ in each period. Formally, the worker’s non-reneging constraint (NR-W) is, in a state where the side opportunity of value $b$ actually arises,\textsuperscript{23}

\begin{equation}
    w + x + \frac{\delta}{1 - \delta} (w + x) \geq w + b + x + \frac{\delta}{1 - \delta} (w + \theta b). \tag{NR-W}
\end{equation}

The wage (the fixed component of the remuneration) drops out.\textsuperscript{24} Simplifying yields

\begin{equation}
    x \geq \frac{b}{\delta} - b (1 - \theta) = \theta b + b \frac{1 - \delta}{\delta}. \tag{4}
\end{equation}

\textsuperscript{21}In some models of relational contracts (e.g., MacLeod and Malcomson (1998)) only one side is tempted in this way, namely, the side that receives a product and then decides whether to pay the promised bonus. In my model, even if only M is tempted in this way, the qualitative message of the model remains the same.

\textsuperscript{22}This is the standard assumption for self-enforcing contracts and has been used both in methodological research (Bull (1987), MacLeod and Malcomson (1989), Levin (2003)) and in applications such as Baker, Gibbons and Murphy (2002). It is motivated by the work of Abreu (1988) who showed that if cooperation is attainable in a repeated game, it is without loss of generality to concentrate on the worst punishment path.

\textsuperscript{23}Note that the non-reneging constraint when no opportunity arises is implied by NR-W.

\textsuperscript{24}That the wage is a transfer and drops out from the non-reneging constraint is a fact common to models of relational contracts with additively separable fixed and bonus components of remuneration. The same occurs, for example, in MacLeod and Malcomson (1998). Therefore, M’s relative productivity $\bar{\alpha}$ has no implications for the feasibility of loyalty (although it will have implications for its desirability).
In other words, the required loyalty rewards are the expected value of outside opportunities plus some extra amount that decreases with the discount factor and increases with the value of $b$. Let $\bar{x}(\theta) = \theta b + b \frac{1 - \delta}{\delta}$ denote the minimal rewards that induce loyalty from type $\theta$.

In addition, the contract that M offers must also satisfy W’s participation constraint

$$w + x \geq \bar{u}(\theta). \tag{5}$$

But any worker who is in spot interaction with M earns $w = \bar{E}y - \theta b$, and M would surely not pay a higher wage to a loyal worker. Thus, (5) simplifies to

$$x \geq \theta b. \tag{6}$$

But because the self-enforcing contract needs to create surplus for the loyal worker, (6) is implied by (4), W’s non-reneging constraint. This implies:

**Lemma 1** Whenever the worker’s non-reneging constraint is fulfilled, his participation constraint is also satisfied.

Thus, in principle, every worker can be induced to be loyal – as long as the rewards $x$ are high enough. But that is, of course, precisely the constraint the manager faces. The higher the promised $x$, the higher is the temptation to renege on the contract. For some realized output today, M’s non-reneging constraint requires that obtaining loyalty benefits $v$ net of loyalty rewards $x$ in each period is preferred to receiving loyalty today without paying for it, but never receiving nor paying for loyalty in the future. In other words,

$$y + v - x - w + \frac{\delta}{1 - \delta}[Ey + v - x - w] \geq y + v - w + \frac{\delta}{1 - \delta}[Ey - w] \tag{NR-M}$$

which conveniently collapses to

$$x \leq \delta v. \tag{7}$$

Thus, the manager will set $x$ as small as possible, but she needs to take into consideration that the worker’s NR is still fulfilled. This task is depicted graphically in Figure 1. In the Figure, M must choose an $x$ that is below the NR-M line, but above the NR-W line. The
latter rotates with the level of competence. Assuming that the parameter values are such that both constraints can hold, there are (infinitely) many divisions of the surplus that work (see Theorem 1 in Levin (2003)).

Even without taking a stance on the division of this surplus, however, we can see that loyalty may dictate less than full competence. By combining the non-reneging constraints (4) and (7), we see that loyalty is feasible if and only if \( \theta b + b \frac{1-\delta}{\delta} \leq \delta v \), i.e.,

\[
\theta \leq 1 - \frac{b - \delta^2 v}{\delta b}.
\]

Denote with \( \theta^{L*} = 1 - \frac{b - \delta^2 v}{\delta b} \) the maximal level of competence compatible with loyalty.\(^{25}\) For example, for some given side opportunities \( \tilde{b} \), that maximal level is presented in Figure 2 as \( \tilde{\theta} \).

\(^{25}\)One noteworthy aspect is that \( \theta^{L*} \) is independent of output levels \( y \) and \( \bar{y} \). The reason is that the market knows W’s competence. Therefore, the competence level is fully internalized in the fixed part of the remuneration \( w \).
Full competence is not compatible with loyalty in that Figure since it would require loyalty rewards in an amount that M cannot credibly promise. In fact, full competence is only compatible with loyalty if the gains to loyalty sufficiently outweigh the value of side opportunities, i.e., if \( v > b/\delta^2 \). The key point thus is that it is not sufficient for them to be merely greater.

As the ratio between the two becomes less favorable for loyalty, only less competent individuals can successfully be induced to loyalty. But if M is productive, such workers reduce M’s welfare because of their lower expected value of production. Therefore, there will be a cutoff value for the gains from loyalty, below which productive M’s prefer spot interaction and full competence to loyalty with lower competence. When, in addition, competence is unimportant for production (compared to W’s alternative employment opportunities), higher levels of competence are unattractive for still another reason: M now must pay a premium for employing a competent worker. Therefore, even though a \( \theta^{L*} > 0 \) may be compatible with loyalty, M will be better off with a completely incompetent worker.

Summarizing the economically significant result, and leaving the details to be spelled out in the proofs, we have a two-fold finding: first, there are relevant ranges of parameter values where loyalty is desired but not feasible for given competence (Proposition 2), and, second, there are regions where, even though loyalty is valuable and feasible, it is not desired by M (Proposition 3). Proposition 2 thus establishes that there is a tradeoff between competence and loyalty because \( \theta^{L*} \leq 1 \). Perhaps the most important insight is that Proposition 3 is not simply due to the fact that loyalty’s absolute cost increases with competence. Instead, it arises because higher competence can be achieved in spot interaction than in loyalty.

**Proposition 2 (Feasibility of loyalty with rewards)** Non-contractible loyalty may not be feasible, even when loyalty would be desired if it were contractible.

**Proof.** See the Appendix. ■

**Corollary 1** A higher value of outside opportunities, b, a lower discount factor, \( \delta \), and a lower value of loyalty to the manager, \( v \), all decrease the maximum level of competence compatible with loyalty, \( \theta^{L*} \).
Proof. These comparative statics follow from observation of the expression for $\theta^{L*}$. □

Proposition 2 implies that non-contractible loyalty is less feasible than contractible loyalty. This simply follows from the fact that loyalty is always feasible when it is contractible. As for the comparative statics in the Corollary, it is worthwhile to state the underlying economic forces. First, note that as $v$ becomes large, even non-contractible loyalty allows full competence. I will return to the implication of this observation below. It is also intuitive that as $\delta$ increases, $\theta^{L*}$ increases. As is well-known from the Folk-Theorems, more patient individuals are less likely to renege on their promise or, conversely, the temptation $b$ weighs comparatively less. We would therefore expect project-based organizations – which have a shorter time horizon – to have difficulties in acquiring competent and loyal partners for projects.\(^{26}\) By contrast, organizations that establish long-horizon mentoring systems are more likely to achieve “affective commitment” (see, e.g., Payne and Huffman (2005) in the context of the US Army). Finally, better side opportunities (higher $b$) for $W$ exacerbate the tradeoff between competence and loyalty. In a small alpine village, hiring the most competent person does not cause great concerns for lack of loyalty, whereas in New York City the tradeoff between competence and loyalty should be more pronounced.\(^{27}\)

Proposition 2 is an important case of the more general notion that loyalty can only be obtained where $M$ can credibly threaten the worker with severe punishments if he is not loyal. The point here is that, in practice, one key determinant of $M$’s punishment ability is $W$’s competence.

The insights of Proposition 2 lead to the following result about the desirability of loyalty in Proposition 3. Consider the case where $M$ picks the optimal worker in each regime to maximize expected profits. I use the terminology that the net value of loyalty for a given type is $v - \tilde{x}(\theta)$. Importantly, this definition does not take into account $M$’s non-reneging

\(^{26}\)However, there may be recurring projects. For example, although the film industry is very much characterized by projects, loyalty of the form discussed here plays a role because the players expect to meet again. (Personal conversation with George Lucas, Chicago, June 12, 2004.)

\(^{27}\)A related result is that of Ramey and Watson (2001) who find that a fall in market friction, as reflected by a rise in the probability of locating a new trading partner in the matching market, makes less onerous the threat of severing an existing trading relationship, thereby tightening the effort constraint.
constraint. In other words, even if there is a positive net value of loyalty, M may not necessarily be able to implement it.

**Proposition 3 (Desirability of loyalty with rewards)** Even when loyalty is feasible and has positive net value, the manager may prefer spot interaction with a more competent worker.

**Proof.** See the Appendix. ■

To understand Proposition 3, consider the case where M is productive, $\bar{\alpha} < 0$. Then, for M to prefer loyalty with competence $\theta_{L^*}$ to spot interaction with competence $\theta^S = 1$, the Appendix shows that we must have $v - x (\theta_{L^*}) > (b - \bar{\alpha}) (\theta^S - \theta_{L^*}) > 0$.

This can be understood as follows. When either M is very productive (and thus $(b - \bar{\alpha})$ is large) or there is a tight bound on the feasibility of loyalty in terms of the allowed competence levels (and thus $(1 - \theta_{L^*})$ is large), loyalty must be very valuable in order to be preferred over spot interaction, where full competence is feasible.

A graphic representation of the case of M being relatively productive is contained in Figure 3. Recall that in this case, the contractible loyalty (static multi-tasking) model prescribes full competence and loyalty whenever loyalty has positive social value, $v > b$. Figure 3 describes the different implications that arise for the case of noncontractibility.

The graph has four relevant regions. Begin on the top left of the graph. In region I, we have $b - \delta^2 v < 0$, i.e., we are to the left and above of the $v = b/\delta^2$ line. The expression for $\theta_{L^*}$ tells us that we can obtain full competence and loyalty. Obviously, this is preferred to full competence alone.

As $b - \delta^2 v$ turns positive, we enter region II. As we move down and to the right of the graph (as $b$ increases relative to $v$), $\theta_{L^*}$ must become smaller than unity. At some point, the losses in competence required to still obtain loyalty become too great, and M is better off not trying to obtain loyalty, but focus instead on full competence under spot interaction. The proof of Proposition 1.3 shows that $\hat{v} (b, \bar{\alpha}, \delta) = b \frac{b - \bar{\alpha}}{\delta^2 \bar{\alpha}}$ is the minimum level of loyalty gains required to make loyalty desired for given other parameter values. It is intuitive and easy to
verify that \( \hat{v}(b, \bar{\alpha}, \delta) \) is increasing in \( b \) and decreasing in \( \delta \) and \( \bar{\alpha} \); for further properties, see the Appendix.

**Figure 3:** The tradeoff between competence and loyalty and its dependence on M’s gains from loyalty and W’s opportunity cost of loyalty. The graph approximately assumes \( \delta = 0.75, \bar{\alpha} = -0.4 \), and plots \( b \) between 0 and 3.

Below \( \hat{v}(b, \bar{\alpha}, \delta) \), in region III of the graph in Figure 1.3, full competence is preferred because the optimal choice of competence in order to retain loyalty is unattractively low. There are two interesting sub-regions, though. Far below \( \hat{v}(b, \bar{\alpha}, \delta) \), namely, where \( v < b \) (in region IV), competence dominates loyalty also under contractible loyalty. However, between the 45-degree line and \( \hat{v}(b, \bar{\alpha}, \delta) \), if M could contract on loyalty, she would be better off doing so. But non-contractible loyalty forces her to give up too much competence and thus induces her to forego loyalty altogether. This area thus shows the losses due to the need for
a self-enforcing loyalty contract.

The implied comparative statics for when loyalty becomes less preferred, i.e., when the set of games where loyalty is dominated by spot-interaction becomes larger, are summarized in the following Corollary.

**Corollary 2** Non-contractible loyalty becomes less preferred when M’s productivity or W’s outside opportunities increase, and when patience or M’s gains from loyalty decrease.

**Proof.** See the Appendix. ■

Finally, note that we have assumed that M fully internalizes reduced competence of W. That is, I have analyzed only "good" loyalty, and reducing competence was a means to achieve a second-best result in the presence of non-contractibility of loyalty. Much as a less productive M more strongly prefers loyalty and is willing to sacrifice more competence, so does a manager who does not reap the full benefits of a productive worker. I continue to concentrate on the case where M is the full residual claimant of output in order to emphasize that "bad" loyalty, i.e., loyalty benefiting only M, but not the firm’s ultimate owners, is not needed to explain why M trades off competence against loyalty. A model of executive compensation might take the loyalty-competence-tradeoff into account when endogenizing the degree to which M’s welfare is tied to production.

### 2.2.5 The optimal competence level and the value of loyalty to M

The model implies some at first nonintuitive, but ultimately straightforward predictions for the relationship between the optimal competence level and the value of loyalty to M.\(^{28}\) One might think that the tradeoff between competence and loyalty implies that where loyalty is more important, there will be less competence. But in fact, the opposite is true in a sense.

Recall Figure 3. Above a certain threshold – namely, the level of \(v\) above which loyalty is optimal, \(\hat{v}(b, \bar{\alpha}, \delta)\) – further increases in the value of loyalty, \(v\), lead to increases in the optimal

\(^{28}\)I focus on relatively productive M’s or M’s who maximize competence for other reasons. For a relatively unproductive M, zero competence is preferred to full competence whenever \(v - b \frac{1 - \delta}{\sigma} > b - \bar{\alpha}\).
competence level. The reason is that the manager can now more credibly promise loyalty rewards. However, at values of $v$ below $\hat{v}(b, \bar{x}, \delta)$, spot interaction and full competence is preferred, and there is therefore a (discrete) downward jump in competence as we move from statically optimal behavior to loyalty.

This is depicted in Figure 4 and summarized in Corollary 3.

**Corollary 3** In equilibrium, a worker who acts loyally is (weakly) less competent than a worker who acts independently, but as loyalty becomes more valuable to the manager ($v$ increases), the maximum level of competence compatible with loyalty increases. Thus, there is a non-monotone relationship between the importance of loyalty to the manager and the optimal level of competence.

The result that the importance of loyalty and (optimal) competence are in fact positively correlated over the range of parameters where establishing a loyalty relationship is optimal is consistent with arguments in the management literature (see, for example, Skinner (1981)).
2.2.6 Intermediate degrees of loyalty

I have so far modeled loyalty in a very stark way: Either W is loyal or not. In practice, W’s loyalty is likely to be a matter of degree. Fortunately, the basic results extend to this more general setting. Specifically, M and W could agree on a self-enforcing contract that (following some public randomization device) requires W to forego his side opportunities in \( \phi \) percent of those cases when a side opportunity appeared. In other words, \( \phi \) would be the degree of loyalty. This leads to the following results. First, for a given level of competence, the maximal sustainable degree of loyalty may be less than 100\%. That maximal degree is decreasing in the level of side opportunities, but increasing in the value of loyalty and the time horizon. Conversely, a given degree of loyalty is only feasible up to a certain level of competence. As the required degree of loyalty increases, the maximum level of competence compatible with it will decrease as long as side opportunities are sufficiently large compared to the discounted value of competence.\(^{29}\) This parallels the result concerning the feasibility of loyalty (Proposition 2 and Corollary 1). Second, the optimal degree of loyalty may be lower than the maximal level of loyalty that could be achieved given the economic environment. This extends the prior result concerning the desirability of loyalty (Proposition 3 and Corollary 2). The intuition is that M may be better off implementing a lower degree of loyalty but secure higher competence.\(^{30}\) Finally, Corollary 3 generalizes in that with continuous degrees of loyalty, there is a u-shaped relationship between the value of loyalty,

\[^{29}\text{In particular, generalizing the above non-reneging constraints, the maximum level of competence compatible with a degree of loyalty } \phi \text{ is given by } \hat{\theta} = \frac{\delta^2 v - (1-\delta)b}{\phi \delta v + (1-\phi)\delta^2 v}. \text{ Therefore, when } v > b > \delta v, \text{ as the required degree of loyalty increases, competence must be decreased, i.e., } \partial \hat{\theta} / \partial \phi < 0. \text{ When } \delta v > b > \delta^2 v, \text{ increasing the required degree of loyalty actually allows increasing the level of competence, i.e., } \partial \hat{\theta} / \partial \phi > 0. \text{ Thus, if any loyalty is optimal, full loyalty is preferred. However, only less-than-full competence secures full loyalty in this case. Only when } \delta^2 v > b \text{ are 100\% loyalty and full competence compatible.} \]

\[^{30}\text{This result applies when } v > b > \delta v. \text{ Note that } \hat{\theta} \text{ is convex in } \phi. \text{ In other words, M is maximizing a concave objective function (with the choice or input variables being the degree of loyalty and competence) over a non-convex technology. Thus, intermediate degrees of loyalty are optimal when M’s indifference curves in } \phi - \theta - \text{space are more strongly curved than the boundary of the production possibility set, given by } \hat{\theta}. \text{ Otherwise, a corner solution with full or zero loyalty is optimal.} \]
$v$, and optimal competence over some range. This non-monotone, but smooth relationship between $v$ and optimal competence is more easily amenable to empirical tests than the discontinuous regime change in Figure 4. The intuition is that at low levels of $v$, low degrees of loyalty (or no loyalty) are optimal, thus allowing high competence levels. As $v$ increases, higher degrees of loyalty become preferred, but initially, M has to sacrifice some competence to achieve them, and she optimally chooses to do so. After some point, higher degrees of loyalty and competence in fact become complements, for the same reason as before: Both M and W are aware that M’s reneging on the loyalty contract would trigger significant losses for her, and so she can credibly promise to reward high degrees of loyalty even of highly competent workers. Finally, as $v$ becomes sufficiently large, full loyalty and full competence become achievable (and, of course, optimal).

2.2.7 A competitive labor market

The basic results continue to hold when W’s alternative employment opportunities also include other firms. The situation is a bit more complicated since W may enter into a loyalty regime at another firm as well, even after reneging on a loyalty contract at M or leaving a spot regime at M. To see the intuition, assume that in addition to varying with respect to productivity, firms vary with respect to their preferences for loyalty. Continue to assume that the value of side opportunities available to W is homogenous across firms.

Intuitively, if M and W expect that W will be loyal at his next employer, they expect him to earn rents above what he would get in spot interaction. Because of this, it will not be enough for M to pay such a W the same base wage as she did when W could only go into self-employment. This makes loyalty even less attractive. The Appendix shows that if loyalty is contractible, this assumption has no implications. In contrast, if M and W expect that W will not be loyal at his next employer, we are back to the case we analyzed above. When W leaves M, he expects to receive wages $Ey - \theta b$. This is because the other firms will also take into account that W has his side opportunities, and each firm faces competitive pressures. Thus, M will pay exactly the same wage to W as if she and W were in spot interaction. The following statement summarizes these results.
Proposition 4 Assume that when W leaves M, he is matched randomly to a firm with expected relative productivity $\bar{\alpha}$ and expected gains from loyalty $\bar{v}$. Then

1. If M values loyalty more than the average firm in the market ($v > \bar{v}$) or if all firms are willing to engage only in spot interaction with a W who has reneged, the results are exactly the same as if self-employment is the only alternative employment opportunity.

2. If M is relatively productive and values loyalty less than the average firm in the market ($v < \bar{v}$ and $\bar{\alpha} < 0$), non-contractible loyalty is less attractive than in the case where the market does not value loyalty.

Proof. See the Appendix.

2.3 Discussion

I have analyzed loyalty as a matter of choice not temperament. Loyalty is a form of non-contractible inside effort, but there are a number of differences to the literature on efficiency wages that go substantially beyond terminology. First, observability is not the problem here (unlike in Shapiro and Stiglitz (1984)). Second, the focus on the tradeoff between competence and loyalty is absent in existing models because they do not focus on the manager’s choice.

$^{31}$Loyalty arises endogenously in my model as a consequence of the differing abilities of people to capture the benefits of state-dependent payoffs. A different sort of endogenous loyalty than the one considered here can be seen in the observation that typically relationships that have lasted longer are more difficult to break up. This notion has been formalized by Lindsey, Polak and Zeckhauser (2004). Earlier work had introduced the idea of social conventions in which the value of relationships increases over time (Ghosh and Ray 1996, Ghosh and Ray 2001, Kranton 1996) as well as the notion that there is mutual lock-in due to gift-giving (Akerlof 1982, Carmichael and MacLeod 1996). Other work on endogenous trust includes Casadesus-Masanell (2003), Watson (1999), and Watson (2002). In the Casadesus-Masanell paper, individuals do not differ in competence, however. In Watson’s work the inclination to cooperate is assumed. Finally, note that there are some important differences between the notion of loyalty as developed here and the related idea of firm-specific investment (Grossman and Hart 1986, Hart 1995). First, firm-specific capital need not make a worker loyal to his boss in the sense that he helps him out when he needs it. To the contrary, when internal opportunities for applying the high firm-specific knowledge arise, such a worker may be quite disloyal to his boss, by, for example, taking a job in another division. Second, the only way to ensure loyalty in my model is via self-enforcing contracts, because the right to decide whether to help out or not is inalienable.
of the worker’s quality. Third, and perhaps most importantly, effort costs are state-varying in my model, as loyalty is only costly when there is a side opportunity for W. As a result, in equilibrium, my model predicts that more competent workers find loyalty more costly, while the more standard assumption would be that more competent workers can more cheaply be induced to inside effort.

The primary result of this essay is that even in the absence of any other “usual suspects” like asymmetric information or credit constraints, there is an endogenous tradeoff between competence and loyalty when more competent individuals not only produce more internally but also have better outside options. M may thus prefer to hire a less competent worker. The result that even where loyalty is worth it, M may not be able to get it can essentially be interpreted as a version of the Folk Theorem.

Another surprising result of the analysis is that where M can choose with whom to interact, even where loyalty is worth it and feasible, M may not want it (or may wish to implement a smaller degree of loyalty than is maximally sustainable). There are three reasons

\[ x = \frac{\theta_L}{\delta_L} - b(1 - \theta_L), \]  
\[ x = \frac{\theta_H}{\delta_H} - b(1 - \theta_H), \]

Since \( \theta_L < 1 \) and \( \Delta \theta < 1 \), the high type would need to be more patient than the low type, and this difference is increasing in the divergence in the competence levels.

---

32 Glazer (2002) presents a static model of rent-division that focuses on the optimal mix of internal and external rent-seeking abilities, but does not consider optimal incentives or the worker’s choice to be loyal, i.e., to forego rents. I assume that the worker can fully use his competence in alternative employment opportunities. A no-compete clauses in the employment contract would lower a worker’s reservation utility and the attractiveness of his reneging path. Intuitively, to the extent that no-compete clauses are enforced, it is as if loyalty were contractible.

33 Egorov and Sonin (2006) show that dictators may favor incompetent viziers. Contrary to my model, they focus on asymmetric information. In terms of results, one crucial difference is that my model predicts that where M needs more loyalty, she will in fact be able to and prefer to hire more competent workers (because \( \theta_L^* \) is increasing in \( v \)).

34 Competent people pose additional difficulties for loyalty relations because their competence acts as if they were too impatient. To see this, suppose that for some given loyalty rewards, some type \( \theta_L \) can just be induced to loyalty. That is, \( x = \frac{\theta_L}{\delta_L} - b(1 - \theta_L) \), where I note explicitly that this equation holds for some \( \{ \delta_L; \theta_L \} \) combination. Now suppose \( \theta_L \) is replaced with \( \theta_H \), or M’s competence increases. If M wanted to induce loyalty with the same rewards scheme also from \( \theta_H \), this would require \( x = \frac{\theta_H}{\delta_H} - b(1 - \theta_H) \geq \frac{\theta_L}{\delta_L} - b(1 - \theta_L) \).

Solving for \( \delta_H \) leads to the condition \( \delta_H \geq \frac{\delta_L}{1 - \theta_L} \). Since \( \delta_L < 1 \) and \( \Delta \theta < 1 \), the high type would need to be more patient than the low type, and this difference is increasing in the divergence in the competence levels.
for this result. First, loyalty is more expensive when it is non-contractible than when it is contractible. Second, this increase in loyalty costs may put M in a situation where she cannot credibly promise rewards for loyalty because W knows her incentive to renege is too great. Third, in spot interaction (or in relations with lower degrees of loyalty), M can benefit from competent workers because this allows her to economize on formal wage costs. Instead, if loyalty is contractible, the assumption that inside and outside productivities are correlated has no effect on the optimal hiring decision. The reason is that in that case the better outside opportunities are exactly compensated by the ability of M to economize on wages. Therefore, a contractible loyalty model would make no predictions about optimal competence levels other than those stemming from comparative advantages in technology. Only the combination of relational contracts with correlation of inside and outside productivities yields such predictions.

3 Evidence

The theory can help us to understand the puzzles posed at the beginning of the paper as well as some additional facts. There is considerable anecdotal and empirical evidence for the existence of the phenomenon described by the model.

3.1 The tradeoff

First, it is frequently recognized that more competent workers are less loyal. For example, Williams (1996) discusses the problems managers face with very intelligent software developers who are less easily induced to follow their managers’ plans. Moreover, studies

\[ v > \frac{b}{\delta} \]

loyalty will be achieved and the optimal competence level is 1. To understand this result, note that now there is no bound \( \theta^{L*} \) on competence that is compatible with loyalty; the only bound on when loyalty will arise is due to the higher costs that come from the higher loyalty rewards M must pay. In particular, for a given level of competence \( \theta \), loyalty is preferred to spot interaction under full competence if

\[ v > \frac{b}{\delta} - (1 - \theta) \bar{\alpha} \]

Thus, productive \((\bar{\alpha} < 0)\) organizations that can build reputation for rewarding loyalty will hire very competent workers.

35 Some organizations may be able to develop a reputation for rewarding loyalty, and may thus not be bound by a non-reneging temptation on their part. In this case, whenever \( v > \frac{b}{\delta} \), loyalty will be achieved and the optimal competence level is 1. To understand this result, note that now there is no bound \( \theta^{L*} \) on competence that is compatible with loyalty; the only bound on when loyalty will arise is due to the higher costs that come from the higher loyalty rewards M must pay. In particular, for a given level of competence \( \theta \), loyalty is preferred to spot interaction under full competence if

\[ v > \frac{b}{\delta} - (1 - \theta) \bar{\alpha} \]

Thus, productive \((\bar{\alpha} < 0)\) organizations that can build reputation for rewarding loyalty will hire very competent workers.

27
in organizational behavior have found that those with many outside options (an important determinant of which is competence) are less likely to contribute organizational citizenship behavior, a subset of which is organizational loyalty (Lawler and Yoon 1996, Hui, Law and Chen 1999, Podsakoff et al. 2000, Thau, Bennett, Stahlberg and Werner 2004).

Second, management books routinely point to the practice of managers hiring less than fully competent workers to ensure loyalty. The analysis here suggests that due to the fundamental tradeoff between competence and loyalty, wherever loyalty is regarded as important enough, the efficient solution may indeed be to give up a bit of competence, compared to the full-competence statically optimal outcome. It is of course hard to find direct and quantitative evidence for the dependence of actual hiring and selection behavior by managers on this calculation, but that managers are well aware of this tradeoff and actively consider it is apparent from often-cited quotes like those of Samuel Goldwyn (“I'll take fifty percent efficiency if I can get 100% loyalty”) and others. Consistent with this, Hall, Rose and Subramanian (2001) describe how Richard Auhull fought a multi-year battle to defend the medical device maker he had founded, Circon, against a hostile take-over from U.S. Surgical. Especially towards the end of that battle, Auhull relied strongly on less competent allies who had fewer outside options in his efforts to promote his “strategic plan” (which, according

36 This is true for both absolute and relative (compared to the superior) competence as well as for insufficient training of existing workers (Sample 2003). Existing explanations often appeal to a slightly different notion of loyalty, namely, the idea that too competent subordinates may try to take the manager’s post, like in Friebel and Raith (2004). Friebel and Raith do not allow the worker to choose to be loyal, nor do they allow for the manager’s attempts to foster loyalty. Although these stories may explain part of what is observed, they are a less convincing factual description than appears at first sight. In particular, there is no evidence that the vast majority of companies that do not use upward appraisals, and large majority of companies that have no-replacement-policies (Bettenhausen and Fedor 1997) succeed in hiring more competent workers.

37 For example, W. Michael Blumenthal, former chairman and CEO of Unisys, was concerned with his ability to keep his managers loyal to the moral values he embraced. Sometimes, he had a worker who could not be induced to live by these rules. Reflecting back on his career, Blumenthal said “When did I make my greatest hiring mistakes? When I put intelligence and energy ahead of morality.” The model’s assumption that W gets directly paid by M for loyalty should perhaps not be taken literally in the case of loyalty to M’s moral values.
to most analysts, was designed mostly to keep him in power because “the company was his baby”). For a related example, though with a somewhat more positive flavor of the value of loyalty for the organization, see the case of Ecolab (Nanda 1996).

So far, we have interpreted the model quite literally, namely, as being about hiring. But we can also look at it as describing the selection of allies and informants. Managers frequently do not choose the most competent advisors precisely because they fear they will be less loyal. For example, Joni (2004) describes the case of a regional vice president who was worried that his subordinates knew the territory and the players too well and “feared they would exploit that knowledge for their own purposes” (p. 87).

### 3.2 Family firms

A stylized fact regarding family firms is that for many positions family members are preferred. It is often argued that such dynastic management is inefficient since internal candidates are less qualified to do a particular job.\(^{38}\) Some have argued that the reason why family firms hire internal candidates is a consequence of financial market imperfections (see, e.g., Caselli and Gennaioli (2005)) or the legal system (see, e.g., Panunzi, Burkart and Shleifer (2003)). The analysis in this paper has a different flavor. Consider the decision of a founder whom to employ as the manager: a family member (who, on average over all firms, will be less competent than an outside manager for most common distributions of talent in the population) or a professional manager. Let us assume that when she hires the family member (e.g., her son) she will very likely obtain loyalty, perhaps because the punishment for reneging — e.g., cutting out of the will — is extremely powerful in this case. That is, with the family member, the founder faces one sort of competence-loyalty tradeoff. But when she hires the professional manager, she knows that he will face the endogenous competence-loyalty tradeoff

\(^{38}\)According to Warren Buffett, those firms that pick executives from the pool of family heirs would be “choosing the 2020 Olympic team by picking the eldest sons of the gold-medal winners of the 2000 Olympics” (cited in Pérèz-Gonzalez (2006)). This is not to say that family firms inherently perform poorly; to the contrary: as long as the founder is involved in one way or another, such firms generally perform very well (see Villalonga and Amit (2005)).
that has been the subject of this paper.\textsuperscript{39} That may already be enough to make her son the preferred manager.

### 3.3 Public agencies

Loyalty of executives in public agencies is valued highly. Even in the US and other developed countries, many “agency executives are selected in order to serve the political needs of the president, and these may or may not involve policy considerations” (Wilson 1989, p. 198). The presidential appointment process is frequently regarded as being strongly dominated by the tradeoff between competence and loyalty (Edwards 2001), sometimes leading to “amateur government” (Cohen 1998). Similarly, Heclo (1977) observed that “many of [the agency’s executive’s] selectors [are] more interested in the process of getting their way than in the executive’s eventual output.” (p. 99).\textsuperscript{40} The theory presented in this paper also provides an explanation for why reforming the presidential appointment process with the goal of increasing quality of appointees is so difficult: there is at least partially an inevitable tradeoff between competence and loyalty, and because the public sector is generally characterized by the absence of strong punishment abilities (see Wilson (1989)), reductions in competence need to act as a substitute tool to obtain the required loyalty. The prediction that where there is more loyalty, there will be less competence is also consistent with casual evidence from comparing (the arguably on average relatively low) talent in political parties (which predominantly favor loyalty because of the high costs of unilateral deviations from messages, say) and (relatively high) talent in think tanks (which favor competence).\textsuperscript{41}

\textsuperscript{39}The point that more qualified professionals have less well aligned interests than family heirs has been made by Bhattacharya and Ravikumar (2002), but they do not discuss \textit{why} that is so. See also Davis, Schoorman and Donaldson (1997) and their “stewardship theory of management.”

\textsuperscript{40}On the other hand, the reputation of the Army Corps of Engineers, the Bureau of Prisons, or the Forest Service would be risky to tamper with (Wilson 1989, p. 199).

\textsuperscript{41}For a theory of the supply side of talent in parties see Caselli and Morelli (2004).
4 Loyalty as a signal

We have so far assumed perfect and symmetric information. In reality, managers or the market may not know a worker’s competence with the same precision as he does. To model this situation, let us now suppose that M has one worker who can be of two different types, i.e., levels of competence, \(0 < \theta_L \leq \theta_H < 1\). Let us begin by assuming that M knows the worker’s type (which I will refer to as low or high), but the market does not know which type the worker is.\(^{42}\) Further suppose that the market values loyalty less than M. A natural candidate equilibrium is one where low workers are loyal and high workers are independent. The manager chooses the wages she offers to the high and low types, respectively, \(w_H, w_L\) and loyalty rewards \(x\).\(^{43}\)

The timing is as before, with the following exceptions. The market does not observe wages or output levels. However, at the beginning of each period, conditional on the history of play (namely, whether the worker was independent), the market updates its belief (starting from some prior) about the worker’s competence. Without loss of much generality, consider the simplest case, where in the first period an opportunity of value \(b\) arises for sure, and the market knows this. Thus, the market learns whether \(W\) was independent or whether \(W\) was loyal. This implies that if the market believes that only low types are loyal, separation occurs either immediately or never.

In a perfect Bayesian equilibrium, given the market’s beliefs and the manager’s wages and rewards schedule, both worker types find it optimal to act according to the market’s belief. Given the market’s beliefs and the workers’ behavior, the manager finds it optimal to offer loyalty rewards and wages consistent with beliefs and behavior.

4.1 Analysis

If M attempts to induce less competent workers to loyalty, the potential problem is that if the market’s rewards for independence are too tempting, low-competence workers will

\(^{42}\)We will see that the spirit of the results still applies in the case where M is also asymmetrically informed.

\(^{43}\)In principle, the manager can offer two rewards levels, \(x_H\) and \(x_L\), but in the equilibrium we consider, it will turn out to be optimal to induce only \(\theta_L\) to loyalty, so I omit the index.
try to mimic high-competence workers. Indeed, I begin by assuming that even though M knows the worker is the low type, in order to keep him inside the firm, she would need to offer a wage corresponding to the high types’s wage forever if W acts independently. The reason why separation is nonetheless potentially possible in equilibrium is that the tradeoff between competence and loyalty provides for an endogenous sorting condition: Loyalty is less attractive for high types than for low types. In other words, the implicit costs of loyalty instead of no loyalty are higher for high types than low types. The mechanism employed in establishing separation is thus methodologically interesting because signaling models usually posit a sorting condition as a (more or less plausible) assumption. Here, such an assumption is not necessary. This combination of signaling and relational contracts appears to be a novel mode of analysis. 44

Employing these insights, and recognizing that loyalty payments also need to cover the wage differential a disloyal worker can obtain, we can state the following result for the basic signaling game $\Gamma (\delta, b, v, Y, \{\theta_H, \theta_L\})$ with the following features: (1) An opportunity of value $b$ appears for certain at the beginning of the game. (2) The market values loyalty less than M and $\bar{\alpha} = 0$. (3) Assume $\bar{y} > b$ and $\theta_H - \theta_L > \frac{1 - \delta}{\delta}$. (4) Low’s loyalty has positive net value under perfect information (i.e., $v - \tilde{x}(\theta_L) > 0$) and High’s loyalty has negative net value under perfect information (i.e., $v - \tilde{x}(\theta_H) < 0$).

Proposition 5 Consider the basic signaling game. There exists a separating perfect Bayesian equilibrium that is characterized by the following features:

1) The market believes that $\theta_H$ acts independently and that $\theta_L$ acts loyally. The two types behave accordingly.

2) The loyalty rewards $x$ and wages $w_H, w_L$ satisfy the following conditions:

a) High is at least as well off as Low. That is: $w_H + \theta_H b \geq w_L + x$.

b) Low does not want to be independent. That is: $w_L + x > w_H + \theta_L b$.

c) High exactly receives his reservation utility. That is: $w_H + \theta_H b = \bar{u}(\theta_H)$.

44 Fryer (2003) and Austen-Smith and Fryer (2005) also address the link between signaling and repeated games. However, their central assumption is that there is one signal at the beginning, and then an infinitely repeated game.
d) Low earns a surplus over his reservation utility. That is: \( w_L + x > \bar{u}(\theta_L) \).

(3) The manager never does better than if the market has symmetric information.

**Proof.** See the Appendix. ■

In other words, whenever M wants to separate types under perfect information, she can do so under asymmetric information (as long as the types are sufficiently different and sufficiently patient), but at additional costs. Note that the characterization of wages and loyalty rewards in part (2) allows for a range of equilibria. This multiplicity of equilibria is not due to a problem of specifying beliefs on an off-equilibrium path, but due to the fact that wages and loyalty payments, \( w_L + x \), are essentially perfect substitutes in inducing Low to loyalty.\(^{45}\)

4.2 Discussion

A first important implication of the model is that M cannot cut wages and loyalty rewards for loyal workers too much. By “ratcheting” and offering too low a wage for the loyal worker, the manager risks that the worker is tempted too much by the ability to get the higher wage that must be offered to disloyal workers. Thus, when we interpret wages and loyalty rewards together and assume that when the manager has a choice she will pay out most of the sum in wages, not in informal loyalty rewards, the model provides a resolution to the puzzle that managers are generally reluctant to lower wages in the face of a decline in a worker’s reservation wage (MacLeod 2001).\(^{46}\)

\(^{45}\)Each type has exactly one behavior to exhibit in equilibrium and there are only two possible behaviors. Therefore, no specification of off-equilibrium beliefs is necessary. Thus, this equilibrium is also sequential (Kreps and Wilson 1982). As before, the reason for treating wages and loyalty rewards separately is that M can only renege on loyalty rewards.

\(^{46}\)Note that here, the potential ratchet effect is not due to M learning about W’s competence (as it is, for example, in Hart and Tirole (1988), Laffont and Tirole (1988), and Kanemoto and MacLeod (1992)) but due to the market’s learning. The literature frequently explains the fact that there is little ratcheting by “moral” reasons. Bewley (1999) points to motivators “having to do with generosity” (p. 431). MacLeod suggests that it can be explained by the threat of less effort by a worker whose wage is cut. My model gives precise
Second, we had assumed that Low is able to secure $w_H$ forever by being independent. This similarity between Low and High is arguably too extreme. To the extent that Low can only achieve a lower wage, separation between types is easier.

Third, recall that Proposition 5 is valid for the case where other firms care less about loyalty than M. If the opposite is true, Low has less incentive to mimic High. The reason is that signaling high competence now is in fact a bad signal - no one wants to appear overqualified.\textsuperscript{47}

Finally, what if M does not know W’s competence either?\textsuperscript{48} Even in this case, we can show that the same logic applies, i.e., that M may use the sorting possibility presented to her in the form of the loyalty-competence trade-off to her advantage. In other words, understanding previous or current loyalty as optimizing behavior rather than as a matter of character gives a manager a powerful tool to discern high and low levels of competence.

Proposition 5’

Under the conditions of Proposition 5, a separating equilibrium with the properties of the equilibrium in Proposition 5 exists also in the case where M is asymmetrically informed about W’s competence.

Proof. See the Appendix. ■

4.3 Evidence

It is difficult to test the model’s plausible predictions directly, but there are some potential applications. For example, one might view this paper as the first step towards a more meaningful to this claim, effort here being the honoring of a relational contract. It is conceivable to extend the model to practically relevant cases where current competence is commonly known, but the future is not.\textsuperscript{47} By contrast, career concerns models and models like MacLeod and Malcomson (1988) and Kaarboe and Olsen (2003) predict that the more a worker can impress the outside market with his production, the better for him.\textsuperscript{48} Although it has become standard to model M as having some sort of informational advantage over the market, here, the fastest way for the manager to become informed is not to wait for a series of output levels to materialize from which she can (approximately) infer competence, but to look what the worker is doing in the first period.

34
plete theory of job mobility. An appendix available on request speculates about an extended model where there are three types of workers in terms of competence: incompetent workers, average workers, and highly competent workers. The argument needs to be more fully developed in future work, but one result the model is likely to predict is that incompetents will be laid off frequently and will have low or negative wage growth, highly competent workers will switch frequently in the beginning and less as their careers progress and average workers will have the lowest turnover throughout. If competence is valued highly by the market, early job-hoppers will have fast wage growth. If instead loyalty is valued highly, stayers will do better than those who switch frequently, even though – or, in fact, precisely because – they are not seen as so competent.49

This theory matches up with some stylized facts and helps to inform a new interpretation of otherwise puzzling pieces of evidence. The theoretical argument for the wage effects of mobility is usually framed in terms of adverse selection.50 Greenwald (1986) shows that in this case workers who change jobs are marked by being part of an inferior group of competence. Yet this model and several that followed (Lazear 1986, McCormick 1990, Gibbons and Katz 1991, Laing 1994) they do not quite seem to capture all there is. For example, while in 1989, job-hopping may have been regarded as a “fad [of] plain lunacy” (Brown 1989), a decade later, Forbes magazine announced that “[J]ob-hopping is in, loyalty is out” (Moreno 2000). Because of the unquestioned assumption that firms always want to retain the most competent workers, the models cited have a hard time accommodating these changing views. Consider what studies looking at the importance of job history have found.51 Light and McGarry (1998) and Munasinghe and Sigman (2004) find that previous

---

49 For the case where the market values competence, this argument is close in terms of results to Black and Loewenstein (1991). Importantly, though, the model sketched here can accommodate both directions of predictions, a crucial feature in the light of the existing evidence.

50 Models of turnover do not easily yield predictions about wage behavior (see Burdett (1978), Weiss (1984), where wage effects arise from mobility costs, Jovanovic (1979b), and others). Human capital theory predicts a negative relationship between job mobility and investments in job specific skills (Becker 1975, Jovanovic 1979a). Johnson (1978) analyzes job-shopping, but his analysis does not take into account reputation effects for the worker.

51 For the impact of one-time job changes on short-term and long-term wages, Bartel and Borjas (1981)
mobility is negatively related to wages and positively related to future turnover. By contrast, Mincer and Jovanovic (1981) and Marshall and Zarkin (1987) (also without distinguishing between the various reasons for interfirm mobility) find that prior mobility has no effect on wage offers for new hires, but has a (small) positive effect on wage offers for stayers. These studies are potentially problematic because the explanatory variable they employ are total separations. Instead, my model explains why one would expect voluntary and involuntary separations to have different effects. Indeed, Keith (1993) shows that once these two forms of mobility are disaggregated, a history of more frequent quits increases (log) wages, while a history of involuntary mobility decreases wages. However, without explicitly controlling for the importance of loyalty in different occupations or industries – which, of course, poses a measurement challenge – this conflict cannot be decided. This is what future research should do.

5 Competence and loyalty as complements

The tradeoff between competence and loyalty is partially driven by the assumption that the gains from loyalty, \( v \), do not depend on the worker’s competence. In many instances, this assumption is plausible. On the other hand, it is intuitive that where the production function of loyalty is \( v(\theta) \) and this function is sufficiently strongly increasing, it may be more desirable and also more feasible to induce more competent workers to loyalty, because in this case it is more credible for M to promise high loyalty rewards. To begin formalizing this intuition let production of M now be given by \( y(\theta) \), while production of the average firm in the market is for simplicity again assumed to be \( \bar{\alpha}\theta + y(\theta) \). As usual, I begin with the scenario where others do not value loyalty as much or where a one-time deviation from

\(52\) Light and McGarry (1998) claim that the qualitative results remain unchanged when they substitute voluntary separations for all separations.
the loyalty contract results in spot interaction forever. Side opportunities are an increasing function \( b(\theta) \) and the production function of loyalty is an increasing and concave function \( v(\theta) \). Let us denote the optimal solution under spot interaction with \( \theta^S \). Denote that optimal level of competence under loyalty with \( \theta^L \). Again, there are two non-reneging constraints, one for \( M \) and one for \( W \). Using the same logic as before, we see that they can be written as \( \delta x \geq b(\theta) \) for the worker and \( \delta v(\theta) \geq x \) for the manager. Combining these two conditions leads to one condition for loyalty to be feasible. Maximizing \( M \)'s objective leads to \( W \)'s optimal competence level. Proposition 6 summarizes the results.

**Proposition 6** Under the maintained assumptions,

1. Loyalty is feasible if
   \[
   \frac{b(\theta^L)}{v(\theta^L)} \leq \delta^2.
   \]

2. If \( M \) is very unproductive (\( \bar{\alpha} > v'(0) \)), competence and loyalty will be substitutes, i.e., \( \theta^L \leq \theta^S \). Otherwise, competence and loyalty are substitutes if
   \[
   \frac{b''(\theta^L)}{v''(\theta^L)} < \frac{\delta}{1 - \delta} \leq \frac{b''(\theta^L)}{v''(\theta^L) - \bar{\alpha}}.
   \]

**Proof.** See the Appendix. ■

The first part, due to the two non-reneging constraints, is familiar from our earlier analysis. This is a *levels* condition, and it says that gains from loyalty must sufficiently outweigh the private benefits \( M \) foregoes. Thus, no matter whether competence and loyalty end up being substitutes or complements, loyalty is less feasible under non-contractibility than under contractibility, and Proposition 2 continues to hold.

The second part says that the *relative slopes and curvatures* of the gains of loyalty and outside opportunities are also relevant. The crucial condition that can overturn the tradeoff between competence and loyalty into a complementarity is the slopes condition.\(^{53}\) The condition reveals that complementarity may occur because the agents are sufficiently patient.

\(^{53}\) Under our assumptions, a sufficient condition for the curvatures condition to be fulfilled is \( b'' > 0 \): Here, brilliant workers are very hard to induce to loyalty, but somewhat less competent ones are much more easily induced to be loyal.
More interestingly, if \( v'(\theta) > 0 \), unless \( M \) is very unproductive (\( \bar{\alpha} > v'(0) \)), there is an offsetting effect of increasing competence in the sense that since the gains from loyalty increase, \( M \) can now more credibly promise rewards for loyalty.\(^{54}\) Proposition 6 does not provide a cutoff value for loyalty to be preferred over spot interaction. But the intuition of Proposition 3 remains in place: when competence and loyalty are substitutes, even feasible and socially valuable loyalty may be foregone by \( M \) if spot interaction with a highly competent worker is possible. When competence and loyalty are complements, loyalty with rewards will be the dominant equilibrium.

What do these results mean economically? A plausible hypothesis would seem to be that condition (10) is likely to hold within organizations, but is less likely to hold between organizations. Why? The scope of the activities within the firm is usually relatively limited compared to what a worker might do with his competence in the outside world. For example, a more competent computer expert at an internet company\(^{55}\) may be more apt at applying his superior knowledge to the particular tasks at hand. However, his outside possibilities increase much more strongly with his competence. Not only can he engage in other similar programming tasks. Since knowing how to program in a computer language is a universally applicable tool, even this relatively specific skill makes outside opportunities grow much faster with competence than inside contributions do. The expert can also do general consulting about the software business. He can also inform other firms about insights he has gained about the process at his current firm. More generally, the theory predicts that companies are more likely to engage the services of the most competent suppliers when the supplier has a narrowly defined business, but are more likely to sacrifice competence in relationships with general suppliers.\(^{56}\)

\(^{54}\)Note that productivity in output alone never is enough for complementarity, because it does not affect the feasibility of loyalty. However, if loyalty is feasible, \( \bar{\alpha} < 0 \) pushes in the direction of higher competence.

\(^{55}\)I thank Steve Case, former chairman of AOL, for sharing his insights in a personal conversation in Chicago, June 12, 2004.

\(^{56}\)Note that this is a different source of differences of contracts within firms and between than was emphasized by Baker et al. (2002).
6 Conclusion

Barnard (1938) argued that the “willingness of persons to contribute efforts to the cooperative system is indispensable [...] loyalty [...] is vaguely recognized as an essential condition of organization” (pp. 83-84). This paper has made this notion more precise and has presented a simple theory of loyalty as an optimal behavior. The central thesis of this paper is that in a wide range of plausible circumstances, more productive economic agents will behave less loyally, i.e., will be less willing to bear personal costs for somebody else. The logic applied to derive this result is straightforward: A “manager” decides whether to reward loyalty or not, and a “worker” decides whether to be loyal or act independently. Any loyalty with rewards equilibrium must be sustained by self-enforcing contracts. Higher inside productivity brings with it higher outside productivity as well. Consequently, any threat of punishment the manager might make does not worry competent workers as much. The first result then is that loyalty may not be feasible. The stronger and surprising result is that even when loyalty is feasible and has positive social value, it may be dominated by spot interaction with more competent individuals. Moreover, loyalty can be a signal of low competence.

The theory allows us to better understand some otherwise puzzling facts of life within and between organizations. For example, it provides one explanation for why systems of patronage and nepotism may be more (second-best) efficient than they appear at first sight. It also provides guidance for managerial action. We have seen, for example, that loyalty of highly competent individuals is possible precisely where disloyalty would be costly for the organization. All the insights become only possible by analyzing loyalty as an optimal behavior.

Although the basic thrust of the argument is expected to hold in many richer settings, the analysis of course has its limitations. It is an open question whether ultimately behavior-based or character-based explanations of loyalty capture the most important part of differences in loyalty. Experiments are likely to be the most promising way to make progress on this question. The value of the present work in that context will be that it pinpoints
the relevant variables that allow the researcher to design the experimental situation. The empirical predictions concerning the relationship between optimal competence and the value of loyalty, side opportunities, and the time horizon can in principle be tested in the real world. Finding appropriate data is a challenge. Operationalizing the value of loyalty, for example, is not easy, but a reasonable starting point may be data on previous performance of managers (who may value loyalty when they are in a weak position) as well as measures of the power of the individual supposed to serve loyally. Collecting data from an individual firm (both by direct observation of actions and from surveys) is likely to be informative for such tests. This has to await future research.

Several theoretical extensions also deserve further inquiry. First, since the personal costs of loyalty may consist in not being loyal to somebody else, two managers may recognize that their desires for loyalty compete against each other. The resulting model of common agency with self-enforcing contracts is likely to provide new predictions for the “servant of two masters” problem in matrix organizations, for example. Second, an analysis of delayed rewards rather than ongoing loyalty rewards – in which case much hinges on the manager’s reputation – is expected to bring out yet another factor by which organizations differ. In all of these important extensions, the fundamental effects the analysis in this paper has highlighted are expected to remain the driving forces.

The analysis centered on loyalty of a worker to a manager as the behavior that is sustained by the relational contract. However, the results apply generally to the feasibility and desirability of relational contracts and therefore hopefully also inform future research on partner selection in other cooperation games. For example, the classic analogy between the analysis of labor markets and that of marriage (see Becker, Landes and Michael (1977)) applies here as well. The theory predicts that individuals who are, objectively or subjectively, not attractive to partners of the opposite (or, for some, the same) sex will tend to pay greater attention to being nice and will tend to be more willing to bear personal costs than would beauty pageant winners and millionaires (if we want to adopt a very narrow view of “competence”). Similarly, richer and better educated individuals are less likely to participate in churches or sects, activities that demand a high degree of loyalty to the organization
Importantly, the theoretical and empirical implications this paper has derived arise only from the introduction of a positive correlation between inside and outside productivities into a framework of self-enforcing contracts; they are neither a feature of a static multi-tasking world (where a correlation between inside and outside productivities was seen to hold no predictions for the preferred competence levels) nor of the standard relational contract framework (where the only prediction of existing theories is that contracts are most easily sustained with individuals who have low reservation utilities). This observation suggests that the class of models the model in this paper is representative of should play a greater role in the analysis of settings where both production and relations matter for organizational success.

7 Appendix

Proposition 1

When loyalty is contractible and has positive social (joint) value, it is preferred to spot interaction.

Proof of Proposition 1

When M is relatively productive, then the maximal payoff under both spot interaction and loyalty is obtained with $\theta = 1$. Thus, loyalty is preferred if $v > b$. If M is relatively unproductive then the maximal payoff under loyalty is obtained with $\theta = 0$ and the maximal payoff under spot interaction is still obtained with $\theta = 1$. With the optimal competence levels, loyalty is therefore preferred if $v > b - \bar{\alpha}$.\footnote{The fact that loyalty is implemented even when $v < b$ as long as $v > b - \bar{\alpha}$ should not be interpreted as an inefficiency. This is solely due to technological preferences for high- or low-competence workers.} Combining the two cases, we have the result. $\blacksquare$

Proposition 2

(Feasibility of loyalty with rewards) Non-contractible loyalty may not be feasible, even when loyalty would be desired if it were contractible.
Proof of Proposition 2

Because \( \theta^{L*} = 1 - \frac{b - \delta^2 v}{\delta} \), when \( b - \delta^2 v \geq 0 \), we cannot have \( \theta^{L*} = 1 \), but loyalty is feasible only for competence levels less than unity. To see that the area of non-feasibility extends into the area where loyalty would be desired if it were contractible, begin with the case where \( M \) is relatively productive (\( \bar{\alpha} < 0 \)). Here, loyalty is desired if it is contractible if \( v > b \). If loyalty is non-contractible, however, for any production technology, it is immediate to see that there exist triplets \( \{b', v', \theta'\} \) such that \( v > b \), but \( v < \frac{b}{\delta} \) and therefore \( \theta' > 1 - \frac{v' - \delta^2 v'}{\delta b'} \).

Next, consider the case where \( M \) is relatively unproductive (\( \bar{\alpha} > 0 \)). Here, we just need to show that there exists a triplet \( \{b_0, v_0, \theta_0\} \) such that \( v_0 > b_0 - \bar{\alpha} \), but \( \theta_0 > 1 - \frac{b_0 - \delta^2 v_0}{\delta b_0} \). This conclusion follows a fortiori from the first case.

Proposition 3

(Desirability of loyalty with rewards) Even when loyalty is feasible and has positive net value, the manager may prefer spot interaction with a more competent worker.

Proof of Proposition 3

Consider first the case where \( M \) is relatively productive. In this case, it is optimal for \( M \) to choose the maximal competence level compatible with loyalty, \( \theta^{L*} = 1 - \frac{b - \delta^2 v}{\delta} \). The payoff for \( M \) then is

\[
U^L_M(\theta^{L*}) = E_{\theta^{L*}y} - w(\theta^{L*}) + v - x(\theta^{L*}) = E_{\theta^{L*}y} - \bar{E}_{\theta^{L*}y} - \theta^{L*}b + v - \theta^{L*}b - b \frac{1 - \delta}{\delta} = v - b \frac{1 - \delta}{\delta} - \theta^{L*} \bar{\alpha}.
\]

Plugging in \( \theta^{L*} = 1 - \frac{b - \delta^2 v}{\delta} = \frac{v}{\delta} - \frac{1 - \delta}{\delta} \), we get

\[
U^L_M(\theta^{L*}) = v - b \frac{1 - \delta}{\delta} - \frac{\bar{\alpha} v}{\delta} + \frac{1 - \delta}{\delta} \bar{\alpha} = v \left( 1 - \frac{\bar{\alpha} \delta}{b} \right) - \left( b - \bar{\alpha} \right) \frac{1 - \delta}{\delta} = v \left( 1 - \frac{\bar{\alpha} \delta}{b} \right) - \frac{b - \bar{\alpha}}{\delta} + b - \bar{\alpha}.
\]

Now we need to compare this payoff to the one obtainable under spot interaction. There, we know that \( \theta = 1 \) is optimal, leading to a payoff of

\[
U^S_M(\theta = 1) = b - \bar{\alpha}.
\]

Therefore,

\[
U^L_M(\theta^{L*}) > U^S_M(\theta = 1) \iff v \left( 1 - \frac{\bar{\alpha} \delta}{b} \right) > \frac{b - \bar{\alpha}}{\delta}
\]
or

\[ v > b \frac{b - \bar{\alpha}}{\delta b - \delta^2 \bar{\alpha}} = \hat{v} (b, \bar{\alpha}, \delta). \]

This is the term used in Figure 1.3. Note that for the case of a relatively productive M, we can easily verify that \( \hat{v} (b, \bar{\alpha}, \delta) > b \), i.e., that \( \frac{b - \bar{\alpha}}{\delta b - \delta^2 \bar{\alpha}} > 1 \) for \( \bar{\alpha} < 0.58 \)

To derive the conclusion in the proposition, it is helpful to restate this calculation somewhat. In particular, rewrite the condition that loyalty is preferred to spot interaction as

\[ v - b \left( \frac{1 - \delta}{\delta} - \theta^{L*} \bar{\alpha} \right) > b - \bar{\alpha} \]

\[ v - b \frac{1 - \delta}{\delta} - \theta^{L*} b > b - \theta^{L*} b + \theta^{L*} \bar{\alpha} \]

\[ v - x (\theta^{L*}) > (b - \bar{\alpha}) (1 - \theta^{L*}) = \frac{(b - \bar{\alpha}) (b - \delta^2 v)}{\delta b} > 0 \]

where the inequality follows from the premise \( b - \delta^2 v > 0 \). Thus, loyalty needs to have strictly positive net value; merely surpassing zero by some small amount \( \varepsilon \) is not enough.

In the case of a relatively unproductive M, the best the manager can do when she wants to implement loyalty is to hire an incompetent individual; in this case, her payoff is \( U^{L*}_M (\theta = 0) = v - b \frac{1 - \delta}{\delta} \). For spot interaction, we know from Proposition 1 that \( \theta = 1 \) is still optimal, leading to a payoff of \( U^{S*}_M (\theta = 1) = b - \bar{\alpha} \). Thus, loyalty is preferred to spot if

\[ v - b \frac{1 - \delta}{\delta} = v - x (\theta = 0) > b - \bar{\alpha} > 0 \]

or

\[ v > b \frac{1}{\delta} - \bar{\alpha}. \]

In other words, loyalty again needs to have strictly positive net value in order to be preferred.

\[ \blacksquare \]

**Corollary 2**

Non-contractible loyalty becomes less preferred when M’s productivity or W’s outside opportunities increase, and when patience or M’s gains from loyalty decrease.

**Proof of Corollary 2**

\[ \frac{58}{58} \text{We can also check that as } \delta \text{ goes to unity, the area where } \theta^{L*} < 1 \text{ is required vanishes and gets } \text{“squeezed out” between the } v = b/\delta^2 \text{ and } \hat{v} (b, \bar{\alpha}, \delta) \text{ schedules. Another way to look at the } \hat{v} (b, \bar{\alpha}, \delta) \text{ schedule is to ask what happens to competence as we move up the graph and to the right along } \hat{v} (b, \bar{\alpha}, \delta). \text{ It is easy to check that } \theta^{L*} \text{ decreases, as indicated by the arrows along } \hat{v} (b, \bar{\alpha}, \delta) \text{ in the Figure. Along } \hat{v} (b, \bar{\alpha}, \delta), \text{ as } b \text{ goes to infinity, } \theta^{L*} \text{ approaches } 2 - \frac{1}{\varepsilon}. \]
We need to check whether the cutoff value that \( v \) needs to surpass in order for loyalty to be preferred is decreasing in \( \bar{\alpha} \). Within the range \( \bar{\alpha} < 0 \), it is easy to verify that

\[
\frac{\partial}{\partial \bar{\alpha}} \left( \frac{b - \bar{\alpha}}{b - \bar{\alpha} - \delta b} \right) = \frac{\delta b^2 (\delta - 1)}{(b - \bar{\alpha} - \delta b) ^2} < 0.
\]

As \( \bar{\alpha} \) turns positive, we must check that

\[
\frac{b - \bar{\alpha}}{b - \bar{\alpha} - \delta b} > \frac{b}{\bar{\alpha}'} - \bar{\alpha} \quad \text{for} \quad \bar{\alpha} < 0 \quad \text{and} \quad \bar{\alpha}' > 0.
\]

That this condition holds is easily verified. Also, \( \frac{b}{\bar{\alpha}} - \bar{\alpha} \) is decreasing in \( \bar{\alpha} \) in the range where \( \bar{\alpha} > 0 \). Overall, we obtain the negative correlation between the attractiveness of non-contractible loyalty and M’s productivity claimed in the Corollary. The other claims are similarly checked. □

**Proposition 4**

Assume that when W leaves M, he is matched randomly to a firm with expected relative productivity \( \bar{\alpha} \) and expected gains from loyalty \( \bar{v} \). Then

1. If M values loyalty more than the average firm in the market \( (v > \bar{v}) \) or if all firms are willing to engage only in spot interaction with a W who has reneged, the results are exactly the same as if self-employment is the only alternative employment opportunity.
2. If M is relatively productive and values loyalty less than the average firm in the market \( (v < \bar{v} \quad \text{and} \quad \bar{\alpha} < 0) \), non-contractible loyalty is less attractive than in the case where the market does not value loyalty.

**Proof of Proposition 4**

Assume that benefits from loyalty at firm \( i \), denoted \( \tilde{v}_i \), are distributed between \( v_{\text{min}} \) and \( v_{\text{max}} \) with mean \( \bar{v} \). Part (1) was discussed in the text. Recall that the reason why this result holds is that the maximum level of competence at a firm with a valuation of loyalty \( v, \theta L^*(v, \cdot) \), is greater than the maximum level of competence compatible with loyalty at a firm with a valuation of loyalty \( \bar{v} < v, \theta L^*(\bar{v}, \cdot) \). Thus, there are no additional costs in obtaining loyalty from \( \theta L^*(v, \cdot) \), which vindicates M’s decision to hire him. For part (2), the text already showed that loyalty is now more expensive under non-contractibility because M has to pay higher wages in expectation just to keep W around. (Note that there is not an infinitely-growing surplus for W. This is because when W expects to leave M – who has a value of loyalty \( v \) – for another M – who is expected to have a value of loyalty \( \bar{v} \) – he expects to be with M’s of loyalty \( \bar{v} \) forever, even if he does switch employers.) However, it could be that under contractible loyalty, these costs increase similarly, thus eliminating any differential in attractiveness. In fact, however, under contractible loyalty, it does not matter whether other firms also care about loyalty. To see this, consider a given \( \theta \) and assume that loyalty is contractible. Firm \( i \) will induce a worker of competence \( \theta \) to loyalty if and only if \( v_i > \theta b \). In other words, if W leaves M, he can expect to be in a loyalty regime if \( \bar{v} > \theta b \) and he can expect to be in a spot regime otherwise. In the first case, he will not reap the benefits of side opportunities, but will get paid \( \theta b \) in expected loyalty rewards. Thus, since
reservation utility is given by \( u(\theta) = \bar{E}y \), his wage must be \( \bar{E}y - \theta b \). But this is exactly the same wage he would expect to receive if he lived in a spot regime later. Therefore, since there is no problem of contracting on loyalty, \( W \) gets the same reservation utility no matter whether he expects to be loyal or independent at his next employer. Ultimately, this is not surprising - after all, a surplus over \( W \)'s reservation utility is only necessary when there is a contracting problem – but it shows that an \( M \) who wants to induce loyalty has the very same maximization problem as before, namely,

\[
\begin{align*}
\max_{\theta} U_M &= y + \theta \Delta y + v - (w + \theta b) \\
s.t. U_W &= w + \theta b \geq \bar{u}(\theta).
\end{align*}
\]

Thus, again we find that the optimal competence levels are zero (if \( \bar{\alpha} > 0 \)) or unity (if \( \bar{\alpha} < 0 \)). Therefore, for the optimal solution, in contractible loyalty it does not matter whether others also care about loyalty or not.

\[\blacksquare\]

**Proposition 5**

Consider a game \( \Gamma(\delta, b, v, Y, \{\theta_H, \theta_L\}) \) with an opportunity of value \( b \) appearing for certain at the beginning of the game. Suppose that the market values loyalty less than \( M \), that \( \bar{\alpha} = 0 \), that \( \bar{y} > b \), and that \( \theta_H - \theta_L > \frac{1 - \delta}{\delta} \).

If \( \theta_L \)'s loyalty has positive net value under perfect information (i.e., \( v - \bar{x}(\theta_L) > 0 \)) and \( \theta_H \)'s loyalty has negative net value under perfect information (i.e., \( v - \bar{x}(\theta_H) < 0 \)) then there exists a separating perfect Bayesian equilibrium that is characterized by the following features:

1) The market believes that \( \theta_H \) is independent and that \( \theta_L \) is loyal. The two types behave accordingly.

2) The loyalty rewards \( x \) and wages \( w_H, w_L \) satisfy the following conditions:

a) High is at least as well off as Low. That is: \( w_H + \theta_H b \geq w_L + x \).

b) Low does not want to be independent. That is: \( w_L + x > w_H + \theta_L b \).

c) High exactly receives his reservation utility. That is: \( w_H + \theta_H b = \bar{u}(\theta_H) \).

\[d)\] Low earns a surplus over his reservation utility. That is: \( w_L + x > \bar{u}(\theta_L) \).

3) The manager never does better than if the market has symmetric information.

**Proof of Proposition 5**

To provide a better overview, the proof is divided into several steps.
Optimality of Low’s strategy. Suppose that the market has the beliefs as stated, i.e., that \( \theta_L \) will be loyal and \( \theta_H \) will be independent. Then, we require two similar but conceptually distinct conditions for Low. First, we require him to be loyal rather than disloyal. By being loyal, W knows that the market will believe him to be \( \theta_L \), and, consistent with equilibrium, it believes that it is optimal for M to consequently offer wage \( w_L \) and loyalty rewards \( x \). By contrast, independence allows the worker to convince the market that he is \( \theta_H \). Recall from the text that we assume that even though M knows that the worker is the low type, in order to keep him inside the firm she would need to offer a wage \( w_H \) forever.

Low is not really a high type, though. Therefore, in spot interaction, he will expect to earn \( \theta_L b \) in side opportunities. In other words, Low prefers loyalty to independence if and only if

\[
w_L + x + \frac{\delta}{1 - \delta} (w_L + x) \geq w_H + b + \frac{\delta}{1 - \delta} (w_H + \theta_L b).
\] (IC-L)

Second, Low must not have an incentive to renege, in addition to “just” mimicing. Using the standard logic, we have

\[
w_L + x + \frac{\delta}{1 - \delta} (w_L + x) \geq w_L + b + x + \frac{\delta}{1 - \delta} (w_H + \theta_L b).
\] (NR-L)

Thus, a Low type knows that reneging on loyalty (or choosing independence over loyalty) has two effects: On the one hand, he is able to secure a higher wage. On the other hand, he loses loyalty rewards \( x \). Both weigh more for incompetent guys, so it is a priori unclear which effect dominates. For separation to become possible we need to find values of \( x \) and the wages that are consistent with M’s optimizing behavior.

To do this, we begin by noting that an immediate implication of NR-L is \( w_L + x \geq w_H + \theta_L b \), since \( \frac{1 - \delta}{\delta} > 0 \). This proves part 2b. Moreover, we have the following relationship between Low’s two constraints.

**Lemma 5.1**

If Low does not renege on loyalty, he chooses loyalty from the beginning. That is, NR-L implies IC-L.

**Proof of Lemma 5.1**

46
To see this, rewrite NR-L as

\[ x \geq (w_H - w_L) + \theta_L b + b \frac{1 - \delta}{\delta}. \]

Therefore, \( w_L + x > w_H \) is guaranteed, which is a sufficient condition for IC-L to be implied by NR-L. The intuition is clear: If Low does not even have an incentive to renege, he surely does not have an incentive to choose independence from the beginning.

Optimality of High’s strategy. High needs to prefer independence in this candidate equilibrium. But in fact M can always easily make that constraint hold. She can just never reward loyalty by High. In other words, she can just set rewards High gets for loyalty very small or negative. If instead IC-H does have to hold, and the rewards to loyalty are constrained to be the same for all types, a separating equilibrium may fail to exist. See Proposition 5’ below.

Optimality of M’s strategy, part 1. M’s behavior also needs to obey a non-reneging constraint of the usual form. The slight difference to the constraint in the case where the market knows the worker’s competence is that once M has reneged on her promise of rewards for loyalty and W is never loyal again, the market will believe that W is \( \theta_H \), thus forcing M to pay \( w_H \). Therefore, M’s non-reneging constraint is

\[ y + v - x - w_L + \frac{\delta}{1 - \delta} [Ey + v - x - w_L] \geq y + v - w_L + \frac{\delta}{1 - \delta} [Ey - w_H] \quad \text{(NR-M)} \]

Rewriting the constraint yields

\[ x \leq \delta (w_H - w_L) + \delta v. \]

Thus, as before, the two non-reneging constraints NR-L and NR-M impose bounds on loyalty rewards. The lower bound is due to the W’s demands for rewards; the upper bound is due to the lack of full commitment by M.

Utility levels. Intuitively, separation is meaningful if we can induce High and Low to act according to the market’s belief (and according to what M found optimal under perfect information) but still retain High’s utility advantage. Indeed, we can show parts 2a and 2c of the Proposition in the following

**Lemma 5.2**
High earns exactly his reservation utility and is better off than Low. That is, \( w_H + \theta_H b \geq w_L + x \).

**Proof of Lemma 5.2**

First, we can note that the wage M will pay High will be such that together with his side opportunities, he exactly reaches his reservation utility. There would be no point in paying more (nor less, because High would leave M). Thus, we know that

\[
W_H = \bar{u} (\theta_H) - \theta_H b
\]

which is part 2c of Proposition 5.

Second, making use of NR-L, we can substitute for the right side of the condition to be shown in order to have

\[
w_H + \theta_H b \geq w_H + \theta_L b + b \frac{1 - \delta}{\delta}.
\]

Rewriting, we get

\[
\left( \theta_H - \theta_L - \frac{1 - \delta}{\delta} \right) b > 0
\]

which always holds under our restrictions on parameters. Intuitively, what this condition says is that the types we want to separate must not be “too close”, relative to a measure of impatience. ■

Moreover, we can show that Low earns a surplus over his reservation wage (part 2d of Proposition 5).

**Lemma 5.3**

Low earns a surplus over his reservation wage: \( w_L + x \geq \bar{u} (\theta_L) \).

**Proof of Lemma 5.3**

To show this, we combine Low’s NR with the insight that High earns exactly his reservation wage. Formally,

\[
w_L + x \geq w_H + \theta_L b + b \frac{1 - \delta}{\delta} = \bar{u} (\theta_H) - \theta_H b + \theta_L b + b \frac{1 - \delta}{\delta}.
\]

This term is greater than \( \bar{u} (\theta_L) = \theta_L \bar{y} + (1 - \theta_L) \bar{y} \), if and only if

\[
(\theta_H - \theta_L) (\bar{y} - b) + b \frac{1 - \delta}{\delta} > 0.
\]

The assumed parameter restriction \( \bar{y} - b > 0 \) is therefore a sufficient condition for the Lemma to hold. ■
Optimality of M’s strategy, part 2. Finally, we need to check whether M indeed “wants” \( \theta_L \) to be loyal and \( \theta_H \) to be disloyal, i.e., whether W’s behavior and the market’s beliefs are consistent with optimizing behavior on M’s part. Employing the fact that \( w_H = y + \theta_H \Delta y - \theta_H b \), we know that the low type, when he pretends to be \( \theta_H \), actually receives less utility than a true high type, because he only realizes opportunities of expected value \( \theta_L b \). The payoffs to M are easily determined as follows:

\[
U_M (\theta_L \text{ is loyal}) = v + b \left( \theta_H - \theta_L - \frac{1 - \delta}{\delta} \right) - (\theta_H - \theta_L) \Delta y
\]

\[
U_M (\theta_L \text{ is independent}) = \theta_H b - (\theta_H - \theta_L) \Delta y
\]

Thus, loyalty is preferred for the low type if and only if

\[
v > b \left( \theta_L + \frac{1 - \delta}{\delta} \right) = x^P (\theta_L).
\]

Similarly,

\[
U_M (\theta_H \text{ is independent}) = \theta_H b
\]

\[
U_M (\theta_H \text{ is loyal}) = v - b \frac{1 - \delta}{\delta}.
\]

Thus, M wants the high type not to be loyal if and only if

\[
v < b \left( \theta_H + \frac{1 - \delta}{\delta} \right) = x^P (\theta_H).
\]

But these two conditions are precisely what it means for High’s loyalty not to be worth it under perfect information, but for Low’s loyalty to be worth it. In other words, if and only if loyalty pays under symmetric information, M is also able to find a way to separate High and Low in behavior when the market sends incentives to Low to mirror High. M’s ability to do this derives endogenously from the tradeoff between competence and loyalty. However, M incurs extra costs in doing this to the extent that Low, by acting independently, expects to be able to receive \( w_H \) for an extended period of time.

This completes the proof of Proposition 5. ■

Consider now the case where M does not know W’s competence either. Even in this case, M may use the sorting possibility presented to him in form of the competence-loyalty
tradeoff to her advantage. One might think that separation now becomes more difficult. After all, recall that in the case so far we were able to avoid High envying Low because M could simply give very negative rewards for loyalty to High. But here, if M has to offer two packages of compensation – one with and one without loyalty rewards – High may envy Low. Intuitively, the two packages must be sufficiently different for separation to be possible, but if the types are too close to each other, this might not be possible. However, Proposition 5’ demonstrates that in fact the conditions for Proposition 5 are enough to allow us to obtain separation even in this case.

**Proposition 5’**

Under the conditions of Proposition 5, a separating equilibrium with the properties of the equilibrium in Proposition 5 exists also in the case where M is asymmetrically informed about W’s competence.

**Proof of Proposition 5’**

Similar conditions as in Proposition 5 need to hold for Low’s and M’s behavior. In addition, now High needs to prefer independence in this candidate equilibrium. For High, there is no non-reneging constraint, because there is no contract to renege on. His incentive-compatibility constraint is

\[
w_H + b + \frac{\delta}{1 - \delta} (w_H + \theta_H b) \geq w_L + x + \frac{\delta}{1 - \delta} (w_L + x)
\]

or

\[
x \leq (w_H - w_L) + b (1 - \delta) + \delta \theta_H b.
\]

Intuitively, if loyalty rewards are too attractive, even high types choose loyalty over spot interaction. A necessary condition for both NR-L and IC-H to hold is that

\[
\delta \theta_H - \theta_L \geq \delta + \frac{1}{\delta} - 2 \geq 0
\]

This condition can be written as

\[
\theta_H > \left( \frac{1}{\delta} - 1 + \theta_L \right) - (1 - \delta - \theta_H + \delta \theta_H).
\]

The term \((1 - \delta - \theta_H + \delta \theta_H)\) is always positive. Therefore, if \(\theta_H > \left( \frac{1}{\delta} - 1 + \theta_L \right)\), the premise of Proposition 5, separation is also possible in the case where M does not know the worker’s competence. ■

**Proposition 6**

50
Under the maintained assumptions,

1. Loyalty is feasible if
   \[ \frac{b(\theta_L)}{v'(\theta_L)} \leq \delta^2. \]  (9)

2. If M is very unproductive \((\bar{\alpha} > v'(0))\), competence and loyalty are always substitutes, i.e., \(\theta_L \leq \theta^S\). Otherwise, competence and loyalty are substitutes if
   \[ \frac{b'(\theta_L)}{v''(\theta_L)} \leq \frac{\delta}{1 - \delta} \leq \frac{b'(\theta_L)}{v'(\theta_L) - \bar{\alpha}}. \]  (10)

Proof of Proposition 6

Part (1) follows directly from combining the two constraints in the text.

For part (2), note that the manager who wants to implement the spot interaction equilibrium solves

\[ \max_{\theta} -\bar{\alpha}\theta + b(\theta) \]

and thus sets \(\theta\) such that \(b'(\theta) = \bar{\alpha}\). As \(b\) is increasing, this implies that productive managers always choose \(\theta = 1\).

If instead, the manager wants to implement loyalty (which for the moment we assume to be feasible), she solves

\[ \max_{\theta} -\bar{\alpha}\theta + b(\theta) + v(\theta) - \frac{b(\theta)}{\delta} \]

If this maximand is concave, and the first derivative of this term is negative for the solution from spot interaction, we know that competence under loyalty will be chosen to be smaller than under spot interaction. First note that a highly positive value of \(\bar{\alpha}\), namely, \(\bar{\alpha} > v'(0)\), implies that the tradeoff between competence and loyalty is guaranteed to hold. If instead \(\bar{\alpha} < v'(0)\), the derivative is nonpositive if, for

\[ -\bar{\alpha} + b' + v' - \frac{b'}{\delta} \leq 0 \]

or

\[ \frac{\delta}{1 - \delta} \leq \frac{b'}{v' 1 - \frac{b'}{\delta}} \]

and the objective function is concave if

\[ \frac{\delta}{1 - \delta} > \frac{b''}{v''} \]

Combining these two conditions yields the claim in the Proposition. ■
References


