MECS 475: Organizational Economics

Michael Powell

June, 2016

Contents

[1] This chapter was written by Dan Barron.

Disclaimer

These lecture notes were written for a second-year Ph.D. course in Organizational Economics. They are a work in progress and may therefore contain errors or misunderstandings. Any comments or suggestions would be greatly appreciated.

Introduction

Neoclassical economics traditionally viewed a firm as a production set—a collection of feasible input and output vectors. Given market prices, the firm chooses a set of inputs to buy, turns them into outputs, and then sells those outputs on the market in order to maximize profits. This "black box" view of the firm captures many important aspects of what a firm does: a firm transforms inputs into outputs, it behaves optimally, and it responds to market prices. And for many of the issues that economists were focused on in the past (such as: what is a competitive equilibrium? do competitive equilibria exist? is there more than one? who gets what in a competitive equilibrium?), this was perhaps the ideal level of abstraction.

But this view is inadequate as a descriptive matter (what do managers do? why do firms often appear dysfunctional?), and it leads to the following result:

Representative Firm Theorem (Acemoglu, 2008) Let $\mathcal{F}$ be a countable set of firms, each with a convex production-possibilities set $Y^f \subset \mathbb{R}^N$. Let $p \in \mathbb{R}_+^N$ be the price vector in this economy, and denote the profit-maximizing net supplies of firm $f \in \mathcal{F}$ by $\hat{Y}^f(p) \subset Y^f$. Then there exists a representative firm with production possibilities set $Y \subset \mathbb{R}^N$ and set of profit-maximizing net supplies $\hat{Y}(p)$ such that for any $p \in \mathbb{R}^N_+$, $\hat{y} \in \hat{Y}(p)$ if and only if $\hat{y}(p) = \sum_{f \in \mathcal{F}} \hat{y}^f$ for some $\hat{y}^f \in \hat{Y}^f(p)$ for each $f \in \mathcal{F}$.

That is, by abstracting from many of the interesting and complex things that happen within firms, we are also left with a simplistic perspective of the production side of the economy as a whole—in particular, we can think of the entire production side as a single (price-taking) firm. This view therefore is also inadequate as a model of firm behavior for many of the questions economists are currently interested in (why do inefficient and efficient firms coexist? should we care about their coexistence? when two firms merge, should this be viewed as a bad thing?).

The purpose of this course is to move beyond the Neoclassical view of the firm and to provide you with a set of models that you can use as a first step when thinking about contemporary economic issues. In doing so, we will recognize the fact that organizations consist of many individuals who almost always have conflicting objectives, and we will see that these conflicting objectives can result in production sets that are determined as equilibrium objects rather than as exogenously specified sets of technological constraints. In the first part of the course, we will think about how these incentive issues affect the set Y^f. That is, given what is technologically feasible, how do different sources of contracting frictions (limits on monetary transfers or transfers of control) affect what is actually feasible and what firms will actually do?

In the second part of the course, we will study theories of the boundary of the firm. We will revisit the representative-firm theorem and ask under what conditions is there a difference between treating two firms, say firm 1 and firm 2, separately or as a single firm. If we denote the characteristics of the environment as θ, and we look at the following object:

$$\Delta(\theta) = \max_{y \in Y^1 + Y^2} \pi(y) - \left[\max_{y^1 \in Y^1} \pi_1(y_1) + \max_{y^2 \in Y^2} \pi_2(y_2) \right],$$

we will ask when it is the case that $\Delta(\theta) \geq 0$ or $\Delta(\theta) \leq 0$. The representative-firm theorem shows that under some conditions, $\Delta(\theta) = 0$. Theories of firm boundaries based solely on technological factors necessarily run into what Oliver Williamson refers to as the "selective intervention puzzle"—why can't a single large firm do whatever a collection of two small firms could do and more (by internalizing whatever externalities these two small firms impose on each other)? That is, shouldn't it always be the case that $\Delta(\theta) \geq 0$? And theories of the firm based solely on the idea that "large organizations suffer from costs of bureaucracy" have to contend with the equally puzzling question—why can't two small firms contractually internalize whatever externalities they impose on each other and remain separate, thereby avoiding bureaucracy costs? That is, shouldn't it be the case that $\Delta(\theta) \leq 0$?

We will then focus on the following widespread phenomenon. If we take any two firms i and j, we almost always see that $\pi_i^* > \pi_j^*$. Some firms are just more productive than others. This is true even within narrowly defined industries, and it is true not just at a point in time, but over time as well—the same firms that outperform their competitors today are also likely to outperform their competitors tomorrow. Understanding the underlying source of profitability is essentially the fundamental question of strategy, so we will spend some time on this question. Economists outside of strategy have also recently started to focus on the implications of these performance differences and have pointed to a number of mechanisms under which (essentially) $\pi_i^* > \pi_j^*$ implies that $\pi_i^* + \pi_j^* < \max_{y \in Y^i + Y^j} \pi(y)$. That is, it may be the case that performance differences are indicative of misallocation of resources across different productive units within an economy, and there is some evidence that this may be especially true in developing countries. The idea that resources may be misallocated in equilibrium has mouth-watering implications, since it suggests that it may be possible to improve living standards for people in a country simply by shifting around existing resources.

Because the literature has in no way settled on a “correct” model of the firm (for reasons that will become clear as the course progresses), much of our emphasis will be on understanding the individual elements that go into these models and the "art" of combining these elements together to create new insights. This will, I hope, provide you with an applied-theoretic tool kit that will be useful both for studying new phenomena within organizations as well as for studying issues in other fields. As such, the course will be primarily theoretical. But in the world of applied theory, a model is only as good as its empirical implications, so we will also spend time confronting evidence both to see how our models stack up to the data and to get a sense for what features of reality our models do poorly at explaining.

Part I

Internal Organization

Chapter 1

Formal and Informal Incentives

In order to move away from the Neoclassical view of a firm as a single individual pursuing a single objective, different strands of the literature have proposed different approaches. The first is what is now known as "team theory" (going back to the 1972 work of Marschak and Radner). Team-theoretic models focus on issues that arise when all members of an organization have the same preferences—these models typically impose constraints on information transmission between individuals and information processing by individuals and look at questions of task and attention allocation.

The alternative approach, which we will focus on in the majority of the course, asserts that different individuals within the organization have different preferences (that is, "People (i.e., individuals) have goals; collectivities of people do not." (Cyert and March, 1963: 30)) and explores the implications that these conflicts of interest have for firm behavior. In turn, this approach examines how limits to formal contracting restrict a firm's ability to resolve these conflicts of interest and how unresolved conflicts of interest determine how decisions are made. We will talk about several different sources of limits to formal contracts and the trade-offs they entail.

We will then think about how to motivate individuals in environments where formal contracts are either unavailable or they are so incomplete that they are of little use. Individuals can be motivated out of a desire to convince "the market" that they are intrinsically productive in the hopes that doing so will attract favorable outside offers in the future—that is, they are motivated by their own career concerns. Additionally, individuals may form long-term attachments with an organization. In such long-term relationships, goodwill can arise as an equilibrium phenomenon, and fear of shattering this goodwill can motivate individuals to perform well and to reward good performance.

1.1 Formal Incentive Contracts

We will look at several different sources of frictions that prevent individuals from writing contracts with each other that induce the same patterns of behavior they would choose if they were all acting as a single individual receiving all the payoffs. The first will be familiar from core microeconomics—individual actions chosen by an agent are not observed but determine the distribution of a verifiable performance measure. The agent is risk-averse, so writing a high-powered contract on that noisy performance measure subjects him to costly risk. As a result, there is a trade-off between incentive provision (and therefore the agent's effort choice) and inefficient risk allocation. This is the famous **risk–incentives trade-off**.

The second contracting friction that might arise is that an agent is either liquidity-constrained or is subject to a limited-liability constraint. As a result, the principal is unable to extract all the surplus the agent generates and must therefore provide the agent with **incentive rents** in order to motivate him. That is, offering the agent a higher-powered contract induces him to exert more effort and therefore increases the total size of the pie, but it also leaves the agent with a larger share of that pie. The principal then, in choosing a contract, chooses one that trades off the creation of surplus with her ability to extract that surplus. This is the **motivation–rent extraction trade-off.**

The third contracting friction that might arise is that the principal's objective simply cannot be written into a formal contract. Instead, the principal has to rely on imperfectly aligned performance measures. Increasing the strength of a formal contract that is based on imperfectly aligned performance measures may increase the agent's efforts toward the principal's objectives, but it may also motivate the agent to exert costly effort towards objectives that either hurt the principal or at least do not help the principal. Since the principal ultimately has to compensate the agent for whatever effort costs he incurs in order to get him to sign a contract to begin with, even the latter proves costly for the principal. Failure to account for the effects of using distorted performance measures is sometimes referred to as **the folly of rewarding A while hoping for B** (Kerr, 1975) or the **multi-task problem** (Holmstrom and Milgrom, 1991).

All three of these sources of contractual frictions lead to similar results—under the optimal contract, the agent chooses an action that is not jointly optimal from his and the principal's perspective. But in different applied settings, different assumptions regarding what is contractible and what is not are more or less plausible. As a result, it is useful to master at least elementary versions of models capturing these three sources of frictions, so that you are well-equipped to use them as building blocks.

In the elementary versions of models of these three contracting frictions that we will look at, the effort level that the Principal would induce if there were no contractual frictions would solve:

$$\max_e pe - \frac{c}{2}e^2,$$

so that $e^{FB} = p/c$. All three of these models yield *equilibrium* effort levels $e^* < e^{FB}$.

Risk-Incentives Trade-off

The exposition of an economic model usually begins with a rough (but accurate and mostly complete) description of the players, their preferences, and what they do in the course of the game. The exposition should also include a precise treatment of the timing, which includes spelling out who does what and when and on the basis of what information, and a description of the solution concept that will be used to derive predictions. Given the description of the economic environment, it is then useful to specify the program(s) that players are solving.

I will begin with a pretty general description of the standard principal-agent model, but I will shortly afterwards specialize the model quite a bit in order to focus on a single point—the risk–incentives trade-off.

Description There is a risk-neutral Principal (P) and a risk-averse Agent (A). The Agent chooses an effort level $e \in \mathbb{R}_+$ at a private cost of $c(e)$, with $c'', c' > 0$, and this effort level affects the distribution over output $y \in Y$, with y distributed according to cdf $F(\cdot|\, e)$. This output can be sold on the product market at price p. The Principal can write a contract $w \in W \subset \{w : Y \to \mathbb{R}\}$ that determines a transfer $w(y)$ that she is compelled to pay the Agent if output y is realized. The Agent has an outside option that provides utility $\bar{u}$ to the Agent and $\bar{\pi}$ to the Principal. If the outside option is not exercised, the Principal's and Agent's preferences are, respectively,

$$\begin{aligned}
\Pi(w,e) &= \int_{y\in Y} (py - w(y))\, dF(y|\, e) = E_y[py - w|\, e] \\
U(w,e) &= \int_{y\in Y} u(w(y) - c(e))\, dF(y|\, e) = E_y[u(w - c(e))|\, e].
\end{aligned}$$

Timing The timing of the game is:

1. P offers A a contract $w(y)$, which is commonly observed.

2. A accepts the contract $(d = 1)$ or rejects it $(d = 0)$ and receives $\bar{u}$, and the game ends. This decision is commonly observed.

3. If A accepts the contract, A chooses effort level e and incurs cost $c(e)$. e is only observed by A.

4. Output y is drawn from distribution with cdf $F(\cdot| e)$. y is commonly observed.

5. P pays A an amount $w(y)$. This payment is commonly observed.

A couple remarks are in order at this point. First, behind the scenes, there is an implicit assumption that there is a third-party contract enforcer (a judge or arbitrator) who can costlessly detect when agreements have been broken and costlessly exact harsh punishments on the offender. Second, it is not necessarily important that e is unobserved by the Principal—given that the Principal takes no actions after the contract has been offered, as long as the contract cannot be conditioned directly on effort, the outcome of the game will be the same whether or not the Principal observes e. Put differently, one way of viewing the underlying source of moral-hazard problems is that contracts cannot be conditioned on relevant variables, not that the relevant variables are unobserved by the Principal. We will return to this issue when we discuss the Property Rights Theory of firm boundaries.

Equilibrium A **pure-strategy subgame-perfect equilibrium** is a contract $w^* \in W$, an acceptance decision $d^* : W \to \{0,1\}$, and an effort choice $e^* : W \times \{0,1\} \to \mathbb{R}_+$ such that, given the contract w^*, the Agent optimally chooses d^* and e^*, and given d^* and e^*, the Principal optimally offers contract w^*. We will say that the optimal contract **induces** effort e^*.

The Program The principal offers a contract $w \in W$ and proposes an effort level e in order to solve

$$\max_{w \in W, e} \int_{y \in Y} (py - w(y))\, dF(y|\, e)$$

subject to two constraints. The first constraint is that the agent actually prefers to choose effort level e rather than any other effort level $\hat{e}$. This is the standard **incentive-compatibility constraint**:

$$e \in \underset{\hat{e} \in \mathbb{R}_+}{\operatorname{argmax}} \int_{y \in Y} u(w(y) - c(\hat{e}))\, dF(y|\, \hat{e}).$$

The second constraint is that, given that the agent knows he will choose e if he accepts the contract, he prefers to accept the contract rather than to reject it and receive his outside utility $\bar{u}$. This is the standard **individual-rationality constraint** or **participation constraint**:

$$\int_{y \in Y} u(w(y) - c(e))\, dF(y|\, e) \geq \bar{u}.$$

CARA-Normal Case with Affine Contracts In order to establish a straightforward version of the risk-incentives trade-off, we will make a number of simplifying assumptions.

Assumption 1. The Agent has CARA preferences over wealth and effort costs, which are quadratic:

$$u\left(w\left(y\right)-c\left(e\right)\right)=-\exp\left\{-r\left(w\left(y\right)-\frac{c}{2}e^{2}\right)\right\},$$

and his outside option yields utility $-\exp\left\{-r\bar{u}\right\}$.

Assumption 2. Effort shifts the mean of a normally distributed random variable. That is, $y\sim N\left(e,\sigma^{2}\right)$.

Assumption 3. $W=\left\{w:Y\to\mathbb{R},w\left(y\right)=s+by\right\}$. That is, the contract space permits only affine contracts.

Discussion. In principle, there should be no exogenous restrictions on the functional form of $w\left(y\right)$. Applications, however, often restrict attention to affine contracts: $w\left(y\right)=s+by$. In many environments, an optimal contract does not exist if the contracting space is sufficiently rich, and situations in which the agent chooses the first-best level of effort, and the principal receives all the surplus can be arbitrarily approximated with a sequence of sufficiently perverse contracts (Mirrlees, 1999; Moroni and Swinkels, 2014). In contrast, the optimal affine contract often results in an effort choice that is lower than the first-best effort level, and the principal receives a lower payoff.

There are then at least three ways to view the exercise of solving for the optimal affine contract.

1. From an applied perspective, many pay-for-performance contracts in the world are affine in the relevant performance measure—franchisees pay a franchise fee and receive a constant fraction of the revenues their store generates, windshield installers receive a base wage and a constant piece rate, fruit pickers are paid per kilogram of fruit they pick. And so given that many practitioners seem to restrict attention to this class of contracts, why don't we just make sure they are doing what they do optimally? Put differently, we can brush aside global optimality on purely pragmatic grounds.

2. Many pay-for-performance contracts in the world are affine in the relevant performance measure. Our models are either too rich or not rich enough in a certain sense and therefore generate optimal contracts that are inconsistent with those we see in the world. Maybe the aspects that, in the world, lead practitioners to use affine contracts are orthogonal to the considerations we are focusing on, so that by restricting attention to the optimal affine contract, we can still say something about how real-world contracts ought to vary with changes in the underlying environment. This view presumes a more positive (as opposed to normative) role for the modeler and hopes that the theoretical equivalent of the omitted variables bias is not too severe.

3. Who cares about second-best when first-best can be attained? If our

models are pushing us toward complicated, non-linear contracts, then maybe our models are wrong. Instead, we should focus on writing down models that generate affine contracts as the optimal contract, and therefore we should think harder about what gives rise to them. (And indeed, steps have been made in this direction—see Holmstrom and Milgrom (1987), Diamond (1998) and, more recently, Carroll (2015)) This perspective will come back later in the course when we discuss the Property Rights Theory of firm boundaries.

Given the assumptions, for any contract $w(y) = s+by$, the income stream the agent receives is normally distributed with mean $s+be$ and variance $b^2\sigma^2$. His expected utility over monetary compensation is therefore a moment-generating function for a normally distributed random variable, (recall that if $X \sim N(\mu, \sigma^2)$, then $E[\exp\{tX\}] = \exp\left\{\mu t + \frac{1}{2}\sigma^2 t^2\right\}$), so his preferences can be written as

$$E[-\exp\{-r(w(y)-c(e))\}] = -\exp\left\{-r(s+be) + \frac{r^2}{2}b^2\sigma^2 + r\frac{c}{2}e^2\right\}.$$

We can take a monotonic transformation of his utility function ($-\frac{1}{r}\log(-x)$) and represent his preferences as:

$$\begin{aligned} U(e,w) &= E[w(y)] - \frac{r}{2}Var(w(y)) - \frac{c}{2}e^2 \\ &= s + be - \frac{r}{2}b^2\sigma^2 - \frac{c}{2}e^2. \end{aligned}$$

The Principal's program is then

$$\max_{s,b,e} pe - (s + be)$$

subject to incentive-compatibility

$$e \in \operatorname*{argmax}_{\hat{e}} b\hat{e} - \frac{c}{2}\hat{e}^2$$

and individual-rationality

$$s + be - \frac{r}{2}b^2\sigma^2 - \frac{c}{2}e^2 \geq \bar{u}.$$

Solving this problem is then relatively straightforward. Given an affine contract $s + be$, the agent will choose an effort level $e(b)$ that satisfies his first-order conditions

$$e(b) = \frac{b}{c},$$

and the Principal will choose the value s to ensure that the agent's individual-rationality constraint holds with equality (for if it did not hold with equality, the Principal could reduce s, making herself better off without affecting the Agent's incentive-compatibility constraint, while still respecting the Agent's individual-rationality constraint). That is,

$$s + be(b) = \frac{c}{2}e(b)^2 + \frac{r}{2}b^2\sigma^2 + \bar{u}.$$

In other words, the Principal has to ensure that the Agent's total expected monetary compensation, $s + be(b)$, fully compensates him for his effort costs, the risk costs he has to bear if he accepts this contract, and his opportunity cost. Indirectly, then, the Principal bears these costs when designing an optimal contract.

The Principal's remaining problem is to choose the incentive slope b to solve

$$\max_b pe(b) - \frac{c}{2}e(b)^2 - \frac{r}{2}b^2\sigma^2 - \bar{u}.$$

This is now an unconstrained problem with proper convexity assumptions, so the Principal's optimal choice of incentive slope solves her first-order condition

$$\begin{aligned} 0 &= pe'(b^*) - ce^*(b^*)e'(b^*) - rb^*\sigma^2 \\ &= \frac{p}{c} - c\frac{b^*}{c}\frac{1}{c} - rb^*\sigma^2 \end{aligned}$$

and therefore

$$b^* = \frac{p}{1 + rc\sigma^2}.$$

Also, given b^* and the individual-rationality constraint, we can back out s^*.

$$s^* = \bar{u} + \frac{1}{2}\left(rc\sigma^2 - 1\right)\frac{(b^*)^2}{c}.$$

Depending on the parameters, it may be the case that $s^* < 0$. That is, the Agent would have to pay the Principal if he accepts the job and does not produce anything.

In this setting, if the Principal could contract directly on effort, she would choose a contract that ensures that the Agent's individual-rationality constraint binds and therefore would solve

$$\max_e pe - \frac{c}{2}e^2,$$

so that

$$e^{FB} = \frac{p}{c}.$$

If the Principal wanted to implement this same level of effort using a contract on output, y, she would choose $b = p$ (since the Agent would choose $\frac{b}{c} = \frac{p}{c}$).

Why, in this setting, does the Principal not choose such a contract? Let us go back to the Principal's problem of choosing the incentive slope b.

$$\max_b pe\left(b\right) - \frac{c}{2}e\left(b\right)^2 - \frac{r}{2}b^2\sigma^2 - \bar{u}$$

Many fundamental points in models in the Organizational Economics literature can be seen as a comparison of first-order losses or gains against second-order gains or losses. Suppose the Principal chooses $b = p$, and consider a marginal reduction in b away from this value. The change in the

Principal's profits would be

$$\begin{aligned} & \frac{d}{db}\left(pe\left(b\right)-\frac{c}{2}e\left(b\right)^{2}-\frac{r}{2}b^{2}\sigma^{2}\right)\Big|_{b=p} \\ = \; & \underbrace{\frac{d}{db}\left(pe\left(b\right)-\frac{c}{2}e\left(b\right)^{2}\right)\Big|_{b=p}}_{=0}-rp\sigma^{2}<0 \end{aligned}$$

This first term is zero, because $b=p$ in fact maximizes $pe\left(b\right)-\frac{c}{2}e\left(b\right)^{2}$ (since it induces the first-best level of effort). The second term is strictly negative. That is, relative to the contract that induces first-best effort, a reduction in the slope of the incentive contract yields a first-order gain resulting from a decrease in the risk costs the Agent bears, while it yields a second-order loss in terms of profits resulting from moving away from the effort level that maximizes revenues minus effort costs. The optimal contract balances the incentive benefits of higher-powered incentives with these risk costs.

This trade-off seems first-order in some settings (e.g., insurance contracts in health care markets, some types of sales contracts in industries in which individual sales are infrequent, large, and unpredictable) and for certain types of output. There are many other environments in which contracts provide less-than-first-best incentives, but the first-order reasons for these low-powered contracts seem completely different, and we will turn to these environments shortly.

Before doing so, it is worth pointing out that many models in this course will involve trade-offs that determine the optimal way of organizing a firm.

In many of the settings these models examine, results that take the form of "X is organized according to Y, because player A is more risk-averse than player B" often seem intuitively unappealing. For example, suppose a model of hierarchies generated the result that less risk-averse individuals should be at the top of an organization, and more risk-averse individuals should be at the bottom. This sounds somewhat sensible—maybe richer individuals are better able to diversify their wealth, and they can therefore behave as if they are less risk averse with respect to the income stream they derive from a particular organization. But it sounds less appealing as a general rule for who should be assigned to what role in an organization—a model that predicts that more knowledgeable or more experienced workers should be assigned to higher positions seems more consistent with experience.

Finally, connecting this analysis back to the Neoclassical view of the firm, what does the risk-incentives trade-off imply about the firm's production set? Let Y^f denote the firm's **technological possibilities set**, in which the firm's input is labor costs C, and its expected output is y. This is the set of input-output vectors that would be feasible if there were no contracting frictions.

We can write

$$C = c(e) = \frac{c}{2}e^2,$$

and since expected output is just equal to the Agent's effort choice, we have that

$$y(C) = e = \left(\frac{2C}{c}\right)^{1/2}.$$

The technological possibilities set is therefore

$$Y^f = \{(y, -C) : y \leq y(C)\}.$$

We will now augment the technological possibilities set with the contractual considerations we have just derived. Because the Principal can increase s without bound, the **contract-augmented possibilities set** can be characterized by its frontier, which is the highest level of expected output the firm can produce for a given level of costs. If the Principal puts in place an optimal contract with incentive slope b (in which case the Agent's effort choice will be $e^*(b) = \frac{b}{c}$) and an s that pins the Agent to his individual-rationality constraint, the firm's profits are

$$p\frac{b}{c} - \frac{c}{2}\left(\frac{b}{c}\right)^2 - \frac{r}{2}\sigma^2 b^2 = p\frac{b}{c} - \frac{1}{2}\frac{1 + rc\sigma^2}{c}b^2.$$

Therefore, producing expected output $\frac{b}{c}$ costs the firm

$$C = \frac{1}{2}\frac{1 + rc\sigma^2}{c}b^2.$$

Finally, we can rearrange this equation to solve for the b such that the total costs to the Principal are C:

$$b = \left(\frac{2Cc}{1+rc\sigma^2}\right)^{1/2}.$$

In this case, the firm produces expected output $y = \tilde{y}(C)$, which is given by

$$\tilde{y}(C) = \left(\frac{2C}{c}\right)^{1/2}\left(\frac{1}{1+rc\sigma^2}\right)^{1/2} = \left(\frac{1}{1+rc\sigma^2}\right)^{1/2} y(C).$$

The contract-augmented possibilities set is therefore

$$\tilde{Y}^f = \{(y, -C) : y \leq \tilde{y}(C)\}.$$

Because of contractual frictions, we have that $\tilde{Y}^f \subset Y^f$, and any change in the parameters of the model for which the divergence between e^{FB} and e^* grows (such as an increase in the Agent's risk aversion or an increase in output uncertainty) will tend to increase the difference between $y(C)$ and $\tilde{y}(C)$ and therefore will shrink the contract-augmented possibilities set.

In this elementary version of this model, the contract-augmented possibilities set is a convex set. More generally, given a production-possibilities set that is convex, it need not be the case that the contract-augmented possibilities set is convex.

Further Reading Many papers restrict attention to linear contracts, even in environments in which the optimal contract (if it exists) is not linear.

Holmstrom and Milgrom (1987) examines an environment in which the principal and the agent have CARA preferences and the agent controls the drift of a Brownian motion for a finite time interval. An optimal contract conditions payments only on the value of the Brownian motion at the end of the time interval. Diamond (1998) considers an environment in which the agent can choose the mean of the output distribution as well as the entire distribution itself and shows (essentially by a convexification argument) that linear contracts are optimal. Carroll (2015) shows that linear contracts can be max-min optimal when the Principal is sufficiently uncertain about the class of actions the Agent can take.

A key comparative static of the risk–incentives moral-hazard model is that incentives are optimally weaker when there is more uncertainty in the mapping between effort and contractible output, but this comparative static is inconsistent with a body of empirical work suggesting that in more uncertain environments, agency contracts tend to involve higher-powered incentives. Prendergast (2002) resolves this discrepancy by arguing that in more uncertain environments, it is optimal to assign greater responsibility to the agent and to complement this greater responsibility with higher-powered incentives. Holding responsibilities fixed, the standard risk–incentives trade-off would arise, but the empirical studies that fail to find this relationship do not control for workers' responsibilities. Raith (2003) argues that these empirical studies examine the relationship between the risk the firm faces and the strength of the agent's incentives, while the theory is about the relationship between the risk the *agent* faces and his incentives. For an examination of several channels through which uncertainty can impact an agent's incentives, see Rantakari (2008).

1.1.1 Limited Liability

We saw in the previous model that the optimal contract sometimes involved up-front payments from the Agent to the Principal. To the extent that the Agent is unable to afford such payments (or legal restrictions prohibit such payments), the Principal will not be able to extract all the surplus that the Agent creates. Further, in order to extract surplus from the Agent, the Principal may have to put in place contracts that reduce the total surplus created. In equilibrium, the Principal may therefore offer a contract that induces effort below the first-best.

Description Again, there is a risk-neutral Principal (P). There is also a **risk-neutral** Agent (A). The Agent chooses an effort level $e \in \mathbb{R}_+$ at a private cost of $c(e)$, with $c'', c' > 0$, and this effort level affects the distribution over outputs $y \in Y$, with y distributed according to cdf $F(\cdot|\, e)$. These outputs can be sold on the product market for price p. The Principal can write a contract $w \in W \subset \{w : Y \to \mathbb{R}, w(y) \geq \underline{w} \text{ for all } y\}$ that determines a transfer $w(y)$ that she is compelled to pay the Agent if output y is realized. The Agent has an outside option that provides utility $\bar{u}$ to the Agent and $\bar{\pi}$ to the Principal. If the outside option is not exercised, the Principal's and

Agent's preferences are, respectively,

$$\begin{aligned}
\Pi(w,e) &= \int_{y\in Y}(py-w(y))\,dF(y|\,e)=E_y[py-w|\,e]\\
U(w,e) &= \int_{y\in Y}(w(y)-c(e))\,dF(y|\,e)=E_y[w-c(e)|\,e].
\end{aligned}$$

There are two differences between this model and the model in the previous subsection. The first difference is that the Agent is risk-neutral (so that absent any other changes, the equilibrium contract would induce first-best effort). The second difference is that the wage payment from the Principal to the Agent has to exceed, for each realization of output, a value $\underline{w}$. Depending on the setting, this constraint is described as a liquidity constraint or a limited-liability constraint. In repeated settings, it is more naturally thought of as the latter—due to legal restrictions, the Agent cannot be legally compelled to make a transfer (larger than $-\underline{w}$) to the Principal. In static settings, either interpretation may be sensible depending on the particular application—if the Agent is a fruit picker, for instance, he may not have much liquid wealth that he can use to pay the Principal.

Timing The timing of the game is exactly the same as before.

1. P offers A a contract $w(y)$, which is commonly observed.

2. A accepts the contract $(d=1)$ or rejects it $(d=0)$ and receives $\bar{u}$, and the game ends. This decision is commonly observed.

3. If A accepts the contract, A chooses effort level e and incurs cost $c(e)$. e is only observed by A.

4. Output y is drawn from distribution with cdf $F(\cdot|\, e)$. y is commonly observed.

5. P pays A an amount $w(y)$. This payment is commonly observed.

Equilibrium The solution concept is the same as before. A **pure-strategy subgame-perfect equilibrium** is a contract $w^* \in W$, an acceptance decision $d^* : W \to \{0,1\}$, and an effort choice $e^* : W \times \{0,1\} \to \mathbb{R}_+$ such that given the contract w^*, the Agent optimally chooses d^* and e^*, and given d^* and e^*, the Principal optimally offers contract w^*. We will say that the optimal contract induces effort e^*.

The Program The principal offers a contract $w \in W$ and proposes an effort level e in order to solve

$$\max_{w\in W, e} \int_{y\in Y} (py - w(y))\, dF(y|\, e)$$

subject to three constraints: the incentive-compatibility constraint

$$e \in \underset{\hat{e}\in\mathbb{R}_+}{\operatorname{argmax}} \int_{y\in Y} (w(y) - c(\hat{e}))\, dF(y|\, \hat{e}),$$

the individual-rationality constraint

$$\int_{y \in Y} (w(y) - c(e))\, dF(y|\, e) \geq \bar{u},$$

and the limited-liability constraint

$$w(y) \geq \underline{w} \text{ for all } y.$$

Binary-Output Case Jewitt, Kadan, and Swinkels (2008) solves for the optimal contract in the general environment above (and even allows for agent risk aversion). Here, I will instead focus on an elementary case that highlights the main trade-off.

Assumption 1. Output is $y \in \{0, 1\}$, and given effort e, its distribution satisfies $\Pr[y = 1|\, e] = e$.

Assumption 2. The agent's costs have a non-negative third derivative: $c''' \geq 0$, and they satisfy conditions that ensure an interior solution: $c'(0) = 0$ and $c'(1) = +\infty$. Or for comparison across models in this module, $c(e) = \frac{c}{2}e^2$, where $p \leq c$ to ensure that $e^{FB} < 1$.

Finally, we can restrict attention to affine, nondecreasing contracts

$$\begin{aligned} W &= \{w(y) = (1-y)\, w_0 + y w_1, w_0, w_1 \geq 0\} \\ &= \{w(y) = s + by, s \geq \underline{w}, b \geq 0\}. \end{aligned}$$

When output is binary, this restriction to affine contracts is without loss of generality. Also, the restriction to nondecreasing contracts is not restrictive (i.e., any optimal contract of a relaxed problem in which we do not impose that contracts are nondecreasing will also be the solution to the full problem). This result is something that needs to be shown and is not in general true, but in this case, it is straightforward.

In principal-agent models, it is often useful to break the problem down into two steps. The first step takes a target effort level, e, as given and solves for the set of cost-minimizing contracts implementing effort level e. Any cost-minimizing contract implementing effort level e results in an expected cost of $C(e)$ to the principal. The second step takes the function $C(\cdot)$ as given and solves for the optimal effort choice.

In general, the cost-minimization problem tends to be a well-behaved convex-optimization problem, since (even if the agent is risk-averse) the objective function is weakly concave, and the constraint set is a convex set (since given an effort level e, the individual-rationality constraint and the limited-liability constraint define convex sets, and each incentive constraint ruling out effort level $\hat{e} \neq e$ also defines a convex set, and the intersection of convex sets is itself a convex set). The resulting cost function $C(\cdot)$ need not have nice properties, however, so the second step of the optimization problem is only well-behaved under restrictive assumptions. In the present case, assumptions 1 and 2 ensure that the second step of the optimization problem is well-behaved.

Cost-Minimization Problem Given an effort level e, the cost-minimization problem is given by

$$C(e, \bar{u}, \underline{w}) = \min_{s,b} s + be$$

subject to the agent's incentive-compatibility constraint

$$e \in \underset{\hat{e}}{\operatorname{argmax}} \{s + b\hat{e} - c(\hat{e})\},$$

his individual-rationality constraint

$$s + be - c(e) \geq \bar{u},$$

and the limited-liability constraint

$$s \geq \underline{w}.$$

I will denote a **cost-minimizing contract implementing effort level** e by (s_e^*, b_e^*).

The first step in solving this problem is to notice that the agent's incentive-compatibility constraint implies that any cost-minimizing contract implementing effort level e must have $b_e^* = c'(e)$.

If there were no limited-liability constraint, the principal would choose s_e^* to extract the agent's surplus. That is, given $b = b_e^*$, s would solve

$$s + b_e^* e = \bar{u} + c(e).$$

That is, s would ensure that the agent's expected compensation exactly equals his expected effort costs plus his opportunity cost. The resulting s, however, may not satisfy the limited-liability constraint. The question then is: given $\bar{u}$ and $\underline{w}$, for what effort levels e is the principal able to extract all the agent's surplus (i.e., for what effort levels does the limited-liability constraint not bind?), and for what effort levels is she unable to do so? Figure 1 below shows cost-minimizing contracts for effort levels e_1 and e_2. Any contract can be represented as a line in this figure, where the line represents the expected pay the agent will receive given an effort level e. The cost-minimizing contract for effort level e_1 is tangent to the $\bar{u} + c(e)$ curve at e_1 and its intercept is $s^*_{e_1}$. Similarly for e_2. Both $s^*_{e_1}$ and $s^*_{e_2}$ are greater than $\underline{w}$, which implies that for such effort levels, the limited-liability constraint is not binding.

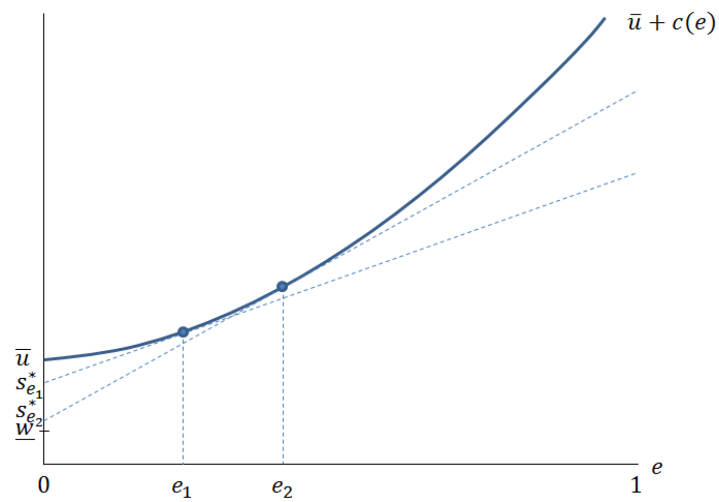


Figure 1

For effort sufficiently high, the limited-liability constraint will be binding in a cost-minimizing contract, and it will be binding for all higher effort levels. Define the threshold $\bar{e}(\bar{u}, \underline{w})$ to be the effort level such that for all $e \geq \bar{e}(\bar{u}, \underline{w})$, $s^*_e = \underline{w}$. Figure 2 illustrates that $\bar{e}(\bar{u}, \underline{w})$ is the effort level at which the contract tangent to the $\bar{u} + c(e)$ curve at $\bar{e}(\bar{u}, \underline{w})$ intersects the vertical axis at exactly $\underline{w}$. That is, $\bar{e}(\bar{u}, \underline{w})$ solves

$$c'(\bar{e}(\bar{u}, \underline{w})) = \frac{\bar{u} + c(\bar{e}(\bar{u}, \underline{w})) - \underline{w}}{\bar{e}(\bar{u}, \underline{w})}.$$

Figure 2 also illustrates that for all effort levels $e > \bar{e}(\bar{u}, \underline{w})$, the cost-minimizing contract involves giving the agent strictly positive surplus. That is, the cost to the principal of getting the agent to choose effort $e > \bar{e}(\bar{u}, \underline{w})$ is equal to the agent's opportunity costs $\bar{u}$ plus his effort costs $c(e)$ plus **incentive costs** $IC(e, \bar{u}, \underline{w})$.

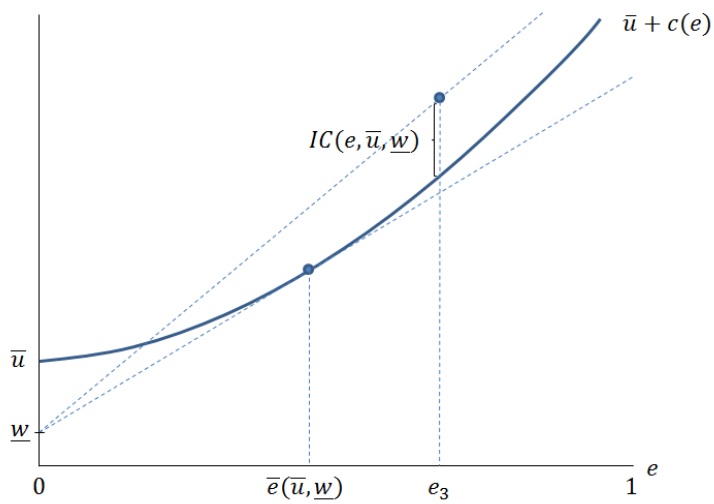


Figure 2

The incentive costs $IC(e, \bar{u}, \underline{w})$ are equal to the agent's expected compensation given effort choice e and cost-minimizing contract (s_e^*, b_e^*) minus his costs:

$$
\begin{aligned}
IC(e, \bar{u}, \underline{w}) &= \begin{cases} 0 & e \leq \bar{e}(\bar{u}, \underline{w}) \\ \underline{w} + c'(e)\, e - c(e) - \bar{u} & e \geq \bar{e}(\bar{u}, \underline{w}) \end{cases} \\
&= \max\{0, \underline{w} + c'(e)\, e - c(e) - \bar{u}\}
\end{aligned}
$$

where I used the fact that for $e \geq \bar{e}(\bar{u}, \underline{w})$, $s_e^* = \underline{w}$ and $b_e^* = c'(e)$. This incentive-cost function $IC(\cdot, \bar{u}, \underline{w})$ is the key object that captures the main contracting friction in this model. I will sometimes refer to $IC(e, \bar{u}, \underline{w})$ as the **incentive rents** required to get the agent to choose effort level e. Putting these results together, we see that

$$C(e, \bar{u}, \underline{w}) = \bar{u} + c(e) + IC(e, \bar{u}, \underline{w}).$$

That is, the principal's total costs of implementing effort level e are the sum of the agent's costs plus the incentive rents required to get the agent to choose effort level e.

Since $IC(e, \bar{u}, \underline{w})$ is the main object of interest in this model, I will describe some of its properties. First, it is continuous in e (including, in particular, at $e = \bar{e}(\bar{u}, \underline{w})$). Next, $\bar{e}(\bar{u}, \underline{w})$ and $IC(e, \bar{u}, \underline{w})$ depend on $(\bar{u}, \underline{w})$ only inasmuch as $(\bar{u}, \underline{w})$ determines $\bar{u} - \underline{w}$, so I will abuse notation and write these expressions as $\bar{e}(\bar{u} - \underline{w})$ and $IC(e, \bar{u} - \underline{w})$. Also, given that $c'' > 0$, IC is increasing in e (since $\underline{w} + c'(e)\, e - c(e) - \underline{u}$ is strictly increasing in e, and IC is just the max of this expression and zero). Further, given that $c''' \geq 0$, IC is convex in e. For $e \geq \bar{e}(\bar{u} - \underline{w})$, this property follows, because

$$\frac{\partial^2}{\partial e^2} IC = c''(e) + c'''(e)\, e \geq 0.$$

And again, since IC is the max of two convex functions, it is also a convex function. Finally, since $IC(\cdot, \bar{u} - \underline{w})$ is flat when $e \leq \bar{e}(\bar{u} - \underline{w})$ and it is strictly increasing (with slope independent of $\bar{u}-\underline{w}$) when $e \geq \bar{e}(\bar{u}-\underline{w})$, the slope of IC with respect to e is (weakly) decreasing in $\bar{u}-\underline{w}$, since $\bar{e}(\bar{u}-\underline{w})$ is increasing in $\bar{u}-\underline{w}$. That is, $IC(e,\bar{u}-\underline{w})$ satisfies decreasing differences in $(e,\bar{u}-\underline{w})$.

Motivation-Rent Extraction Trade-off The second step of the optimization problem takes as given the function

$$C(e,\bar{u}-\underline{w}) = \bar{u} + c(e) + IC(e,\bar{u}-\underline{w})$$

and solves for the optimal effort choice by the principal:

$$\begin{aligned} & \max_e pe - C(e,\bar{u}-\underline{w}) \\ = \ & \max_e pe - \bar{u} - c(e) - IC(e,\bar{u}-\underline{w}). \end{aligned}$$

Note that total surplus is given by $pe-\bar{u}-c(e)$, which is therefore maximized at $e=e^{FB}$ (which, if $c(e)=ce^2/2$, then $e^{FB}=p/c$). Figure 3 below depicts the principal's expected benefit line pe, and her expected costs of implementing effort e at minimum cost, $C(e,\bar{u}-\underline{w})$. The first-best effort level, e^{FB} maximizes the difference between pe and $\bar{u}+c(e)$, while the equilibrium effort level e^* maximizes the difference between pe and $C(e,\bar{u}-\underline{w})$.

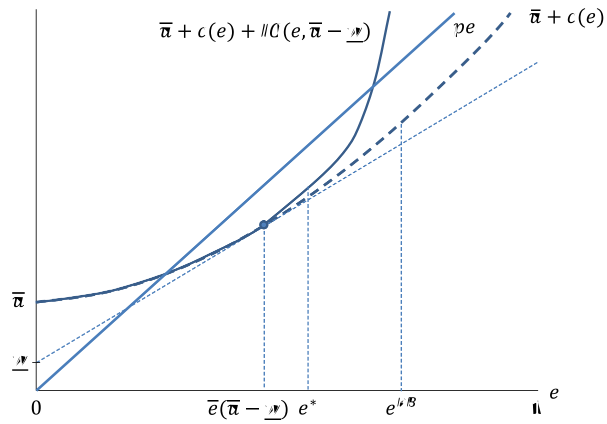


Figure 3

If $c(e) = ce^2/2$, we can solve explicitly for $\bar{e}(\bar{u} - \underline{w})$ and for $IC(e, \bar{u} - \underline{w})$ when $e > \bar{e}(\bar{u} - \underline{w})$. In particular,

$$\bar{e}(\bar{u} - \underline{w}) = \left(\frac{2(\bar{u} - \underline{w})}{c}\right)^{1/2}$$

and when $e > \bar{e}(\bar{u} - \underline{w})$,

$$IC(e, \bar{u} - \underline{w}) = \underline{w} + \frac{1}{2}ce^2 - \bar{u}.$$

If $\underline{w} < 0$ and p is sufficiently small, we can have $e^* = e^{FB}$ (i.e., these are the conditions required to ensure that the limited-liability constraint is not binding for the cost-minimizing contract implementing $e = e^{FB}$). If p is sufficiently large relative to $\bar{u} - \underline{w}$, we will have $e^* = \frac{1}{2}\frac{p}{c} = \frac{1}{2}e^{FB}$. For p

somewhere in between, we will have $e^* = \bar{e}(\bar{u} - \underline{w}) < e^{FB}$. In particular, $C(e, \bar{u} - \underline{w})$ is kinked at this point.

As in the risk–incentives model, we can illustrate through a partial characterization why (and when) effort is less-than first-best. Since we know that e^{FB} maximizes $pe - \bar{u} - c(e)$, we therefore have that

$$\frac{d}{de}\left[pe - \bar{u} - c(e) - IC(e, \bar{u} - \underline{w})\right]_{e=e^{FB}} = -\frac{\partial}{\partial e} IC\left(e^{FB}, \bar{u} - \underline{w}\right) \leq 0,$$

with strict inequality if the limited-liability constraint binds at the cost-minimizing contract implementing e^{FB}. This means that, even though e^{FB} maximizes total surplus, if the principal has to provide the agent with rents at the margin, she may choose to implement a lower effort level. Reducing the effort level away from e^{FB} leads to second-order losses in terms of total surplus, but it leads to first-order gains in profits for the principal. In this model, there is a tension between total-surplus creation and rent extraction, which yields less-than first-best effort in equilibrium.

In my view, liquidity constraints are extremely important and are probably one of the main reasons for why many jobs do not involve first-best incentives. The Vickrey-Clarke-Groves logic that first-best outcomes can be obtained if the firm transfers the entire profit stream to each of its members in exchange for a large up-front payment seems simultaneously compelling, trivial, and obviously impracticable. In for-profit firms, in order to make it worthwhile to transfer a large enough share of the profit stream to an individual worker to significantly affect his incentives, the firm would require a large up-front transfer that most workers cannot afford to pay. It is therefore not surprising that we do not see most workers' compensation tied directly to the firm's overall profits in a meaningful way. One implication of this logic is that firms have to find alternative instruments to use as performance measures, which we will turn to next. In principle, models in which firms do not motivate their workers by writing contracts directly on profits should include assumptions under which the firm optimally chooses not to write contracts directly on profits, but they almost never do.

Exercise. *Let* $\Delta \equiv (2c(\bar{u} - \underline{w}))^{1/2}$. *Part 1: Show that when* $p \leq \Delta$, *the contract-augmented possibilities set is* $\tilde{Y}^f = \left\{(y, -C) : y \leq \left(\frac{2C}{c}\right)^{1/2}\left(\frac{C-\underline{w}}{2C}\right)^{1/2}\right\}$. *Part 2: Show that when* $p \geq 2\Delta$, *the contract-augmented possibilities set is* $\tilde{Y}^f = \left\{(y, -C) : y \leq \left(\frac{2C}{c}\right)^{1/2}\right\}$. *Part 3: Solve for* $\tilde{Y}^f$ *for* $\Delta < p < 2\Delta$. *(This part is somewhat more complicated.) Part 4: In this model, the contract-augmented possibilities depend on the equilibrium price level, which implies that in a competitive-equilibrium framework, the firm's production possibilities are endogenous to the equilibrium. This was not the case for the risk-incentives trade-off model. If we define* $\tilde{Y}^f(p)$ *as the contract-augmented possibilities set given price level* p, *how does* $\tilde{Y}^f(p)$ *vary in* p? *(Note that since* $\tilde{Y}^f(p)$ *is a set, you will have to think about what it means for a set to vary in a parameter.)*

Exercise. *Holmstrom (1979) shows that in the risk-incentives model in the previous subsection, if there is a costless additional performance measure* m

that is informative about e*, then an optimal formal contract should always put some weight on* m *unless* y *is a sufficient statistic for* y *and* m*. This is known as Holmstrom's "informativeness principle" and suggests that optimal contracts should always be extremely sensitive to the details of the environment the contract is written in. Suppose instead that the agent is risk-neutral but liquidity-constrained, and suppose there is a performance measure* $m \in \{0,1\}$ *such that* $\Pr[m=1|\,e]=e$ *and conditional on* e*,* m *and* y *are independent. Suppose contracts of the form* $w(y,m) = s + b_y y + b_m m + b_{ym} ym$ *can be written but must satisfy* $w(y,m) \geq \bar{w}$ *for each realization of* (y,m)*. Is it again always the case that* $b_m \neq 0$ *and/or* $b_{ym} \neq 0$*?*

Further Reading Jewitt, Kadan, and Swinkels (2008) derive optimal contracts in a broad class of environments with risk-averse agents and bounded payments (in either direction). Chaigneau, Edmans, and Gottlieb (2015) provide necessary and sufficient conditions for additional informative signals to have strictly positive value to the Principal. Wu (forthcoming) shows that firms' contract-augmented possibilities sets are endogenous to the competitive environment they face when their workers are subject to limited-liability constraints.

1.1.2 Multiple Tasks and Misaligned Performance Measures

In the previous two models, what the Principal cared about was output, and output, though a noisy measure of effort, was perfectly measurable. This assumption seems sensible when we think about overall firm profits (ignoring basically everything that accountants think about every day), but as we alluded to in the discussion above, overall firm profits are generally too blunt of an instrument to use to motivate individual workers within the firm if they are liquidity-constrained. As a result, firms often try to motivate workers using more specific performance measures, but while these performance measures are informative about what actions workers are taking, they may be less useful as a description of how the workers' actions affect the objectives the firm cares about. And paying workers for what is measured may not get them to take actions that the firm cares about. This observation underpins the title of the famous 1975 paper by Steve Kerr called "On the Folly of Rewarding A, while Hoping for B."

Description Again, there is a risk-neutral Principal (P) and a risk-neutral Agent (A). The Agent chooses an effort vector $e = (e_1, e_2) \in \mathbb{R}_+^2$ at a private cost of $\frac{c}{2}\left(e_1^2 + e_2^2\right)$. This effort vector affects the distribution of output

$y \in Y = \{0, 1\}$ and a performance measure $m \in M = \{0, 1\}$ as follows:

$$\Pr[y = 1|\, e] = f_1 e_1 + f_2 e_2$$
$$\Pr[m = 1|\, e] = g_1 e_1 + g_2 e_2,$$

where it may be the case that $f = (f_1, f_2) \neq (g_1, g_2) = g$. Assume that $f_1^2 + f_2^2 = g_1^2 + g_2^2 = 1$ (i.e., the norms of the f and g vectors are unity). The output can be sold on the product market for price p. The Principal can write a contract $w \in W \subset \{w : M \to \mathbb{R}\}$ that determines a transfer $w(m)$ that she is compelled to pay the Agent if performance measure m is realized. Since the performance measure is binary, contracts take the form $w = s + bm$. The Agent has an outside option that provides utility $\bar{u}$ to the Agent and $\bar{\pi}$ to the Principal. If the outside option is not exercised, the Principal's and Agent's preferences are, respectively,

$$\Pi(w, e) = f_1 e_1 + f_2 e_2 - E[w(m)|\, e]$$
$$U(w, e) = s + b(g_1 e_1 + g_2 e_2) - \frac{c}{2}\left(e_1^2 + e_2^2\right).$$

Timing The timing of the game is exactly the same as before.

1. P offers A a contract $w(m)$, which is commonly observed.

2. A accepts the contract $(d = 1)$ or rejects it $(d = 0)$ and receives $\bar{u}$ and the game ends. This decision is commonly observed.

3. If A accepts the contract, A chooses effort vector e and incurs cost $c(e)$. e is only observed by A.

4. Performance measure m is drawn from distribution with pdf $f(\cdot|\, e)$ and output y is drawn from distribution with pdf $g(\cdot|\, e)$. m is commonly observed.

5. P pays A an amount $w(m)$. This payment is commonly observed.

Equilibrium The solution concept is the same as before. A **pure-strategy subgame-perfect equilibrium** is a contract $w^* \in W$, an acceptance decision $d^* : W \to \{0,1\}$, and an effort choice $e^* : W \times \{0,1\} \to \mathbb{R}_+^2$ such that given the contract w^*, the Agent optimally chooses d^* and e^*, and given d^* and e^*, the Principal optimally offers contract w^*. We will say that the optimal contract induces effort e^*.

The Program The principal offers a contract $w = s + bm$ and proposes an effort level e in order to solve

$$\max_{s,b,e} p(f_1 e_1 + f_2 e_2) - (s + b(g_1 e_1 + g_2 e_2))$$

subject to the incentive-compatibility constraint

$$e \in \underset{\hat{e} \in \mathbb{R}_+}{\operatorname{argmax}}\, s + b(g_1 \hat{e}_1 + g_2 \hat{e}_2) - \frac{c}{2}\left(\hat{e}_1^2 + \hat{e}_2^2\right)$$

and the individual-rationality constraint

$$s + b\left(g_1 e_1 + g_2 e_2\right) - \frac{c}{2}\left(e_1^2 + e_2^2\right) \geq \bar{u}.$$

Equilibrium Contracts and Effort Given a contract $s + bm$, the Agent will choose efforts

$$\begin{aligned} e_1^*(b) &= \frac{b}{c} g_1 \\ e_2^*(b) &= \frac{b}{c} g_2. \end{aligned}$$

The Principal will choose s so that the individual-rationality constraint holds with equality

$$s + b\left(g_1 e_1^*(b) + g_2 e_2^*(b)\right) = \bar{u} + \frac{c}{2}\left(e_1^*(b)^2 + e_2^*(b)^2\right).$$

Since contracts send the Agent off in the "wrong direction" relative to what maximizes total surplus, providing the Agent with higher-powered incentives by increasing b sends the agent farther of in the wrong direction. This is costly for the Principal, because in order to get the Agent to accept the contract, she has to compensate him for his effort costs, even if they are in the wrong direction.

The Principal's unconstrained problem is therefore

$$\max_b p\left(f_1 e_1^*(b) + f_2 e_2^*(b)\right) - \frac{c}{2}\left(e_1^*(b)^2 + e_2^*(b)^2\right) - \bar{u}.$$

Taking first-order conditions,

$$pf_1 \frac{\partial e_1^*}{\partial b} + pf_2 \frac{\partial e_2^*}{\partial b} = ce_1^*(b^*) \frac{\partial e_1^*}{\partial b} + ce_2^*(b^*) \frac{\partial e_2^*}{\partial b},$$

or

$$\begin{aligned} pf_1 g_1 + pf_2 g_2 &= b^* g_1 g_1 + b^* g_2 g_2 \\ b^* &= p \frac{f_1 g_1 + f_2 g_2}{g_1^2 + g_2^2} = p \frac{f \cdot g}{g \cdot g} = p \frac{||f||}{||g||} \cos\theta = p \cos\theta, \end{aligned}$$

where $\cos\theta$ is the angle between the vectors f and g. That is, the optimal incentive slope depends on the relative magnitudes of the f and g vectors (which in this model were assumed to be the same, but in a richer model this need not be the case) as well as how well-aligned they are. If m is a perfect measure of what the firm cares about, then g is a linear transformation of f and therefore the angle between f and g would be zero, so that $\cos\theta = 1$. If m is completely uninformative about what the firm cares about, then f and g are orthogonal, and therefore $\cos\theta = 0$. As a result, this model is often referred to as the **"cosine of theta model."**

Another way to view this model is as follows. Since formal contracts allow for unrestricted lump-sum transfers between the Principal and the Agent, the Principal would optimally like efforts to be chosen in such a way that they maximize total surplus:

$$\max_{e} p\left(f_1 e_1 + f_2 e_2\right) - \frac{c}{2}\left(e_1^2 + e_2^2\right),$$

or $e_1^* = \frac{p}{c} f_1$ and $e_2^* = \frac{p}{c} f_2$. That is, the Principal would like to choose a vector of efforts that is collinear with the vector f :

$$\left(e_1^*, e_2^*\right) = \frac{p}{c} \cdot \left(f_1, f_2\right).$$

Since contracts can only depend on m and not directly on y, the Principal has only limited control over the actions that the Agent chooses. That is, given a contract specifying incentive slope b, the Agent chooses $e_1^*(b) = \frac{b}{c} g_1$ and $e_2^*(b) = \frac{b}{c} g_2$. Therefore, the Principal can only (indirectly) choose a vector of efforts that is collinear with the vector g:

$$\left(e_1^*(b), e_2^*(b)\right) = \frac{b}{c} \cdot \left(g_1, g_2\right).$$

The question is then: which such vector maximizes total surplus (which the Principal will extract with an ex-ante lump-sum transfer)? That is, which point along the $k \cdot (g_1, g_2)$ ray minimizes the mean-squared error distance to $\frac{p}{c} \cdot (f_1, f_2)$?

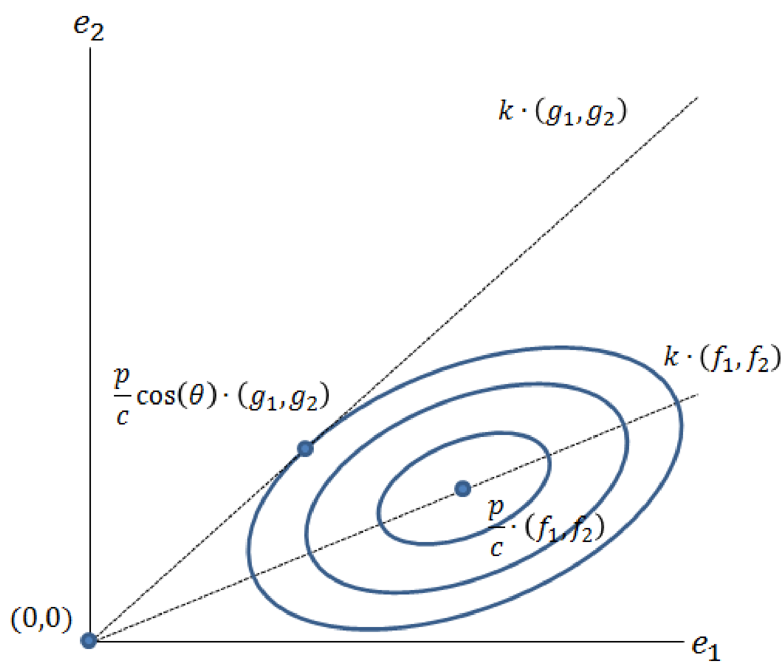


Figure 4

This is a more explicit "incomplete contracts" model of motivation. That is, we are explicitly restricting the set of contracts that the Principal can offer the Agent in a way that directly determines a subset of the effort space that the Principal can induce the Agent to choose among. And it is founded not on the idea that certain measures (in particular, y) are unobservable, but rather that they cannot be contracted upon.

Exercise. *Suppose there are N tasks rather than 2 (i.e., $e = (e_1, \ldots, e_N)$ and $c(e) = \frac{1}{2}(e_1^2 + \cdots + e_N^2)$) and $M < N$ linearly independent performance mea-*

sures rather than 1 (i.e., $m_j = g_{1j}a_1 + \cdots + g_{Nj}a_N$ for $j = 1, \ldots, M$). Show that the optimal incentive slope vector is equal to the regression coefficient that would be obtained if one ran the regression $f_i = \alpha + \beta_1 g_{i1} + \cdots + \beta_M g_{iM} + \varepsilon_i$.

Finally, we can derive the contract-augmented possibilities set and compare it to the technological possibilities set in this setting. First, let us derive the technological possibilities set. We can write

$$y(C) = \max_{e_1, e_2} f_1 e_1 + f_2 e_2$$

subject to

$$\frac{c}{2}\left(e_1^2 + e_2^2\right) \leq C.$$

The Lagrangian for this problem is

$$\mathcal{L} = p\left(f_1 e_1 + f_2 e_2\right) + \lambda\left(C - \frac{c}{2}\left(e_1^2 + e_2^2\right)\right),$$

and its first-order conditions are

$$\begin{aligned} p f_1 &= \lambda c e_1 \\ p f_2 &= \lambda c e_2, \end{aligned}$$

which implies that the optimum must always satisfy

$$\frac{e_1^*}{e_2^*} = \frac{f_1}{f_2}.$$

If we plug this condition into the constraint, which will hold with equality, we get

$$C = \frac{c}{2}\left(\left(\frac{f_1}{f_2}e_2\right)^2 + e_2^2\right) = \frac{c}{2}\frac{f_1^2 + f_2^2}{f_2^2}e_2^2 = \frac{c}{2}\frac{||f||^2}{f_2^2}e_2^2,$$

so that

$$e_2 = \left(\frac{2C}{c}\right)^{1/2}\frac{f_2}{||f||} \text{ and } e_1 = \left(\frac{2C}{c}\right)^{1/2}\frac{f_1}{||f||}.$$

The frontier of the technological possibilities set is therefore

$$y(C) = \left(\frac{2C}{c}\right)^{1/2}\frac{f_1^2 + f_2^2}{||f||} = \left(\frac{2C}{c}\right)^{1/2},$$

and the technological possibilities set is

$$Y^f = \{(y, -C) : y \leq y(C)\}.$$

To solve for the contract-augmented possibilities set, note that given b, the Agent chooses

$$(e_1(b), e_2(b)) = \frac{b}{c}(g_1, g_2).$$

At cost C, maximum production solves

$$\hat{y}(C) = \max_b f_1 e_1(b) + f_2 e_2(b)$$

subject to

$$\frac{c}{2}\left(e_1(b)^2 + e_2(b)^2\right) \leq C.$$

This cost constraint will hold with equality, which gives us

$$\frac{c}{2}\left(\frac{b}{c}\right)^2\left(g_1^2+g_2^2\right)=C$$

or

$$b\left(C\right)=\left(2Cc\right)^{1/2}\frac{1}{||g||}.$$

The frontier of the contract-augmented possibilities set is therefore

$$\begin{aligned}\hat{y}\left(C\right) &= f_1\frac{b\left(C\right)}{c}g_1+f_2\frac{b\left(C\right)}{c}g_2=\frac{b\left(C\right)}{c}f\cdot g=\left(\frac{2C}{c}\right)^{1/2}\frac{1}{||g||}\,||f||\,||g||\cos\theta \\ &= \left(\frac{2C}{c}\right)^{1/2}\cos\theta=\cos\theta\cdot y\left(C\right),\end{aligned}$$

and therefore the contract-augmented possibilities set is given by

$$\hat{Y}^f=\left\{\left(y,-C\right):y\leq\hat{y}\left(C\right)\right\}.$$

Further Reading Holmstrom and Milgrom (1991, 1994) explore many interesting organizational implications of misaligned performance measures in multi-task settings. In particular, they show that when performance measures are misaligned, it may be optimal to put in place rules that restrict the actions an agent is allowed to perform, it may be optimal to split up activities across agents (job design), and it may be optimal to adjust the boundaries of the firm. Job restrictions, job design, boundaries of the firm, and incentives should be designed to be an internally consistent system. The model described in this section is formally equivalent to Baker's (1992) model in which the agent receives noncontractible private information about the effectiveness of his (single) task before making his effort decision, since his contingent plan of effort choices can be viewed as a vector of effort choices that differentially affect his expected pay. This particular specification was spelled out in Baker's (2002) article, and it is related to Feltham and Xie's (1994) model.

1.1.3 Contracts with Externalities

Before moving on to consider environments in which no formal contracts are available, we will briefly examine another source of contractual frictions that can arise and prevent parties from taking first-best actions. So far, we have considered what happens when certain states of nature or actions were impossible to contract upon or where there were legal or practical restrictions on the form of the contract. Here, we will consider limits on the number of parties that can be part of the same contract. We refer to these situations as "contracts with externalities," following Segal (1999). We will highlight, in the context of two separate models, some of the problems that can arise when there are multiple principals offering contracts to a single Agent.

In the first model, I show that when there are otherwise no contracting frictions, so that if the Principals could jointly offer a single contract to the Agent, they would be optimally choose a contract that induces first-best effort, there may be **coordination failures**. There are equilibria in which the Principals offer contracts that do not induce first-best effort, and there are equilibria in which they offer contracts that do induce first-best effort. In the second model, I show that when there are direct costs associated with higher-powered incentives (as is the case when the Agent is risk-averse or when Principals have to incur a setup cost to put in place higher-powered incentive schemes, as in Battigalli and Maggi (2002)). In this setting, if the Principals could jointly offer a single contract to the Agent, they would optimally choose a contract that induces an effort level e^C lower than the first-best effort level, because of a contracting costs-incentives trade-off (analogous to the risk-incentives trade-off). If they cannot jointly offer a single contract, there will be a unique equilibrium in which the Principals offer contracts that induce effort $e^* < e^C$.

Description of Coordination-Failure Version There is a risk-neutral Agent (A) and two risk-neutral Principals (P_1 and P_2). The Agent chooses an effort $e \in \{0, 1\}$ at cost ce. This effort determines outputs $y_1 = e$ and $y_2 = e$ that accrue to the Principals. These outputs can be sold on the product market for prices p_1 and p_2, respectively, and let $p \equiv p_1 + p_2$. Principals simultaneously offer contracts $w_1, w_2 \in W = \{w : \{0, 1\} \to \mathbb{R}\}$ to the Agent. Denote Principal i's contract offer by $w_i = s_i + b_i e$. The Agent has an outside option that yields utility $\bar{u}$ to the Agent and 0 to each Principal. If the outside option is not exercised, players' payoffs are:

$$\begin{aligned}\Pi_1(w_1, w_2, e) &= p_1 e - w_1\\ \Pi_2(w_1, w_2, e) &= p_2 e - w_2\\ U(w_1, w_2, e) &= w_1 + w_2 - ce.\end{aligned}$$

Timing The timing of the game is exactly the same as before.

1. P_1 and P_2 simultaneously offer contracts w_1, w_2 to A. Offers are commonly observed.

2. A accepts both contracts $(d = 1)$ or rejects both contracts $(d = 0)$ and receives $\bar{u}$ and the game ends. This decision is commonly observed.

3. If A accepts the contracts, A chooses effort $e \in \{0, 1\}$ at cost ce. e is commonly observed.

4. P_1 and P_2 pay A amounts $w_1(e), w_2(e)$. These payments are commonly observed.

Equilibrium A **pure-strategy subgame-perfect equilibrium** is a pair of contracts $w_1^*, w_2^* \in W$, an acceptance strategy $d^* : W^2 \to \{0, 1\}$, an effort strategy $e^* : W^2 \to \{0, 1\}$ such that given contracts w_1^*, w_2^*, A optimally chooses $d^*(w_1^*, w_2^*)$ and $e^*(w_1^*, w_2^*)$. Given d^*, e^* and w_2^*, P_1 optimally chooses w_1^*, and given d^*, e^*, and w_1^*, P_2 optimally chooses w_2^*.

The Program Given contracts w_1 and w_2 specifying (s_1, b_1) and (s_2, b_2), if the Agent accepts these contracts, he will choose $e = 1$ if $b_1 + b_2 \geq c$, and he will choose $e = 0$ if $b_1 + b_2 \leq c$. Define

$$e(b_1, b_2) = \begin{cases} 1 & b_1 + b_2 \geq c \\ 0 & b_1 + b_2 \leq c, \end{cases}$$

and he will accept these contracts if

$$s_1 + b_1 e(b_1, b_2) + s_2 + b_2 e(b_1, b_2) - ce(b_1, b_2) \geq \bar{u}.$$

Suppose P_1 believes P_2 will offer contract (s_2, b_2). Then P_1 will choose $\hat{s}_1, \hat{b}_1$ to solve

$$\max_{\hat{s}_1, \hat{b}_1} p_1 e\left(\hat{b}_1, b_2\right) - \left(\hat{s}_1 + \hat{b}_1 e\left(\hat{b}_1, b_2\right)\right)$$

subject to the Agent's individual-rationality constraint

$$\hat{s}_1 + \hat{b}_1 e\left(\hat{b}_1, b_2\right) + s_2 + b_2 e\left(\hat{b}_1, b_2\right) - ce\left(\hat{b}_1, b_2\right) \geq \bar{u}.$$

P_1 will choose $\hat{s}_1$ so that this individual-rationality constraint holds with equality:

$$\hat{s}_1 + \hat{b}_1 e\left(\hat{b}_1, b_2\right) = \bar{u} - s_2 - b_2 e\left(\hat{b}_1, b_2\right) + ce\left(\hat{b}_1, b_2\right).$$

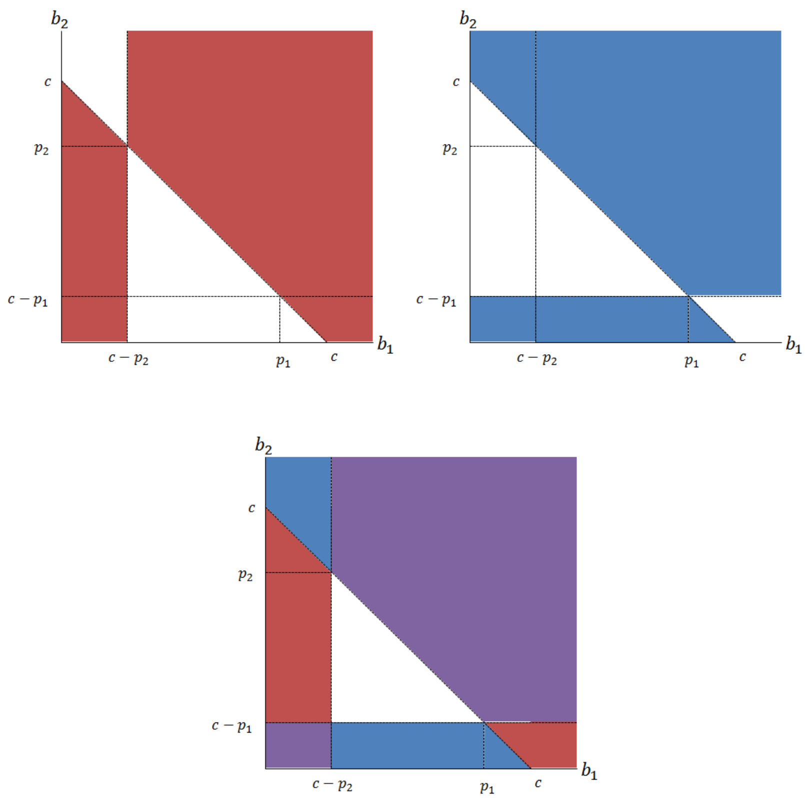
P_1's unconstrained problem is then

$$\max_{\hat{b}_1} p_1 e\left(\hat{b}_1, b_2\right) + b_2 e\left(\hat{b}_1, b_2\right) - ce\left(\hat{b}_1, b_2\right) - \bar{u} + s_2.$$

Since $\hat{b}_1$ only affects P_1's payoff inasmuch as it affects the Agent's effort choice, and since given any b_2 and any effort choice e, P_1 can choose a $\hat{b}_1$ so that the Agent will choose that effort choice. In other words, we can view Principal 1's problem as:

$$\max_{e \in \{0,1\}} (p_1 + b_2 - c)\, e.$$

When $p_1 + b_2 \geq c$, P_1 will choose b_1 to ensure that $e^*(b_1, b_2) = 1$. That is, when $b_2 \geq c - p_1$, it is a best response for P_1 to offer any contract (s_1, b_1) with $b_1 \geq c - b_2$ (and with s_1 such that the Agent's individual-rationality constraint holds with equality). When $p_1 + b_2 \leq c$, b_1 will be chosen to ensure that $e^*(b_1, b_2) = 0$. That is, when $b_2 \leq c - p_1$, it is a best response for P_1 to offer any contract (s_1, b_2) with $b_1 < c - b_2$. The following figures show best-response correspondences in this (b_1, b_2) space. In the first figure, the red regions represent the optimal choices of b_2 given a choice of b_1. In the second figure, the blue regions represent the optimal choices of b_1 given a choice of b_2. The third figure puts these together–the purple regions represent equilibrium contracts. Those equilibrium contracts in the upper-right region induce $e^* = 1$, while those in the lower-left region do not.



The set of equilibrium contracts is therefore any (s_1, b_1) and (s_2, b_2) such that either:

1. $b_1 \geq c - p_1$, $b_2 \geq c - p_2$ and $b_1 + b_2 \geq c$.

2. $0 \leq b_1 < c - p_1$ and $0 \leq b_2 < c - p_2$.

The first set of equilibrium contracts implement $e^* = 1$, while the second set of equilibrium contracts implement $e^* = 0$. Equilibrium contracts with $\{0 \leq b_i \leq c - p_i\}$ therefore represent a **coordination failure**.

Description of Free-Rider Version There is a risk-neutral Agent (A) and two risk-neutral Principals (P_1 and P_2). The Agent chooses an effort $e \in [0,1]$ at cost $\frac{c}{2}e^2$. Output is $y \in Y = \{0,1\}$ with $\Pr[y = 1|\, e] = e$. Principals 1 and 2 receive revenues $p_1 y$ and $p_2 y$, respectively. The Principals simultaneously offer contracts $w_1, w_2 \in W = \{w : Y \to \mathbb{R}\}$. Denote Principal i's contract offer by $w_i = s_i + b_i y$. If the Agent accepts a pair of contracts with total incentives $b = b_1 + b_2$, he incurs an additional cost $k \cdot b$. These costs are reduced-form, but we can think of them either as risk costs associated with higher-powered incentives or, if they were instead borne by the Principals, we could think of them as setup costs associated with writing higher-powered contracts (as in Battigalli and Maggi, 2002). The analysis would be similar in this latter case. The Agent has an outside option that yields utility $\bar{u}$ to the Agent and 0 to each Principal. If the outside option is not exercised, players' expected payoffs are:

$$
\begin{aligned}
\Pi_1(w_1, w_2, e) &= p_1 e - w_1 \\
\Pi_2(w_1, w_2, e) &= p_2 e - w_2 \\
U(w_1, w_2, e) &= w_1 + w_2 - \frac{c}{2}e^2 - k \cdot (b_1 + b_2)
\end{aligned}
$$

Timing The timing of the game is exactly the same as before.

1. P_1 and P_2 simultaneously offer contracts w_1, w_2 to A. Offers are commonly observed.

2. A accepts both contracts ($d = 1$) or rejects both contracts ($d = 0$) and receives $\bar{u}$ and the game ends. This decision is commonly observed.

3. If A accepts the contracts, he incurs cost $k \cdot (b_1 + b_2)$ and then chooses effort $e \in [0, 1]$ at cost $\frac{c}{2}e^2$. e is commonly observed.

4. Output $y \in Y$ is realized with $\Pr[y = 1| e] = e$. Output is commonly observed.

5. P_1 and P_2 pay A amounts $w_1(y), w_2(y)$. These payments are commonly observed.

Equilibrium A **pure-strategy subgame-perfect equilibrium** is a pair of contracts $w_1^*, w_2^* \in W$, an acceptance strategy $d^* : W^2 \to \{0, 1\}$, an effort strategy $e^* : W^2 \to \{0, 1\}$ such that given contracts w_1^*, w_2^*, A optimally chooses $d^*(w_1^*, w_2^*)$ and $e^*(w_1^*, w_2^*)$. Given d^*, e^* and w_2^*, P_1 optimally chooses w_1^*, and given d^*, e^*, and w_1^*, P_2 optimally chooses w_2^*.

The Program Given total incentives $b = b_1 + b_2$, A chooses effort e to solve

$$\max_e be - \frac{c}{2}e^2,$$

or $e^*(b) = \frac{b}{c}$. Suppose P_1 believes P_2 will offer contract (s_2, b_2). Then P_1's problem is to

$$\max_{\hat{b}_1, \hat{s}_1} p_1 e^*\left(\hat{b}_1 + b_2\right) - \hat{s}_1 - \hat{b}_1 e^*\left(\hat{b}_1 + b_2\right)$$

subject to A's individual-rationality constraint

$$\hat{s}_1 + \hat{b}_1 e^* \left(\hat{b}_1 + b_2\right) + s_2 + b_2 e^* \left(\hat{b}_1 + b_2\right) - \frac{c}{2} e^* \left(\hat{b}_1 + b_2\right)^2 - k \cdot \left(\hat{b}_1 + b_2\right) \geq \bar{u}.$$

As in the previous models, P_1 will choose s_1 so that this constraint holds with equality:

$$\hat{s}_1 + \hat{b}_1 e^* \left(\hat{b}_1 + b_2\right) = \bar{u} + \frac{c}{2} e^* \left(\hat{b}_1 + b_2\right)^2 + k \cdot \left(\hat{b}_1 + b_2\right) - s_2 - b_2 e^* \left(\hat{b}_1 + b_2\right).$$

P_1's unconstrained problem is then to

$$\max_{\hat{b}_1} p_1 e^* \left(\hat{b}_1 + b_2\right) + b_2 e^* \left(\hat{b}_1 + b_2\right) - \frac{c}{2} e^* \left(\hat{b}_1 + b_2\right)^2 - k \cdot \left(\hat{b}_1 + b_2\right),$$

which yields first-order conditions

$$\begin{aligned} 0 &= p_1 \frac{\partial e^*}{\partial \hat{b}_1} + b_2 \frac{\partial e^*}{\partial \hat{b}_1} - c e^* \left(b_1 + b_2\right) \frac{\partial e^*}{\partial \hat{b}_1} - k \\ &= \left(p_1 + b_2 - \left(b_1^* + b_2\right)\right) \frac{1}{c} - k \end{aligned}$$

so that $b_1^* = p_1 - ck$. This choice of b_1 is independent of b_2. Analogously, P_2 will choose a contract with $b_2^* = p_2 - ck$. The Agent's equilibrium effort will satisfy

$$e^* \left(b_1^* + b_2^*\right) = \frac{p}{c} - 2k.$$

If the two Principals could collude and offer a single contract $w = s + by$

to the agent, they would offer a contract that solves:

$$\max_b pe(b) - \frac{c}{2}e(b)^2 - kb,$$

where $e(b) = \frac{b}{c}$. The associated first-order conditions are

$$\frac{p}{c} - \frac{b^C}{c} = k$$

or

$$b^C = p - ck$$

and therefore equilibrium effort would be

$$e^*\left(b^C\right) = \frac{p}{c} - k.$$

In particular, $e^*\left(b^C\right) = e^*(b^*) + k > e^*(b^*)$. This effect is often referred to as the free-rider effect in common-agency models.

1.2 No Contracts

In many environments, contractible measures of performance may be so bad as to render them useless. Yet, aspects of performance that are relevant for the firm's objectives may be observable, but for whatever reason, they cannot be written into a formal contract that the firm can commit to. These aspects of performance may then form the basis for informal reward schemes. We will discuss two classes of models that build off this insight.

1.2.1 Career Concerns

An Agent's performance within a firm may be observable to outside market participants—for example, fund managers' returns are published in prospectuses, academics post their papers online publicly, a CEO's performance is partly announced in quarterly earnings reports. Holmstrom (1999) developed a model to show that in such an environment, even when formal performance-contingent contracts are impossible to write, workers may be motivated to work hard out of a desire to convince "the market" that they are intrinsically productive in the hopes that doing so will attract favorable outside offers in the future—that is, they are motivated by their own career concerns.

Description There are two risk-neutral Principals, whom we will denote by P_1 and P_2, and a risk-neutral Agent (A) who interact in periods $t = 1, 2$. The Agent has ability θ, which is drawn from a normal distribution, $\theta \sim N\left(m_0, h_0^{-1}\right)$. θ is unobservable by all players, but all players know the distribution from which it is drawn. In each period, the Agent chooses an effort level $e_t \in E$ at cost $c(e_t)$ (with $c(0) = c'(0) = 0 < c', c''$) that, together with his ability and luck (denoted by ε_t), determine his output $y_t \in Y$ as follows:

$$y_t = \theta + e_t + \varepsilon_t.$$

Luck is also normally distributed, $\varepsilon_t \sim N\left(0, h_\varepsilon^{-1}\right)$ and is independent across periods and independent from θ. This output accrues to whichever Principal employs the Agent in period t. At the beginning of each period, each Principal i offers the Agent a short-term contract $w_i \in W \subset \{w_i : M \to \mathbb{R}\}$, where M is the set of outcomes of a performance measure. The Agent has to accept one of the contracts, and if he accepts Principal i's contract in period t, then Principal $j \neq i$ receives 0 in period t. For now, we will assume that there are no available performance measures, so short-term contracts can only take the form of a constant wage.

Comment on Assumption. Do you think the assumption that the Agent does not know more about his own productivity than the Principals do is sensible?

If Principal P_i employs the Agent in period t, the agent chooses effort e_t, and output y_t is realized, payoffs are given by

$$\begin{aligned}
\pi_i\left(w_{it}, e_t, y_t\right) &= p y_t - w_{it} \\
\pi_j\left(w_{it}, e_t, y_t\right) &= 0 \\
u_i\left(w_{it}, e_t, y_t\right) &= w_{it} - c\left(e_t\right).
\end{aligned}$$

Players share a common discount factor of $\delta < 1$.

Timing There are two periods $t = 1, 2$. In each period, the following stage game is played:

1. P_1 and P_2 propose contracts w_{1t} and w_{2t}. These contracts are commonly observed.

2. A chooses one of the two contracts. The Agent's choice is commonly observed. If A chooses contract offered by P_i, denote his choice by $d_t = i$. The set of choices is denoted by $D = \{1, 2\}$.

3. A receives transfer w_{it}. This transfer is commonly observed.

4. A chooses effort e_t and incurs cost $c(e_t)$. e_t is only observed by A.

5. Output y_t is realized and accrues to P_i. y_t is commonly observed.

Equilibrium The solution concept is Perfect-Bayesian Equilibrium. A **Perfect-Bayesian Equilibrium** of this game consists of a strategy profile $\sigma^* = \left(\sigma^*_{P_1}, \sigma^*_{P_2}, \sigma^*_A\right)$ and a belief profile μ^* (defining beliefs of each player about the distribution of θ at each information set) such that σ^* is sequentially rational for each player given his beliefs (i.e., each player plays the best response at each information set given his beliefs) and μ^* is derived from σ^* using Bayes's rule whenever possible.

It is worth spelling out in more detail what the strategy space is. By doing so, we can get an appreciation for how complicated this seemingly simple environment is, and how different assumptions of the model contribute to simplifying the solution. Further, by understanding the role of the different assumptions, we will be able to get a sense for what directions the model could be extended without introducing great complexity.

Each Principal i chooses a pair of contract-offer strategies $w_{i1}^{*} : \Delta(\Theta) \to \mathbb{R}$ and $w_{i2}^{*} : W \times D \times Y \times \Delta(\Theta) \to \mathbb{R}$. The first-period offers depend only on each Principal's beliefs about the Agent's type (as well as their equilibrium conjectures about what the Agent will do). The second-period offer can also be conditioned on the first-period contract offerings, the Agent's first-period contract choice, and the Agent's first-period output. In equilibrium, it will be the case that these variables determine the second-period contract offers only inasmuch as they determine each Principal's beliefs about the Agent's type.

The Agent chooses a set of acceptance strategies in each period, $d_1 : W^2 \times \Delta(\Theta) \to \{1, 2\}$ and $d_2 : W^4 \times D \times E \times Y \times \Delta(\Theta) \to \{1, 2\}$ and a set of effort strategies $e_1 : W^2 \times D \times \Delta(\Theta) \to \mathbb{R}_+$ and $e_2 : W^4 \times D \times E \times Y \times \Delta(\Theta) \to \mathbb{R}_+$. In the first period, the agent chooses which contract to accept based on which ones are offered as well as his beliefs about his own type. In the present model, the contract space is not very rich (since it is only the set of scalars), so it will turn out that the Agent does not want to condition his acceptance decision on his beliefs about his own ability. This is not necessarily the case in richer models in which Principals are allowed to offer contracts involving performance-contingent payments. The Agent then chooses effort on the basis of which contracts were available, which one he chose, and his beliefs about his type. In the second period, his acceptance decision and effort choice can also be conditioned on events that occurred in the first period.

It will in fact be the case that this game has a unique Perfect-Bayesian

Equilibrium, and in this Perfect-Bayesian equilibrium, both the Principals and the Agent will use **public** strategies in which $w_{i1}^* : \Delta(\Theta) \rightarrow \mathbb{R}$, $w_{i2}^* : \Delta(\Theta) \rightarrow \mathbb{R}$, $d_1 : W^2 \rightarrow \{1,2\}$, $d_2 : W^2 \rightarrow \{1,2\}$, $e_1 \in \mathbb{R}_+$ and $e_2 \in \mathbb{R}_+$.

The Program Sequential rationality implies that the Agent will choose $e_2^* = 0$ in the second period, no matter what happened in previous periods. This is because no further actions or payments that the Agent will receive are affected by the Agent's effort choice in the second period. Given that the agent knows his effort choice will be the same no matter which contract he chooses, he will choose whichever contract offers him a higher payment.

In turn, the Principals will each offer a contract in which they earn zero expected profits. This is because they have the same beliefs about the Agent's ability. This is the case since they have the same prior and have seen the same public history, and in equilibrium, they have the same conjectures about the Agent's strategy and therefore infer the same information about the Agent's ability. As a result, if one Principal offers a contract that will yield him positive expected profits, the other Principal will offer a contract that pays the Agent slightly more, and the Agent will accept the latter contract. The second-period contracts offered will therefore be

$$w_{12}^*\left(\hat{\theta}(y_1)\right) = w_{22}^*\left(\hat{\theta}(y_1)\right) = w_2^*\left(\hat{\theta}(y_1)\right) = pE\left[y_2 | y_1, \sigma^*\right] = pE\left[\theta | y_1, \sigma^*\right],$$

where $\hat{\theta}(y_1)$ is the equilibrium conditional distribution of θ given realized output y_1.

If the agent chooses e_1 in period 1, first-period output will be $y_1 = \theta + e_1 + \varepsilon_1$. Given conjectured effort e_1^*, the Principals' beliefs about the Agent's ability will be based on two signals: their prior, and the signal $y_1 - e_1^* = \theta + \varepsilon_1$, which is also normally distributed with mean m_0 and variance $h_0^{-1} + h_\varepsilon^{-1}$. The joint distribution is therefore

$$\begin{bmatrix} \theta \\ \theta + \varepsilon_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \theta \\ \varepsilon_1 \end{bmatrix} \sim N\left(\begin{bmatrix} m_0 \\ m_0 \end{bmatrix}, \begin{bmatrix} h_0^{-1} & h_0^{-1} \\ h_0^{-1} & h_0^{-1} + h_\varepsilon^{-1} \end{bmatrix} \right)$$

Their beliefs about θ conditional on these signals will therefore be normally distributed:

$$\theta | \, y_1 \sim N\left(\varphi y_1 + (1 - \varphi)\, m_0, \frac{1}{h_\varepsilon + h_0} \right),$$

where $\varphi = \frac{h_\varepsilon}{h_0 + h_\varepsilon}$ is the signal-to-noise ratio. Here, we used the normal updating formula, which just to jog your memory is stated as follows. If X is a $K \times 1$ random vector and Y is an $N - K$ random vector, then if

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma'_{XY} & \Sigma_{YY} \end{bmatrix} \right),$$

then

$$X | \, Y = y \sim N\left(\mu_X + \Sigma_{XY} \Sigma_{YY}^{-1} (y - \mu_Y), \Sigma_{XX} - \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma'_{XY} \right).$$

Therefore, given output y_1, the Agent's second-period wage will be

$$w_2^*\left(\hat{\theta}\left(y_1\right)\right) = p\left[\varphi\left(y_1 - e_1^*\right) + \left(1-\varphi\right)m_0\right] = p\left[\varphi\left(\theta + e_1 + \varepsilon_1 - e_1^*\right) + \left(1-\varphi\right)m_0\right].$$

In the first period, the Agent chooses a non-zero effort level, even though his first-period contract does not provide him with performance-based compensation. He chooses a non-zero effort level, because doing so affects the distribution of output, which the Principals use in the second period to infer his ability. In equilibrium, of course, they are not fooled by his effort choice.

Given an arbitrary belief about his effort choice, $\hat{e}_1$, the signal the Principals use to update their beliefs about the Agent's type is $y_1 - \hat{e}_1 = \theta + \varepsilon_1 + e_1 - \hat{e}_1$. The agent's incentives to exert effort in the first period to shift the distribution of output are therefore the same no matter what the Principals conjecture his effort choice to be. He will therefore choose effort e_1^* in the first period to solve

$$\max_{e_1} -c\left(e_1\right) + \delta E_{y_1}\left[w_2^*\left(\hat{\theta}\left(y_1\right)\right)\middle| e_1\right] = \max_{e_1} -c\left(e_1\right) + \delta p\left(\varphi\left(\theta + e_1 - e_1^*\right) + \left(1-\varphi\right)m_0\right),$$

so that he will choose

$$c'\left(e_1^*\right) = p\delta\frac{h_\varepsilon}{h_0 + h_\varepsilon},$$

and if $c\left(e\right) = \frac{c}{2}e^2$,

$$e_1^* = \frac{p}{c}\delta\frac{h_\varepsilon}{h_0 + h_\varepsilon}.$$

This second-period effort choice is, of course, less than first-best, since first-best effort satisfies $c'\left(e_1^{FB}\right) = 1$ or $e_1^{FB} = p/c$. He will choose a higher effort level in the first period the less he discounts the future (δ larger), the more prior uncertainty there is about his type (h_0 small), and the more informative output is about his ability (h_ε large). Finally, given that the Agent will choose e_1^*, the first-period wages will be

$$w_{11}^* = w_{21}^* = pE\left[y_1\right] = p\left(m_0 + e_1^*\right).$$

This model has a number of nice features. First, despite the fact that the Agent receives no formal incentives, he still chooses a positive effort level, at least in the first period. Second, he does not choose first-best effort (indeed, in versions of the model with three or more periods, he may initially choose excessively high effort), even though there is perfect competition in the labor market for his services. When he accepts an offer, he cannot commit to choose a particular effort level, so competition does not necessarily generate efficiency when there are contracting frictions.

The model is remarkably tractable, despite being quite complicated. This is largely due to the fact that this is a symmetric-information game, so players neither infer nor communicate information about the agent's type when making choices. The functional-form choices are also aimed at ensuring that it not only starts out as a symmetric information game, but it also remains one as it progresses. At the end of the first period, if one of the Principals (say the one that the Agent worked for in the first period) learned more about the Agent's type than the other Principal did, then there would be asymmetric information at the wage-offering stage in the second period.

This model extends nicely to three or more periods. In such an extension, however, if the Agent's effort affected the variance of output, he would have more information about his type at the beginning of the second period than the Principals would. This is because he would have more information about the conditional variance of his own ability, because he knows what effort he chose. In turn, his choice of contract in the second period would be informative about what effort level he would be likely to choose in the second period, which would in turn influence the contract offerings. If ability and effort interact, and their interaction cannot be separated out from the noise with a simple transformation (e.g., if $y_t = \theta e_t + \varepsilon_t$), then the Agent would acquire private information about his marginal returns to effort, which would have a similar effect. For these reasons, the model has seen very little application to environments with more than two periods, except in a couple special cases (see Bonatti and Horner (forthcoming) for a recent example with public all-or-nothing learning).

Finally, if the Agent's effort choice affects the informativeness of the public signal (e.g., $\varepsilon_t \sim N\left(0, h_\varepsilon\left(e_t\right)^{-1}\right)$), then the model may generate multiple equilibria. In particular, the equilibrium condition for effort in the first period will be

$$c'\left(e_1^*\right) = p\delta \frac{h_\varepsilon\left(e_1^*\right)}{h_0 + h_\varepsilon\left(e_1^*\right)},$$

which may have multiple solutions if $h'_{\varepsilon}(e_t) > 0$. Intuitively, if the Principals believe that the Agent will not put in effort in $t = 1$, then they think the signal is not very informative, which means that they will not put much weight on it in their belief formation. As a result, the Agent indeed has little incentive to put in effort in period 1. In contrast, if the Principals believe the Agent will put in lots of effort in $t = 1$, then they think the signal will be informative, so they will put a lot of weight on it, and the Agent will therefore have strong incentives to exert effort.

Exercise. *Can the above model be extended in a straightforward way to environments with more than* 3 *periods if the Agent has imperfect recall regarding the effort level he chose in past periods?*

Further Reading Dewatripont, Jewitt, and Tirole (1999b) shows that when there are complementarities between effort and the informativeness of the agent's output, there may be multiple equilibria. Dewatripont, Jewitt, and Tirole (1999a) explore a more-general two-period model and examine the relationship between the information structure and the incentives the agent faces. They also highlight the difficulties in extending the model beyond two periods with general distributions, since, in general, asymmetric information arises on the equilibrium path. Bonatti and Horner (forthcoming) explore an alternative setting in which effort and the agent's ability are non-separable, but nevertheless, asymmetric information does not arise on the equilibrium path, in particular because their information structure features all-or-nothing learning. Cisternas (2016) sets up a tractable environment in which asymmetric information in fact arises on the equilibrium path.

The contracting space in the analysis above was very limited—principals could only offer short-term contracts specifying a fixed wage. Gibbons and Murphy (1992) allow for principals to offer (imperfect) short-term performance-based contracts. Such contracts are substitutes for career-concerns incentives and become more important later in a worker's career, as the market becomes less impressionable. In principle, we can think of the model above as characterizing the agent's incentives for a particular long-term contract—the contract implicitly provided by market competition when output is publicly observed. He, Wei, Yu, and Gao (2014) characterize the agent's incentives for general long-term contracts in a continuous-time version of this setting and derives optimal long-term contracts.

1.2.2 Relational Incentive Contracts

If an Agent's performance is commonly observed only by other members of his organization, or if the market is sure about his intrinsic productivity, then the career concerns motives above cannot serve as motivation. However, individuals may form long-term attachments with an organization. In such long-term relationships, goodwill can arise as an equilibrium phenomenon, and fear of shattering this goodwill can motivate individuals to perform well and to reward good performance. This intuition is captured in models of relational contracts (informal contracts enforced by relationships). An entire section of this course will be devoted to studying many of the issues that arise in such models, but for now we will look at the workhorse model in the literature to get some of the more general insights.

The workhorse model is an infinitely repeated Principal-Agent game with publicly observed actions. We will characterize the "optimal relational contract" as the equilibrium of the repeated game that either maximizes the Principal's equilibrium payoffs or the Principal and Agent's joint equilibrium payoffs. A couple comments are in order at this point. First, these are applied models of repeated games and therefore tend to focus on situations where the discount factor is not close to 1, asking questions like "how much effort can be sustained in equilibrium?"

Second, such models often have many equilibria, and therefore we will be taking a stance on equilibrium selection in their analysis. The criticism that such models have no predictive power is, as Kandori puts it "... misplaced if we regard the theory of repeated games as a theory of informal contracts. Just as anything can be enforced when the party agrees to sign a binding contract, in repeated games [many outcomes can be] sustained if players agree on an equilibrium. Enforceability of a wide range of outcomes is the essential property of effective contracts, formal or informal." (Kandori, 2008, p. 7) Put slightly differently, focusing on optimal contracts when discussing formal contract design is analogous to focusing on optimal relational contracts when discussing repeated principal-agent models. Our objective, therefore, will be to derive properties of *optimal* relational contracts.

Description A risk-neutral Principal and risk-neutral Agent interact repeatedly in periods $t = 0, 1, 2, \ldots$. In period t, the Agent chooses an effort level $e_t \in E$ at cost $c(e_t) = \frac{c}{2}e_t^2$ that determines output $y_t = e_t \in Y$, which accrues to the Principal. The output can be sold on the product market for price p. At the beginning of date t, the Principal proposes a compensation package to the agent. This compensation consists of a fixed salary s_t and a contingent payment $b_t : E \to \mathbb{R}$ (with positive values denoting a transfer from the Principal to the Agent and negative values denoting a transfer from the Agent to the Principal), which can depend on the Agent's effort choice. The Agent can accept the proposal (which we denote by $d_t = 1$) or reject it (which we denote by $d_t = 0$) in favor of an outside option that yields per-period utility $\bar{u}$ for the Agent and $\bar{\pi}$ for the Principal. If the Agent accepts the proposal, the Principal is legally compelled to pay the transfer s_t, but she is not legally compelled to pay the contingent payment b_t.

Timing The stage game has the following five stages

1. P makes A a proposal (b_t, s_t).
2. A accepts or rejects in favor of outside opportunity yielding $\bar{u}$ to A and $\bar{\pi}$ to P.
3. P pays A an amount s_t.
4. A chooses effort $\hat{e}_t$ at cost $c(\hat{e}_t)$, which is commonly observed.
5. P pays A a transfer $\hat{b}_t$.

Equilibrium The Principal is not legally required to make the promised payment b_t, so in a one-shot game, she would always choose $\hat{b}_t = 0$ (or analogously, if $b_t < 0$, the Agent is not legally required to pay b_t, so he would choose $\hat{b}_t = 0$). However, since the players are engaged in a long-term relationship and can therefore condition future play on this transfer, nonzero transfers can potentially be sustained as part of an equilibrium.

Whenever we consider repeated games, we will always try to spell out explicitly the variables that players can condition their behavior on. This exercise is tedious but important. Let $h_0^t = \left\{s_0, d_0, \hat{e}_0, \hat{b}_0, \ldots, s_{t-1}, d_{t-1}, \hat{e}_{t-1}, \hat{b}_{t-1}\right\}$ denote the history up to the beginning of date t. In this game, all variables are commonly observed, so the history up to date t is a public history. We will also adopt the notation $h_s^t = h^t \cup \{s_t\}$, $h_d^t = h_s^t \cup \{d_t\}$, and $h_e^t = h_d^t \cup \{\hat{e}_t\}$, so we can cleanly keep track of within-period histories. (If we analogously defined h_b^t, it would be the same as h_0^{t+1}, so we will refrain from doing so.) Finally, let $\mathcal{H}_0^t, \mathcal{H}_s^t, \mathcal{H}_d^t$, and $\mathcal{H}_e^t$ denote, respectively, the sets of such histories.

Following Levin (2003), we define a **relational contract** to be a complete plan for the relationship. It describes (1) the salary that the Principal should offer the Agent ($h_0^t \mapsto s_t$), (2) whether the Agent should accept the offer ($h_s^t \mapsto d_t$), (3) what effort level the Agent should choose ($h_d^t \mapsto \hat{e}_t$), and (4) what bonus payment the Principal should make $\left(h_e^t \mapsto \hat{b}_t\right)$. A relational contract is **self-enforcing** if it describes a subgame-perfect equilibrium of the repeated game. An **optimal relational contract** is a self-enforcing relational contract that yields higher equilibrium payoffs for the Principal than any other self-enforcing relational contract. It is important to note that a relational contract describes behavior on and off the equilibrium path.

Comment. *Early papers in the relational-contracting literature (Bull, 1987; MacLeod and Malcomson, 1989; Baker, Gibbons, and Murphy, 1994) referred to the equilibrium of the game instead as an implicit (as opposed to relational) contract. More recent papers eschew the term implicit, because the term "implicit contracts" has a connotation that seems to emphasize whether agreements are common knowledge, whereas the term "relational contracts" more clearly focuses on whether agreements are enforced formally or must be self-enforcing.*

The Program Though the stage game is relatively simple, and the game has a straightforward repeated structure, solving for the optimal relational contract should in principle seem like a daunting task. There are tons of things that the Principal and Agent can do in this game (the strategy space is quite rich), many of which are consistent with equilibrium play—there are lots of equilibria, some of which may have complicated dynamics. Our objective is to pick out, among all these equilibria, those that maximize the Principal's equilibrium payoffs.

Thankfully, there are several nice results (many of which are contained in Levin (2003) but have origins in the preceding literature) that make this task achievable. We will proceed in the following steps:

1. We will argue, along the lines of Abreu (1988), that the unique stage

game SPNE is an optimal punishment.

2. We will show that optimal reward schedules are "forcing." That is, they pay the Agent a certain amount if he chooses a particular effort level, and they revert to punishment otherwise. An optimal relational contract will involve an optimal reward scheme.

3. We will then show that distribution and efficiency can be separated out in the stage game. Ex ante transfers have to satisfy participation constraints, but they otherwise do not affect incentives or whether continuation payoffs are self-enforcing.

4. We will show that an optimal relational contract is sequentially optimal on the equilibrium path. Increasing future surplus is good for ex-ante surplus, which can be divided in any way, according to (3), and it improves the scope for incentives in the current period. Total future surplus is always maximized in an optimal relational contract, and since the game is a repeated game, this implies that total future surplus is therefore constant in an optimal relational contract.

5. We will then argue that we can restrict attention to stationary relational contracts. By (4), the total future surplus is constant in every period. Contemporaneous payments and the split of continuation payoffs are perfect substitutes for motivating effort provision and bonus payments and for participation. Therefore, we can restrict attention

to agreements that "settle up" contemporaneously rather than reward and punish with continuation payoffs.

6. We will then solve for the set of stationary relational contracts, which is not so complicated. This set will contain an optimal relational contract.

In my view, while the restriction to stationary relational contracts is helpful for being able to tractably characterize optimal relational contracts, the important economic insights are actually that the relational contract is sequentially optimal and how this result depends on the separation of distribution and efficiency. The separation of distribution and efficiency in turn depends on several assumptions: risk-neutrality, unrestricted and costless transfers, and a simple information structure. Later in the course, we will return to these issues and think about settings where one or more of these assumptions is not satisfied.

Step 1 is straightforward. In the unique SPNE of the stage game, the Principal never pays a positive bonus, the Agent exerts zero effort, and he rejects any offer the Principal makes. The associated payoffs are $\bar{u}$ for the Agent and $\bar{\pi}$ for the Principal. It is also straightforward to show that these are also the Agent's and Principal's maxmin payoffs, and therefore they constitute an optimal penal code (Abreu, 1988). Define $\bar{s} = \bar{u} + \bar{\pi}$ to be the outside surplus.

Next, consider a relational contract that specifies, in the initial period, payments w and $b(\hat{e})$, an effort level e, and continuation payoffs $u(\hat{e})$ and

$\pi(\hat{e})$. The equilibrium payoffs of this relational contract, if accepted are:

$$\begin{aligned} u &= (1-\delta)(w - c(e) + b(e)) + \delta u(e) \\ \pi &= (1-\delta)(p \cdot e - w - b(e)) + \delta \pi(e). \end{aligned}$$

Let $s = u + \pi$ be the equilibrium contract surplus. This relational contract is self-enforcing if the following four conditions are satisfied.

1. Participation:
$$u \geq \bar{u}, \pi \geq \bar{\pi}$$

2. Effort-IC:
$$e \in \underset{\hat{e}}{\operatorname{argmax}} \{(1-\delta)(-c(\hat{e}) + b(\hat{e})) + \delta u(\hat{e})\}$$

3. Payment:
$$\begin{aligned} (1-\delta)(-b(e)) + \delta \pi(e) &\geq \delta \bar{\pi} \\ (1-\delta) b(e) + \delta u(e) &\geq \delta \bar{u} \end{aligned}$$

4. Self-enforcing continuation contract: $u(e)$ and $\pi(e)$ correspond to a self-enforcing relational contract that will be initiated in the next period.

Step 2: Define the Agent's **reward schedule** under this relational contract by

$$R(\hat{e}) = b(\hat{e}) + \frac{\delta}{1-\delta} u(\hat{e}).$$

The Agent's no-reneging constraint implies that $R(\hat{e}) \geq \frac{\delta}{1-\delta}\bar{u}$ for all $\hat{e}$. Given a proposed effort level e, suppose there is some other effort level $\hat{e}$ such that $R(\hat{e}) > \frac{\delta}{1-\delta}\bar{u}$. Then we can define an alternative relational contract in which everything else is the same, but $\tilde{R}(\hat{e}) = R(\hat{e}) - \varepsilon$ for some $\varepsilon > 0$. The payment constraints remain satisfied, and the effort-IC constraint becomes easier to satisfy. Therefore, such a change makes it possible to weakly improve at least one player's equilibrium payoff. Therefore, it has to be that $R(\hat{e}) = \frac{\delta}{1-\delta}\bar{u}$ for all $\hat{e} \neq e$.

Step 3: Consider an alternative relational contract in which everything else is the same, but $\tilde{w} = w - \varepsilon$ for some $\varepsilon \neq 0$. This changes the equilibrium payoffs u, π to $\tilde{u}, \tilde{\pi}$ but not the joint surplus s. Further, it does not affect the effort-IC, the payment, or the self-enforcing continuation contract conditions. As long as $\tilde{u} \geq \bar{u}$ and $\tilde{\pi} \geq \bar{\pi}$, then the proposed relational contract is still self-enforcing.

Define the value s^* to be the maximum total surplus generated by any self-enforcing relational contract. The set of possible payoffs under a self-enforcing relational contract is then $\{(u, \pi) : u \geq \bar{u}, \pi \geq \bar{\pi}, u + \pi \leq s^*\}$. For a given relational contract to satisfy the self-enforcing continuation contract condition, it then has to be the case that for any equilibrium effort e,

$$(u(e), \pi(e)) \in \{(u, \pi) : u \geq \bar{u}, \pi \geq \bar{\pi}, u + \pi \leq s^*\}.$$

Step 4: Suppose the continuation relational contract satisfies $u(e)+\pi(e) < s^*$. Then $\pi(e)$ can be increased in a self-enforcing relational contract, holding everything else the same. Increasing $\pi(e)$ does not affect the effort-IC constraint, it relaxes both the Principal's participation and payment constraints, and it increases equilibrium surplus. The original relational contract is then not optimal. Therefore, any optimal relational contract has to satisfy $s(e) = u(e) + \pi(e) = s^*$.

Step 5: Suppose the proposed relational contract is optimal and generates surplus $s(e)$. By the previous step, it has to be the case that $s(e) = e - c(e) = s^*$. This in turn implies that optimal relational contracts involve the same effort choice, e^*, in each period. Now we want to construct an optimal relational contract that provides the same incentives for the agent to exert effort, for both players to pay promised bonus payments, and also yields continuation payoffs that are equal to equilibrium payoffs (i.e., not only is the action that is chosen the same in each period, but so are equilibrium payoffs). To do so, suppose an optimal relational contract involves reward scheme $R(\hat{e}) = \frac{\delta}{1-\delta}\bar{u}$ for $\hat{e} \neq e^*$ and

$$R(e^*) = b(e^*) + \frac{\delta}{1-\delta} u(e^*).$$

Now, consider an alternative reward scheme $\tilde{R}(e^*)$ that provides the same incentives to the agent but leaves him with a continuation payoff of u^*:

$$\tilde{R}(e^*) = \tilde{b}(e^*) + \frac{\delta}{1-\delta}u^* = R(e^*).$$

This reward scheme also leaves him with an equilibrium utility of u^*

$$\begin{aligned} u^* &= (1-\delta)(w - c(e^*) + b(e^*)) + \delta u(e^*) = (1-\delta)(w - c(e^*) + R(e^*)) \\ &= (1-\delta)\left(w - c(e^*) + \tilde{R}(e^*)\right) = (1-\delta)\left(w - c(e^*) + \tilde{b}(e^*)\right) + \delta u^*. \end{aligned}$$

Since $\bar{u} \leq u^* \leq s^* - \bar{\pi}$, this alternative relational contract also satisfies the participation constraints.

Further, this alternative relational contract also satisfies all payment constraints, since by construction,

$$\tilde{b}(e^*) + \frac{\delta}{1-\delta}u^* = b(e^*) + \frac{\delta}{1-\delta}u(e^*),$$

and this equality also implies the analogous equality for the Principal (since $s^* = u^* + \pi^*$ and $s^* = u(e^*) + \pi(e^*)$):

$$-\tilde{b}(e^*) + \frac{\delta}{1-\delta}\pi^* = -b(e^*) + \frac{\delta}{1-\delta}(\pi(e^*)).$$

Finally, the continuation payoffs are (u^*, π^*), which can themselves be part of this exact same self-enforcing relational contract initiated the following period.

Step 6: The last step allows us to set up a program that we can solve to find an optimal relational contract. A stationary effort level e generates total surplus $s = e - c(e)$. The Agent is willing to choose effort level e if he expected to be paid a bonus b satisfying

$$b + \frac{\delta}{1-\delta}(u - \bar{u}) \geq c(e).$$

That is, he will choose e as long as his effort costs are less than the bonus b and the change in his continuation payoff that he would experience if he did not choose effort level e. Similarly, the Principal is willing to pay a bonus b if

$$\frac{\delta}{1-\delta}(\pi - \bar{\pi}) \geq b.$$

A necessary condition for both of these inequalities to be satisfied is that

$$\frac{\delta}{1-\delta}(s - \bar{s}) \geq c(e).$$

This condition is also sufficient for an effort level e to be sustainable in a stationary relational contract, since if it is satisfied, there is a b such that the preceding two inequalities are satisfied. This pooled inequality is referred to as the **dynamic-enforcement constraint.**

The Program: Putting all this together, then, an optimal relational contract will involve an effort level that solves

$$\max_e pe - \frac{c}{2}e^2$$

subject to the dynamic-enforcement constraint:

$$\frac{\delta}{1-\delta}\left(pe - \frac{c}{2}e^2 - \bar{s}\right) \geq \frac{c}{2}e^2.$$

The first-best effort level $e^{FB} = \frac{p}{c}$ solves this problem as long as

$$\frac{\delta}{1-\delta}\left(pe^{FB} - \frac{c}{2}\left(e^{FB}\right)^2 - \bar{s}\right) \geq \frac{c}{2}\left(e^{FB}\right)^2,$$

or

$$\delta \geq \frac{p^2}{2p^2 - 2c\bar{s}}.$$

Otherwise, the optimal effort level e^* is the larger solution to the dynamic-enforcement constraint, when it holds with equality:

$$e^* = \frac{p}{c}\left(\delta + \sqrt{\frac{p^2\delta^2 - 2\delta\bar{s}c}{p^2}}\right).$$

For all $\delta < \frac{p^2}{2p^2-2c\bar{s}}$, $\delta + \sqrt{\frac{p^2\delta^2-2\delta\bar{s}c}{p^2}} < 1$, so $e^* < e^{FB}$.

Comment. *People not familiar or comfortable with these models often try to come up with ways to artificially generate commitment. For example, they might propose something along the lines of, "If the problem is that the*

Principal doesn't have the incentives to pay a large bonus when required to, why doesn't the Principal leave a pot of money with a third-party enforcer that she will lose if she doesn't pay the bonus?" This proposal seems somewhat compelling, except for the fact that it would only solve the problem if the third-party enforcer could withhold that pot of money from the Principal if and only if the Principal breaks her promise to the Agent. Of course, this would require that the third-party enforcer condition its behavior on whether the Principal and the Agent cooperate. If the third-party enforcer could do this, then the third-party enforcer could presumably also enforce a contract that conditions on these events as well, which would imply that cooperation is contractible. On the other hand, if the third-party enforcer cannot conditionally withhold the money from the Principal, then the Principal's reneging temptation will consist of the joint temptation to (*a*) *not pay the bonus she promised the agent and* (*b*) *recover the pot of money from the third-party enforcer.*

Further Reading The analysis in this section specializes Levin's (2003) analysis to a setting of perfect public monitoring and no private information about the marginal returns to effort. Levin (2003) shows that in a fairly general class of repeated environments with imperfect public monitoring, if an optimal relational contract exists, there is a stationary relational contract that is optimal. Further, the players' inability to commit to payments enters the program only through a dynamic enforcement constraint. Using these results, he is able to show how players' inability to commit to payments shapes optimal incentive contracts in moral-hazard settings and settings in which the agent has private information about his marginal returns to effort.

MacLeod and Malcomson (1998) show that the structure of payments in an optimal relational contract can take the form of contingent bonuses or efficiency wages. Baker, Gibbons, and Murphy (1994) show that formal contracts can complement relational contracts, but they can also crowd out relational contracts. We will explore a number of further issues related to relational-incentive contracts later in the course.

The motivation I gave above begins with the premise that formal contracts are simply not enforceable and asks what *equilibrium* arrangement is best for the parties involved. Another strand of the relational-contracting literature begins with the less-stark premise that formal contracts are costly (but not infinitely so) to write, and informal agreements are less costly (but again, are limited because they must be self-enforcing). Under this view, relational contracts are valuable, because they give parties the ability to adapt to changing circumstances without having to specify in advance just how they will adapt (Macaulay, 1963). Baker, Gibbons, and Murphy (2011) and Barron, Gibbons, Gil, and Murphy (2015) explore implications of relational *adaptation*, and the former paper also considers the question of when adaptation should be governed by a formal contract and when it should be governed through informal agreements.

Chapter 2

Decision Making in Organizations

In the first couple weeks of the class, we considered environments in which one party (the Agent) chose actions that affected the payoffs of another party (the Principal), and we asked how the Principal could motivate the Agent to choose actions that were more favorable for the Principal. We considered the use of formal incentive contracts, and we also looked at environments in which such contracts were unavailable. Throughout, however, we took as given that only the agent was able to make those decisions. In other words, **decision rights** were **inalienable**. In this part of the class, we will allow for the possibility that control can be transferred from one party to the other, and we will ask when one party or the other should possess these decision rights. Parts of this discussion will echo parts of the discussion on the boundaries of the firm, where asset allocation was tantamount to decision-rights allocation, but the trade-offs we will focus on here will be different. For now, the discussion will be focused on the topic of delegation, but in the future, I will also discuss hierarchies and decision-making processes.

2.1 Delegation

In the first couple weeks of the class, we considered environments in which one party (the Agent) chose actions that affected the payoffs of another party (the Principal), and we asked how the Principal could motivate the Agent to choose actions that were more favorable for the Principal. We considered the use of formal incentive contracts, and we also looked at environments in which such contracts were unavailable. Throughout, however, we took as given that only the agent was able to make those decisions. In other words, **decision rights** were **inalienable**. This week, we will allow for the possibility that control can be transferred from one party to the other, and we will ask when one party or the other should possess these decision rights. Parts of this discussion will echo parts of the discussion on the boundaries of the firm, where asset allocation was tantamount to decision-rights allocation, but the trade-offs we will focus on here will be different.

If in principle, important decisions could be made by the Principal, why would the Principal ever want to delegate such decisions to an Agent? In his book on the design of bureaucracies, James Q. Wilson concludes that

"In general, authority [decision rights] should be placed at the lowest level at which all essential elements of information are available." A Principal may therefore want to delegate to a better-informed Agent who knows more about what decisions are available or what their payoff consequences are. But delegation itself may be costly as well, because the Principal and the Agent may disagree about the ideal decision to be made. This conflict is resolved in different ways in different papers in the literature.

First, if the Principal can commit to a decision rule as a function of an announcement by the Agent, then the formal allocation of control is irrelevant. This mechanism-design approach to delegation (Holmstrom, 1984; Alonso and Matouschek, 2008; Frankel, *Forthcoming*) focuses on the idea that while control is irrelevant, implementable decision rules can be implemented via constrained delegation: the Principal delegates to the Agent, but the Agent is restricted to making decisions from a restricted "delegation set." The interesting results of these papers is their characterization of optimal delegation sets.

If the Principal cannot commit to a decision rule, then the allocation of control matters. The optimal allocation of control is determined by one of several trade-offs identified in the literature. The most direct trade-off that a Principal faces is the trade-off between a loss of control under delegation (since the Agent may not necessarily make decisions in the Principal's best interest) and a loss of information under centralization (since the Principal may not be able to act upon the Agent's information). This trade-off occurs even if the Agent is able to communicate his information to the Principal in a cheap-talk manner (Dessein, 2002). Next, if the Agent has to exert non-contractible effort in order to become informed, then his incentives to do so are greater if he is able to act upon that information: delegation improves incentives for information acquisition. There is therefore a trade-off between loss of control under delegation and loss of initiative under centralization (Aghion and Tirole, 1997).

The previous two trade-offs are only relevant if the preferences of the Principal and the Agent are at least somewhat well-aligned. Even if they are not, however, delegation can serve a role. It may be beneficial to promise the Agent future control as a reward for good decision making today in order to get the Agent to use his private information in a way that is beneficial for the Principal. There is therefore a dynamic trade-off between loss of information today and loss of control in the future (Li, Matouschek, and Powell, forthcoming; Lipnowski and Ramos, 2015).

2.1.1 Mechanism-Design Approach to Delegation

Description There is a Principal (P) and an Agent (A) and a single decision $d \in \mathbb{R}$ to be made. Both P and A would like the decision to be tailored to the state of the world, $s \in S$, which is privately observed only by A. The Principal selects (and commits to) a control-rights allocation $g \in \{P, A\}$, a mechanism (M, d), which consists of a message space M and a deterministic decision rule $d : M \to \mathbb{R}$, which selects a decision $d(m)$ as a function of a message $m \in M$ sent by the Agent, and a delegation set $D \subset \mathbb{R}$. If $g = P$, then P makes decisions according to $d(\cdot)$. If $g = A$, then A makes decision $d_A \in D \subset \mathbb{R}$. Players' preferences are given by

$$\begin{aligned} u_P(d, s) &= -(d-s)^2 \\ u_A(d, s) &= -(d - y_A(s))^2, \end{aligned}$$

where $y_A(\cdot)$ is strictly increasing in s. Given state of the world s, P would like the decision to be $d = s$, and A would like the decision to be $d = y_A(s)$. There are no transfers.

Timing The timing of the game is:

1. P chooses control-rights allocation $g \in \{P, A\}$, mechanism (M, d), and delegation set D. g, M, d, and D are commonly observed.

2. A privately observes s.

3. A sends message $m \in M$ and chooses $d_A \in D$, which are commonly observed.

4. If $g = P$, the resulting decision is $d = d(m)$. If $g = A$, the resulting decision is $d = d_A$.

Equilibrium A **pure-strategy subgame-perfect equilibrium** is a control-rights allocation g^*, a mechanism (M^*, d^*), a delegation set D^*, an announcement function $m^* : S \to M^*$, and a decision rule $d_A^* : S \to D^*$ such that given g^* and (M^*, d^*), the Agent optimally announces $m^*(s)$ and chooses $d_A^*(s)$ when the state of the world is s, and the Principal optimally chooses control-rights allocation g^*, mechanism (M^*, d^*), and delegation set D^*.

The Program The Principal chooses (g, M, d, D) to solve

$$\max_{g,M,d,D} \int_s \left[u_P\left(d\left(m^*(s)\right), s\right) 1_{g=P} + u_P\left(d_A^*(s), s\right) 1_{g=A}\right] dF(s)$$

subject to

$$m^*(s) \in \underset{m \in M}{\operatorname{argmax}} \int_s u_A\left(d(m), s\right) dF(s)$$

and

$$d_A^*(s) \in \underset{d \in D}{\operatorname{argmax}} \int_s u_A(d, s)\, dF(s).$$

Functional-Form Assumptions We will assume that $s \sim U[-1, 1]$ and $y_A(s) = \beta s$, where $\beta > 1/2$.

Outline of the Analysis I will begin by separating out the problem of choosing a mechanism (M, d) from the problem of choosing a delegation set D. Define

$$V^P = \max_{M,d} \int_s u_P\left(d\left(m^*(s)\right), s\right) dF(s)$$

subject to

$$m^*(s) \in \underset{m \in M}{\operatorname{argmax}} \int_s u_A\left(d(m), s\right) dF(s)$$

and define

$$V^{A} = \max_{D} \int_{s} u_{P}\left(d_{A}^{*}\left(s\right), s\right) dF\left(s\right)$$

subject to

$$d_{A}^{*}\left(s\right) \in \underset{d \in D}{\operatorname{argmax}} \int_{s} u_{A}\left(d, s\right) dF\left(s\right).$$

The Coasian program can then be written as

$$\max_{g} V^{g}.$$

I will now proceed in several steps, for the most part following the analysis of Alonso and Matouschek (2008).

1. First, I will show that under $g = P$, there is an analog of the revelation principle that simplifies the search for an optimal mechanism: it is without loss of generality to set $M = S$ and focus on incentive-compatible decision rules $d\left(s\right)$ that satisfy

$$u_{A}\left(d\left(s\right), s\right) \geq u_{A}\left(d\left(s'\right), s\right) \text{ for all } s', s \in S.$$

2. I will then show that all incentive-compatible decision rules have some nice properties.

3. Further, each incentive-compatible decision rule $d\left(s\right)$ is associated with a range $\tilde{D} = \left\{d\left(s\right) : s \in S\right\}$, and the incentive-compatibility condition

is equivalent to

$$u_A\left(d\left(s\right),s\right)\geq u_A\left(d',s\right)\text{ for all }d'\in\tilde{D}.$$

This result implies that **the allocation of control is irrelevant**. For any incentive-compatible direct mechanism (Θ, d), there is a delegation set D such that under either control-rights allocation g, the decision rule is the same: $d\left(s\right)=d_A\left(s\right)$, which implies that $V^A=V^P$. It is therefore without loss of generality to solve for the optimal delegation set D.

4. I will restrict attention to **interval delegation sets** $D=[d_L,d_H]$, which under the specific functional-form assumptions I have made, is indeed without loss of generality. The Principal's problem will then be to

$$\max_{d_L,d_H}\int_s u_P\left(d_A^*\left(s\right),s\right)dF\left(s\right)$$

subject to

$$u_A\left(d_A^*\left(s\right),s\right)\geq u_A\left(d,s\right)\text{ for all }d_L\leq d\leq d_H.$$

Step 1: Revelation Principle Given $g=P$, any choice (M,d) by the Principal implements some distribution over outcomes $\sigma\left(s\right)$, which may be a nontrivial distribution, since the Agent might be indifferent between sending two different messages that induce two different decisions. Since $y_A\left(s\right)$ is strictly increasing in s, it follows that $\sigma(s)$ must be increasing in s in the sense that if $d \in \operatorname{supp}\sigma(s)$ and $d' \in \operatorname{supp}\sigma(s')$ for $s > s'$, then $d > d'$. This distribution determines some expected payoffs (given state s) for the Principal:

$$\pi(s) = E_{\sigma(s)}\left[u_P(d(m), s)\right],$$

where the expectation is taken over the distribution over messages that induces $\sigma(s)$. For each s, take $\hat{d}(s) \in \operatorname{supp}\sigma(s)$ such that

$$u_P\left(\hat{d}(s), s\right) \geq \pi(s).$$

The associated direct mechanism (S, d) is well-defined, incentive-compatible, and weakly better for the Principal, so it is without loss of generality to focus on direct mechanisms.

Step 2: Properties of Incentive-Compatible Mechanisms The set of incentive-compatible direct mechanisms $d : S \to \mathbb{R}$ satisfies

$$u_A(d(s), s) \geq u_A(d(s'), s) \text{ for all } s, s' \in S.$$

or

$$|d(s) - y_A(s)| \leq |d(s') - y_A(s)| \text{ for all } s, s'.$$

This condition implies a couple properties of $d(\cdot)$, but the proofs establishing these properties are fairly involved (which correspond to Proposition 1 in

Melumad and Shibano (1991)), so I omit them here. First, $d(\cdot)$ must be weakly increasing, since $y_A(\cdot)$ is increasing. Next, if it is strictly increasing and continuous on an open interval (s_1, s_2), it must be the case that $d(s) = y_A(s)$ for all $s \in (s_1, s_2)$. Finally, if d is not continuous at s', then there must be a jump discontinuity such that

$$\lim_{s \uparrow s'} u_A(d(s), s') = \lim_{s \downarrow s'} u_A(d(s), s'),$$

and $d(s)$ will be flat in an interval to the left and to the right of s'.

Step 3: Control-Rights Allocation is Irrelevant For any direct mechanism d, we can define the range of the mechanism to be $\tilde{D} = \{d(s) : s \in S\}$. The incentive-compatibility condition is then equivalent to

$$u_A(d(s), s) \geq u_A(d', s) \text{ for all } d' \in \tilde{D}.$$

That is, given a state s, the Agent has to prefer decision $d(s)$ to any other decision that he could induce by any other announcement s'. Under $g = P$, choosing a decision rule $d(s)$ therefore amounts to choosing its range $\tilde{D}$ and allowing the Agent to choose his ideal decision $d \in \tilde{D}$. The Principal's problem is therefore identical under $g = P$ as under $g = A$, so that $V^P = V^A$. Therefore, the allocation of control rights is irrelevant when the Principal has commitment either to a decision rule or to formal constraints on the delegation set. It is therefore without loss of generality to solve for the optimal delegation set D, so the Principal's problem becomes

$$\max_{D} \int u_P\left(d_A^*\left(s\right), s\right) dF\left(s\right)$$

subject to

$$u_A\left(d_A^*\left(s\right), s\right) \geq u_A\left(d', s\right) \text{ for all } s \text{ and for all } d' \in D.$$

Step 4: Optimal Interval Delegation Under the specific functional-form assumptions I have made, it is without loss of generality to focus on interval delegation sets of the form $D = [d_L, d_H]$, where $d_L \leq d_H$ and d_L can be $-\infty$ and d_H can be $+\infty$ (this result is nontrivial and follows from Proposition 3 in Alonso and Matouschek (2008)). Any interval $[d_L, d_H]$ will be associated with an interval of states $[s_L, s_H] = [d_L/\beta, d_H/\beta]$ such that

$$d_A^*\left(s\right) = \begin{cases} d_L & s \leq s_L \\ \beta s & s_L < s < s_H \\ d_H & s \geq s_H. \end{cases}$$

The Principal's problem will then be to

$$\max_{d_L, d_H} \int_{-1}^{s_L} u_P\left(d_L, s\right) dF\left(s\right) + \int_{s_L}^{s_H} u_P\left(\beta s, s\right) dF\left(s\right) + \int_{s_H}^{1} u_P\left(d_H, s\right) dF\left(s\right)$$

or since $dF(s) = 1/2ds$, $s_L = d_L/\beta$ and $s_H = d_H/\beta$,

$$\max_{d_L, d_H} -\frac{1}{2}\left[\int_{-1}^{d_L/\beta} (d_L - s)^2 \, ds + \int_{d_L/\beta}^{d_H/\beta} (\beta s - s)^2 \, ds + \int_{d_H/\beta}^{1} (d_H - s)^2 \, ds\right].$$

Applying the Kuhn-Tucker conditions (using Leibniz's rule), with some effort, we get

$$d_L^* = \max\left\{-\frac{\beta}{2\beta - 1}, -1\right\}, d_H^* = \min\left\{\frac{\beta}{2\beta - 1}, 1\right\},$$

if interior.

It is worth noting that if $\beta = 1$, so that P and A are perfectly aligned, then $d_L^* = -1$ and $d_H^* = 1$. That is, the Principal does not constrain the Agent's choices if their ideal decisions coincide. If $\beta > 1$, $d_L^* > -1$ and $d_H^* < 1$. In this case, the Agent's ideal decision is more responsive to the state of the world than the Principal would like, and the only instrument the Principal has to reduce the sensitivity of the Agent's decision rule is to constrain his decision set.

Finally, if $\beta < 1$, then again $d_L^* = -1$ and $d_H^* = 1$. In this case, the Agent's ideal decision is not as responsive to the state of the world as the Principal would like, but the Principal cannot use interval delegation to make the Agent's decision rule more responsive to the state of the world. Alonso and Matouschek (2008) provide conditions under which the Principal may like to remove points from the Agent's delegation set precisely in order to make the Agent's decision rule more sensitive to the state of the world.

Exercise If in addition to a message-contingent decision rule, the Principal is able to commit to a set of message-contingent transfers, it will still be the case that the allocation of control is irrelevant. Show that this is the case. In doing so, assume that the Agent has an outside option that yields utility $\bar{u}$ and that the Principal makes a take-it-or-leave-it offer of a mechanism (M, d, t), where $d : M \to \mathbb{R}$ is a decision rule and $t : M \to \mathbb{R}$ is a set of transfers from the Principal to the Agent.

Further Reading Melumad and Shibano (1991) characterize the set of incentive-compatible mechanisms when transfers are not feasible. Alonso and Matouschek (2008) provide a complete characterization of optimal delegation sets in the model above with more general distributions and preferences. Optimal delegation sets need not be interval-delegation sets. Frankel (Forthcoming) and Frankel (2014) explore optimal delegation mechanisms when the Principal has to make many decisions. Frankel (Forthcoming) shows that simple "cap" mechanisms can be approximately optimal. Frankel (2014) shows that seemingly simple mechanisms can be optimal in a max-min sense when the Principal is uncertain about the Agent's preferences.

2.1.2 Loss of Control vs. Loss of Information

The result that the allocation of control rights is irrelevant under the mechanism-design approach to delegation depends importantly on the Principal's ability to commit. The picture changes significantly if the Principal is unable to commit to a message-contingent decision rule and she is unable to restrict the Agent's decisions through formal rules (i.e., she cannot force A to choose from a restricted delegation set). When this is the case, there will be a trade-off between the "loss of control" she experiences when delegating to the Agent who chooses his own ideal decision and the "loss of information" associated with making the decision herself. This section develops an elemental model highlighting this trade-off in a stark way.

Description There is a Principal (P) and an Agent (A) and a single decision $d \in \mathbb{R}$ to be made. Both P and A would like the decision to be tailored to the state of the world, $s \in S$, which is privately observed only by A. The Principal chooses a control-rights allocation $g \in \{P, A\}$. Under allocation g, player g makes the decision. Players' preferences are given by

$$\begin{aligned} u_P(d, s) &= -(d - s)^2 \\ u_A(d, s) &= -(d - y_A(s))^2, \end{aligned}$$

where $y_A(s) = \alpha + s$. Given state of the world s, P would like the decision to be $d = s$, and A would like the decision to be $d = \alpha + s$. There are no transfers. Assume $s \sim U[-1, 1]$.

Timing The timing of the game is:

1. P chooses control-rights allocation $g \in \{P, A\}$, which is commonly observed.

2. A privately observes s.

3. Under allocation g, player g chooses d.

Equilibrium A **pure-strategy subgame-perfect equilibrium** is a control-rights allocation g^*, a decision by the Principal, d_P^*, and a decision rule $d_A^* : S \to \mathbb{R}$ by the Agent such that given g, d_g^* is chosen optimally by player g.

The Program The Principal's problem is to

$$\max_{g \in \{P,A\}} E\left[u_P\left(d_g^*(s), s\right)\right],$$

where I denote $d_P^*(s) \equiv d_P^*$. It remains to calculate d_P^* and $d_A^*(s)$.

Under $g = A$, given s, A solves

$$\max_d -(d - (\alpha + s))^2,$$

so that $d_A^*(s) = \alpha + s$. Under $g = P$, P solves

$$\max_d E\left[-(d-s)^2\right],$$

so that $d_P^* = E[s] = 0$.

The Principal's payoffs under $g = P$ are

$$E\left[u_P\left(d_P^*, s\right)\right] = -E\left[s^2\right] = -Var\left(s\right).$$

When the Principal makes a decision without any information, she faces a loss that is related to her uncertainty about what the state of the world is. Under $g = A$, the Principal's payoffs are

$$E\left[u_P\left(d_A^*, s\right)\right] = -E\left[\left(\alpha + s - s\right)^2\right] = -\alpha^2.$$

When the Principal delegates, she can be sure that the Agent will tailor the decision to the state of the world, but given the state of the world, he will always choose a decision that differs from the Principal's ideal decision.

The Principal then wants to choose the control-rights allocation that leads to a smaller loss: she will make the decision herself if $Var\left(s\right) < \alpha^2$, and she will delegate to the Agent if $Var\left(s\right) > \alpha^2$. She therefore faces a **trade-off between "loss of control" under delegation the "loss of information" under centralization**.

In this model, if the Agent is not making the decision, he has no input into the decision-making process. If the Agent is informed about the decision, he will clearly have incentives to try to convey some of his private information to the Principal, since he could benefit if the Principal made some use of that information. Centralization with communication would therefore always dominate Centralization without communication (since the Principal could always ignore the Agent's messages). Going further, if the Agent perfectly reveals his information to the Principal through communication, then Centralization with communication would also always be better for the Principal than Delegation. This leaves open the question of whether allowing for communication by the Agent undermines the trade-off we have derived.

Dessein (2002) explores this question by developing a version of this model in which under $g = P$, the Agent is able to send a cheap-talk message about s to the Principal. As in Crawford and Sobel (1982), fully informative communication is not an equilibrium if $\alpha > 0$, but as long as α is not too large, some information can be communicated in equilibrium. When α is larger, the most informative cheap-talk equilibrium becomes less informative, so decision making under centralization becomes less sensitive to the Agent's private information. However, when α is larger, the costs associated with the loss of control under delegation are also higher.

It turns out that whenever α is low, so that decision making under centralization would be very responsive to the state of the world, delegation performs even better than centralization. When α is high so that decision making under centralization involves throwing away a lot of useful information, delegation performs even worse than centralization. In this sense, from the Principal's perspective, delegation is optimal when players are well-aligned, and centralization is optimal when they are not.

When communication is possible, there is still a nontrivial trade-off between "loss of control" under delegation and "loss of information" under centralization, but it holds for more subtle reasons. In particular, at $\alpha = 0$, the Principal is indifferent between centralization and decentralization. Increasing α slightly leads to a second-order "loss of control" cost under delegation since the Agent still makes nearly optimal decisions from the Principal's perspective. However, it leads to a first-order "loss of information" cost under centralization in the most informative cheap-talk equilibrium. This is why for low values of α, delegation is optimal. For sufficiently high values of α, there can be no informative communication. At this point, an increase in α increases the "loss of control" costs under delegation, but it does not lead to any additional "loss of information" costs under centralization (since no information is being communicated at that point). At some point, the former costs become sufficiently high that centralization is preferred.

Further Reading Alonso, Dessein, Matouschek (2008) and Rantakari (2008) explore a related trade-off in multidivisional organizations: the optimal decision-rights allocation trades off divisions' ability to adapt to their local state with their ability to coordinate with other divisions. This trade-off occurs even when divisions are able to communicate with each other (horizontal communication) and with a headquarters that cares about the sum of their payoffs (vertical communication). When coordinating the activities of the two divisions is very important, both horizontal communication and vertical communication improve, so it may nevertheless be optimal to decentralize control.

2.1.3 Loss of Control vs. Loss of Initiative

Model Description There is a risk-neutral Principal and a risk-neutral Agent who are involved in making a decision about a new project to be undertaken. The Principal decides who will have formal authority, $g \in G \equiv \{P, A\}$, for choosing the project. There are four potential projects the players can choose from, which I will denote by $k = 0, 1, 2, 3$. The $k = 0$ project is the **status-quo project** and yields low, known payoffs (which I will normalize to 0). Of the remaining three projects, one is a **third-rail project** (don't touch the third rail) that yields $-\infty$ for both players. The remaining two projects are **productive projects** and yield positive payoffs for both players. The projects can be summarized by four payoff pairs: (u_{P0}, u_{A0}), (u_{P1}, u_{A1}), (u_{P2}, u_{A2}), and (u_{P3}, u_{A3}). Assume $(u_{P0}, u_{A0}) = (0, 0)$ is commonly known by both players. With probability α the remaining three projects yield payoffs $(-\infty, -\infty)$, (B, b), and $(0, 0)$, and with probability $(1 - \alpha)$, they yield payoffs $(-\infty, -\infty)$, $(B, 0)$, and $(0, b)$. The players do not initially know which projects yield which payoffs. α is referred to as the **congruence parameter**, since it indexes the probability that players' ideal projects coincide.

The Agent chooses an effort level $e \in [0, 1]$ at cost $c(e)$, which is increasing and convex. With probability e, the Agent becomes fully informed about his payoffs from each of the three projects (but he remains uninformed about the Principal's payoffs). That is, he observes a signal $\sigma_A = (u_{A1}, u_{A2}, u_{A3})$. With probability $1 - e$, he remains uninformed about all payoffs from these projects. That is, he observes a null signal $\sigma_A = \emptyset$. The Principal becomes fully informed about her payoffs (observing signal $\sigma_P = (u_{P1}, u_{P2}, u_{P3})$) with probability E, and she is uninformed (observing signal $\sigma_P = \emptyset$) with probability $1 - E$. The players then simultaneously send messages $m_P, m_A \in M \equiv \{0, 1, 2, 3\}$ to each other. And the player with formal authority makes a decision $d \in D \equiv \{0, 1, 2, 3\}$.

Timing The timing is as follows:

1. P chooses the allocation of formal authority, $g \in G$, which is commonly observed.

2. A chooses $e \in [0, 1]$. Effort is privately observed.

3. P and A privately observe their signals $\sigma_P, \sigma_A \in \Sigma$.

4. P and A simultaneously send messages $m_P, m_A \in M$.

5. Whoever has control under g chooses $d \in D$.

Equilibrium A **perfect-Bayesian equilibrium** is set of beliefs μ, an allocation of formal authority, g^*, an effort decision $e^* : G \to [0, 1]$, message functions $m_P^* : G \times [0, 1] \times \Sigma \to M$ and $m_A^* : G \times [0, 1] \times \Sigma \to M$, a decision function $d^* : G \times \Sigma \times \mu \to D$ such that each player's strategy is optimal given their beliefs about project payoffs, and these beliefs are determined by Bayes's rule whenever possible. We will focus on the set of **most-informative equilibria**, which correspond to equilibria in which player j sends message $m_j = k^*$ where $u_{jk^*} > 0$ if player j is informed, and $m_j = 0$ otherwise.

The Program In a most-informative equilibrium in which $g = P$, the Principal makes the decision d that maximizes her expected payoffs given her beliefs. If $\sigma_A \neq \emptyset$, then $m_A = k^*$ where $u_{Ak^*} = b$. If $\sigma_P = \emptyset$, then P receives expected payoff αB if she chooses project k^*, she receives 0 if she chooses project 0, and she receives $-\infty$ if she chooses any other project. She will therefore choose project k^*. That is, even if she possesses formal authority, the Agent may possess **real authority** in the sense that she will rubber stamp a project proposal of his if she is uninformed. If $\sigma_P \neq \emptyset$, then P will choose whichever project yields her a payoff of B. Under P-formal authority, therefore, players' expected payoffs are

$$\begin{aligned} U_P &= EB + (1 - E)\, e\alpha B \\ U_A &= E\alpha b + (1 - E)\, eb - c(e). \end{aligned}$$

In period 2, anticipating this decision rule, A will choose e^{*P} such that

$$c'\left(e^{*P}\right) = (1 - E)\, b.$$

Under P-formal authority, the Principal therefore receives equilibrium payoffs

$$V^P = EB + (1 - E)\, e^{*P} \alpha B.$$

In a most-informative equilibrium in which $g = A$, the Agent makes the decision d that maximizes his expected payoffs given his beliefs. If $\sigma_P \neq \emptyset$, then $m_P = k^*$ where $u_{Pk^*} = B$. If $\sigma_A = \emptyset$, then A receives expected payoff αb if he chooses project k^*, 0 if he chooses project 0, and $-\infty$ if he chooses any other project. He will therefore choose project k^*. If $\sigma_A \neq \emptyset$, then A will choose whichever project yields himself a payoff of b. Under A-formal authority, therefore, players' expected payoffs are

$$\begin{aligned} U_P &= e\alpha B + (1-e) EB \\ U_A &= eb + (1-e) E\alpha b - c(e). \end{aligned}$$

In period 2, anticipating this decision rule, A will choose e^{*A} such that

$$\begin{aligned} c'\left(e^{*A}\right) &= b(1-E\alpha) = (1-E)b + (1-\alpha)Eb \\ &= c'\left(e^{*P}\right) + (1-\alpha)Eb. \end{aligned}$$

The Agent therefore chooses higher effort under A-formal authority than under P-formal authority. This is because under A-formal authority, the Agent is better able to tailor the project choice to his own private information, which therefore increases the returns to becoming informed. This is the sense in which (formal) delegation increases the agent's initiative.

Under A-formal authority, the Principal therefore receives equilibrium payoffs

$$\begin{aligned} V^A &= e^{*A}\alpha B + E\left(1 - e^{*A}\right) B \\ &= EB + (1 - E)\, e^{*A}\alpha B - E e^{*A} B\,(1 - \alpha)\,. \end{aligned}$$

The first two terms correspond to the two terms in V^P, except that e^{*P} has been replaced with e^{*A}. This represents the "increased initiative" gain from delegation. The third term, which is negative is the "loss of control" cost of delegation. With probability $E \cdot e^{*A}$, the Principal is informed about the ideal decision and would get B if she were making the decision, but the Agent is also informed, and since he has formal authority, he will choose his own preferred decision, which yields a payoff of B to the Principal only with probability α.

In period 1, the Principal will therefore choose an allocation of formal authority to

$$\max_{g \in \{P,A\}} V^g,$$

and A-formal authority (i.e., delegation) is preferred if and only if

$$\underbrace{(1 - E)\,\alpha B \left(e^{*A} - e^{*P}\right)}_{\text{increased initiative}} \geq \underbrace{E B e^{*A}\,(1 - \alpha)}_{\text{loss of control}}.$$

That is, the Principal prefers A-formal authority whenever the increase in initiative it inspires outweighs the costs of ceding control to the Agent.

Discussion This paper is perhaps best known for its distinction between formal authority (who has the legal right to make a decision within the firm) and real authority (who is the actual decision maker), which is an interesting and important distinction to make. The model clearly highlights why those with formal authority might cede real authority to others: if our preferences are sufficiently well-aligned, then I will go with your proposal if I do not have any better ideas, because the alternative is inaction or disaster. Real authority is therefore a form of informational authority. Consequently, you have incentives to come up with good ideas and to tell me about them.

One important issue that I have not discussed either here or in the discussion of the "loss of control vs. loss of information" trade-off is the idea that decision making authority in organizations is unlikely to be formally transferable. Formal authority in firms always resides at the top of the hierarchy, and it cannot be delegated in a legally binding manner. As a result, under A-formal authority, it seems unlikely that the Agent will succeed in implementing a project that is good for himself but bad for the Principal if the Principal knows that there is another project that she prefers. That is, when both players are informed, if they disagree about the right course of action, the Principal will get her way. Baker, Gibbons, and Murphy (1999) colorfully point out that within firms, "decision rights [are] loaned, not owned," (p. 56) and they examine to what extent informal promises to relinquish control to an agent (what Li, Matouschek, and Powell (forthcoming) call "power") can be made credible.

2.2 Hierarchies (TBA)

TBA

Chapter 3

Careers in Organizations

As we have seen in the past few weeks, treating the firm as a “black box” has simplistic implications for firm behavior and for the supply side of the economy as a whole. This treatment further has simplistic implications (and in some empirically relevant dimensions, essentially no implications) for the labor side of the economy and in particular, for workers’ careers. In an anonymous spot market for labor, individual workers have upward-sloping labor-supply curves, individual firms have downward-sloping labor-demand curves, and equilibrium wages ensure that the total amount of labor supplied in a given period is equal to the total amount of labor demanded in that period. Workers are indifferent among potential employers at the equilibrium wage, so the approach is silent on worker–firm attachment. Workers’ wages are determined by the intersection of labor supply and labor demand, so the approach predicts that variation over time in a worker’s wage is driven by aggregate changes in labor supply or labor demand. And further, the approach is agnostic about what exactly the workers do for their employers, so this approach cannot capture notions such as job assignment and promotions.

In this note, I will introduce some natural modeling elements that enrich both the labor-demand and labor-supply sides of the equation to generate predictions about the dynamics of workers' careers, job assignments, and wages.

3.1 Internal Labor Markets

Doeringer and Piore (1971) define an **internal labor market** as an administrative unit "within which the pricing and allocation of labor is governed by a set of administrative rules and procedures" rather than being determined solely by market forces. Several empirical studies using firms' personnel data (with Baker, Gibbs, and Holmstrom (1994ab) being the focal study) highlight a number of facts regarding the operation of internal labor markets that would not arise in an anonymous spot market for labor. These facts include:

1. Many workers begin employment at the firm at a small number of positions. Doeringer and Piore refer to such positions as **ports of entry**.

2. Long-term employment relationships are common.

3. Nominal wage decreases and demotions are rare (but real wage de-

creases are not).

4. Workers who are promoted early on in their tenure at a firm are likely to be promoted to the next level quicker than others who were not initially promoted quickly. That is, promotions tend to be serially correlated.

5. Wage increases are serially correlated.

6. Promotions tend to be associated with large wage increases, but these wage differences are small relative to the average wage differences across levels within the firm.

7. Large wage increases early on in a worker's tenure predict promotions.

8. There is a positive relationship between seniority and wages but no relationship between seniority and job performance or wages and contemporaneous job performance.

In addition, there are many other facts regarding the use of particular and peculiar personnel policies. For example, prior to the 1980s in the U.S., many firms made use of mandatory-retirement policies in which workers beyond a certain age were required to retire, and the firms were required to dismiss these workers. Another common policy is the use of up-or-out promotion policies, of which academics are all-too-aware. All of this is to say that the personnel policies that firms put in place are much richer and much more systematic than would be expected in an anonymous spot market for labor, and several of these facts are consistent with workers' careers being managed at the firm-, rather than the individual-worker-, level through firm-wide policies.

There is a large and interesting theoretical literature proposing enrichments of the labor-demand or labor-supply side that in isolation generate predictions consistent with several (but typically not all) of the above features. In this note, I will focus on only a couple of the models from this literature. The models I focus on are not representative, though they do highlight a number of economic forces that are both natural and common in the literature.

3.1.1 Job Assignment and Human Capital Acquisition

The model in this section is based on Gibbons and Waldman (1999), and it introduces a number of important ingredients into an otherwise-standard model in order to capture many of the facts described above. First, in order for the notion of a "promotion" to be well-defined, it has to be the case that the firm has multiple jobs and reasons for assigning different workers to different jobs. In the two models I will describe here, workers in different jobs perform different activities (though in other models, such as Malcomson's (1984), this is not the case). Moreover, the models introduce heterogeneity among workers (i.e., worker "ability") and human-capital acquisition. The two models differ in (1) how firms other than the worker's current employer draw inferences about the worker's ability and (2) the nature of human-capital acquisition.

Description There are two risk-neutral firms, F_0 and F_1, a single risk-neutral agent A, and two periods of production. The worker's ability $\theta \in \{\theta_L, \theta_H\}$, with $\theta_L < \theta_H$ and $\Pr[\theta = \theta_H] = p$, and his work experience ℓ determine his **effective ability in period** t, $\eta_t = \theta f(\ell)$, where $f(\ell) = 1 + g\ell$, $g > 0$, and $\ell = 0$ in the first period of production and $\ell = 1$ in the second period of production. In each period, the agent can perform one of two activities for the firm that employs him. Activity 0 produces output $q^0 = d^0 + b^0(\eta_t + \varepsilon_t)$ and activity 1 produces output $q^1 = d^1 + b^1(\eta_t + \varepsilon_t)$, where $d^0 > d^1 > 0$ and $0 < b^0 < b^1$, so that output in activity 1 is more sensitive to a worker's effective ability η and mean 0 random noise ε_t than is output in activity 0. Denote the agent's activity assignment in period t by $j_t \in \{0, 1\}$. Output in period t is therefore $q_t = (1 - j_t) q^0 + j_t q^1$. The agent's ability is symmetrically unknown, and at the end of the first period of production, both firms observe a signal $\varphi_1 \in \Phi_1 = \{q_1, j_1\}$ from which they draw an inference about η. I further assume that at the beginning of the first period of production, both firms observe $\varphi_0 \in \Phi_0 \subset \{\eta\}$. This formulation allows for the complete-information case (if $\varphi_0 = \eta$), which I will use as a benchmark. The worker's utility is

$$u_A = w_1 + w_2,$$

where w_t is his period-t wage. Firm F_i's profits in period t are

$$\pi_{it} = q_t - w_t$$

if the agent works for F_i and 0 otherwise.

Timing The timing of the model is as follows.

1. $\theta \in \{\theta_L, \theta_H\}$ is drawn and is unobserved. φ_0 is publicly observed.

2. F_0 and F_1 simultaneously offer wages w_1^0, w_1^1 to A.

3. A chooses $d_1 \in \{0, 1\}$, where d_1 is the identity of his first-period employer, and he receives wage $w_1^{d_1}$ from F_{d_1}. Without loss of generality, assume $d_1 = 1$ (or else we can just relabel the firms).

4. F_{d_1} chooses an activity assignment $j_1 \in \{0, 1\}$, output q_1 is realized and accrues to F_1, and both firms observe the public signal φ_1.

5. F_0 and F_1 simultaneously offer wages w_2^0, w_2^1 to A.

6. A chooses $d_2 \in \{0, 1\}$, where d_2 is the identity of his second-period employer, and he receives wage $w_2^{d_2}$ from F_{d_2}. Assume that if A is indifferent, he chooses $d_2 = 1$.

7. F_{d_2} chooses an activity assignment $j_2 \in \{0, 1\}$. Output q_2 accrues to F_{d_2}.

Solution Concept A subgame-perfect equilibrium is a set of first-period wage offers $w_1^{0*} : \Phi_0 \to \mathbb{R}, w_1^{1*} : \Phi_0 \to \mathbb{R}$, a first-period acceptance decision rule $d_1^* : \mathbb{R}^2 \to \{0,1\}$, a first-period job-assignment rule $j_1^{d_1*} : \mathbb{R}^2 \times \{0,1\} \to \{0,1\}$, second-period wage offers $w_2^{1*} : \Phi_0 \times \mathbb{R}^2 \times \{0,1\} \times \Phi_1 \to \mathbb{R}$ and $w_2^{2*} : \Phi_0 \times \mathbb{R}^2 \times \{0,1\} \times \Phi_1 \to \mathbb{R}$, a second-period acceptance decision $d_2^* : \mathbb{R}^2 \times \{0,1\} \times \mathbb{R}^2 \to \{0,1\}$, and a second-period job assignment rule $j_2^{d_2*} : \Phi_0 \times \mathbb{R}^2 \times \{0,1\} \times \Phi_1 \times \mathbb{R}^2 \times \{0,1\} \to \{0,1\}$ such that each player's decision is sequentially optimal. The agent is said to be **promoted** if $j_2^{d_2*} > j_1^{d_1*}$, and he is said to be **demoted** if $j_2^{d_2*} < j_1^{d_1*}$.

Analysis In the second period, the agent optimally chooses to work for whichever firm offers him a higher second-period wage w_2. In fact, both firms will offer the agent the same wage, so the agent will work for F_1 in the second period. This second-period wage will depend on the expected output the agent would produce for F_0, given that F_0 infers something about the agent's ability θ from the public signal $\varphi = (\varphi_0, \varphi_1)$. Define the quantity $\eta_2^e(\varphi) = E[\eta_2 | \varphi]$

$$\begin{aligned} w_2^*(\varphi) &= E\left[\left(1 - j_2^{0*}\right) q^0 + j_2^{0*} q^1 \middle| \varphi\right] \\ &= \left(1 - j_2^{0*}\right)\left(d^0 + b^0 \eta_2^e(\varphi)\right) + j_2^{0*}\left(d^1 + b^1 \eta_2^e(\varphi)\right). \end{aligned}$$

In any subgame-perfect equilibrium, both firms will choose $w_2^{i*} = w_2^*(\varphi)$. To see why, suppose the second-period wage vector $(w_2^1, w_2^2) \neq (w_2^*(\varphi), w_2^*(\varphi))$

is an equilibrium. Then if $w_2^1 < w_2^*(\varphi)$, F_2 can always profitably deviate to some $w \in (w_2^1, w_2^*(\varphi))$. If $w_2^1 > w_2^*(\varphi)$, F_1 can profitably deviate by setting $w = \max\{w_2^*(\varphi), w_2^2\}$.

Given that both firms will choose the same wage in the second period, given the public signal φ, the agent will work for F_1 in the second period. He will be assigned to activity 1 if

$$d^1 + b^1\eta_2^e(\varphi) \geq d^0 + b^0\eta_2^e(\varphi)$$

or if his expected ability is sufficiently high

$$\eta_2^e(\varphi) \geq \bar{\eta}^e \equiv \frac{d^0 - d^1}{b^1 - b^0} > 0,$$

and he will be assigned to activity 0 otherwise. Figure 1 plots $E[q^0]$ and $E[q^1]$ as a function of η^e and depicts why this activity assignment rule is optimal.

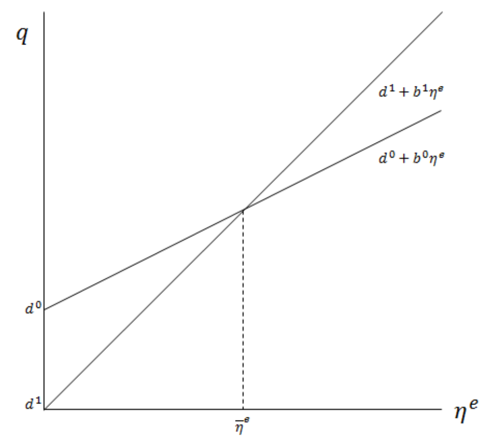


Figure 1

The first period of production is similar to the second. The agent optimally chooses to work for whichever firm offers him a higher first-period wage w_1, and indeed both firms will offer him the same wage, so without loss of generality, we assume he works for F_1. Again, his first-period wage depends on the expected output he would produce for F_0 given firms' prior knowledge about θ. Define $\eta_1^e(\varphi_0) = E[\eta_1 | \varphi_0]$. His first-period wage is given by

$$\begin{aligned} w_1^* &= E\left[\left(1 - j_1^{0*}\right) q^0 + j_1^{0*} q^1\right] \\ &= \left(1 - j_1^{0*}\right)\left(d^0 + b^0 \eta_1^e(\varphi_0)\right) + j_1^{0*}\left(d^1 + b^1 \eta_1^e(\varphi_0)\right), \end{aligned}$$

and again, his first-period employer will optimally assign him to activity 1 if and only if

$$\eta_1^e(\varphi_0) \geq \bar{\eta}^e = \frac{d^0 - d^1}{b^1 - b^0}.$$

Importantly, threshold is the same in each period, even though $E\left[\eta_2^e\left(\varphi\right)\right] = \left(1+g\right)\eta_1^e\left(\varphi_0\right) > \eta_1^e\left(\varphi_0\right)$.

Discussion Slight extensions of this model generate a number of predictions that are consistent with several of the facts I outlined in the discussion above. First, if p is sufficiently low, then all workers begin their employment spell by performing activity 1, which therefore serves as a port of entry into the firm. Long-term employment relationships are common, although this result follows because of the particular tie-breaking rule I have assumed—as we will see in the next model, if human capital acquisition is firm-specific rather than general, long-term employment relationships would arise for other tie-breaking rules as well.

Next, demotions are rare in this model. To see why, suppose that it is optimal to assign the agent to activity 1 in the first period. That is,

$$\hat{p}\left(\varphi_0\right)\theta^H + \left(1-\hat{p}\left(\varphi_0\right)\right)\theta^L \geq \frac{d^0-d^1}{b^1-b^0},$$

where $\hat{p}\left(\varphi_0\right)$ is the conditional probability that $\theta = \theta^H$ given public signal φ_0. In order for the agent to be demoted in period 2, if we denote by $\hat{p}\left(\varphi\right)$ the conditional probability that $\theta = \theta^H$ given public signal φ, it must be the case that

$$\hat{p}\left(\varphi\right)\theta^H + \left(1-\hat{p}\left(\varphi\right)\right)\theta^L \geq \frac{1}{1+g}\frac{d^0-d^1}{b^1-b^0}.$$

If $\varphi_0 = \eta$, so that we are in a complete-information environment, then workers are never demoted, because $\hat{p}(\varphi) = \hat{p}(\varphi_0)$. If $\varphi_0 = \emptyset$, so that we are in a symmetric-learning environment, then demotions are rare, because $E[\hat{p}(\varphi)] = p$, so that in expectation the left-hand side of the second-period cutoff is the same as the left-hand side of the first-period cutoff, but the right-hand side is strictly smaller. Wage cuts are also rare for the same reason.

The model also generates the prediction that promotions tend to be associated with especially large wage increases. This is true for both the complete-information and the symmetric-learning versions of the model. In the complete-information model, the wage increase for a worker conditional on not being promoted (i.e., if the parameters were such that the worker is optimally assigned to activity 0 in both periods) is $\theta b^0 g$ (since $w_2 = d^0 + b^0\theta(1+g)$ and $w_1 = d^0 + b^0\theta$). Analogously, the wage increase for a worker conditional on being promoted is $d^1 - d^0 + \theta(b^1 - b^0) + \theta b^1 g$. Since the worker is optimally being promoted, it has to be the case that $d^1 - d^0 + \theta(b^1 - b^0) > 0$, so this wage increase exceeds $\theta b^1 g$, which is certainly higher than $\theta b^0 g$ conditional on θ. Moreover, for it to be optimal to promote some workers but not all workers, it must be the case that the promoted workers have $\theta = \theta^H$, and the workers who are not promoted have $\theta = \theta^L$, further widening the difference in wage increases. This justification for wage jumps at promotion is a bit unsatisfying, and this is an issue that the model in the next section is partly designed to address.

With only two periods of production and two activities, it is not possible for the model to deliver serially correlated wage increases and promotions, but with more periods and more activities, it is.

3.1.2 Promotions as Signals

Description There are two firms, F_0 and F_1, a single agent A, and two periods of production. In each period, the agent can perform one of two activities for the firm that employs him. Activity 0 produces output that is independent of the agent's ability θ, and activity 1 produces output that is increasing in his ability. Output is sold into a competitive product market at price 1. The agent's ability is $\theta \sim U[0,1]$, and it is symmetrically unknown at the beginning of the game, but it is observed at the end of the first period of production by the agent's first-period employer but not by the other firm. The other firm will infer something about the worker's ability by his first-period employer's decision about his second-period activity assignment: promotions will therefore serve as a signal to the market. The agent acquires firm-specific human capital for his first-period employer.

In the first period, the worker produces an amount $q_1 = x \in (1/2, 1)$ for his employer if he is assigned to activity 0 and $q_1 = \theta$ if he is assigned to activity 1, so that his first-period employer will always assign him to activity 0, since $E[\theta] < 1/2$. In the second period, if he is assigned to activity j, he produces $q_2(j, \theta, d_2) = (1 + s1_{d_2=d_1})[(1-j)x + j\theta]$, where $1_{d_2=d_1}$ is an indicator variable for the event that the worker works for the same firm in both periods and $s \geq 0$ represents firm-specific human capital. The worker's utility is

$$u_A = w_1 + w_2,$$

where w_t is his period-t wage. Firm F_i's profits in period t are

$$\pi_{it} = q_t - w_t$$

if the agent works for F_i and 0 otherwise.

Timing The timing of the model is as follows.

1. F_0 and F_1 simultaneously offer wages w_1^0, w_1^1 to A.

2. A chooses $d_1 \in \{0, 1\}$, where d_1 is the identity of his first-period employer, and he receives wage $w_1^{d_1}$ from F_{d_1}. Without loss of generality, assume $d_1 = 1$ (or else we can just relabel the firms).

3. $\theta \sim U[0, 1]$ is drawn. θ is observed by F_1. Output q_1 is realized and accrues to F_1.

4. F_1 offers A a pair (j^1, w_2^1) consisting of a second-period activity assignment $j^1 \in \{0, 1\}$ and a second-period wage. j^1 is commonly observed, but w_2^1 is not.

5. F_0 offers A a pair (j^0, w_2^0). This offer is observed by A.

6. A chooses $d_2 \in \{0, 1\}$, where d_2 is the identity of his second-period employer, and he receives wage $w_2^{d_2}$ from F_{d_2}. Assume that if A is

indifferent, he chooses $d_2 = 1$.

7. Output $q_2(j, \theta, d_2)$ accrues to F_{d_2}.

Solution Concept A Perfect-Bayesian equilibrium (PBE) is a belief assessment μ, first-period wage offers $w_1^{0*}, w_1^{1*} \in \mathbb{R}$, a first-period acceptance decision rule $d_1^* : \mathbb{R}^2 \to \{0,1\}$, a second-period job assignment $j^{1*} : \mathbb{R}^2 \times \{0,1\} \times [0,1] \to \{0,1\}$ and wage offer $w_2^{1*} : \mathbb{R}^2 \times \{0,1\} \times [0,1] \to \mathbb{R}^2$ by F_1, a second-period offer $(j^{0*}, w_2^{0*}) : \mathbb{R}^2 \times \{0,1\} \times \{0,1\} \to \{0,1\} \times \mathbb{R}$ by F_0, and a second-period acceptance decision $d_2^* : \mathbb{R}^2 \times \{0,1\} \times \{0,1\}^2 \times \mathbb{R}^2 \to \{0,1\}$ such that each player's decision is sequentially optimal, and beliefs are consistent with Bayes's rule whenever possible. A **promotion rule** is a mapping from θ to $\{0,1\}$. Firm F_1's optimal promotion rule will turn out to be a threshold promotion rule, which greatly simplifies the analysis.

Analysis In the second period, the agent optimally chooses to work for whichever firm offers him a higher second-period wage w_2. In fact, in every PBE, both firms will offer the agent the same wage, so the agent will work for F_1 in the second period. This second-period wage will, however, depend on the expected output the agent would produce for F_0, given that F_0 infers something about θ from F_1's second-period activity-assignment decision. Define the quantity

$$w_2^*\left(j^1\right) = E\left[\left(1 - j^{0*}\right)x + j^{0*}\theta \middle| j^{1*}(\theta) = j^1\right].$$

$w_2^*(j^1)$ is equal to the expected output of F_0 if it employs A in the second period, given F_1's equilibrium promotion rule and its outcome j^1.

Result 1. In any PBE $w_2^{1*} = w_2^{2*} = w_2^*(j^1)$.

Proof of Result 1. If $w_2^{1*} < w_2^*(j^1)$, then F_0 will optimally choose some $w_2^0 \in (w_2^{1*}, w_2^*)$. Such a wage offer will ensure that the worker will work for F_0 and that F_0 will earn strictly positive profits. If $w_2^{1*} \geq w_2^*(j^1)$, then it is optimal for F_0 to choose $w_2^{0*} = w_2^*$. In any equilibrium, given a choice j^1, F_1 chooses $w_2^{1*} = w_2^*(j^1)$. If $w_2^{1*} < w_2^*(j^1)$, firm F_0 would optimally choose some wage in between w_2^{1*} and $w_2^*(j^1)$, and F_1 would earn zero profits, but F_1 could guarantee itself strictly positive profits by deviating to $w_2^1 = w_2^*(j^1)$, because A would then choose $d_2 = 1$, and $w_2^*(j^1)$ is strictly less than F_1's expected output in period 2, because of firm-specific human capital. If $w_2^{1*} > w_2^*(j^1)$, then F_1 could increase its profits by deviating to any $w_2^1 \in (w_2^*(j^1), w_2^{1*})$, since F_0 offers at most $w_2^*(j^1)$, and therefore A will still choose $d_2 = 1$.

Firm F_1 has to choose between "promoting" the Agent and offering him $(1, w_2^*(1))$ and not promoting him, offering $(0, w_2^*(0))$. F_1 therefore chooses $j^{1*}(\theta)$ to solve

$$\max_{j^1} \underbrace{\left\{(1+s)\left[\left(1-j^1\right)x + j^1\theta\right] - w_2^*\left(j^1\right)\right\}}_{\pi_1(j^1,\theta)}.$$

The function $\pi_1(j^1, \theta)$ has increasing differences in j^1 and θ, so F_1's optimal promotion rule will necessarily be monotone increasing in θ, and therefore

$j^{1*}(\theta)$ will be a threshold promotion rule.

Result 2. In any PBE, F_1 chooses a threshold promotion rule

$$j^{1*}(\theta) = \begin{cases} 0 & 0 \leq \theta < \hat{\theta} \\ 1 & \hat{\theta} \leq \theta \leq 1, \end{cases}$$

for some threshold $\hat{\theta}$.

It therefore remains to determine the equilibrium threshold $\hat{\theta}^*$. Given a threshold $\hat{\theta} \in (0,1)$, the expected ability of promoted workers is $E\left[\theta \middle| j^1 = 1\right] = \left(1+\hat{\theta}\right)/2$, and the expected ability of non-promoted workers is $E\left[\theta \middle| j^1 = 0\right] = \hat{\theta}/2$. The wages for promoted workers and non-promoted workers are therefore

$$\begin{aligned} w_2^*\left(j^1 = 1; \hat{\theta}\right) &= \max\left\{x, \frac{1+\hat{\theta}}{2}\right\} \\ w_2^*\left(j^1 = 0; \hat{\theta}\right) &= \max\left\{x, \frac{\hat{\theta}}{2}\right\} = x, \end{aligned}$$

where the last equality holds, because $x > 1/2$, and therefore $x > \hat{\theta}/2$. Given these wage levels as a function of the equilibrium threshold $\hat{\theta}^*$, the equilibrium threshold $\hat{\theta}^*$ is the θ that makes F_1 indifferent between promoting the Agent and not:

$$(1+s)\,x - x = (1+s)\,\hat{\theta}^* - \max\left\{x, \frac{1+\hat{\theta}^*}{2}\right\}.$$

The equilibrium threshold $\hat{\theta}^*$ necessarily satisfies $\left(1+\hat{\theta}^*\right)/2 > x$. If this were not the case, then the indifference condition above would imply that $\hat{\theta}^* = x$, which would in turn contradict the presumption that $\left(1 + \hat{\theta}^*\right)/2 < x$, since $x < 1$. The indifference condition therefore uniquely pins down the equilibrium threshold $\hat{\theta}^*$.

Result 3. In any PBE, the promotion threshold is given by

$$\hat{\theta}^* = \frac{1 + 2sx}{1 + 2s},$$

which is strictly greater than x and weakly less than one.

We can now contrast the equilibrium promotion rule to the first-best promotion rule. Under a first-best promotion rule, the Agent would be assigned to activity 1 whenever $\theta \geq x$. In contrast, in any PBE, the Agent is assigned to activity 1 whenever $\theta \geq \hat{\theta}$, where $\hat{\theta} > x$. That is, the firm fails to promote the agent when it would be socially efficient to do so. Indeed, when $s = 0$, the firm promotes the agent with probability zero. When the firm promotes the worker, his outside option increases, because a promotion is a positive signal about his ability, and so the firm has to raise his wage in order to prevent him from going to the other firm. Promoting the worker increases the firm's output by $(1 + s)(\theta - x)$, but it also increases the worker's wage by $\frac{1+\hat{\theta}^*}{2} - x$, which is equal to $(1 + s)\frac{1-x}{1+2s}$.

Further Reading The theoretical literature on the reasons for and properties of internal labor markets is large. Waldman (1984), Bernhardt (1995), Ghosh and Waldman (2010), Bose and Lang (2013), and Bond (2015) provide symmetric- and asymmetric-learning-based models. Lazear and Rosen (1981), Malcomson (1984), Rosen (1986), MacLeod and Malcomson (1988), Milgrom and Roberts (1988), Prendergast (1993), Chan (1996), Zabojnik and Bernhardt (2001), Waldman (2003), Krakel and Schottner (2012), Auriol, Friebel, and von Bieberstein (2016), and Ke, Li, and Powell (2016) provide incentives-based models. Prendergast (1993), Demougin and Siow (1994), Zabojnik and Bernhardt (2001), Camara and Bernhardt (2009), and DeVaro and Morita (2013) provide models based on human-capital acquisition.

Part II

Boundaries of the Firm

Chapter 4

Theories of the Firm

The central question in this part of the literature goes back to Ronald Coase (1937): if markets are so great at coordinating productive activity, why is productive activity carried out within firms rather than by self-employed individuals who transact on a spot market? And indeed it is, as Herbert Simon (1991) vividly illustrated:

> A mythical visitor from Mars... approaches Earth from space, equipped with a telescope that reveals social structures. The firms reveal themselves, say, as solid green areas with faint interior contours marking out divisions and departments. Market transactions show as red lines connecting firms, forming a network in the spaces between them. Within firms (and perhaps even between them) the approaching visitor also sees pale blue lines, the lines of authority connecting bosses with various lev-

> els of workers... No matter whether our visitor approached the United States or the Soviet Union, urban China or the European Community, the greater part of the space below it would be within the green areas, for almost all inhabitants would be employees, hence inside the firm boundaries. Organizations would be the dominant feature of the landscape. A message sent back home, describing the scene, would speak of "large green areas interconnected by red lines." It would not likely speak of "a network of red lines connecting green spots." ...When our visitor came to know that the green masses were organizations and the red lines connecting them were market transactions, it might be surprised to hear the structure called a market economy. "Wouldn't 'organizational economy' be the more appropriate term?" it might ask.

It is obviously difficult to put actual numbers on the relative importance of trade within and between firms, since, I would venture to say, most transactions within firms are not recorded. From dropping by a colleague's office to ask for help finding a reference, transferring a shaped piece of glass down the assembly line for installation into a mirror, getting an order of fries from the fry cook to deliver to the customer, most economic transactions are difficult even to define as such, let alone track. But we do have some numbers. In what I think is one of the best opening sentences of a job-market paper, Pol Antras provides a lower bound: "Roughly one-third of world trade is intrafirm trade."

Of course, it could conceivably be the case that boundaries don't really matter—that the nature of a particular transaction and the overall volume of transactions is the same whether boundaries are in place or not. And indeed, this would exactly be the case if there were no costs of carrying out transactions: Coase's (1960) eponymous theorem suggests, roughly, that in such a situation, outcomes would be the same no matter how transactions were organized. But clearly this is not the case—in 1997, to pick a random year, the volume of corporate mergers and acquisitions was $1.7 trillion dollars (Holmstrom and Roberts, 1998). It is implausible that this would be the case if boundaries were irrelevant, as even the associated legal fees have to ring up in the billions of dollars.

And so, in a sense, the premise of the Coase Theorem's contrapositive is clearly true. Therefore, there must be transaction costs. And understanding the nature of these transaction costs will hopefully shed some light on the patterns we see. And as D.H. Robertson also vividly illustrated, there are indeed patterns to what we see. Firms are "islands of conscious power in this ocean of unconscious co-operation like lumps of butter coagulating in a pail of buttermilk." So the question becomes: what transaction costs are important, and how are they important? How, in a sense, can they help make sense out of the pattern of butter and buttermilk?

The field was basically dormant for the next forty years until the early 1970s, largely because "transaction costs" came to represent essentially "a name for the residual"—any pattern in the data could trivially be attributed to some story about transaction costs. The empirical content of the theory was therefore zero.

Williamson put structure on the theory by identifying specific factors that composed these transaction costs. And importantly, the specific factors he identified had implications about economic objects that at least could, in principle, be contained in a data set. Therefore his causal claims could be, and were, tested. (As a conceptual matter, it is important to note that even if Williamson's causal claims were refuted, this would not invalidate the underlying claim that "transaction costs are important," since as discussed earlier, this more general claim is essentially untestable, because it is impossible to measure, or even conceive of, *all* transaction costs associated with *all* different forms of organization.) The gist of his theory, which we will describe in more detail shortly, is that when contracts are incomplete and parties have disagreements, they may waste resources "haggling" over the appropriate course of action if they transact in a market, whereas if they transact within a firm, these disagreements can be settled by "fiat" by a mediator. Integration is therefore more appealing when haggling costs are higher, which is the case in situations in which contracts are relatively more incomplete and parties disagree more.

But there was a sense in which his theory (and the related work by Klein, Crawford, and Alchian (1978)) was silent on many foundational questions. After all, why does moving the transaction from the market into the firm imply that parties no longer haggle—that is, what is integration? Further, if settling transactions by fiat is more efficient than by haggling, why aren't all transactions carried out within a single firm? Williamson's and others' response was that there are bureaucratic costs ("accounting contrivances," "weakened incentives," and others) associated with putting more transactions within the firm. But surely those costs are also higher when contracts are more incomplete and when there is more disagreement between parties. Put differently, Williamson identified particular costs associated with transacting in the market and other costs associated with transacting within the firm and made assertions about the rates at which these costs vary with the underlying environment. The resulting empirical implications were consistent with evidence, but the theory still lacked convincing foundations, because it treated these latter costs as essentially exogenous and orthogonal. We will discuss the Transaction-Cost Economics (TCE) approach in the first subsection.

The Property Rights Theory, initiated by Grossman and Hart (1986) and expanded upon in Hart and Moore (1990), proposed a theory which (a) explicitly answered the question of "what is integration?" and (b) treated the costs and benefits of integration symmetrically. Related to the first point is an observation by Alchian and Demsetz that

> It is common to see the firm characterized by the power to settle issues by fiat, by authority, or by disciplinary action superior to that available in the conventional market. This is delusion.

> The firm does not own all its inputs. It has no power of fiat, no authority, no disciplinary action any different in the slightest degree from ordinary market contracting between any two people. I can "punish" you only by withholding future business or by seeking redress in the courts for any failure to honor our exchange agreement. This is exactly all that any employer can do. He can fire or sue, just as I can fire my grocer by stopping purchases from him or sue him for delivering faulty products.

What, then, is the difference between me "telling my grocer what to do" and me "telling my employee what to do?" In either case, refusal would potentially cause the relationship to break down. The key difference, according to Grossman and Hart's theory, is in what happens after the relationship breaks down. If I stop buying goods from my grocer, I no longer have access to his store and all its associated benefits. He simply loses access to a particular customer. If I stop employing a worker, on the other hand, the worker loses access to all the assets associated with my firm. I simply lose access to that particular worker.

Grossman and Hart's (1986) key insight is that property rights determine who can do what in the event that a relationship breaks down—property rights determine what they refer to as the residual rights of control. And allocating these property rights to one party or another may change their incentives to take actions that affect the value of this particular relationship. This logic leads to what is often interpreted as Grossman and Hart's main result: property rights (which define whether a particular transaction is carried out "within" a firm or "between" firms) should be allocated to whichever party is responsible for making more important investments in the relationship. We will discuss the Property Rights Theory (PRT) approach in the second subsection.

From a theoretical foundations perspective, Grossman and Hart was a huge step forward—their theory treats the costs of integration and the costs of non-integration symmetrically and systematically analyzes how different factors drive these two costs in a single unified framework. From a conceptual perspective, however, all the action in the theory is related to how organization affects parties' incentives to make relationship-specific investments. As we will see, their theory assumes that conditional on relationship-specific investments, transactions are always carried out efficiently. A manager never wastes time and resources arguing with an employee. An employee never wastes time and resources trying to convince the boss to let him do a different, more desirable task.

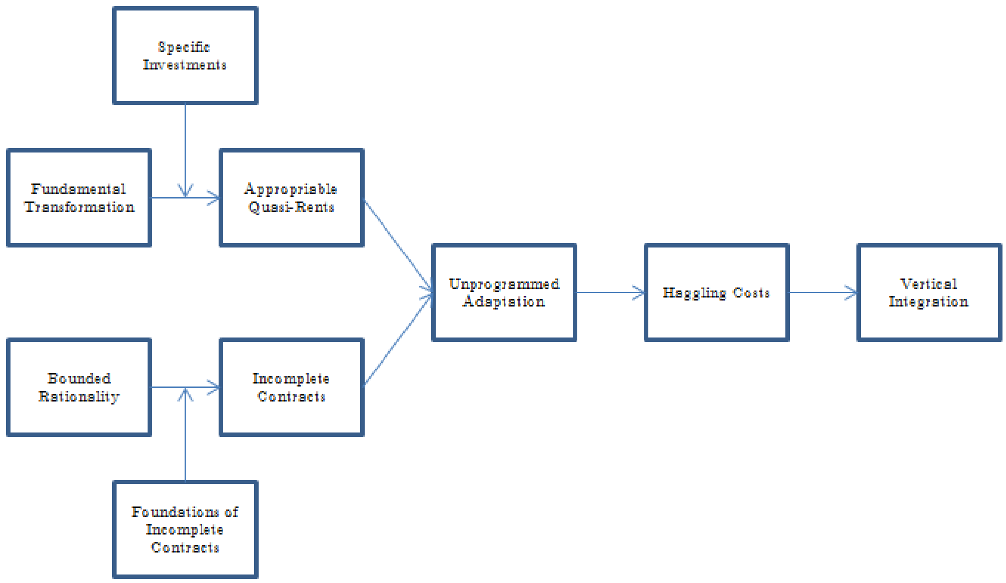
In contrast, in Transaction-Cost Economics, all the action takes place ex post, during the time in which decisions are made. Integration is chosen, precisely because it avoids inefficient haggling costs. We will look at two implications of this observation in the context of two models. The first, which we will examine in the third subsection, will be the adaptation model of Tadelis and Williamson. The second, which we will examine in the fourth subsection, will be a model based on influence activities.

Finally, even the Property Rights Theory does not stand on firm theoretical grounds, since the theory considers only a limited set of institutions the players can put in place to manage their relationship. That is, they focus only on the allocation of control, ignoring the possibility that individuals may write contracts or put in place other types of mechanisms that could potentially do better. In particular, they rule out revelation mechanisms that, in principle, should induce first-best investment. We will address this issue in the sixth subsection.

As a research topic, the theory of the firm can be a bit overwhelming. In contrast to many applied-theory topics, the theory of the firm takes a lot of non-standard variables as endogenous. Further, there is often ambiguity about what variables should be taken as endogenous and what methodology should be used, so the "playing field" is not well-specified. But ultimately, I think that developing a stronger understanding of what determines firm boundaries is important, since it simultaneously tells us what the limitations of markets are. I will try to outline some of the "ground rules" that I have been able to discern from spending some time studying these issues.

4.1 Transaction-Cost Economics

The literature on the boundaries of the firm introduces many new concepts and even more new terms. So we will spend a little bit of time sorting out the terminology before proceeding. The following figure introduces most of the new terms that we will talk about in this section and in the following sections.



As an overview, the basic argument of the Transaction-Cost Economics approach is the following.

Consider a transaction between an upstream manufacturer and a downstream distributor. Should the distributor **buy** from the manufacturer or should it buy the manufacturer and **make** goods itself? The **fundamental transformation** of ex ante perfect competition among manufacturers for the customer's business to ex post small numbers results *from* **specific investments** and results *in* **appropriable quasi-rents**—ex post rents that parties can potentially fight over, because they are locked in with each other. If the parties have written **incomplete contracts** (the **foundations** for which are that **bounded rationality** limits their ability to foresee all relevant contingencies), then they might find themselves in situations that call for **unprogrammed adaptation**. At this point, they may fight over the appropriate course of action, incurring **haggling costs**. These haggling costs can be reduced or eliminated if either the manufacturer purchases the distributor or the distributor purchases the manufacturer, and they become **vertically integrated**.

Every link in this figure is worth discussing. The **fundamental transformation** is, in my view, the most important economic idea to emerge from this theory. We had known since at least Edgeworth that under bilateral monopoly, many problems were possible (Edgeworth focused on indeterminacy) and perhaps inevitable (e.g., the Myerson-Satterthwaite theorem), but that competition among large numbers of potential trading partners would generally (with some exceptions—perfect complementarities between, say, right-shoe manufacturers and left-shoe manufacturers could persist even if the economy became arbitrarily large) push economies towards efficient allocations. Perfect competition is equivalent to the no-surplus condition (Ostroy, 1980; Makowski and Ostroy, 1995)—under perfect competition, you fully appropriate whatever surplus you generate, so everyone else in the economy as a whole is indifferent toward what you do. As a result, your incentives to maximize your own well-being do not come into conflict with others, so this leads to efficient allocations and nothing worth incurring costs to fight over. The underlying intuition for why large numbers of trading partners leads to efficient allocations is that a buyer can always play a seller and her competitors off each other, and in the limit, the next-best seller is just as good as the current one (and symmetrically for buyers). Williamson's observation was that after a trading relationship has been initiated, the buyer and the sellers develop ties to each other (**quasi-rents**), so that one's current trading partner is always discretely better than the next best alternative. In other words, the beneficial forces of perfect competition almost never hold.

Of course, if during the ex ante competition phase of the relationship, potential trading partners competed with each other by offering enforceable, complete, long-term trading contracts, then the fact that ex post, parties are locked in to each other would be irrelevant. Parties would compete in the market by offering each other utility streams that they are contractually obligated to fulfill, and perfect competition ex ante would lead to maximized long-term gains from trade.

This is where **incomplete contracts** comes into the picture. Such contracts are impossible to write, because they would require parties to be able to conceive of and enumerate all possible contingencies. Because parties are **boundedly rational**, they will only be able to do so for a subset of the possible states. As a result, ex ante competition will lead parties to agree to incomplete contracts for which the parties will need to fill in the details as they go along. In other words, they will occasionally need to make **unprogrammed adaptations**. As an example, a legacy airline (say, American Airlines) and a regional carrier (say, American Eagle) may agree on a flight schedule for flights connecting two cities. But when weather-related disruptions occur, the ideal way of shifting around staff and equipment depends to a large extent on where both carrier's existing staff and equipment are, and there are simply too many different potential configurations for this. As a result, airlines typically do not contractually specify what will happen in the event that there are weather-related disruptions, and they therefore have to figure it out on the spot.

The need to make unprogrammed adaptations would also not be a problem if the parties could simply agree to bargain efficiently ex post after an event occurs that the parties had not planned for. (And indeed, if there was perfect competition ex post, they would not even need to bargain ex post.) However, under the TCE view, ex-post bargaining is rarely if ever efficient. The legacy airline will insist that its own staff and equipment are unable to make it, so everything would be better if the regional carrier made concessions, and conversely. Such ex post bargaining inevitably leads either to bad ex post decisions (the carrier with the easier-to-access equipment and staff is not the one who ends up putting it in place) or results in other types of **rent-seeking** costs (time and resources are wasted in the bargaining process). These **haggling costs** could be eliminated if both parties were under the direction of a common headquarters that could issue commands and easily resolve these types of conflicts. This involves setting up a **vertically integrated** organization.

Further, vertically integrated organizations involve **bureaucratic costs**.

Reorganization involves setup costs. Incentives are usually low-powered inside organizations. Division managers engage in **accounting contrivances** in order to alter decision making of other divisions or the headquarters. Finally, the contract law that governs decisions made by one division that affect another division differs from the contract law that governs decisions made by one firm that affect another—in essence, the latter types of contracts are enforceable, whereas the former types of contracts are not. This difference in contract law is referred to as **forbearance**.

When would we be more likely to see vertical integration? When the environment surrounding a particular transaction is especially complex, contracts are more likely to be incomplete, or they are likely to be more incomplete. As a result, the need for unprogrammed adaptations and their associated haggling costs will be greater. When parties are more locked in to each other, their ability to access the outside market either to look for alternatives or to use alternatives to discipline their bargaining process is lessened. As a result, there is more to fight over when unprogrammed adaptations are required, and their associated haggling costs will be greater. Additionally, integration involves setup costs, and these setup costs are only worth incurring if the parties expect to interact with each other often. Finally, integration itself involves other bureaucratic costs, and so vertical integration is more appealing if these costs are low. Put differently, the integration decision involves a trade-off between haggling costs under non-integration and bureaucratic costs under integration. To summarize, the main empirical predictions of the TCE theory are:

1. Vertical integration is more likely for transactions that are more complicated.

2. Vertical integration is more likely when there is more specificity.

3. Vertical integration is more likely when the players interact more frequently.

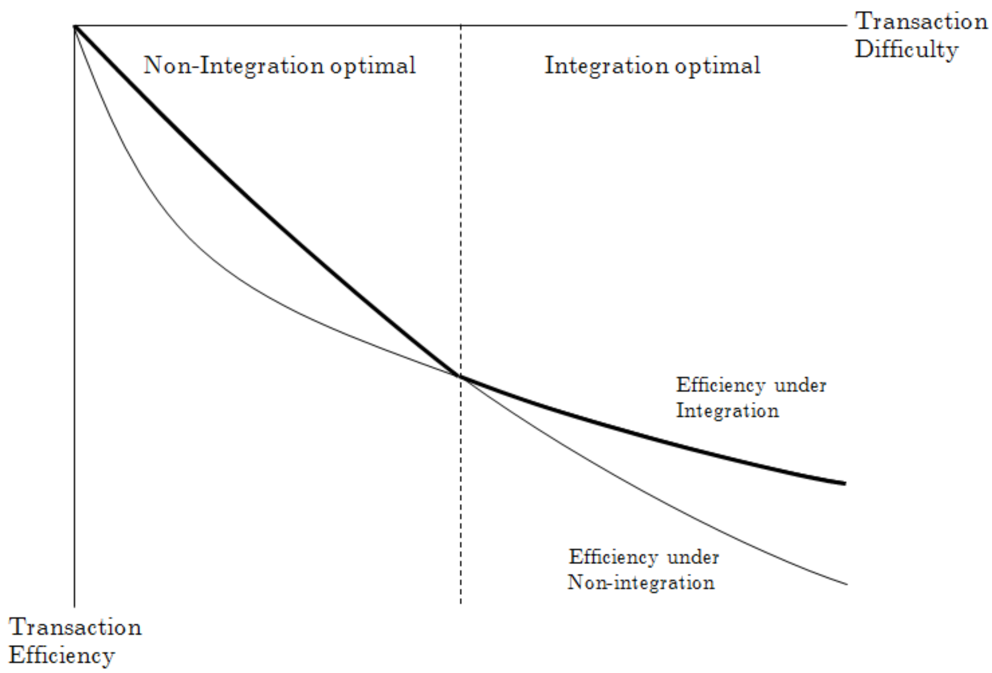
4. Vertical integration is more likely when bureaucratic costs are low.

This discussion of TCE has been informal, because the theory itself is informal. There are at least two aspects of this informal argument that can be independently formalized. When unprogrammed adaptation is required, the associated costs can either come from costly haggling (rent-seeking) or from inefficient ex post decision making (adaptation). I will describe two models that each capture one of these sources of ex post inefficiencies.

There are a couple common themes that arise in the analysis of both of these models. The first theme is that when thinking about the make-or-buy or boundary-of-the-firm question, the **appropriate unit of analysis is at the transaction level**. Another theme is that these models consider **private-ordering solutions** rather than solutions imposed on the transacting parties by a third party such as a government—resolving conflict between parties need not involve government intervention. The question is therefore:

for a given transaction, what institutions should *the interested parties* put in place to manage this transaction?

Since transactions differ in their characteristics, more difficult transactions will be more difficult no matter how they are organized. As a result, looking at the performance of various transactions and relating that performance to how those transactions are organized could lead one inappropriately to the conclusion that integration is actually bad for performance. Or one could dismiss the agenda, as one prominent economics blogger once did: "I view the Coasian tradition as somewhat of a dead end in industrial organization. Internally, firms aren't usually more efficient than markets... ."



As the figure above (which Gibbons (2005) colorfully describes as "Coase meets Heckman") shows, this is *exactly* the type of prediction that this class of theories predicts. And whether integrated transactions are less efficient, because integration is bad for transaction efficiency, or because transactions that are more complicated are more likely to involve integration and are likely to be less efficient matters. This difference matters, because these two different views have the opposite implications. Under the TCE view, discouraging firms from engaging in vertical integration (through, for example, strict antitrust policy) will necessarily be bad for firms' internal efficiency. Under the alternative view, strict antitrust policy would serve not only to facilitate product-market competition, it would also *increase* firms' internal efficiency.

Another theme that arises in these models is that there are many differences between transactions carried out between firms and those carried out within firms. An upstream division manager will typically be on a lower-powered incentive scheme than she would be if she were the owner of an independent upstream firm. Transactions within firms are subject to different legal regimes than transactions between firms. Transactions within firms tend to be characterized by more "bureaucracy" than transactions across firms. There are two ways to look at these bundles of differences. Viewed one way, low-powered incentives and bureaucracy are the baggage associated with integration and therefore a cost of integration. Viewed another way, low-powered incentives and bureaucracy are also optimal choices that complement integration because they help solve other problems that arise under integration.

Finally, I will conclude with a description of one question that I do not think the literature has produced satisfying answers to. First, there are obviously more ways to organize a transaction than "vertical integration" and "non-integration." In particular, the transacting parties could engage in simple spot-market transactions; they could engage in short-run contracting across firm boundaries in which they specify a small number of contingencies; they could engage in long-term contracting across firm boundaries in which perhaps decision rights are contractually reallocated (for example, an upstream firm may have some say over the design specifics for a product that the downstream firm is producing); or one party could buy the other party. The line between integration and non-integration is therefore much blurrier than it seemed at first glance.

4.1.1 Adaptation-Cost Model

This model is adapted from Tadelis and Williamson (2013). It is a reduced-form model that captures some of the aspects of the TCE argument that I outlined above. By doing so in a reduced-form way, the model importantly highlights a set of primitive assumptions that are sufficient for delivering the types of comparative statics predicted by the informal theory.

Description There is a risk-neutral upstream manufacturer U of an intermediate good and a risk-neutral downstream producer, D, who can costlessly transform a unit of the intermediate good into a final good that is then sold into the market at price p. Production of the intermediate good involves a cost of $C(e,g) = \bar{C} - eg \in \mathcal{C}$, where e is an effort choice by U and involves a private cost of $c(e) = \frac{c}{2}e^2$ being borne by U. $g \in G$ denotes the governance structure, which we will describe shortly. There is a state of the world $\theta \in \Theta = \Theta_C \cup \Theta_{NC}$, with $\Theta_C \cap \Theta_{NC} = \emptyset$, where $\theta \in \Theta_C$ is a contractible state and $\theta \in \Theta_{NC}$ is a noncontractible state. The parties can either be integrated ($g = I$) or non-integrated ($g = NI$), and they can also sign a contract $w \in W = \left\{ w : \mathcal{C} \times \Theta \to \{s + (1-b)C\}_{b \in \{0,1\}} \right\}$, which compels D to make a transfer of $s + (1-b)C$ to U in any state $\theta \in \Theta$. Note in particular that b must either be 0 or 1: the contract space includes only cost-plus and fixed-price contracts. Additionally, if $\theta \in \Theta_{NC}$, the contract has to be renegotiated. In this case, D incurs **adaptation costs** of $k(b,g)$, which depends on whether or not the parties are integrated as well as on the cost-sharing characteristics of the contract. The contract is always successfully renegotiated so that trade still occurs, and the same cost-sharing rule as specified in the original contract is obtained. The probability that adaptation is required is $\Pr[\theta \in \Theta_{NC}] = \sigma$.

Timing The timing is as follows:

1. D makes an offer of a governance structure g and a contract w to U.

(g, w) is publicly observed.

2. U can accept the contract $(d = 1)$ or reject it $(d = 0)$ in favor of an outside option that yields utility 0.

3. If $d = 1$, then U chooses effort e at cost $c(e) = \frac{c}{2}e^2$. e is commonly observed.

4. $\theta \in \Theta$ is realized and is commonly observed.

5. If $\theta \in \Theta_{NC}$, parties have to adjust the contract, in which case D incurs adaptation costs $k(b, g)$. Trade occurs, and the final good is sold at price p.

Equilibrium A **subgame-perfect equilibrium** is a governance structure g^*, a contract w^*, an acceptance decision strategy $d^* : G \times W \to \{0, 1\}$, and an effort choice strategy $e^* : G \times W \times D \to \mathbb{R}_+$ such that given g^* and w^*, U optimally chooses $d^*(g^*, w^*)$ and $e^*(g^*, w^*, d^*)$, and D optimally offers governance structure g^* and contract w^*.

The Program The downstream producer makes an offer of a governance structure g and a contract $w = s + (1 - b) C$ as well as a proposed effort level e to maximize his profits:

$$\max_{g \in \{I, NI\}, b \in \{0,1\}, e, s} p - s - (1 - b) C(e, g) - \sigma k(b, g)$$

subject to U's incentive-compatibility constraint

$$e \in \underset{\hat{e}}{\operatorname{argmax}}\, s + (1-b)\, C(\hat{e}, g) - C(\hat{e}, g) - c(\hat{e})$$

and her individual-rationality constraint

$$s + (1-b)\, C(e, g) - C(e, g) - c(e) \geq 0.$$

Since we are restricting attention to linear contracts, U's incentive-compatibility constraint can be replaced by her first-order condition:

$$\begin{aligned} c'(e(b,g)) &= -b \frac{\partial C(e,g)}{\partial e} \\ e(b,g) &= \frac{b}{c} g, \end{aligned}$$

and any optimal contract offer by D will ensure that U's individual-rationality constraint holds with equality. D's problem then becomes

$$\max_{g \in \{I, NI\}, b \in \{0,1\}} p - C(e(b,g), g) - c(e(b,g)) - \sigma k(b,g).$$

That is, D chooses a governance structure $g \in \{I, NI\}$ and an incentive intensity $b \in \{0,1\}$ in order to maximize total ex-ante expected equilibrium surplus. Let

$$W(g, b; \sigma) = p - C(e(b,g), g) - c(e(b,g)) - \sigma k(b,g)$$

be the **Coasian objective**. We will refer to the following problem as the **Coasian program**, the solution to which is the optimal governance structure (g^*, b^*):

$$W^*(\sigma) = \max_{g \in \{I, NI\}, b \in \{0,1\}} W(g, b; \sigma).$$

Assumptions Several assumptions will be important for the main results of this model:

1. Supplier effort is more effective under non-integration than under integration (i.e., $\frac{\partial C(e,NI)}{\partial e} < \frac{\partial C(e,I)}{\partial e}$, which is true if $I < NI$)

2. Adaptation costs are lower under integration than under non-integration (i.e., $k(b, NI) > k(b, I)$)

3. Adaptation costs are lower when cost incentives are weaker (i.e., $\frac{\partial k(b,g)}{\partial b} > 0$)

4. Reducing adaptation costs by weakening incentives is more effective under integration than under non-integration (i.e., $\frac{\partial k(b,NI)}{\partial b} > \frac{\partial k(b,I)}{\partial b}$).

Tadelis and Williamson (2013) outline many ways to justify several of these assumptions, but at the end of the day, these assumptions are quite reduced-form. However, they map nicely into the main results of the model, so at the very least, we can get a clear picture of what a more structured model ought to satisfy in order to get these results.

Solution To solve this model, we will use some straightforward monotone comparative statics results. Recall that if $F(x, \theta)$ is a function of choice variables $x \in X$ and parameters $\theta \in \Theta$, then if $F(x, \theta)$ is supermodular in (x, θ), $x^*(\theta)$ is increasing in θ, where

$$x^*(\theta) = \underset{x \in X}{\operatorname{argmax}} F(x, \theta).$$

Once the Coasian program has been expressed as an unconstrained maximization problem, the key comparative statics are very easy to obtain if the objective function is supermodular. This model's assumptions are purposefully made in order to ensure that the objective function is supermodular.

To see this, let

$$\begin{aligned} W(g, b; \sigma) &= p - C(e(b, g), g) - c(e(b, g)) - \sigma k(b, g) \\ &= p - \bar{C} + e(b, g) \cdot g - \frac{c}{2} e(b, g)^2 - \sigma k(b, g) \end{aligned}$$

We can easily check supermodularity by taking some first-order derivatives and looking at second-order differences:

$$\begin{aligned} \frac{\partial W}{\partial b} &= \frac{\partial e(b, g)}{\partial b} \cdot g - ce(b, g) \frac{\partial e(b, g)}{\partial b} - \sigma \frac{\partial k(b, g)}{\partial b} \\ &= (1 - b) \frac{g^2}{c} - \sigma \frac{\partial k(b, g)}{\partial b} \\ \frac{\partial W}{\partial \sigma} &= -k(b, g) \end{aligned}$$

and

$$\begin{aligned}
\frac{\partial^2 W}{\partial b \partial \sigma} &= -\frac{\partial k(b,g)}{\partial b} \\
\frac{\partial W(NI,b;\sigma)}{\partial b} - \frac{\partial W(I,b;\sigma)}{\partial b} &= \left((1-b)\frac{(NI)^2}{c} - \sigma\frac{\partial k(b,NI)}{\partial b}\right) - \left((1-b)\frac{I^2}{c} - \sigma\frac{\partial k(b,I)}{\partial b}\right) \\
&= \frac{1-b}{c}\left((NI)^2 - I^2\right) + \sigma\left(\frac{\partial k(b,I)}{\partial b} - \frac{\partial k(b,NI)}{\partial b}\right) \\
\frac{\partial W(NI,b;\sigma)}{\partial \sigma} - \frac{\partial W(I,b;\sigma)}{\partial \sigma} &= k(b,I) - k(b,NI).
\end{aligned}$$

By assumption 1, $I < NI$. By assumption 2, $\frac{\partial W(NI,b;\sigma)}{\partial \sigma} - \frac{\partial W(I,b;\sigma)}{\partial \sigma} < 0$. By assumption 3, $\frac{\partial^2 W}{\partial b \partial \sigma} < 0$. By assumptions 1 and 4, $\frac{\partial W(NI,b;\sigma)}{\partial b} - \frac{\partial W(I,b;\sigma)}{\partial b} > 0$. Putting these together, we have that $W(g,b;\sigma)$ is supermodular in $(g,b;-\sigma)$, where we adopt the order that $g = NI$ is greater than $g = I$. We then have some straightforward comparative statics:

1. g^* is increasing in $-\sigma$. That is, when there is more uncertainty, NI becomes more desirable relative to I.

2. b^* is increasing in $-\sigma$. That is, when there is more uncertainty, low-powered incentives become relatively more desirable.

3. (g^*, b^*) are complementary. Incentives are more high-powered under non-integration than under integration.

4.2 Property Rights

Essentially the main result of TCE is the observation that when haggling costs are high under non-integration, then integration is optimal. This result is unsatisfying in at least two senses. First, TCE does not tell us what exactly is the mechanism through which haggling costs are reduced under integration, and second, it does not tell us what the associated costs of integration are, and it therefore does not tell us when we would expect such costs to be high. In principle, in environments in which haggling costs are high under non-integration, then the within-firm equivalent of haggling costs should also be high.

Grossman and Hart (1986) and Hart and Moore (1990) set aside the "make or buy" question and instead begin with the more fundamental question, "What is a firm?" In some sense, nothing short of an answer to *this* question will consistently provide an answer to the questions that TCE leaves unanswered. Framing the question slightly differently, what do I get if I buy a firm from someone else? The answer is typically that I become the owner of the firm's non-human assets.

Why, though, does it matter who owns non-human assets? If contracts are complete, it does not matter. The parties to a transaction will, ex ante, specify a detailed action plan. One such action plan will be optimal. That action plan will be optimal regardless of who owns the assets that support the transaction, and it will be feasible regardless of who owns the assets.

If contracts are incomplete, however, not all contingencies will be specified. The key insight of the PRT is that ownership endows the asset's owner with the right to decide what to do with the assets in these contingencies. That is, ownership confers **residual control rights**. When unprogrammed adaptations become necessary, the party with residual control rights has **power** in the relationship and is protected from expropriation by the other party. That is, control over non-human assets leads to control over human assets, since they provide leverage over the person who lacks the assets. Since she cannot be expropriated, she therefore has incentives to make investments that are specific to the relationship.

Firm boundaries are tantamount to asset ownership, so detailing the costs and benefits of different ownership arrangements provides a complete account of the costs and benefits of different firm-boundary arrangements. Asset ownership, and therefore firm boundaries, determine who possesses power in a relationship, and power determines investment incentives. Under integration, I have all the residual control rights over non-human assets and therefore possess strong investment incentives. Non-integration splits apart residual control rights, and therefore provides me with weaker investment incentives and you with stronger investment incentives. If I own an asset, you do not. Power is scarce and therefore should be allocated optimally.

Methodologically, the PRT makes significant advances over the preceding theory. PRT's conceptual exercise is to hold technology, preferences, information, and the legal environment constant across prospective governance structures and ask, for a given transaction with given characteristics, whether the transaction is best carried out within a firm or between firms. That is, prior theories associated "make" with some vector $(\alpha_1, \alpha_2, \dots)$ of characteristics and "buy" with some other vector $(\beta_1, \beta_2, \dots)$ of characteristics. "Make" is preferred to "buy" if the vector $(\alpha_1, \alpha_2, \dots)$ is preferred to the vector $(\beta_1, \beta_2, \dots)$. In contrast, PRT focuses on a single aspect: α_1 versus β_1. Further differences may arise between "make" and "buy," but to the extent that they are also choice variables, they will arise optimally rather than passively. We will talk about why this is an important distinction to make when we talk about the influence-cost model in the next section.

Description There is a risk-neutral upstream manager U, a risk-neutral downstream manager D, and two assets A_1 and A_2. Managers U and D make investments e_U and e_D at private cost $c_U(e_U)$ and $c_D(e_D)$. These investments determine the value that each manager receives if trade occurs, $V_U(e_U, e_D)$ and $V_D(e_U, e_D)$. There is a state of the world, $s \in S = S_C \cup S_{NC}$, with $S_C \cap S_{NC} = \emptyset$ and $\Pr[s \in S_{NC}] = \mu$. In state s, the identity of the ideal good to be traded is s—if the managers trade good s, they receive $V_U(e_U, e_D)$ and $V_D(e_U, e_D)$. If the managers trade good $s' \neq s$, they both receive $-\infty$. The managers choose an asset allocation, denoted by g, from a set $G = \{UI, DI, NI, RNI\}$. Under $g = UI$, U owns both assets. Under $g = DI$, D owns both assets. Under $g = NI$, U owns asset A_1 and D owns asset A_2. Under $g = RNI$, D owns asset A_1, and U owns asset A_2.

In addition to determining an asset allocation, manager U also offers an incomplete contract $w \in W = \{w : E_U \times E_D \times S_C \rightarrow \mathbb{R}\}$ to D. The contract specifies a transfer $w(e_U, e_D, s)$ to be paid from D to U if they trade good $s \in S_C$. If the players want to trade a good $s \in S_{NC}$, they do so in the following way. With probability $\frac{1}{2}$, U makes a take-it-or-leave-it offer $w_U(s)$ to D, specifying trade and a price. With probability $\frac{1}{2}$, D makes a take-it-or-leave-it offer $w_D(s)$ to U specifying trade and a price. If trade does not occur, then manager U receives payoff $v_U(e_U, e_D; g)$ and manager D receives payoff $v_D(e_U, e_D; g)$, which depends on the asset allocation.

Timing There are five periods:

1. U offers D an asset allocation $g \in G$ and a contract $w \in W$. Both g and w are commonly observed.

2. U and D simultaneously choose investment levels e_U and e_D at private cost $c(e_U)$ and $c(e_D)$. These investment levels are commonly observed by e_U and e_D.

3. The state of the world, $s \in S$ is realized.

4. If $s \in S_C$, D buys good s at price specified by w. If $s \in S_{NC}$, U and D engage in 50-50 take-it-or-leave-it bargaining.

5. Payoffs are realized.

Equilibrium A **subgame-perfect equilibrium** is an asset allocation g^*, a contract w^*, investment strategies $e_U^* : G \times W \to \mathbb{R}_+$ and $e_D^* : G \times W \to \mathbb{R}_+$, and a pair of offer rules $w_U^* : E_D \times E_U \times S_{NC} \to \mathbb{R}$ and $w_D^* : E_D \times E_U \times S_{NC} \to \mathbb{R}$ such that given $e_U^* (g^*, w^*)$ and $e_D^* (g^*, w^*)$, the managers optimally make offers $w_U^* (e_U^*, e_D^*)$ and $w_D^* (e_U^*, e_D^*)$ in states $s \in S_{NC}$; given g^* and w^*, managers optimally choose $e_U^* (g^*, w^*)$ and $e_D^* (g^*, w^*)$; and U optimally offers asset allocation g^* and contract w^*.

Assumptions As always, we will assume $c_U (e_U) = \frac{1}{2} e_U^2$ and $c_D (e_D) = \frac{1}{2} e_D^2$. We will also assume that $\mu = 1$, so that the probability that an ex ante specifiable good is optimal to trade ex post is zero. We will return to this issue later. Let

$$\begin{aligned} V_U (e_U, e_D) &= f_{UU} e_U + f_{UD} e_D \\ V_D (e_U, e_D) &= f_{DU} e_U + f_{DD} e_D \\ v_U (e_U, e_D; g) &= h_{UU}^g e_U + h_{UD}^g e_D \\ v_D (e_U, e_D; g) &= h_{DU}^g e_U + h_{DD}^g e_D, \end{aligned}$$

and define

$$\begin{aligned} F_U &= f_{UU} + f_{DU} \\ F_D &= f_{UD} + f_{DD}. \end{aligned}$$

Finally, outside options are more sensitive to one's own investments the more assets one owns:

$$
\begin{aligned}
h_{UU}^{UI} &\geq h_{UU}^{NI} \geq h_{UU}^{DI}, h_{UU}^{UI} \geq h_{UU}^{RNI} \geq h_{UU}^{DI} \\
h_{DD}^{DI} &\geq h_{DD}^{NI} \geq h_{DD}^{UI}, h_{DD}^{DI} \geq h_{DD}^{RNI} \geq h_{DD}^{UI}.
\end{aligned}
$$

The Program We solve backwards. For all $s \in S_{NC}$, with probability $\frac{1}{2}$, U will offer price $w_U(e_U, e_D)$. D will accept this offer as long as $V_D(e_U, e_D) - w_U(e_U, e_D) \geq v_D(e_U, e_D; g)$. U's offer will ensure that this holds with equality (or else U could increase w_U a bit and increase his profits while still having his offer accepted):

$$
\begin{aligned}
\pi_U &= V_U(e_U, e_D) + w_U(e_U, e_D) = V_U(e_U, e_D) + V_D(e_U, e_D) - v_D(e_U, e_D; g) \\
\pi_D &= V_D(e_U, e_D) - w_U(e_U, e_D) = v_D(e_U, e_D; g).
\end{aligned}
$$

Similarly, with probability $\frac{1}{2}$, D will offer price $w_D(e_U, e_D)$. U will accept this offer as long as $V_U(e_U, e_D) + w_D(e_U, e_D) \geq v_U(e_U, e_D; g)$. D's offer will ensure that this holds with equality (or else D could decrease w_D a bit and increase her profits while still having her offer accepted):

$$
\begin{aligned}
\pi_U &= V_U(e_U, e_D) + w_D(e_U, e_D) = v_U(e_U, e_D; g) \\
\pi_D &= V_D(e_U, e_D) - w_D(e_U, e_D) = V_U(e_U, e_D) + V_D(e_U, e_D) - v_U(e_U, e_D; g).
\end{aligned}
$$

In period 2, manager U will conjecture e_D and solve

$$\max_{\hat{e}_U} \frac{1}{2}\left(V_U\left(\hat{e}_U, e_D\right) + V_D\left(\hat{e}_U, e_D\right) - v_D\left(\hat{e}_U, e_D; g\right)\right) + \frac{1}{2} v_U\left(\hat{e}_U, e_D; g\right) - c\left(\hat{e}_U\right)$$

and manager D will conjecture e_U and solve

$$\max_{\hat{e}_D} \frac{1}{2} v_D\left(e_U, \hat{e}_D; g\right) + \frac{1}{2}\left(V_U\left(e_U, \hat{e}_D\right) + V_D\left(e_U, \hat{e}_D\right) - v_U\left(e_U, \hat{e}_D; g\right)\right) - c\left(\hat{e}_D\right).$$

Substituting in the functional forms we assumed above, these problems become:

$$\max_{\hat{e}_U} \frac{1}{2}\left(F_U \hat{e}_U + F_D e_D\right) + \frac{1}{2}\left(\left(h_{UU}^g - h_{DU}^g\right)\hat{e}_U + \left(h_{UD}^g - h_{DD}^g\right) e_D\right) - \frac{1}{2}\hat{e}_U^2$$

and

$$\max_{\hat{e}_D} \frac{1}{2}\left(F_U e_U + F_D \hat{e}_D\right) + \frac{1}{2}\left(\left(h_{DU}^g - h_{UU}^g\right) e_U + \left(h_{DD}^g - h_{UD}^g\right)\hat{e}_D\right) - \frac{1}{2}\hat{e}_D^2.$$

These are well-behaved objective functions, and in each one, there are no interactions between the managers' investments, so each manager has a dominant strategy, which we can solve for by taking first-order conditions:

$$\begin{aligned} e_U^{*g} &= \frac{1}{2}F_U + \frac{1}{2}\left(h_{UU}^g - h_{DU}^g\right) \\ e_D^{*g} &= \frac{1}{2}F_D + \frac{1}{2}\left(h_{DD}^g - h_{UD}^g\right) \end{aligned}$$

Each manager's incentives to invest are derived from two sources: (1) the marginal impact of investment on total surplus and (2) the marginal impact of investment on the "threat-point differential." The latter point is worth expanding on. If U increases his investment, his outside option goes up by h_{UU}^g, which increases the price that D will have to offer him when she makes her take-it-or-leave-it offer, which increases U's ex-post payoff if $h_{UU}^g > 0$. Further, D's outside option goes up by h_{DU}^g, which increases the price that U has to offer D when he makes his take-it-or-leave-it-offer, which decreases U's ex-post payoff if $h_{DU}^g > 0$.

Ex ante, players' equilibrium payoffs are:

$$\begin{aligned}
\Pi_U^{*g} &= \frac{1}{2}\left(F_U e_U^{*g} + F_D e_D^{*g}\right) + \frac{1}{2}\left(\left(h_{UU}^g - h_{DU}^g\right) e_U^{*g} + \left(h_{UD}^g - h_{DD}^g\right) e_D^{*g}\right) - \frac{1}{2}\left(e_U^{*g}\right)^2 \\
\Pi_D^{*g} &= \frac{1}{2}\left(F_U e_U^{*g} + F_D e_D^{*g}\right) + \frac{1}{2}\left(\left(h_{DU}^g - h_{UU}^g\right) e_U^{*g} + \left(h_{DD}^g - h_{UD}^g\right) e_D^{*g}\right) - \frac{1}{2}\left(e_D^{*g}\right)^2 .
\end{aligned}$$

If we let $\theta = \left(f_{UU}, f_{UD}, f_{DU}, f_{DD}, \{h_{UU}^g, h_{UD}^g, h_{DU}^g, h_{DD}^g\}_{g\in G}\right)$ denote the parameters of the model, the Coasian objective for **governance structure** g is:

$$W^g(\theta) = \Pi_U^{*g} + \Pi_D^{*g} = F_U e_U^* + F_D e_D^* - \frac{1}{2}\left(e_U^{*g}\right)^2 - \frac{1}{2}\left(e_D^{*g}\right)^2 .$$

The **Coasian Program** that describes the optimal governance structure is then:

$$W^*(\theta) = \max_{g\in G} W^g(\theta) .$$

At this level of generality, the model is too rich to provide straightforward insights. In order to make progress, we will introduce the following definitions. If $f_{ij} = h_{ij}^{g} = 0$ for $i \neq j$, we say that investments are **self-investments**. If $f_{ii} = h_{ii}^{g} = 0$, we say that investments are **cross-investments**. When investments are self-investments, the following definitions are useful. Assets A_1 and A_2 are **independent** if $h_{UU}^{UI} = h_{UU}^{NI} = h_{UU}^{RNI}$ and $h_{DD}^{DI} = h_{DD}^{NI} = h_{DD}^{RNI}$ (i.e., if owning the second asset does not increase one's marginal incentives to invest beyond the incentives provided by owning a single asset). Assets A_1 and A_2 are **strictly complementary** if either $h_{UU}^{NI} = h_{UU}^{RNI} = h_{UU}^{DI}$ or $h_{DD}^{NI} = h_{DD}^{RNI} = h_{DD}^{UI}$ (i.e., if for one player, owning one asset provides the same incentives to invest as owning no assets). U's **human capital is essential** if $h_{DD}^{DI} = h_{DD}^{UI}$, and D's human capital is essential if $h_{UU}^{UI} = h_{UU}^{DI}$.

With these definitions in hand, we can get a sense for what features of the model drive the optimal governance-structure choice.

PROPOSITION (Hart 1995). If A_1 and A_2 are independent, then NI or RNI is optimal. If A_1 and A_2 are strictly complementary, then DI or UI is optimal. If U's human capital is essential, UI is optimal. If D's human capital is essential, DI is optimal. If both U's and D's human capital is essential, all governance structures are equally good.

These results are straightforward to prove. If A_1 and A_2 are independent, then there is no additional benefit of allocating a second asset to a single party. Dividing up the assets therefore strengthens one party's investment incentives without affecting the other's. If A_1 and A_2 are strictly complementary, then relative to integration, dividing up the assets necessarily weakens one party's investment incentives without increasing the other's, so one form of integration clearly dominates. If U's human capital is essential, then D's investment incentives are independent of which assets he owns, so UI is at least weakly optimal.

The more general results of this framework are that (a) allocating an asset to an individual strengthens that party's incentives to invest, since it increases his bargaining position when unprogrammed adaptation is required, (b) allocating an asset to one individual has an opportunity cost, since it means that it cannot be allocated to the other party. Since we have assumed that investment is always socially valuable, this implies that assets should always be allocated to exactly one party (if joint ownership means that both parties have a veto right). Further, allocating an asset to a particular party is more desirable the more important that party's investment is for joint welfare and the more sensitive his/her investment is to asset ownership. Finally, assets should be co-owned when there are complementarities between them.

While the actual results of the PRT model are sensible and intuitive, there are many limitations of the analysis. First, as Holmstrom points out in his 1999 JLEO article, "The problem is that the theory, as presented, really is a theory about asset ownership by individuals rather than by firms, at least if one interprets it literally. Assets are like bargaining chips in an entirely autocratic market... Individual ownership of assets does not offer a theory of organizational identities unless one associates individuals with firms." Holmstrom concludes that, "... the boundary question is in my view fundamentally about the distribution of activities: What do firms do rather than what do they own? Understanding asset configurations should not become an end in itself, but rather a means toward understanding activity configurations." That is, by taking payoff functions V_U and V_D as exogenous, the theory is abstracting from what Holmstrom views as the key issue of what a firm really is.

Second, after assets have been allocated and investments made, adaptation is made efficiently. The managers always reach an ex post efficient arrangement in an efficient manner, and all inefficiencies arise ex ante through inadequate incentives to make relationship-specific investments. Williamson (2000) argues that "The most consequential difference between the TCE and GHM setups is that the former holds that maladaptation in the contract execution interval is the principal source of inefficiency, whereas GHM vaporize ex post maladaptation by their assumptions of common knowledge and ex post bargaining." That is, Williamson believes that ex post inefficiencies are the primary sources of inefficiencies that have to be managed by adjusting firm boundaries, while the PRT model focuses solely on ex ante inefficiencies. The two approaches are obviously complementary, but there is an entire dimension of the problem that is being left untouched under this approach.

Finally, in the Coasian Program of the PRT model, the parties are unable to write formal contracts (in the above version of the model, this is true only when $\mu = 1$) and therefore the only instrument they have to motivate relationship-specific investments is the allocation of assets. The implicit assumption underlying the focus on asset ownership is that the characteristics defining what should be traded in which state of the world are difficult to write into a formal contract in a way that a third-party enforcer can unambiguously enforce. State-contingent trade is therefore unverifiable, so contracts written directly or indirectly on relationship-specific investments are infeasible. However, PRT assumes that relationship-specific investments, and therefore the value of different ex post trades, are commonly observable to U and D. Further, U and D can correctly anticipate the payoff consequences of different asset allocations and different levels of investment. Under the assumptions that relationship-specific investments are commonly observable and that players can foresee the payoff consequences of their actions, Maskin and Tirole (1999) show that the players should always be able to construct a mechanism in which they truthfully reveal the payoffs they would receive to a third-party enforcer. If the parties are able to write a contract on these announcements, then they should indirectly be able to write a contract on ex ante investments. This debate over the "foundations of incomplete contracting" mostly played out over the mid-to-late 1990s, but it has attracted some recent attention. We will discuss it in more detail later.

Further Reading See Antras (2003) and Acemoglu, Antras, and Helpman (2007) for applications of the incomplete-contracts framework to international trade; Hart, Shleifer, and Vishny (1997) and Besley and Ghatak

(2001) for applications to the optimal scope of government; and Aghion and Bolton (1992), Dewatripont and Tirole (1994), Hart and Moore (1998) for applications to financial contracting. Halonen (2002) and Baker, Gibbons, and Murphy (2002) explore how long-run relationships between the parties affects the optimal ownership of assets between them.

4.2.1 Foundations of Incomplete Contracts

The Property Rights Theory we discussed in the previous set of notes shows that property rights have value when contracts are incomplete, because they determine who has residual rights of control, which in turn protects that party (and its relationship-specific investments) from expropriation by its trading partners. In this note, I will discuss some of the commonly given reasons for why contracts might be incomplete, and in particular, I will focus on whether it makes sense to apply these reasons as justification for incomplete contracts in the Property Rights Theory.

Contracts may be incomplete for one of three reasons. First, parties might have private information. This is the typical reason given for why, in our discussion of the risk–incentives trade-off in moral hazard models, contracts could only depend on output rather than directly on the agent's effort. But in such models, contracts specified in advance are likely to be just as incomplete as contracts that are filled in at a later date.

Another reason often given is that it may just be costly to write a complicated state-contingent decision rule into a contract that is enforceable by a third party. This is surely important, and several authors have modeled this idea explicitly (Dye, 1985; Bajari and Tadelis, 2001; and Battigalli and Maggi, 2002) and drawn out some of its implications. Nevertheless, I will focus instead on the final reason.

The final reason often given is that parties may like to specify what to do in each state of the world in advance, but some of these states of the world are either unforeseen or indescribable by these parties. As a result, parties may leave the contract incomplete and "fill in the details" once more information has arrived. Decisions may be ex ante non-contractible but ex post contractible (and importantly for applied purposes, tractably derived by the economist as the solution to an efficient bargaining protocol), as in the Property Rights Theory.

I will focus in this note on the third justification, providing some of the arguments given in a sequence of papers (Maskin and Tirole, 1999; Maskin and Moore, 1999; Maskin, 2002) about why this justification alone is insufficient if parties can foresee the payoff consequences of their actions (which they must if they are to accurately assess the payoff consequences of different allocations of property rights). In particular, these papers point out that there exists auxiliary mechanisms that are capable of ensuring truthful revelation of mutually known, payoff-relevant information as part of the unique subgame-perfect equilibrium. Therefore, even though payoff-relevant information may not be directly observable by a third-party enforcer, truthful revelation via the mechanism allows for indirect verification, which implies that any outcome attainable with ex ante describable states of the world is also attainable with ex ante indescribable states of the world.

This result is troubling in its implications for the Property Rights Theory. Comparing the effectiveness of second-best institutional arrangements (e.g., property-rights allocations) under incomplete contracts is moot when a mechanism exists that is capable of achieving, in this setting, first best outcomes. In this note, I will provide an example of the types of mechanisms that have proposed in the literature, and I will point out a couple of recent criticisms of these mechanisms.

An Example of a Subgame-Perfect Implementation Mechanism

I will first sketch an elemental hold-up model, and then I will show that it can be augmented with a subgame-perfect implementation mechanism that induces first-best outcomes.

Hold-Up Problem There is a Buyer (B) and a Seller (S). S can choose an effort level $e \in \{0, 1\}$ at cost ce, which determines how much B values the good that S produces. B values this good at $v = v_L + e(v_H - v_L)$. There are no outside sellers who can produce this good, and there is no external market on which the seller could sell his good if he produces it. Assume $(v_H - v_L)/2 < c < (v_H - v_L)$.

There are three periods:

1. S chooses e. e is commonly observed but unverifiable by a third party.

2. v is realized. v is commonly observed but unverifiable by a third party.

3. With probability 1/2, B makes a take-it-or-leave-it offer to S, and with probability 1/2, S makes a take-it-or-leave-it offer to B.

This game has a unique subgame-perfect equilibrium. At $t = 3$, if B gets to make the offer, B asks for S to sell him the good at price $p = 0$. If S gets to make the offer, S demands $p = v$ for the good. From period 1's perspective, the expected price that S will receive is $E[p] = v/2$, so S's effort-choice problem is

$$\max_{e \in \{0,1\}} \frac{1}{2} v_L + \frac{1}{2} e (v_H - v_L) - ce.$$

Since $(v_H - v_L)/2 < c$, S optimally chooses $e^* = 0$. In this model, ex ante effort incentives arise as a by-product of ex post bargaining, and as a result, the trade price may be insufficiently sensitive to S's effort choice to induce him to choose $e^* = 1$. This is the standard hold-up problem. Note that the assumption that v is commonly observed is largely important, because it simplifies the ex post bargaining problem.

Subgame-Perfect Implementation Mechanism While effort is not verifiable by a third-party court, public announcements can potentially be used in legal proceedings. Thus, the two parties can in principle write a contract that specifies trade as a function of announcements $\hat{v}$ made by B. If B always tells the truth, then his announcements can be used to set prices that induce S to choose $e = 1$. One way of doing this is to implement a mechanism that allows announcements to be challenged by S and to punish B any time he is challenged. If S challenges only when B has told a lie, then the threat of punishment will ensure truth telling.

The crux of the implementation problem, then, is to give S the power to challenge announcements, but to prevent "he said, she said" scenarios wherein S challenges B's announcements when he has in fact told the truth. The key insight of SPI mechanisms is to combine S's challenge with a test that B will pass if and only if he in fact told the truth.

To see how these mechanisms work, and to see how they could in principle solve the hold-up problem, let us suppose the players agree ex-ante to subject themselves to the following multi-stage mechanism.

1. B and S write a contract in which trade occurs at price $p(\hat{v})$. $p(\cdot)$ is commonly observed and verifiable by a third party.

2. S chooses e. e is commonly observed but unverifiable by a third party.

3. v is realized. v is commonly observed but unverifiable by a third party.

4. B announces $\hat{v} \in \{v_L, v_H\}$. $\hat{v}$ is commonly observed and verifiable by a third party.

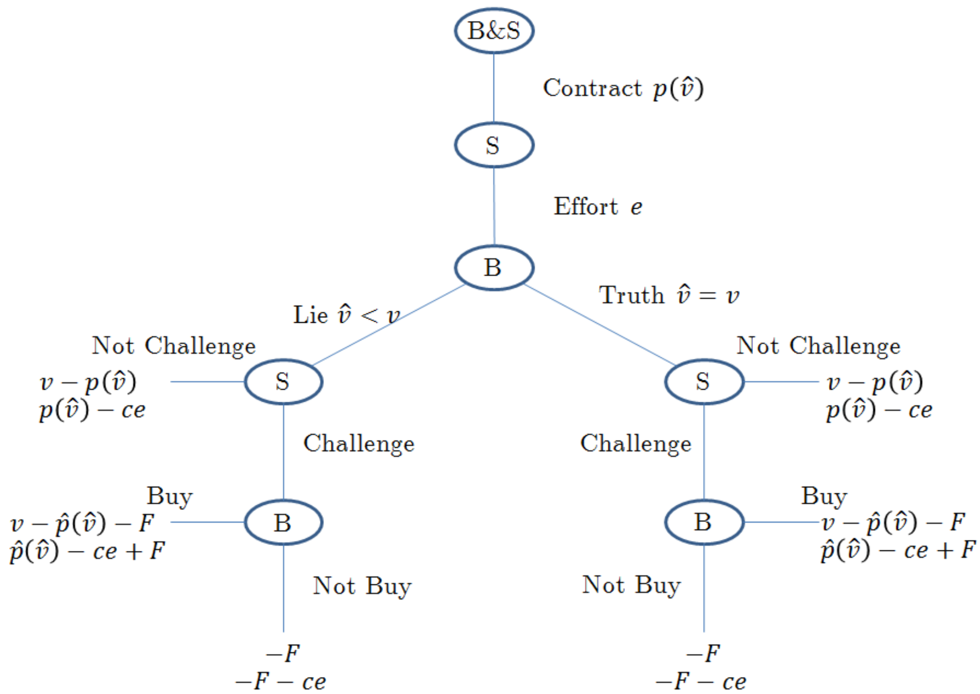
5. S can challenge B's announcement or not. The challenge decision is

commonly observed and verifiable by a third party. If S does not challenge the announcement, trade occurs at price $p(\hat{v})$. Otherwise, play proceeds to the next stage.

6. B pays a fine F to a third-party enforcer and is presented with a counter offer in which he can purchase the good at price $\hat{p}(\hat{v}) = \hat{v} + \varepsilon$. B's decision to accept or reject the counter off is commonly observed and verifiable by a third party.

7. If B accepts the counter offer, then S receives F from the third-party enforcer. If B does not, then S also has to pay F to the third-party enforcer.

The game induced by this mechanism seems slightly complicated, but we can sketch out the game tree in a relatively straightforward manner.



If the fine F is large enough, the unique SPNE of this game involves the following strategies. If B is challenged, he accepts the counter offer and buys the good at the counter-offer price if $\hat{v} < v$ and he rejects it if $\hat{v} \geq v$. S challenges B's announcement if and only if $\hat{v} < v$, and B announces $\hat{v} = v$. Therefore, B and S can, in the first stage, write a contract of the form $p(\hat{v}) = \hat{v} + k$, and as a result, S will choose $e^* = 1$.

To fix terminology, the mechanism starting from stage 4, after v has been realized, is a special case of the mechanisms introduced by Moore and Repullo (1988), so I will refer to that mechanism as the Moore and Repullo mechanism. The critique that messages arising from Moore and Repullo mechanisms can be used as a verifiable input into a contract to solve the hold-up problem (and indeed to implement a wide class of social choice functions) is known as the Maskin and Tirole (1999) critique. The main message of this criticism is that complete information about payoff-relevant variables and common knowledge of rationality implies that verifiability is not an important constraint to (uniquely) implement most social choice functions, including those involving efficient investments in the Property Rights Theory model.

The existence of such mechanisms is troubling for the Property Rights Theory approach. However, the limited use of implementation mechanisms in real-world environments with observable but non-verifiable information has led several recent authors to question the Maskin and Tirole critique itself. As Maskin himself asks: "To the extent that [existing institutions] do not replicate the performance of [subgame-perfect implementation mechanisms], one must ask why the market for institutions has not stepped into the breach, an important unresolved question." (Maskin, 2002)

Recent theoretical work by Aghion, Fudenberg, Holden, Kunimoto, and Tercieux (2012) demonstrates that the truth-telling equilibria in Moore and Repullo mechanisms are fragile. By perturbing the information structure slightly, they show that the Moore and Repullo mechanism does not yield even approximately truthful announcements for any setting in which multi-stage mechanisms are necessary to obtain truth-telling as a unique equilibrium of an indirect mechanism. Aghion, Fehr, Holden, and Wilkening

(2016) take the Moore and Repullo mechanism into the laboratory and show that indeed, when they perturb the information structure away from common knowledge of payoff-relevant variables, subjects do not make truthful announcements.

Relatedly, Fehr, Powell, and Wilkening (2016) take an example of the entire Maskin and Tirole critique into the lab and ensure that there is common knowledge of payoff-relevant variables. They show that in the game described above, there is a strong tendency for B's to reject counter offers after they have been challenged following small lies, S's are reluctant to challenge small lies, B's tend to make announcements with $\hat{v} < v$, and S's often choose low effort levels.

These deviations from SPNE predictions are internally consistent: if indeed B's reject counter offers after being challenged for telling a small lie, then it makes sense for S to be reluctant to challenge small lies. And if S often does not challenge small lies, then it makes sense for B to lie about the value of the good. And if B is not telling the truth about the value of the good, then a contract that conditions on B's announcement may not vary sufficiently with S's effort choice to induce S to choose high effort.

The question then becomes: why do B's reject counter offers after being challenged for telling small lies if it is in their material interests to accept such counter offers? One possible explanation, which is consistent with the findings of many laboratory experiments, is that players have preferences for negative reciprocity. In particular, after B has been challenged, B must immediately pay a fine of F that he cannot recoup no matter what he does going forward. He is then asked to either accept the counter offer, in which case S is rewarded for appropriately challenging his announcement; or he can reject the counter offer (at a small, but positive, personal cost), in which case S is punished for inappropriately challenging his announcement.

The failure of subjects to play the unique SPNE of the mechanism suggests that at least one of the assumptions of Maskin and Tirole's critique is not satisfied in the lab. Since Fehr, Powell, and Wilkening are able to design the experiment to ensure common knowledge of payoff-relevant information, it must be the case that players lack common knowledge of preferences and rationality, which is also an important set of implicit assumptions that are part of Maskin and Tirole's critique. Indeed, Fehr, Powell, and Wilkening provide suggestive evidence that preferences for reciprocity are responsible for their finding that B's often reject counter offers.

The findings of Aghion, Fehr, Holden and Wilkening and of Fehr, Powell, and Wilkening do not necessarily imply that it is impossible to find mechanisms in which in the unique equilibrium of the mechanisms, the hold-up problem can be effectively solved. What they do suggest, however, is that if subgame-perfect implementation mechanisms are to be more than a theoretical curiosity, they must incorporate relevant details of the environment in which they might be used. If people have preferences for reciprocity, then the mechanism should account for this. If people are concerned about whether their trading partner is rational, then the mechanism should account for this.

If people are concerned that uncertainty about what their trading partner is going to do means that the mechanism imposes undue risk on them, then the mechanism should account for this. Framing the implementation problem in the presence of these types of "behavioral" considerations and proving possibility or impossibility results could potentially be a fruitful direction for the implementation literature to proceed.

4.3 Influence Costs

At the end of the discussion of the Transaction-Cost Economics approach to firm boundaries, I mentioned that there are two types of costs that can arise when unprogrammed adaptation is required: costs associated with inefficient ex post decision making (adaptation costs) and costs associated with rent-seeking behavior (haggling costs). The TCE view is that when these costs are high for a particular transaction between two independent firms, it may make sense to take the transaction in-house and vertically integrate. I then described a model of adaptation costs in which this comparative static arises. I will now describe Powell's (2015) model of rent-seeking behavior in which similar comparative statics arise.

This model brings together the TCE view of haggling costs between firms as the central costs of market exchange and the Milgrom and Roberts (1988) view that influence costs—costs associated with activities aimed at persuading decision makers—represent the central costs of internal organization.

Powell asserts that the types of decisions that managers in separate firms argue about typically have analogues to the types of decisions that managers in different divisions within the same firm argue about (e.g., prices versus transfer prices, trade credit versus capital allocation) and that there is no reason to think a priori that the ways in which they argue with each other differ across different governance structures. They may in fact argue in different ways, but this difference should be derived, not assumed.

The argument that this model puts forth is the following. Decisions are ex post non-contractible, so whoever has control will *exercise* control (this is in contrast to the Property Rights Theory in which ex post decisions arise as the outcome of ex post efficient bargaining). As a result, the party who does not have control will have the incentives to try to influence the decision(s) of the party with control.

Control can be allocated via asset ownership, and therefore you can take away someone's right to make a decision. However, there are **position-specific private benefits**, so you cannot take away the fact that they care about that decision. In principle, the firm could replace them with someone else, but that person would also care about that decision. Further, while you can take away the rights to make a decision, you cannot take away the ability of individuals to try to influence whoever has decision rights, at least not unless you are willing to incur additional costs. As a result, giving control to one party reduces that party's incentives to engage in influence activities, but it intensifies the now-disempowered party's incentives to do so.

As in the Property Rights Theory, decision-making power affects parties' incentives. Here, it affects their incentives to try to influence the other party. This decision-making power is therefore a scarce resource that should be allocated efficiently. In contrast to the Property Rights Theory, decisions are ex post non-contractible. Consequently, whoever has control will exercise their control and will make different decisions ex post. So allocating control also affects the quality of ex post decision making. There may be a tension between allocating control to improve ex post decision making and allocating control to reduce parties' incentives to engage in influence activities.

Yet control-rights allocations are not the only instrument firms have for curtailing influence activities—firms can also put in place rigid organizational practices that reduce parties' incentives to engage in influence activities, but these practices may have costs of their own. Powell's model considers the interaction between these two substitute instruments for curtailing influence activities, and he shows that unified control and rigid organizational practices may complement each other.

Description Two managers, L and R, are engaged in a business relationship, and two decisions, d_1 and d_2 have to be made. Managers' payoffs for a particular decision depends on an underlying state of the world, $s \in S$. s is unobserved; however, L and R can potentially commonly observe an informative but manipulable signal σ. Managers bargain ex ante over a **control structure**, $c \in \mathcal{C} = \{I_L, I_R, NI, RNI\}$ and an **organizational practice**,

$p \in \mathcal{P} = \{O, C\}$. Under I_i, manager i controls both decisions; under NI, L controls d_1, and R controls d_2; and conversely under RNI. Under an **open-door organizational practice**, $p = O$, the signal σ is commonly observed by L and R, and under a **closed-door organizational practice**, $p = C$, it is not. A bundle $g = (c, p) \in \mathcal{G} \equiv \mathcal{C} \times \mathcal{P}$ is a **governance structure**. Assume that in the ex ante bargaining process, L makes an offer to R, which consists of a proposed governance structure g and a transfer $w \in \mathbb{R}$ to be paid to R. R can accept the offer or reject it in favor of outside option yielding utility 0.

Given a governance structure, each manager chooses a level of **influence activities**, λ_i, at private cost $k(\lambda)$, which is convex, symmetric around zero, and satisfies $k(0) = k'(0) = 0$. Influence activities are chosen prior to the observation of the public signal without any private knowledge of the state of the world, and they affect the conditional distribution of σ_p given the state of the world s. The managers cannot bargain over a signal-contingent decision rule ex ante, and they cannot bargain ex post over the decisions to be taken or over the allocation of control.

Timing The timing of the model is as follows:

1. L makes an offer of a governance structure $g \in \mathcal{G}$ and a transfer $w \in \mathbb{R}$ to R. g and w are publicly observed. R chooses whether to accept $(d = 1)$ or reject $(d = 0)$ this offer in favor of outside option yielding utility 0. $d \in D = \{0, 1\}$ is commonly observed.

2. L and R simultaneously choose influence activities $\lambda_L, \lambda_R \in \mathbb{R}$ at cost $k(\lambda)$; λ_i is privately observed by i.

3. L and R publicly observe signal σ_p.

4. Whoever controls decision ℓ chooses $d_\ell \in \mathbb{R}$.

5. Payoffs are realized.

Functional-Form Assumptions The signal under $p = O$ is linear in the state of the world, the influence activities, and noise: $\sigma_O = s + \lambda_L + \lambda_R + \varepsilon$. All random variables are independent and normally distributed with mean zero: $s \sim N(0, h^{-1})$ and $\varepsilon \sim N(0, h_\varepsilon^{-1})$. The signal under $p = C$ is uninformative, or $\sigma_C = \emptyset$. For the purposes of Bayesian updating, the signal-to-noise ratio of the signal is $\varphi_p = h_\varepsilon / (h + h_\varepsilon)$ under $p = O$ and, abusing notation, can be thought of as $\varphi_p = 0$ under $p = C$. Influence costs are quadratic, $k(\lambda_i) = \lambda_i^2/2$, and each manager's payoffs gross of influence costs are

$$U_i(s, d) = \sum_{\ell=1}^{2} \left[-\frac{\alpha_i}{2} (d_\ell - s - \beta_i)^2 \right], \ \alpha_i > 0, \beta_i \in \mathbb{R}.$$

Both managers prefer each decision to be tailored to the state of the world, but given the state of the world, manager i prefers that $d_1 = d_2 = s + \beta_i$, so there is disagreement between the two managers. Define $\Delta \equiv \beta_L - \beta_R > 0$ to be the **level of disagreement,** and assume that $\alpha_L \geq \alpha_R$: manager L cares more about the losses from not having her ideal decision implemented.

Further, assume that **managers operate at similar scales**: $\alpha_R \leq \alpha_L \leq \sqrt{3}\alpha_R$.

Although there are four possible control-rights allocations, only two will ever be optimal: unifying control with manager L or dividing control by giving decision 1 to L and decision 2 to R. Refer to unified control as **integration,** and denote it by $c = I$, and refer to divided control as **non-integration**, and denote it by $c = NI$. Consequently, there are effectively four governance structures to consider:

$$\mathcal{G} = \{(I, O), (I, C), (NI, O), (NI, C)\}.$$

Solution Concept A governance structure $g = (c, p)$ induces an extensive-form game between L and R, denoted by $\Gamma(g)$. A **Perfect-Bayesian Equilibrium** of $\Gamma(g)$ is a belief profile μ^*, an offer $(g^*, \theta^*), w^*$ of a governance structure and a transfer, a pair of influence-activity strategies $\lambda_L^* : \mathcal{G} \times \mathbb{R} \times D \times \Delta(s) \to \mathbb{R}$ and $\lambda_R^* : \mathcal{G} \times \mathbb{R} \times D \times \Delta(s) \to \mathbb{R}$, and a pair of decision rules $d_\ell^* : \mathcal{G} \times \mathbb{R} \times D \times \mathbb{R} \times \Sigma \times \Delta(s) \to \mathbb{R}$ such that the influence-activity strategies and the decision rules are sequentially optimal for each player given his/her beliefs, and μ^* is derived from the equilibrium strategy using Bayes's rule whenever possible.

This model is a signal-jamming game, like the career concerns model earlier in the class. Further, the assumptions we have made will ensure that players want to choose relatively simple strategies. That is, they will choose public influence-activity strategies $\lambda_L^* : \Delta(s) \to \mathbb{R}$ and $\lambda_R^* : \Delta(s) \to \mathbb{R}$ and decision rules $d_\ell^* : \mathcal{G} \times \Sigma \times \mathbb{R} \times \Delta(s) \to \mathbb{R}$.

The Program Take a governance structure g as given. Suppose manager i has control of decision ℓ under governance structure g. Let $\lambda^{g*} = (\lambda_L^{g*}, \lambda_R^{g*})$ denote the equilibrium level of influence activities. Manager i chooses d_ℓ to minimize her expected loss given her beliefs about the vector of influence activities, which I denote by $\hat{\lambda}(i)$. She therefore chooses d_ℓ^* to solve

$$\max_{d_\ell} E_s \left[-\frac{\alpha_i}{2} (d_\ell - s - \beta_i)^2 \middle| \sigma_p, \hat{\lambda}(i) \right].$$

She will therefore choose

$$d_\ell^{g*}\left(\sigma_p; \hat{\lambda}(i)\right) = E_s\left[s \middle| \sigma_p, \hat{\lambda}(i)\right] + \beta_i.$$

The decision that manager i makes differs from the decision manager $j \neq i$ would make if she had the decision right for two reasons. First, $\beta_i \neq \beta_j$, so for a given set of beliefs, they prefer different decisions. Second, out of equilibrium, they may differ in their beliefs about λ. Manager i knows λ_i but only has a conjecture about λ_j. These differences in out-of-equilibrium beliefs are precisely the channel through which managers might hope to change decisions through influence activities.

Since random variables are normally distributed, we can make use of the normal updating formula to obtain an expression for $E_s\left[s \middle| \sigma_P, \hat{\lambda}(i)\right]$. In particular, it will be a convex combination of the prior mean, 0, and the modified signal $\hat{s}(i) = \sigma_p - \hat{\lambda}_L(i) - \hat{\lambda}_R(i)$, which of course must satisfy $\hat{\lambda}_i(i) = \lambda_i$. The weight that i's preferred decision strategy attaches to the signal is given by the φ_p, so

$$d_\ell^{g*}\left(\sigma_p; \hat{\lambda}(i)\right) = \varphi_p \cdot \hat{s}(i) + \beta_i.$$

Given decision rules $d_\ell^{g*}(\sigma_p; \lambda^{g*})$, we can now set up the program that the managers solve when deciding on the level of influence activities to engage in. Manager j chooses λ_j to solve

$$\max_{\lambda_j} E_{s,\varepsilon}\left[\sum_{\ell=1}^{2} -\frac{\alpha_j}{2}\left(d_\ell^{g*}(\sigma_p; \lambda^{g*}) - s - \beta_j\right)^2\right] - k(\lambda_j).$$

Taking first-order conditions, we get:

$$\begin{aligned} \left|k'\left(\lambda_j^{g*}\right)\right| &= \left|E_{s,\varepsilon}\left[\sum_{\ell=1}^{2} -\alpha_j \underbrace{\left(d_\ell^{g*}(\sigma_p; \lambda^{g*}) - s - \beta_j\right)}_{=0 \text{ if } j \text{ controls } g;\ =\Delta \text{ otherwise}} \underbrace{\frac{\partial d_\ell^{g*}}{\partial \sigma}}_{\varphi_p} \underbrace{\frac{\partial \sigma}{\partial \lambda_j}}_{=1}\right]\right| \\ &= N_{-j}^c \alpha_j \Delta \varphi_p, \end{aligned}$$

where N_{-j}^c is the number of decisions manager j does not control under control structure c.

Finally, at $t = 1$, L will make an offer g, w to

$$\max_{g,w} E_{s,\varepsilon}\left[\sum_{\ell=1}^{2} -\frac{\alpha_L}{2}\left(d_\ell^{g*}\left(\sigma_p;\lambda^{g*}\right) - s - \beta_L\right)^2\right] - k\left(\lambda_L^{g*}\right) - w$$

subject to R's participation constraint:

$$E_{s,\varepsilon}\left[\sum_{\ell=1}^{2} -\frac{\alpha_R}{2}\left(d_\ell^{g*}\left(\sigma_p;\lambda^{g*}\right) - s - \beta_R\right)^2\right] - k\left(\lambda_R^{g*}\right) + w \geq 0.$$

w will be chosen so that the participation constraint holds with equality, so that L's problem becomes:

$$\max_{g} \underbrace{E_{s,\varepsilon}\left[\sum_{i\in\{L,R\}}\sum_{\ell=1}^{2} -\frac{\alpha_i}{2}\left(d_\ell^{g*}\left(\sigma_p;\lambda^{g*}\right) - s - \beta_i\right)^2\right] - \sum_{i\in\{L,R\}} k\left(\lambda_i^{g*}\right)}_{W(g)}.$$

The **Coasian Program** is then

$$\max_{g\in\mathcal{G}} W\left(g\right).$$

Solution Managers' payoffs are quadratic. The first term can therefore be written as the sum of the mean-squared errors of d_1^{g*} and d_2^{g*} as estimators of the **ex post surplus-maximizing decision**, which is

$$s + \frac{\alpha_L}{\alpha_L + \alpha_R}\beta_L + \frac{\alpha_R}{\alpha_L + \alpha_R}\beta_R$$

for each decision. As a result, the first term can be written as the sum of a bias term and a variance term (recall that for two random variables X and Y, $Var(Y) = E[Var(Y|X)] + Var(E[Y|X])$):

$$W(c,p) = -(ADAP(p) + ALIGN(c) + INFL(c,p)),$$

where after several lines of algebra, the expressions for these terms are:

$$\begin{aligned} ADAP(p) &= (\alpha_L + \alpha_R)\frac{1-\varphi_p}{h} \\ ALIGN(c) &= \frac{\alpha_R}{2}\Delta^2 + \frac{\alpha_L}{2}\Delta^2\mathbf{1}_{c=NI} + \frac{\alpha_R}{2}\Delta^2\mathbf{1}_{c=I} \\ INFL(c,p) &= \left(\frac{1}{2}\left(\alpha_R\Delta\varphi_p\right)^2 + \frac{1}{2}\left(\alpha_L\Delta\varphi_p\right)^2\right)\mathbf{1}_{c=NI} + \frac{1}{2}\left(2\alpha_R\Delta\varphi_p\right)^2\mathbf{1}_{c=I}. \end{aligned}$$

$ADAP(p)$ represents the costs associated with basing decision making on a noisy signal. $ADAP(p)$ is higher for $p = C$, because under $p = C$, even the noisy signal is unavailable. $ALIGN(c)$ represents the costs associated with the fact that ex post, decisions will always be made in a way that are not ideal for someone. Whether they are ideal for manager L or R depends on the control structure c. Finally, $INFL(c,p)$ are the influence costs, $k(\lambda_L^{g*}) + k(\lambda_R^{g*})$. When $p = C$, these costs will be 0, since there is no signal to manipulate. When $p = O$, these costs will depend on the control structure.

There will be two trade-offs of interest.

Influence-cost–alignment-cost trade-off First, let us ignore $ADAP(p)$ and look separately at $ALIGN(c)$ and $INFL(c,p)$. To do so, let us begin with $INFL(c,p)$. When $p=C$, these costs are clearly 0. When $p=O$, they are:

$$\begin{aligned} INFL(I,O) &= \frac{1}{2}(2\alpha_R\Delta\varphi_O)^2 \\ INFL(NI,O) &= \frac{1}{2}(\alpha_L\Delta\varphi_O)^2+\frac{1}{2}(\alpha_R\Delta\varphi_O)^2. \end{aligned}$$

Divided control minimizes influence costs, as long as managers operate at similar scale:

$$INFL(I,O)-INFL(NI,O)=\frac{1}{2}\left(3(\alpha_R)^2-(\alpha_L)^2\right)(\Delta\varphi_O)^2>0.$$

Next, let us look at $ALIGN(c)$. When $c=I$, manager L gets her ideal decisions on average, but manager R does not:

$$ALIGN(I)=\alpha_R\Delta^2.$$

When $g=NI$, each manager gets her ideal decision correct on average for one decision but not for the other decision:

$$ALIGN(NI)=\frac{\alpha_L+\alpha_R}{2}\Delta^2.$$

When $\alpha_L=\alpha_R$, so that $ALIGN(I)=ALIGN(NI)$, we have that

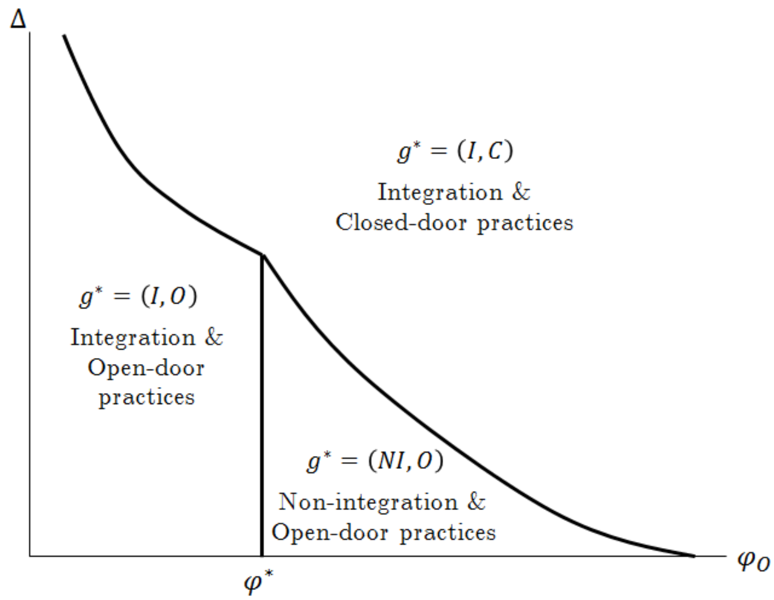
$INFL(I,O) - INFL(NI,O) > 0$, so that influence costs are minimized under non-integration. When $p = C$, so that there are no influence costs, and $\alpha_L > \alpha_R$, $ALIGN(I) < ALIGN(NI)$, so that alignment costs are minimized under integration. Unified control reduces ex post alignment costs and divided control reduces influence costs, and there is a trade-off between the two.

Influence-cost–adaptation-cost trade-off Next, let us ignore $ALIGN(c)$ and look separately at $ADAP(p)$ and $INFL(c,p)$. Since

$$ADAP(p) = (\alpha_L + \alpha_R)\frac{1-\varphi_p}{h},$$

adaptation costs are higher under closed-door practices, $p = C$. But when $p = C$, influence costs are reduced to 0. Closed-door practices therefore eliminate influence costs but reduce the quality of decision making, so there is a trade-off here as well. Finally, it is worth noting that when $p = C$, influence activities are eliminated, so the parties might as well unify control, since doing so reduces ex post alignment costs. That is, closed-door policies and integration are complementary.

The following figure illustrates optimal governance structures for different model parameters. The figure has three boundaries, each of which correspond to different results from the literature on influence activities and organizational design. The vertical boundary between (I,O) and (NI,O) is the "Meyer, Milgrom, and Roberts boundary": a firm rife with politics should perhaps disintegrate.



The diagonal boundary between (I, O) and (I, C) is the "Milgrom and Roberts boundary": rigid decision-making rules should sometimes be adopted within firms. These two boundaries highlight the idea that non-integration and rigid organizational practices are substitute instruments for curtailing influence costs: sometimes a firm prefers to curtail influence activities with the former and sometimes with the latter. Finally, the boundary between (NI, O) and (I, C) is the "Williamson boundary." If interactions across firm boundaries, which are characterized by divided control and open lines of communication, invite high levels of influence activities, then it may be optimal instead to unify control *and* adopt closed-door practices.

At the end of the day, any theory of the firm has to contend with two polar questions. First, why are all transactions not carried out in the market? Second, why are all transactions not carried out within a single large firm? TCE identifies "haggling costs" as an answer to the first question and "bureaucratic costs of hierarchy" as an answer to the second. Taking a parallel approach focused on the costs of internal organization, Milgrom and Roberts identify "influence costs" as an answer to the second question and "bargaining costs" between firms as an answer to the first. The model presented above blurs the distinction between TCE's "haggling costs" and Milgrom and Roberts's "influence costs" by arguing that the types of decisions over which parties disagree across firm boundaries typically have within-firm analogs, and the methods parties employ to influence decision makers within firms are not exogenously different than the methods they employ between firms.

This perspective implies, however, that unifying control *increases* influence costs, in direct contrast to Williamson's claim that "fiat [under integration] is frequently a more efficient way to settle minor conflicts": modifying firm boundaries without adjusting practices does not solve the problem of haggling. However, adopting rigid organizational practices in addition to unifying control provides a solution. Fiat (unified control) appears effective at eliminating haggling, precisely because it is coupled with bureaucracy. This influence-cost approach to firm boundaries therefore suggests that bureaucracy is not a cost of integration. Rather, it is an endogenous response to the actual cost of integration, which is high levels of influence activities.

Finally, we can connect the implications of this model to the empirical implications of the TCE approach. As with most theories of the firm, directly testing the model's underlying causal mechanisms is inherently difficult, because many of the model's dependent variables, such as the levels of influence activities and the optimality of ex post decision making, are unlikely to be observed by an econometrician. As a result, the model focuses on predictions regarding how potentially observable independent variables, such as environmental uncertainty and the level of ex post disagreement, relate to optimal choices of potentially observable dependent variables, such as the integration decision or organizational practices.

In particular, the model suggests that if interactions across firm boundaries involve high levels of influence costs, it may be optimal to unify control and adopt closed-door practices. This may be the case when the level of ex post disagreement (Δ) is high and when the level of ex ante uncertainty (h) is low. The model therefore predicts a positive relationship between integration and measures of ex post disagreement and a negative relationship between integration and measures of ex ante uncertainty.

The former prediction is consistent with the TCE hypothesis and is consistent with the findings of many empirical papers, which we will soon discuss. The second prediction contrasts with the TCE hypothesis that greater environmental uncertainty leads to more contractual incompleteness and more scope for ex post haggling, and therefore makes integration a relatively more appealing option. This result is in line with the failure of empirical TCE papers to find consistent evidence in favor of TCE's prediction that integration and uncertainty are positively related.

Chapter 5

Evidence on Firm Boundaries (TBA)

TBA

Part III

Assorted Topics in Organizations

Chapter 6

Competition and Organization

For much of the class so far, we have focused on how exogenous external factors shape firm-level organization decisions. In our discussion of incentives, we took as given the contracting space and the information structure and derived the optimal action that the firm wants the agent to take as well as the contract designed to get him to do so. In our discussion of firm boundaries, we again took as given the contracting space (which was necessarily less complete than the parties would have preferred) and other characteristics of the firm's environment (such as the returns to managers' investments, costs of adapting to unforeseen contingencies, and the informativeness of manipulable public signals) and derived optimal control-rights allocations and other complementary organizational variables. This week, we will look at how external factors shape firm-level organization decisions *through their effects on product-market competition and the price mechanism*. We will begin with the treatment of a classic topic: the effects of product-market competition on managerial incentives. We will then discuss the interplay between firm boundaries and the competitive environment and between relational incentive contracts and the competitive environment. We will be interested in particular in the question of how the firms' competitive environment and firm-level productivity interact.

6.1 Competition and Managerial Incentives

The claim that intense product-market competition disciplines firms and forces them to be more productive seems straightforward and obviously true. Hicks (1935) described this intuition evocatively as "The best of all monopoly profits is a quiet life." (p. 8) Product-market competition requires firm owners and firm managers to work hard to remove slack (or as Leibenstein (1966) describes it, "X-inefficiency") from their production processes in order to survive. In addition, recent empirical work (Backus, 2014) suggests that the observed correlation between competition and productivity is driven by within-firm productivity improvements in more-competitive environments. As straightforward as this claim may seem, it has been remarkably problematic to provide conditions under which it holds. In this section, I will sketch a high-level outline of a model that nests many of the examples from the literature, and I will hopefully provide some intuition about why this claim has been difficult to pin down.

There are $N \geq 1$ firms that compete in the product market. In what follows, I will look at monopoly markets ($N = 1$) and duopoly markets ($N = 2$), and I will focus on the incentives a single firm has to reduce its marginal costs. Firms are ex ante identical and can produce output at a constant marginal cost of c. Prior to competing in the product market, firm 1 can reduce its marginal cost of production to $c - e$ at cost $C_1(e)$ for $e \in E = [0, c]$. Given e, firm 1 earns gross profits $\pi_1(e)$ on the product market. Firm 1 therefore chooses e^* to solve

$$\max_e \pi_1(e) - C_1(e),$$

and the ultimate question in this literature is: when does an increase in product-market competition lead to an increase in e^*?

As you might expect, the reason why this question is difficult to answer at such a high level is that it is not clear what "an increase in product-market competition" *is*. And different papers in this literature present largely different notions of what it means for one product market to be more competitive than another. Further, some papers (Hart, 1983; Scharfstein, 1988; Hermalin, 1992; Schmidt, 1997) focus specifically on how product-market competition affects the *costs* of implementing different effort levels, $C_1(e)$, while others (Raith, 2003; Vives, 2008) focus on how product-market competition affects the *benefits* of implementing different effort levels, $\pi_1(e)$.

To see how these fit together, note that we can define firm 1's *product-*

market problem as

$$\pi_1(e) = \max_{p_1}(p_1 - (c - e))\, q_1(p_1),$$

where $q_1(p_1)$ is either the market demand curve if $N = 1$ or, if $N = 2$, firm 1's residual demand curve given firm 2's equilibrium choice of its competitive variable. Throughout, we will assume that for any e, there is a unique Nash equilibrium of the product-market competition game. Otherwise, we would have to choose (and, importantly, justify) a particular equilibrium-selection rule.

If the firm's manager is its owner, $C_1(e)$ captures the effort costs associated with reducing the firm's marginal costs. If the firm's manager is not its owner, $C_1(e)$ additionally captures the agency costs associated with getting the manager to choose effort level e. Let $W \subset \{w : Y \to \mathbb{R}\}$, where $y \in Y$ is a contractible outcome. As we described in our discussion of incentives, the case where the firm's manager is its owner can be captured by a model in which $Y = E$, so that effort is directly contractible. Under this formulation, we can define firm 1's *agency problem* as

$$C_1(e) = \min_{w \in W} \int w(y)\, dF(y|\, e)$$

subject to

$$\int u(w(y) - c(e))\, dF(y|\, e) \geq \int u(w(y) - c(e'))\, dF(y|\, e')$$

for all $e' \in E$. In the remarks below, I provide expressions for $C_1(e)$ for the three elemental models we discussed in the first week of class.

We can now see that the original problem,

$$\max_e \pi_1(e) - C_1(e),$$

which looked simple, actually masks a great deal of complication. In particular, it is an optimization problem built upon two sub-problems, and so the question then is how changes in competition affect either or both of these sub-problems. Despite these complications, we can still make some progress.

In particular, focusing on the benefits side, we can apply the envelope theorem to the product-market competition problem to get

$$\pi_1'(e) = q_1^*(e),$$

where $q_1^*(e) = q_1(p_1^*(e))$, and again, we can write

$$q_1^*(e) = q_1^*(0) + \int_0^e \frac{dq_1^*(s)}{ds} ds = q_1^*(0) + \int_0^e \eta_1^*(s)\, ds,$$

where $\eta_1^*(\cdot)$ is the quantity pass-through of firm 1's residual demand curve. That is, $\eta_1^*(e) = q_1'(p_1^*(e))\,\rho(e)$, where $\rho(e)$ is the pass-through of firm 1's residual demand curve:

$$\rho(e) = -\frac{1}{1 - \left(q(p_1^*(e)) / q_1'(p_1^*(e))\right)'}.$$

See Weyl and Fabinger (2013) for an excellent discussion on the role of pass-through for many comparative statics in industrial organization. Further, we can write

$$\begin{aligned}\pi_1(e) &= \pi_1(0) + \int_0^e q_1^*(s)\,ds = \pi_1(0) + \int_0^e \left[q_1^*(0) + \int_0^t \eta_1^*(s)\,ds\right] dt \\ &= \pi_1(0) + e q_1^*(0) + \int_0^e (e-s)\,\eta_1^*(s)\,ds,\end{aligned}$$

where the last equality can be derived by integrating by parts.

Next, for the three elemental models of incentives we discussed in the first week of class, we can derive explicit expressions for $C_1(e)$, which I do in the remarks below.

Remark 1 (Limited Liability) *Suppose* $Y = \{0,1\}$, $\Pr[y=1|\,e] = e$, $c(e) = \frac{c}{2}e^2$, $W = \{w : Y \to \mathbb{R} : w(1) \geq w(0) \geq 0\}$, $u(x - c(e)) = x - c(e)$, *and* $\bar{u} = 0$*. Then*

$$C_1(e) = c(e) + R(e),$$

where $R(e) = ce$ *are the incentive rents that must be provided to the agent to induce him to choose effort level* e*.*

Remark 2 (Risk-Incentives) *Suppose* $Y = \mathbb{R}$, $y = e + \varepsilon$, $\varepsilon \sim N(0,\sigma^2)$, $c(e)$ *is increasing, convex, and differentiable,* $W = \{w : Y \to \mathbb{R} : w(y) = s + by, s, b \in \mathbb{R}\}$, *and* $u(x - c(e)) = -\exp\{-r[x - c(e)]\}$, *and* $\bar{u} = 0$*. Then*

$$C_1(e) = c(e) + r(e),$$

where $r(e) = \frac{1}{2} r\sigma^2 e^2$ *is the risk premium that the agent must be paid in order to provide him with strong enough incentives to choose effort level* e*.*

Remark 3 (Misalignment) *Suppose* $Y = \{0, 1\}$, $\Pr[y = 1 | e_1, e_2] = f_1 e_1 + f_2 e_2 \equiv e$, $P = \{0, 1\}$, $\Pr[p = 1 | e_1, e_2] = g_1 e_1 + g_2 e_2$, $c(e_1, e_2) = \frac{1}{2}(e_1^2 + e_2^2)$, $W = \{w : P \to \mathbb{R}\}$, *and* $u(x - c(e_1, e_2)) = x - c(e_1, e_2)$, *and* $\bar{u} = 0$*. Then*

$$C_1(e) = c(e) + m(e),$$

where $m(e) = \frac{1}{2}(\tan\theta)^2 e^2$, *where* $\tan\theta$ *is the tangent of the angle between* (f_1, f_2) *and* (g_1, g_2)*. If* $(f_1, f_2) \neq (g_1, g_2)$, *in order to get the agent to choose a particular probability of high output,* e, *the principal has to provide him with incentives that send him off in the "wrong direction," which implies that his effort costs are higher than they would be if* $(f_1, f_2) = (g_1, g_2)$, *which the principal must compensate the agent for.* $m(e)$ *represents the costs due to the misalignment of the performance measure. If* $(f_1, f_2) = (g_1, g_2)$, *then* $\tan\theta = 0$*. If* $(f_1, f_2) \perp (g_1, g_2)$, *then* $\tan\theta = \infty$*.*

If we restrict attention to one of the three elemental models of incentive provision described in the remark above, we have:

$$C_1(e) = c(e) + R(e) + r(e) + m(e),$$

so that the original problem can now be written as

$$\max_e eq_1^* (0) + \int_0^e (e - s) \eta_1^* (s) \, ds - c(e) - R(e) - r(e) - m(e).$$

We therefore have that $q_1^* (0), \eta_1^* (\cdot)$, and $c(\cdot) + R(\cdot) + r(\cdot) + m(\cdot)$ constitutes a set of *sufficient statistics* for firm 1's product-market problem and its agency problem, respectively. In other words, in order to figure out what the optimal effort choice e^* by the firm is, we only need to know a couple things. First, we need to know how effort choices e map into the quantity of output the firm will sell in the product market, $q_1^* (e) = q_1^* (0) + \int_0^e \eta_1^* (s) \, ds$. This schedule fully determines the *benefits* of choosing different effort levels. Second, we need to know what the expected wage bill associated with implementing effort e at minimum cost is. This schedule fully determines the *costs* to the firm of choosing different effort levels.

The motivating question then becomes: how does an increase in product-market competition affect $q_1^* (0)$, $\eta_1^* (\cdot)$, and $R(\cdot) + r(\cdot) + m(\cdot)$. (We can ignore the effects of competition on $c(\cdot)$, since the agent's cost function is usually taken to be an exogenous parameter of the model.) The point that now needs to be clarified is: what is an "increase in product-market competition?" Different papers in the literature take different approaches to addressing this point. Hart (1983), Nalebuff and Stiglitz (1983), and Scharfstein (1988) view an increase in competition as providing a firm with additional information about industry-wide cost shocks. Hermalin (1992) and Schmidt

(1997) view an increase in competition as a reduction in firm profits, conditional on a given effort level by the agent. Raith (2003) and Vives (2008) view an increase in competition as either an exogenous increase in the number of competitors in a market, or if entry is endogenous, an increase in competition can be viewed as either an increase in product substitutability across firms, an increase in the market size, or a decrease in entry costs.

We are now in a position to describe the laundry list of intuitions that each of these papers provides. Going down the list, we can view Hart (1983) and Nalebuff and Stiglitz (1983) as showing that an increase in competition reduces required risk premia $r(e)$, since Principals will be able to use the additional information provided through the market price by more competition as part of an optimal contract. By Holmstrom's informativeness principle, since this additional information is informative about the agent's action, the risk premium necessary to induce any given effort level e is reduced. They therefore conclude that an increase in competition increases e^* because of this effect. However, as Scharfstein (1988) points out, reducing $r(e)$ is not the same as reducing $r'(e)$. In particular, the $r(\cdot)$ schedule can fall by more for lower effort levels than for higher effort levels, implying that an increase in competition could actually *reduce* effort e^*. Alternatively, one could think of competition as increasing the alignment between the contractible performance measure and the firm's objectives. In this case, an increase in competition would decrease $\tan\theta$ and therefore decrease $m'(e)$, which would in turn lead to an increase in e^*.

Hermalin (1992) emphasizes the role of negative profit shocks when the agent has some bargaining power, and there are income effects (as there would be if $u(z)$ satisfied decreasing absolute risk aversion). If competition reduces firm profits, if the agent has bargaining power, it also reduces the agent's expected wage. This in turn makes the agent less willing to substitute out of actions that increase expected wages (i.e., high effort in this context) and into actions that increase private benefits (i.e., low effort in this context). Under this view, an increase in competition effectively reduces $r'(e)$, thereby increasing e^*.

Schmidt (1997) argues that an increase in competition increases the likelihood that the firm will go bankrupt. If an agent receives private benefits from working for the firm, and the firm is unable to capture these private benefits from the agent (say because of a limited-liability constraint), then the agent will be willing to work harder (under a given contract) following an increase in competition if working harder reduces the probability that the firm goes bankrupt. This intuition therefore implies that competition reduces $R'(e)$, the marginal incentive rents required to induce the agent to work harder. In turn, under this "increased threat of avertable bankruptcy risk," competition can lead to an increase in e^*.

In each of these papers so far, the emphasis has been on how an increase in competition impacts marginal agency costs: $R'(e) + r'(e) + m'(e)$ and therefore how it impacts the difference between what the firm would like the agent to do (i.e., the first-best effort level) and what the firm optimally gets the agent to do (i.e., the second-best effort level). However, putting agency costs aside, it is not necessarily clear how an increase in competition affects the firm's *first-best* level of effort. If we ignore agency costs, the problem becomes

$$\max_e eq_1^*(0) + \int_0^e (e-s)\,\eta_1^*(s)\,ds - c(e).$$

The question is therefore: how does an increase in competition affect $q_1^*(0)$ and $\eta_1^*(e)$? Raith (2003) and Vives (2008) argue that an increase in competition affects the firm's optimal *scale of operations* (which corresponds to $q_1^*(0)$) and the firm's residual-demand elasticity (which is related to but is not the same as $\eta_1^*(e)$). In my view, these are the first-order questions that should have been the initial focus of the literature. First, develop an understanding of how an increase in competition affects what the firm would like the agent to do; *then*, think about how an increase in competition affects what the firm optimally gets the agent to do.

Raith (2003) provides two sets of results in a model of spatial competition. First, he shows that an exogenous increase in the number of competitors reduces $q_1^*(e)$ for each e and therefore always reduces e^*. He then shows that, in a model with endogenous firm entry, an increase in parameters that foster additional competition affects e^* in different ways, because they affect $q_1^*(e)$ in different ways. An increase in product substitutability has the effect of reducing the profitability of the industry and therefore reduces entry into the industry. Raith assumes that the market is covered, so aggregate sales remain the same. This reduction in the number of competitors therefore increases $q_1^*(e)$ for each e and therefore increases e^*. An increase in the market size leads to an increase in the profitability of the industry and therefore an increase in entry. However, the increased entry does not (under the functional forms he assumes) fully offset the increased market size, so $q_1^*(e)$ nevertheless increases for each e, and therefore an increase in market size increases e^*. A reduction in entry costs, however, leads to an increase in firm entry, reducing the sales per firm ($q_1^*(e)$) and therefore reduces e^*.

Raith's results are intuitively plausible and insightful in part because they focus on the $q_1^*(\cdot)$ schedule, which is indeed the appropriate sufficient statistic for the firm's problem absent agency costs. However, his results are derived under a particular market structure, so a natural question to ask is whether they are also relevant under alternative models of product-market competition. This is the question that Vives (2008) addresses. In particular, he shows that while some of the effects that Raith finds do indeed depend on his assumptions about the nature of product-market competition, most of them hold under alternative market structures as well. His analysis focuses on the scale effect (i.e., how does an increase in competition affect $q_1^*(0)$) and the elasticity effect (i.e., how does an increase in competition affect the elasticity of firm 1's residual demand curve?), but as pointed out above, the latter effect should be replaced with a quantity pass-through effect (i.e., how does an increase in competition affect the quantity pass-through of firm 1's residual demand curve?)

To illustrate how competition could affect the quantity pass-through in different ways depending on the nature of competition, suppose there are two firms, and the market demand curve is $D(p) = A - Bp$. This market demand curve is a constant quantity pass-through demand curve $\eta(p) = B/2$. Suppose firms compete by choosing supply functions, and firm 2 chooses supply functions of the form $S_2(p) = a_2 + b_2 p$. An increase in a_2 or an increase in b_2 can be viewed as an aggressive move by firm 2. I will think of an increase in either of these parameters as an increase in competition. Given firm 2 chooses supply function $S_2(p) = a_2 + b_2 p$, firm 1's residual demand curve is $q_1(p) = \tilde{A} - \tilde{B}p$, where $\tilde{A} = A - a_2$ and $\tilde{B} = B + b_2$. The quantity pass-through of firm 1's residual demand curve is

$$\eta_1(p) = \frac{B + b_2}{2}.$$

Firm 1 solves

$$\max_p (p - (c - e)) q_1(p),$$

which yields the solution

$$\begin{aligned} p^*(e) &= \frac{\tilde{A}}{2\tilde{B}} + \frac{1}{2}(c - e) \\ q_1(p^*(e)) &= \underbrace{\frac{\tilde{A} - \tilde{B}c}{2}}_{q(p^*(0))} + \int_0^e \underbrace{\frac{\tilde{B}}{2}}_{\eta_1^*(s)} ds. \end{aligned}$$

Two polar forms of competition will highlight the key differences I want to stress. The first form of competition I will consider is standard *Cournot competition*, in which firm 2 chooses supply function parameter a_2 and fixes $b_2 = 0$. A higher value of a_2 is a more aggressive move by firm 2, and we can see that

$$q_1\left(p^*\left(e\right)\right) = \frac{A - Bc}{2} - \frac{a_2}{2} + \int_0^e \frac{B}{2} ds.$$

If we interpret an increase in competition as a more aggressive move by firm 1's competition, then an increase in competition decreases $q_1\left(p^*\left(e\right)\right)$ for all e, which in turn implies a decrease in firm 1's optimal choice e^*.

The other form of competition I will consider is *rotation competition*, in which firm 2 chooses supply function parameters a_2 and b_2 such that $\left(A - a_2 - \left(B + b_2\right)c\right)$ is held constant. That is, firm 2 can only choose a_2 and b_2 such that $a_2 + b_2 c = 0$. Firm 2 therefore chooses b_2, which yields a supply function $S_2\left(p\right) = b_2\left(p - c\right)$. A higher value of b_2 is a more aggressive move by firm 2. Further we can see that

$$q_1\left(p^*\left(e\right)\right) = \frac{A - Bc}{2} + \int_0^e \frac{B + b_2}{2} ds.$$

In this case, an increase in competition increases the quantity pass-through of firm 1's residual demand curve and therefore increases $q_1\left(p^*\left(e\right)\right)$ for all e, which in turn implies an increase in e^*.

6.2 Competition and Firm Boundaries (TBA)

TBA

6.3 Productivity Measures

In this note, I will discuss two commonly used measures of total factor productivity that are used in the literature. Which one is used in a particular application is typically determined by data availability, but we will see that the two measures have significantly different interpretations. A firm chooses capital K and labor L at constant unit costs r and w respectively to produce quantity according to a Cobb-Douglas production function $q(K, L) = AK^{\alpha}L^{1-\alpha}$, which it sells on a product market at price p.

The first measure of productivity we will be concerned with is **quantity total factor productivity** (referred to as $TFPQ$), which is given by:

$$TFPQ = \frac{q(K, L)}{K^{\alpha}L^{1-\alpha}} = A.$$

That is, $TFPQ$ is a ratio of physical output to physical inputs, appropriately weighted according to their production elasticities. (i.e., $\alpha = \frac{d \log q}{d \log K}$) Differences in $TFPQ$ across firms correspond to variations in output across firms that are not explained by variation in inputs. In other words, it is a measure of our ignorance about the firm's underlying production process. One objective of organizational economics is to improve our understanding of firms' production processes.

The second measure of productivity we will discuss is **revenue total factor productivity** (referred to as $TFPR$), which is given by

$$TFPR = \frac{p \cdot q(K,L)}{(rK)^{\alpha}(wL)^{1-\alpha}} = \frac{p}{r^{\alpha}w^{1-\alpha}} \cdot A$$

or sometimes

$$TFPR = \frac{p \cdot q(K,L)}{K^{\alpha}L^{1-\alpha}} = p \cdot A.$$

That is, $TFPR$ is a ratio of revenues to input costs, appropriately weighted according to their production elasticities. Differences in $TFPR$ across firms correspond to variations in revenues across firms that are not explained by variation in measured costs. Since $TFPR$ depends on unit costs and output prices, it may depend on the market conditions that determine them. Under some assumptions, $TFPR$ may depend exclusively on market conditions, and variations in $TFPR$ across firms is indicative of misallocation of productive resources across firms. (Hsieh and Klenow, 2009) In this note, I will focus primarily on $TFPR$.

TFPR Under Constraints and Market Power A firm produces according to a Cobb-Douglas production function $q(K,L) = AK^{\alpha}L^{1-\alpha}$ and is subject to constraints in either labor or capital. Suppose a firm is constrained to produce with $L \leq \bar{L}$ and $K \leq \bar{K}$. Let $p(q(K,L))$ denote the firm's residual inverse demand curve, and let $\varepsilon_{qp}(q) = \frac{1}{\frac{dp(q)}{dq}\frac{q}{p(q)}}$. The firm's problem is to

$$\max p\left(q\left(K,L\right)\right)q\left(K,L\right)-wL-rK$$

subject to $L\leq\bar{L}$ and $K\leq\bar{K}$. The Lagrangian is

$$\mathcal{L}=p\left(q\left(K,L\right)\right)q\left(K,L\right)-wL-rK+\lambda_{K}\left(\bar{K}-K\right)+\lambda_{L}\left(\bar{L}-L\right).$$

Taking first-order conditions, we get

$$\begin{aligned}MRP_{K}^{*}&=p'\left(q^{*}\right)q_{K}q^{*}+p\left(q^{*}\right)q_{K}=r+\lambda_{K}\\MRP_{L}^{*}&=p'\left(q^{*}\right)q_{L}q^{*}+p\left(q^{*}\right)q_{L}=w+\lambda_{L}\end{aligned}$$

We can rearrange these expressions and derive

$$\begin{aligned}\left(p'\left(q^{*}\right)\frac{q^{*}}{p\left(q^{*}\right)}+1\right)p\left(q^{*}\right)q_{K}&=r+\lambda_{K}\\\left(p'\left(q^{*}\right)\frac{q^{*}}{p\left(q^{*}\right)}+1\right)p\left(q^{*}\right)q_{L}&=w+\lambda_{L}\end{aligned}$$

or

$$\begin{aligned}p\left(q^{*}\right)q_{K}&=\frac{\left|\varepsilon_{qp}\left(q^{*}\right)\right|}{\left|\varepsilon_{qp}\left(q^{*}\right)\right|-1}\left(r+\lambda_{K}\right)\\p\left(q^{*}\right)q_{L}&=\frac{\left|\varepsilon_{qp}\left(q^{*}\right)\right|}{\left|\varepsilon_{qp}\left(q^{*}\right)\right|-1}\left(w+\lambda_{L}\right).\end{aligned}$$

Under Cobb-Douglas, we know that $q_K = \alpha \frac{q}{K}$ and $q_L = (1-\alpha)\frac{q}{L}$. We therefore have

$$\begin{aligned} p(q^*)\frac{q^*}{rK^*} &= \frac{1}{\alpha}\frac{|\varepsilon_{qp}(q^*)|}{|\varepsilon_{qp}(q^*)|-1}\frac{MRP_K^*}{r} \\ p(q^*)\frac{q^*}{wL^*} &= \frac{1}{1-\alpha}\frac{|\varepsilon_{qp}(q^*)|}{|\varepsilon_{qp}(q^*)|-1}\frac{MRP_L^*}{w} \end{aligned}$$

Revenue total factor productivity is defined as revenue divided by a geometric average of capital expenditures and labor expenditures, and is therefore given by

$$\begin{aligned} TFPR^* &= p(q^*)\frac{q^*}{(rK^*)^{\alpha}(wL^*)^{1-\alpha}} = \left[p(q^*)\frac{q^*}{rK^*}\right]^{\alpha}\left[p(q^*)\frac{q^*}{wL^*}\right]^{1-\alpha} \\ &= \frac{1}{\alpha^{\alpha}(1-\alpha)^{1-\alpha}}\frac{|\varepsilon_{qp}(q^*)|}{|\varepsilon_{qp}(q^*)|-1}\left(\frac{MRP_K^*}{r}\right)^{\alpha}\left(\frac{MRP_L^*}{w}\right)^{1-\alpha} \end{aligned}$$

In this model, heterogeneity in $TFPR$ arises from heterogeneity in α, $|\varepsilon_{qp}(q^*)|$, $\frac{MRP_K^*}{r}$, or $\frac{MRP_L^*}{w}$. Heterogeneity in α arises from differences in technology. Heterogeneity in $|\varepsilon_{qp}(q^*)|$ can result from idiosyncratic demand shocks. Heterogeneity in $\frac{MRP_K^*}{r}$ or $\frac{MRP_L^*}{w}$ results from either heterogeneity in the capital and labor constraints or heterogeneity in A.

If $TFPR$ is defined as revenue divided by a geometric average of capital and labor inputs (rather than expenditures), it is given by

$$\begin{aligned} TFPR^* &= p(q^*)\frac{q^*}{(K^*)^{\alpha}(L^*)^{1-\alpha}} = \left[p(q^*)\frac{q^*}{K^*}\right]^{\alpha}\left[p(q^*)\frac{q^*}{L^*}\right]^{1-\alpha} \\ &= \frac{1}{\alpha^{\alpha}(1-\alpha)^{1-\alpha}}\frac{|\varepsilon_{qp}(q^*)|}{|\varepsilon_{qp}(q^*)|-1}(MRP_K^*)^{\alpha}(MRP_L^*)^{1-\alpha}. \end{aligned}$$

In this case, heterogeneity can arise from differences in the price of inputs.

Why doesn't this equal $p \cdot A$ when $|\varepsilon_{qp}| \to \infty$ and $\lambda_K, \lambda_L \to 0$? In fact it does. The Kuhn-Tucker conditions can only be replaced by the first-order conditions I derived above when $p, r,$ and w are such that this expression equals $p \cdot A$.

There are two special cases of this expression in the literature:

Hsieh and Klenow '09 Hsieh and Klenow (2009) measure $TFPR^*$ as value-added (revenues) divided by a share-weighted geometric average of the net book value of fixed capital of a firm, net of depreciation (rK^*) and labor compensation, which is the sum of wages, bonuses, and benefits. They derive an expression for $TFPR^*$ under the assumption of constant elasticity demand of the form $q(p) = \frac{D}{p^{-\varepsilon}}$, so that $\varepsilon_{qp} = \varepsilon$. In this case,

$$TFPR^* \propto \left(\frac{MRP_K^*}{r}\right)^{\alpha}\left(\frac{MRP_L^*}{w}\right)^{1-\alpha}.$$

Heterogeneity in $TFPR^*$ is then interpreted as firm-specific wedges (i.e. heterogeneity in MRP_K^*/r or MRP_L^*/w, which should both be equal to 1 at the non-distorted optimum).

Foster, Haltiwanger, Syverson '08 Foster, Haltiwanger, and Syverson measure $TFPR^*$ as plant-level prices times $TFPQ^*$, which uses physical output data, labor measured in hours, capital as plant's book values of equipment and structures deflated to 1987 dollars, and materials expenditures (which I will ignore). They assume firms are not constrained, so $MRP_K^* = r$ and $MRP_L^* = w$ for all firms. Their definition of $TFPR^*$ corresponds to the second measure listed above, and therefore

$$TFPR^* \propto \frac{|\varepsilon_{qp}(q^*)|}{|\varepsilon_{qp}(q^*)| - 1}$$

They interpret differences in $TFPR^*$ as arising from differences in $\frac{|\varepsilon_{qp}(q^*)|}{|\varepsilon_{qp}(q^*)|-1}$.

Alternative Interpretation of TFPR Let $\tilde{\alpha} = \frac{rK^*}{rK^*+wL^*}$ and $1 - \tilde{\alpha} = \frac{wL^*}{rK^*+wL^*}$ denote the realized cost shares of capital and labor, respectively. These need not be equal to α and $1 - \alpha$, because constraints may tilt the optimal input mix. Note that

$$\begin{aligned} \frac{(rK^*)^\alpha (wL^*)^{1-\alpha}}{rK^* + wL^*} &= (1-\tilde{\alpha})\,\tilde{\alpha}\left(\frac{wL^*}{rK^*}\right)^{1-\alpha} + \tilde{\alpha}\,(1-\tilde{\alpha})\left(\frac{rK^*}{wL^*}\right)^{\alpha} \\ &= \tilde{\alpha}^\alpha (1-\tilde{\alpha})^{1-\alpha} \\ rK^* + wL^* &= \frac{(rK^*)^\alpha (wL^*)^{1-\alpha}}{\tilde{\alpha}^\alpha (1-\tilde{\alpha})^{1-\alpha}} \end{aligned}$$

Then,

$$\begin{aligned} TFPR &= \frac{p\left(q^*\right)q^*}{\left(rK^*\right)^{\alpha}\left(wL^*\right)^{1-\alpha}} = \frac{1}{\tilde{\alpha}^{\alpha}\left(1-\tilde{\alpha}\right)^{1-\alpha}}\frac{p\left(q^*\right)q^*}{rK^*+wL^*} \\ &= \frac{1}{\tilde{\alpha}^{\alpha}\left(1-\tilde{\alpha}\right)^{1-\alpha}}\frac{REV}{TVC} \end{aligned}$$

In particular, we see TFPR is proportional to the revenue (REV) to total variable cost (TVC) ratio. This ratio is given by

$$\frac{REV}{TVC} = \frac{\left|\varepsilon_{qp}\left(q^*\right)\right|}{\left|\varepsilon_{qp}\left(q^*\right)\right|-1}\left(\frac{\tilde{\alpha}}{\alpha}\frac{MRP_K^*}{r}\right)^{\alpha}\left(\frac{1-\tilde{\alpha}}{1-\alpha}\frac{MRP_L^*}{w}\right)^{1-\alpha}.$$

The profit/cost ratio is this expression plus 1. Average profits per dollar of inputs are therefore increasing in markups and distortions (corrected by distortions in input mix).

Heterogeneous Returns-to-Scale Suppose a firm faces a downward-sloping residual demand curve $p\left(q\right)$ for its product, and it has a Cobb-Douglas production function $q = AL^{\beta}$, where L is labor inputs, and β is the elasticity of production with respect to labor. Further, suppose the firm faces a labor constraint $L \leq \bar{L}$. The firm's Lagrangian is

$$\mathcal{L} = p\left(AL^{\beta}\right)AL^{\beta} - wL + \lambda_L\left(\bar{L}-L\right)$$

and its first-order conditions are given by:

$$\begin{aligned} w+\lambda_L^* &= MRP_L^* = \left(p'\left(q^*\right)\frac{q^*}{p\left(q^*\right)}+1\right)p\left(q^*\right)q_L^* \\ MRP_L^* &= \frac{\left|\varepsilon^*\left(q^*\right)\right|-1}{\left|\varepsilon^*\left(q^*\right)\right|}\beta\frac{pq^*}{L^*}, \end{aligned}$$

where $|\varepsilon|$ is the elasticity of the firm's (strategic) residual demand curve. Average labor productivity is given by

$$ALP = \frac{p^*\left(q^*\right)q^*}{wL^*} = \frac{|\varepsilon|}{|\varepsilon|-1}\frac{MRP_L^*/w}{\beta}$$

Heterogeneity in average labor productivity is driven by heterogeneity in either MRP_L/w (i.e. labor wedges), heterogeneity in β (i.e. heterogeneous technologies), or heterogeneity in $|\varepsilon|$ (which could be due to idiosyncratic demand shocks). Under perfect competition, $|\varepsilon| = \infty$, so this becomes

$$ALP = \frac{MRP_L^*/w}{\beta}. \tag{6.1}$$

Here, prices are exogenous to the model, which should eliminate the concerns about the differences between TFP and $TFPR$. However, we see from this expression that heterogeneity in ALP still does not depend on TFP. In fact, all heterogeneity in average labor productivity is driven by heterogeneous returns to scale (and if so-desired, labor constraints).

Mismeasured Scale Effects Let us maintain the assumption that $|\varepsilon| = \infty$, so that $TFPR$ should just be a constant multiple of TFP (and should therefore reflect A). The real issue is that ALP does not correct for scale effects. If it did, then it would not be driven by β, and it would reflect $TFPR$ (and hence TFP, since prices are exogenous):

$$TFPR^* = \frac{pq^*}{(wL^*)^\beta} = p\frac{A\,(L^*)^\beta}{(wL^*)^\beta} = p\frac{A}{w^\beta}. \tag{6.2}$$

More generally, suppose the scale is assumed (by the econometrician) to be $\gamma \geq \beta$. Then

$$\frac{pq}{(wL)^\gamma} = ALP\,(wL)^{1-\gamma} \tag{6.3}$$

We know that for $\gamma = \beta$, $p\frac{A}{w^\beta} = ALP\,(wL)^{1-\beta}$. Solving this for $(wL)^{1-\gamma}$, we have

$$(wL)^{1-\gamma} = \left(p\frac{A}{w^\beta}\frac{1}{ALP}\right)^{\frac{1-\gamma}{1-\beta}}$$

Plugging this into (6.3), we get

$$\frac{pq}{(wL)^\gamma} = \left(p\frac{A}{w^\beta}\right)^{1-\frac{\gamma-\beta}{1-\beta}}\left(\frac{MRP_L/w}{\beta}\right)^{\frac{\gamma-\beta}{1-\beta}}.$$

Thus, when TFP is calculated using the incorrect returns to scale, the result is a geometric average of (6.2), which depends on actual TFP (A), and (6.1), which depends on labor constraints (MRP_L/w) and the actual returns to scale (β).

When is $TFPR$ increasing in TFP? Let ρ denote the pass-through rate characterized by the demand system. That is, $\rho = \frac{dp}{d(C')}$, where $C' = \frac{c}{TFP}$ is the marginal cost of production. In the no-constraints case, we can derive the following expression:

$$TFPR'(TFP) = \left(\frac{|\varepsilon(p^*(TFP))|}{|\varepsilon(p^*(TFP))| - 1} - \rho \right) \frac{c}{TFP},$$

which is positive whenever $\rho < \frac{|\varepsilon(p^*(TFP))|}{|\varepsilon(p^*(TFP))|-1}$. For the case of linear demand (as in Foster, Haltiwanger, and Syverson), $\rho < 1 < \frac{|\varepsilon(p^*(TFP))|}{|\varepsilon(p^*(TFP))|-1}$, so $TFPR$ is increasing in TFP. For the case of constant elasticity demand, $\rho = \frac{|\varepsilon(p^*(TFP))|}{|\varepsilon(p^*(TFP))|-1}$, so $TFPR'(TFP) = 0$, and therefore $TFPR$ is independent of TFP (which is emphasized in Hsieh and Klenow). If pass-through is sufficiently high (i.e. significantly greater than one-for-one), then it can in fact be the case that $TFPR$ is decreasing in TFP. Intuitively, this would happen if prices fell by more than TFP increased following an increase in TFP.

Chapter 7

Dynamic Inefficiencies[1]

A vast literature in contract theory considers a question that is fundamental to organizational economics: "what contracting imperfections might lead firms to act differently from the 'black box' production functions we typically assume?" This is an important question if we want to understand how firms operate within an economy, which in turn is important if we want to advise managers or policymakers.

These notes will focus on a particular departure from 'black box' production functions highlighted by contract theory. In dynamic settings, an optimal contract may not be *sequentially optimal*: there may be histories at which the *continuation* contract looks very different than a contract that would be written if the relationship started at that history. In other words, history matters: past actions and outcomes can affect firm performance, even

[1] This chapter was written by Dan Barron.

if they do not affect the technology or options available to the firm. To borrow a phrase from Li, Matouschek, and Powell (forthcoming), these dynamic contracts carry the "burden of past promises": actions are chosen not to maximize surplus in the future, but to fulfill past obligations to the players.

These lecture notes will cover four papers. We will begin by discussing a classic paper in contract theory by Fudenberg, Holmstrom, and Milgrom (1990). We will then link this seminal analysis to three recent papers:

1. Fuchs (2007), "Contracting with Repeated Moral Hazard with Private Evaluations,"
2. Halac (2012), "Relational Contracts and the Value of Relationships," and
3. Board (2011), "Relational Contracts and the Value of Loyalty."

As time permits, I will also discuss Barron and Powell (2016), "Policies in Relational Contracts."

7.1 Fudenberg, Holmstrom, and Milgrom (1990)

Setup Consider a very general dynamic contracting problem. The Principal and Agent interact for $T < \infty$ periods with a common discount factor $\delta \in [0, 1]$. In each period $t \geq 0$, the sequence of events is:

1. The Agent learns some information $\theta_t \in \Theta$.

2. The Agent chooses effort $e_t \in \mathbb{R}$.

3. Output y_t is realized according to the distribution $F\left(\cdot|\left\{\theta_{t'}, e_{t'}\right\}_{t' \leq t}\right)$.

4. The Principal pays $w_t \in W$ to the Agent, where $W \subset \mathbb{R}_+$.

5. The Agent chooses to consume c_t and save $s_t = s_{t-1} + w_t - c_t$.

Denote $y = \{y_1, \ldots, y_T\}$ and similarly for θ, e, w, and s. Payoffs for the Principal and the Agent are, respectively $\Pi(y, w)$ and $U(\theta, c, e, s_T)$.

We assume that the Principal can commit to a long-term contract. Define $x_t \in X_t$ as the set of contractible variables in period t, and define $X^t = \times_{t'=0}^{t} X_{t'}$ as the contractible histories of length t. In general, the sequence of outcomes $\{y_{t'}\}_{t'=0}^{T}$ will be contractible, but x_t might contain other variables as well. A formal contract is a mapping $w : \cup_{t=0}^{T} X^t \rightarrow W$ that gives a payment in period t for each possible history of contractible variables $x^t \in X^t$. We leave unspecified how this contract is offered. The Principal and Agent simultaneously accept or reject the contract, with outside options yielding payoffs $\bar{\pi}$ and $\bar{u}$, respectively.

Definitions: Incentive Compatibility, Individual Rationality, and Sequential Efficiency Let $\mathcal{H}^t$ be the set of full histories in period t. The Agent's effort and consumption plans map $e, c : \cup_{t=0}^{T} \mathcal{H}^t \times \Theta \rightarrow \mathbb{R}$. Given a contract w, (c_w, e_w) is *incentive compatible* (IC) if at each history h^t,

$$(c_w, e_w) \in \underset{c,e}{\operatorname{argmax}}\, E\left[U(\theta, c, e, s_T)|\, h^t, w\right].$$

Define U_w and Π_w as the Agent and Principal's total payoffs from contract w and the IC (c_w, e_w) that maximize the Principal's payoff. Then w is *individually rational* (IR) if

$$\begin{aligned} U_w &\geq \bar{u} \\ \Pi_w &\geq \bar{\pi}. \end{aligned}$$

A contract w is *efficient* if there exists no other contract w' such that

$$\begin{aligned} U_w &\leq U_{w'} \\ \Pi_w &\leq \Pi_{w'} \end{aligned}$$

with at least one inequality strict.

Define $U_w(x^t)$ and $\Pi_w(x^t)$ as the expected payoffs for the Agent and Principal given contractible history x^t. Then w is *sequentially efficient* if for every x^t, there exists no other contract w' such that

$$\begin{aligned} U_w\left(x^t\right) &\leq U_{w'}\left(x^t\right) \\ \Pi_w\left(x^t\right) &\leq \Pi_{w'}\left(x^t\right) \end{aligned}$$

with at least one inequality strict. That is, a contract w' is sequentially efficient if at the start of each period, the continuation contract, effort plan, and message plan are efficient with respect to the information partition given by X^t.

When is the Efficient Contract not Sequentially Efficient? If an efficient contract is sequentially efficient, then the contract at the start of each period resembles a contract that could have been written if the game were just starting. The relationship between the Principal and the Agent does not "develop inefficiencies" over time, except to the extent that the production technology itself changes.

Why might dynamic inefficiencies arise in an efficient contract? Fudenberg, Holmstrom, and Milgrom (1990) outline (at least) three reasons why an efficient contract is not necessarily sequentially efficient

1. The payoff frontier between the Principal and the Agent is not downward-sloping. Given contractible history X^t, if $U_w(x^t) \leq U_{w'}(x^t)$, then $\Pi_w(x^t) \geq \Pi_{w'}(x^t)$.

    - If this holds, then the only way to punish the Agent may be to also punish the Principal. But simultaneously punishing both Principal and Agent is inefficient.

2. The Principal learns information about the Agent's past effort over time: y_t is not a sufficient statistic for (y_t, e_t).

    - Suppose that $y_{t+t'}$ contains information about e_t that y_t does not. Then the optimal contract would motivate e_t by making payments contingent on $y_{t+t'}$ in a way that might not be sequentially efficient.

3. At the start of each period, not all variables that determine future preferences and production technology are contractible.

    - If future payoffs depend on noncontractible information, then adverse selection might lead to sequential inefficiencies.

The next sections consider a series of examples that illustrate why sequential inefficiencies might arise if any of these three conditions hold.

7.1.1 Payoff Frontier not Downward-Sloping

Consider the following two-period example: in each period $t \in \{0, 1\}$,

1. The Agent chooses effort $e_t \in \{0, 1\}$ at cost ke_t.

2. Output $y_t \in \{0, H\}$ realized, with

$$\Pr[y_t = H] = q + (p - q)\, e_t$$

   and $0 < q < p < 1$.

3. Payoffs are

$$\begin{aligned} \Pi &= y_0 + y_1 - w_0 - w_1 \\ U &= c_0 + c_1 - ke_0 - ke_1. \end{aligned}$$

4. The agent has limited liability: $w_0, w_1 \geq 0$. He receives $-\infty$ if $s_T < 0$.

The Principal can write a long-term contract as a function of realized output: $w_0(y_0)$ and $w_1(y_0, y_1)$. However, the Agent has limited liability, so $w_0, w_1 \geq 0$. It is easy to show that saving plays no role in this contract, so $c_t = w_t$ without loss of generality.

Let $w_1 = w_1^{y_1}$ following output y_1. In period $t = 1$, the IC constraint is

$$\frac{k}{p-q} \leq w_1^H - w_1^0$$

which implies that $w_1^H \geq k/(p-q)$. Thus, the Agent's utility if $e_1 = 1$ can be no less than

$$\frac{k}{p-q}p - k = \frac{q}{p-q}k > 0$$

if $q > 0$. In other words, the Agent must *earn a rent* to be willing to work hard. Suppose that motivating high effort in $t = 1$ is efficient, or

$$H - \frac{p}{p-q}k > 0.$$

Now consider period $t = 0$. In any sequentially efficient contract, the Agent must choose $e_1 = 1$, regardless of y_0. Therefore, the Principal can earn no more than

$$2\left(H - \frac{p}{p-q}k\right)$$

in a sequentially efficient contract.

Consider the following alternative: if $y_0 = 0$, then $w_1^H = w_1^0 = 0$ and $e_1 = 0$. Intuitively, following low output in $t = 0$, the Agent is not motivated in $t = 1$ (and earns no rent). This alternative is clearly sequentially inefficient. However, the Agent is now willing to work hard in $t = 0$, since he loses both a bonus *and* a future rent if output is low.

Therefore, the Principal can motivate the Agent to work hard in $t = 0$ if

$$w_0^H + \frac{q}{p-q}k \geq \frac{k}{p-q}$$

or

$$w_0^H \geq \frac{1-q}{p-q}k.$$

Relative to the sequentially efficient payoff, the alternative with low effort leads to a higher payoff for the Principal if

$$H - \frac{p\,(1-q)}{p-q}k + p\left(H - \frac{p}{p-q}k\right) \geq 2\left(H - \frac{p}{p-q}k\right)$$

or

$$p\,(H-k) \geq H - \frac{p}{p-q}k.$$

Both sides are strictly positive. This inequality holds if, for example, $p \approx 1$ and $1 - k/H \approx q$. So in some circumstances, the Principal would like to implement a contract that hurts *both* the Agent *and* herself following low output in order to provide incentives to the Agent in period 0. But such a contract is clearly not sequentially efficient.

7.1.2 Information about Past Performance Revealed Over Time

Consider the following two-period example, which is adapted from Fudenberg and Tirole (1990):

1. In period 0, the Agent chooses $e_0 \in \{0, 1\}$ at cost ke_0. Output is $y_0 = 0$ with probability 1.

2. In period 1, the Agent has no effort choice: $e_1 = 0$. Output is $y_1 \in \{0, H\}$, with $y_1 = H$ with probability pe_0.

3. Payoffs are

$$\begin{aligned} \Pi &= y_1 - w_0 - w_1 \\ U &= u(c_0) + u(c_1) - ke_0 \end{aligned}$$

   with $u(\cdot)$ strictly concave. The Agent receives a payoff of $-\infty$ if $s_T < 0$.

As in the first example, output y_1 is contractible. It is easy to see that $w_0 = c_0 = 0$ and $c_1 = w_1$ in this setting.

The optimal contract will condition only on y_1. In order to motivate the Agent in $t = 0$, the payment w^{y_1} for output y_1 must satisfy

$$pu\left(w^H\right) + (1-p)\, u\left(w^0\right) - k \geq u\left(w^0\right)$$

or

$$u\left(w^H\right) - u\left(w^0\right) \geq \frac{k}{p}.$$

Now, consider the contract starting at the beginning of $t = 1$. The effort e_0 has already been chosen in this contract. Because u is strictly concave,

$$pu\left(w^H\right) + (1-p)\,u\left(w^0\right) < u\left(pw^H + (1-p)\,w^0\right).$$

The Principal would earn strictly higher profits if she instead offered a contract with $w = pw^H + (1-p)\,w^0$. So if $e_0 = 1$, then any sequentially efficient contract must have a constant payment in $t = 1$. But then $e_0 = 1$ is not incentive-compatible.

In this example, the efficient contract requires the payment in $t = 1$ to vary in output in order to motivate the Agent to work hard in $t = 0$. After the Agent has worked hard, however, the parties have an incentive to renegotiate in order to shield the Agent from risk (since u is strictly concave). The efficient contract is not sequentially efficient, because y_1 contains information about e_0 that is not also contained in y_0.

7.1.3 Players have Private Information about the Future

Consider the following two-period example, which is adapted from Fudenberg, Holmstrom, and Milgrom (1990). In each $t \in \{0, 1\}$,

1. In period $t = 0$, the Agent does not exert effort ($e_0 = 0$) and produces no output ($y_0 = 0$).

2. In period $t = 1$, the Agent chooses $e_1 \in \{0, 1\}$ at cost ke_1.

3. Output in $t = 1$ is $y_1 \in \{0, H\}$, with $\Pr[y_1 = H] = q + e(p - q)$ for $0 < q < p < 1$.

4. Payoffs are

$$\begin{aligned} \Pi &= y_1 - w_0 - w_1 \\ U &= u(c_0) + u(c_1) - ke_1. \end{aligned}$$

The Agent receives a payoff of $-\infty$ if $s_T < 0$.

Importantly, note that the Agent works only in period t but consumes in both periods 0 and 1. The agent is able to borrow money to consume early (so $c_0 > w_0$). The Principal can write a formal contract on y_1, but not on consumption or savings.

Suppose the agent consumes c_0 in $t = 0$ and chooses $e_1 = 1$. The Agent could always deviate by choosing some other consumption $\tilde{c}_0$ in $t = 0$ and then shirking: $e_1 = 0$. Therefore, the Agent's IC constraint is:

$$\begin{aligned} & u(c_0) + pu\left(w^H - c_0\right) + (1-p)\,u\left(w^0 - c_0\right) \\ \geq\ & u(\tilde{c}_0) + qu\left(w^H - \tilde{c}_0\right) + (1-q)\,u\left(w^0 - \tilde{c}_0\right). \end{aligned}$$

Any IC contract must satisfy $w^H - w^0 > 0$. Because u is strictly concave, it must be that $c_0 > \tilde{c}_0$. That is, the Agent consumes more in $t = 0$ if he expects to work hard in $t = 1$. He does so, because he expects higher monetary compensation in $t = 1$.

Now, suppose the Agent chooses to consume c_0, because he anticipates working hard. Once he has consumed this amount, his IC constraint becomes

$$\begin{aligned} & pu\left(w^H - c_0\right) + (1-p)\,u\left(w^0 - c_0\right) - k \\ \geq\; & qu\left(w^H - c_0\right) + (1-q)\,u\left(w^0 - c_0\right). \end{aligned}$$

This constraint is slack, because $c_0 \neq \tilde{c}_0$. Therefore, the parties could renegotiate the original contract to expose the Agent to less risk (by making w^H and w^0 closer to each other). So the efficient contract is sequentially inefficient.

The key for this inefficiency is that the contract cannot condition on consumption c_0. But consumption in period 0 determines the Agent's willingness to work hard in $t = 1$. That is, c_0 is effectively "private information" about the Agent's utility in $t = 1$.

7.2 Recent Papers that Apply FHM

7.2.1 Upward-Sloping Payoff Frontier: Board (2011)

In many real-world environments, a Principal interacts with several Agents. For example, Toyota allocates business among its suppliers. The government interacts with multiple companies in procurement auctions. Bosses oversee multiple workers. And so on.

This paper, along with Andrews and Barron (forthcoming) and Barron and Powell (2016), focus on dynamics in *multilateral* relationships.

Setup A single principal P interacts repeatedly with N agents A with discount factor δ. In each period of the interaction:

1. For each A $i \in \{1, \ldots, N\}$, the cost of investing in i, $c_{i,t}$, is publicly observed.

2. P invests in one agent i, pays $c_{i,t}$. Let $Q_{i,t}$ be the probability of investing in Agent i.

3. Chosen Agent creates value v and chooses an amount $p_t \in [0, v]$ to keep. The remainder goes to P.

Payoffs are $u_{i,t} = p_t Q_{i,t}$ for Agent i and $\pi_t = \sum_{i=}^{N} Q_{i,t} (v - p_t - c_{i,t}) Q_{i,t}$ for P.

A note about the model: this problem has a limited-liability constraint, which is built into the requirement that $p_t \in [0, v]$.

Limited Liability leads to Sequential Inefficiencies Suppose that the Principal can commit to an investment scheme. That is, $Q_{i,t}$ can be made conditional on any past variables. However, the Agent cannot commit to repay the Principal.

If this game is played once, then $p_t = v$ and so P chooses not to invest. Suppose the game is played repeatedly, but P invests in Agent i only once. Then again, i has no incentive to give money to the Principal, $p_t = v$, and the Principal prefers not to invest in i. So **repeated interaction with a single agent** is key to providing incentives.

Define

$$U_{i,t} = E\left[\sum_{s=t}^{\infty} \delta^{s-t} p_s Q_{i,s}\right]$$

as Agent i's continuation surplus. Define

$$\Pi_t = \sum_{i=1}^{N} E\left[\sum_{s=t}^{\infty} \delta^{s-t} \left(v - p_s - c_{i,s}\right) Q_{i,s}\right]$$

as P's continuation surplus.

Dynamic Enforcement: Agent i is only willing to follow a strategy if

$$\left(U_{i,t} - v\right) Q_{i,t} \geq 0$$

for all t. Otherwise, if P invests in i, then i can run away with the money and earn v. P can always choose not to invest in i, so if i does run away then he earns 0 in the continuation game.

Principal's Problem: At time $t = 0$, P chooses $\{Q_{i,t}\}$ and $\{p_t\}$ to maximize his profit subject to the dynamic-enforcement constraint. **Principal profit**

at time 0 **equals total surplus minus agent payoff at time** 0:

$$\Pi_0 = E\left[\sum_{t=0}^{\infty} \delta^t \left(v - c_{i,t}\right) Q_{i,t}\right] - \sum_{i=1}^{N} U_{i,0}.$$

The key observation is that an Agent can be motivated by **promises of future rent**. In particular, promising future rent motivates an agent **in every period before that rent is paid**. Therefore, the Principal only really needs to give Agent i rent **once**.

More precisely, define $\tau_i(t) \geq t$ as the period on or after period t in which $Q_{i,\tau_i(t)} > 0$. Agent i earns 0 surplus if P does not invest in i, so

$$U_{i,t} = E\left[U_{i,\tau_i(t)} \delta^{\tau_i(t)-t}\right].$$

i's dynamic-enforcement constraint is only satisfied if

$$U_{i,\tau_i(t)} \geq v,$$

so $U_{i,t} \geq E\left[v\delta^{\tau_i(t)-t}\right]$.

Can we make this inequality bind? **Yes**. One way is to ask the Agent to keep just enough so that he earns v continuation surplus. For example,

$$p_t = vE_t\left[1 - \delta^{\tau_i(t+1)}\right]$$

would work. To see why, suppose $U_{i,\tau_i(t+1)} = v$. Then if $Q_{i,t} = 1$, Agent i's continuation surplus is

$$p_t + \delta^{\tau_i(t+1)} v = v.$$

Biases in Investment Decisions: We know that $U_{i,\tau_i(0)} = v$. So

$$\Pi_0 = E\left[\sum_{t=0}^{\infty} \delta^t \left(v - c_{i,t}\right) Q_{i,t}\right] - \sum_{i=1}^{N} E\left[v\delta^{\tau_i(0)}\right].$$

So P's objective is to maximize **total surplus minus a "rent cost"** v **that is incurred the first time** P **trades with a new agent.**

Main Result: Define $\mathcal{I}_t$ as the set of Agents with whom P has already traded in period t. Then in each period:

1. If P invests in $i \in \mathcal{I}$, then $i \in \mathcal{I}$ has the lowest cost **among agents in** $\mathcal{I}$.

2. If $i \notin \mathcal{I}$ and there exists $j \in \mathcal{I}$ with

$$\left(c_{j,t} - c_{i,t}\right) \leq \left(1 - \delta\right) v,$$

then P never invests in i.

If costs are i.i.d. across agents and over time, then there exists a unique integer n^* such that the optimal contract entails at most n^* insiders.

Principal Dynamic Enforcement: What if the Principal cannot commit to an investment plan? If P is punished by reversion to static Nash following a deviation, then it is easy to sustain the contract above.

But what if Agents have trouble coordinating their punishments? More precisely, suppose that P loses no more than

$$\Pi_{i,t} = \sum_{s=t}^{\infty} \delta^{s-t} Q_{i,s} \left(v - p_s - c_{i,t}\right)$$

if he deviates by not investing in Agent i. Intuitively, Agent i stops repaying P, but the other Agents keep on repaying as before.

Suppose $c \in [\underline{c}, \bar{c}]$ is i.i.d. across Agents and periods. Suppose $v > \bar{c}$ or $\underline{c} > 0$. Then $\Pi_{i,t}$ increases in δ. This is not *a priori* obvious, because the **number of insiders** increases in δ. However, as the number of insiders increases, it is increasingly likely that an insider has costs very close to $\underline{c}$. Therefore, P does not gain much by including an additional insider. This effect dominates as $\delta \to 1$.

The upshot: as $\delta \to 1$, the Principal is willing to follow the optimal contract **even if punishment is bilateral**.

7.2.2 Information about Past Outcomes is Revealed Over Time: Fuchs (2007)

Setup The following is a simplified version of Fuchs (2007).

Consider a repeated game with a Principal P and Agent A who share a discount factor $\delta \in (0, 1)$. At the start of the game, P offers a long-term contract to A that maps verifiable information into payments and termination decisions. If A rejects, the parties earn 0. Otherwise, the following stage game is repeatedly played:

1. A chooses effort $e_t \in \{0, 1\}$ at cost ce_t.

2. P privately observes output $y_t \in \{L, H\}$. If $e_t = 1$, $y_t = H$ with probability p. If $e_t = 0$, $y_t = H$ with probability $q < p$.[10]

3. P sends a public message m_t. A public randomization device x_t is realized after P's message is sent.

4. The formal contract determines transfers: wage w_t, bonus b_t, and burnt money B_t.

5. The parties decide whether to continue the relationship or not. Outside options are 0.

Payoffs are $\pi_t = y_t - w_t - b_t - B_t$ for P and $u_t = w_t + b_t - ce_t$ for A, with discounted continuation payoffs $\sum \delta^t (1 - \delta) \pi_t$ and $\sum \delta^t (1 - \delta) u_t$. For the moment, assume that parties are locked into the contract and cannot choose to terminate the relationship.

What is Verifiable? The wage w_t can depend only on past realizations of the public randomization device $x_{t'}$. The bonus b_t and burnt money B_t can depend both on past $x_{t'}$ and on past messages $m_{t'}$ (including the message from the current period).

[10] In the full model, P also observes the outcome of a random variable ϕ_t.

Note that if parties are locked into the contract, then this is not really a "relational contract." Everything observable is verifiable. We will be altering that assumption in a bit.

Intuition - Formal Contract Set $\delta = 1$ for simplicity.

One-shot Game: Suppose the game is played once, and suppose moreover that y is verifiable. Then the parties can easily attain first-best: P pays b_H following $y = H$ and b_L following $y = L$, where

$$pb_H + (1-p)\, b_L - c \geq qb_H + (1-q)\, b_L$$

or

$$b_H - b_L \geq \frac{c}{p-q}.$$

The wage is set to satisfy the Agent's outside option.

Why doesn't this simple contract work if P privately observes y? The short answer is that P **would have an incentive to report** $m = L$ **regardless of** y. The Principal must pay $b_H > b_L$ if he reports $m = H$, which *ex post* he would rather not do.

In order for P to have incentives to tell the truth, it must be that

$$b_L + B_L = b_H + B_H$$

which immediately implies that

$$B_L - B_H \geq \frac{c}{p-q}.$$

Following low effort, P must burn some money in order to "convince" A that he is telling the truth.

Two-shot Game: Now, suppose the game is played **twice**. What contracts motivate the Agent to work hard?

One easy option is to simply repeat the one-shot contract twice. In this case, $c/(p-q)$ surplus is burnt **whenever** $y_t = L$. So this contract isn't very efficient.

An alternative is to make the contract **history-dependent**. For example, suppose the contract allowed the Agent to avoid punishment if he produces L in the first period but H in the second period. In that case, the money burnt is 0 following (H, H), $c/(p-q)$ following (H, L), 0 following (L, H), and $\frac{c}{p-q} + \frac{c}{(p-q)(1-p)}$ following (L, L). One can show that this alternative scheme also induces high effort (check it as an exercise!). It also leads to the **same expected efficiency loss**.

Both of these contracts assume that P reports A's output after **each period**. However, P could instead **keep silent** until the very end of the game. In that case, the Agent does not know whether he produced high output or not in period 1, and so his second-period IC constraint is satisfied so long as

$$E\left[b\right| e_1 = 1, y_2 = H] - E\left[b\right| e_1 = 1, y_2 = L] \geq \frac{c}{p-q}.$$

This is clearly easier to satisfy than the outcome-by-outcome IC constraints if the Principal reveals information. **So the principal will not reveal information until the very end of the game in the optimal contract**.

Let $b_{y_1 y_2}$ be the bonus following (y_1, y_2). One can show that $b_{HH} \geq b_{HL} = b_{LH} \geq b_{LL}$ in the optimal contract. Moreover, suppose $b_{HH} > b_{HL}$. Consider increasing b_{HL} by $\varepsilon > 0$ and decreasing b_{LL} by $\frac{2p}{1-p}\varepsilon$. This change is rigged to ensure that the same amount of money is burnt under the new scheme. A receives the same payoff if he works hard.

If A shirks in a single period, his payoff is now

$$pqb_{HH} + \left[(1-p)\,q + (1-q)\,p\right](b_{HL} + \varepsilon) + (1-q)\,(1-p)\left(b_{LL} - \frac{2p}{1-p}\varepsilon\right) - c.$$

The coefficient on ε in this expression is

$$(1-p)\,q - p\,(1-q) < 0.$$

Therefore, **holding on-path surplus fixed,** A's **surplus following a deviation is strictly lower under this alternative contract**. So deviations are easier to deter.

Hence, the optimal contract has $b_{HH} = b_{HL} > b_{LL}$. P **does not com-**

municate until the end of the game, and A is only punished if he produces low output in both periods.

What if the Principal Cannot Commit? The intuition outlined above extends to any number of periods. P does not communicate, and A is punished only if he produces low output in every period.

However, this is not a terribly realistic solution. If the game is played repeatedly, then the "optimal contract" would entail an infinitely large punishment infinitely far in the future, accompanied by an infinitely large amount of burnt surplus. Instead, we might think that the amount of surplus that can be burnt equals the **future value of the relationship**.

In other words, the Principal **cannot commit** to burn money, so the worst that could happen is that the Agent leaves the relationship. This is a relational contract that "burns money" by termination: because termination is inefficient (it hurts both the Principal and the Agent), it can be used to induce the Principal to truthfully report output.

The main result of the paper argues that any optimal relational contract is equivalent to a relational contract in which:

1. A is paid a constant wage w that is strictly above his outside option, until he is fired.

2. A exerts $e = 1$ until fired.

3. P sends no messages to A until A is fired.

The upshot? Efficiency wages are an optimal contract. Note that neither this result nor the paper pin down **when** firing occurs, although it must occur with positive probability on the equilibrium path.

7.2.3 There is Private Information about Payoff-Relevant Variables: Halac (2012)

Consider Levin's (2003) relational contract. In this relational contract:

1. **Total** surplus is independent of how **rent is split**. Therefore, Levin has nothing to say about bargaining between players.

2. Surplus depends on **outside options**. Recall the dynamic enforcement constraint:

$$c(y^*) \leq \frac{\delta}{1-\delta}(S^* - \bar{u} - \bar{\pi}).$$

   The **larger** the outside options, the **lower** the output.

In Levin's world, the parties would love to **decrease** their outside options, which would increase total surplus on the equilibrium path. In particular, if the Principal could **pretend to have a worse outside option, then he would**.

In the real world, the Principal might be wary of small outside options, because he is afraid he will be **held up by the other player**. Suppose relationship rents are split according to Nash bargaining. The Principal has bargaining weight λ and so earns

$$(1-\lambda)\,\bar{\pi}+\lambda\,(S^{*}-\bar{u})\,.$$

If $\lambda = 1$, then the Principal would like to misreport his type to be **smaller** in order to increase S^*. If $\lambda = 0$, then the Principal would like to pretend his type is **larger** in order to capture more rent.

Halac (2012) formalizes this loose intuition by consider a model in which the Principal has persistent private information about his outside option.

Setup A Principal P and Agent A interact repeatedly with common discount factor $\delta \in (0,1)$. At the beginning of the game, P learns her type $\theta \in \{l, h\}$, which determines her outside option r_θ with $r_h > r_l$. This type is private and constant over time. $\Pr[\theta = l] = \mu_0$.

In each period of the game:

1. With probability λ, P makes an offer of a wage w_t and promised bonus $b_t\,(y_t)$. Otherwise, A makes the offer.

2. The party that did not make the offer accepts or rejects.

3. If accept, A chooses effort $e_t \in [0,1)$.

4. Output $y_t \in \left\{\underline{y}, \bar{y}\right\}$ is realized, with $\Pr\left[\,y_t = \bar{y}\,|\,e_t\right] = e_t$.

5. Payments: the fixed wage w_t is enforced by a court. The bonus b_t is

discretionary.[14]

Denote P and A's payoffs by π_t and u_t, respectively. If the offer is accepted, $u_t = w_t + b_t - c(e_t)$ and $\pi_t = y_t - w_t - b_t$. If the offer is rejected, $u_t = r_A$ and $\pi_t = r_\theta$.

To highlight the intuition outlined at the start of this section, the paper makes several restrictions to equilibrium. The paper looks for a Perfect Public Bayesian Equilibrium that is on the Pareto frontier. Moreover:

1. Once a posterior assigns probability 1 to a type, it forever assigns probability 1 to that type.

2. If a party reneges on a **payment**, then the relationship either breaks down or remains on the Pareto frontier.

3. If a party deviates **in any other way**, then the relationship remains on the Pareto frontier.

Note that "Pareto frontier" is a little strange here, since parties have different beliefs about payoffs. What is assumed is that the equilibrium is Pareto efficient **given the (commonly known) beliefs of the Agent**.

The paper also makes assumptions about the **symmetric-information** relational contract. Regardless of r_θ, the discount factor δ is such that:

1. If both players know that $\theta = l$, then first-best is not attainable in a relational contract.

[14] Parties simultaneously choose nonnegative payments to make to one another.

2. If both players know that $\theta = h$, then some positive effort is attainable in a relational contract.

Sketch of Results We begin with the first proposition.

Proposition 1: Suppose that the two types of P choose different actions in period t. Then it must be that either (i) one type rejects an offer by A, or (ii) one type reneges on a bonus.

Why? Types could separate in one of four ways: either (i) or (ii) above; or (iii) they accept different contracts from A; or (iv) they offer different contracts to A. **By assumption**, once types separate play is on the symmetric-information Pareto frontier. In particular, players never separate in the future on the equilibrium path.

Suppose (iii). Then there is no rejection in the current period. So P's on-path payoff doesn't depend on type. So on-path payoffs must be equal. But then $\theta = h$ can imitate $\theta = l$ and then take his outside option or renege to earn a strictly larger profit.

Suppose (iv). If the l-type is supposed to reject, then the h-type also wants to reject, because the l-type has a better symmetric info contract, and the h-type has a better outside option. If the h-type is supposed to reject, then the h-type can deviate and offer his symmetric-info contract immediately. This contract is an equilibrium for **any** agent beliefs. The deviation generates a strictly higher payoff, because it doesn't involve any breakdown. Moreover, whenever the Agent offers a payoff below r_h in future periods, the principal can simply reject.

If P has bargaining power, $\lambda = 1$: Separation occurs only if P defaults on a payment (because A never makes a contract offer). P is the residual claimant, so the h-type wants to imitate the l-type.

How do the parties separate? If P is supposed to pay a large bonus and is threatened with breakdown, then h-type is less willing to pay. So the **cost of separation is the probability of breakdown**. The **benefit of breakdown is that the l-type can credibly induce higher effort than the h-type**.

As a result, separation is optimal if **l-types are sufficiently likely:**

Proposition 2: There exists $\hat{\mu}_0$ such that if $\mu_0 > \hat{\mu}_0$, the optimal contract entails separation. Otherwise, the optimal contract pools on the h-type symmetric-info contract.

Under further restrictions on equilibrium, the paper characterizes the **speed of separation**. Because the l-type must compensate A for the possibility of default, l-type's payoff is larger when separation is **slower**.

If A has bargaining power, $\lambda = 0$: Separation occurs only if P rejects a contract (because P can never credibly promise a positive payoff). P earns his outside option, so the l-type wants to imitate the h-type.

After separation, each type earns his outside option (on-path). However, l-type can imitate h-type and earn r_h in all future periods. So it must be that the h-type rejects the contract while the l-type accepts. Let v_l be **today's**

payoff from accepting the contract.

	Today's Payoff	Tomorrow's Payoff
h playing h	$(1-\delta)\, r_h$	δr_h
h playing l	$(1-\delta)\, v_l$	δr_h
l playing l	$(1-\delta)\, v_l$	δr_l
l playing h	$(1-\delta)\, r_l$	δr_h

Players are only willing to follow their specified roles if there exists a v_l that satisfies:

$$\begin{aligned}(1-\delta)\, v_l + \delta r_l &\geq (1-\delta)\, r_l + \delta r_h \\ r_h &\geq (1-\delta)\, v_l + \delta r_h\end{aligned}$$

So $r_h \geq v_l$ and hence $\delta \leq 1/2$. So **separation is only feasible if players are impatient**. If P is patient, then all types are willing to mimic h-type to get a better continuation payoff. The paper shows that separation occurs immediately if it occurs at all. Separation is only optimal if r_l is sufficiently likely.

Bibliography

[1] Abreu, Dilip. 1988. On the Theory of Infinitely Repeated Games with Discounting. *Econometrica*, 56(2): 383-396.

[2] Acemoglu, Daron. 2008. *Modern Economic Growth*. Princeton University Press.

[3] Acemoglu, Daron, Pol Antras, and Elhanan Helpman. 2007. Contracts and Technology Adoption. *American Economic Review*, 97(3): 916-943.

[4] Aghion, Philippe and Patrick Bolton. 1992. An Incomplete Contracts Approach to Financial Contracting. *Review of Economic Studies*, 59(3): 473-494.

[5] Aghion, Philippe, Ernst Fehr, Richard Holden, and Tom Wilkening. 2015. The Role of Bounded Rationality and Imperfect Information in Subgame Perfect Implementation: An Empirical Investigation. Mimeo.

[6] Aghion, Philippe, Drew Fudenberg, Richard Holden, Takashi Kunimoto, and Oliver Tercieux. 2012. Subgame-Perfect Implementation Under Information Perturbations. *Quarterly Journal of Economics*, 127(4): 1843-1881.

[7] Aghion, Philippe and Jean Tirole. 1997. Formal and Real Authority in Organizations. *Journal of Political Economy*, 105(1): 1-29.

[8] Alchian, Armen and Harold Demsetz. 1972. Production, Information Costs, and Economic Organization. *American Economic Review*, 62(5): 777-795.

[9] Alonso, Ricardo and Niko Matouschek. 2008. Optimal Delegation. *Review of Economic Studies*, 75(1): 259-293.

[10] Andrews, Isaiah, and Daniel Barron. Forthcoming. The Allocation of Future Business: Dynamic Relational Contracts with Multiple Agents. *American Economic Review*.

[11] Antras, Pol. 2003. Firms, Contracts, and Trade Structure. *Quarterly Journal of Economics*, 118(4): 1375-1418.

[12] Auriol, Emmanuelle, Guido Friebel, and Frauke von Bieberstein. 2016. The Firm as a Locus of Social Comparisons: Standard Promotion Practices Versus Up-or-Out. *Journal of Economic Behavior and Organization*, 121(1): 41-49

[13] Backus, Matthew. 2014. Why is Productivity Correlated with Competition? Mimeo.

[14] Bajari, Patrick and Steven Tadelis. Incentives versus Transaction Costs: A Theory of Procurement Contracts. *RAND Journal of Economics*, 32(3): 387-407.

[15] Baker, George. 1992. Incentive Contracts and Performance Measurement. *Journal of Political Economy*, 100(3): 598-614.

[16] Baker, George. 2002. Distortion and Risk in Optimal Incentive Contracts. *Journal of Human Resources*, 37(4): 728-751.

[17] Baker, George, Michael Gibbs, and Bengt Holmstrom. 1994a. The Internal Economics of the Firm: Evidence from Personnel Data. *Quarterly Journal of Economics*, 109(4): 881-919.

[18] Baker, George, Michael Gibbs, and Bengt Holmstrom. 1994b. The Wage Policy of a Firm. *Quarterly Journal of Economics*, 109(4): 921-955.

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

[20] Baker, George, Robert Gibbons, and Kevin Murphy. Informal Authority in Organizations. *Journal of Law, Economics, and Organization*, 15(1): 56-73.

[21] Baker, George, Robert Gibbons, and Kevin Murphy. Relational Contracts and the Theory of the Firm. *Quarterly Journal of Economics*, 117(1): 39-84.

[22] Barron, Daniel, Robert Gibbons, Ricard Gil, and Kevin Murphy. 2015. Relational Contracts under Reel Authority. Mimeo.

[23] Barron, Daniel and Michael Powell. 2016. Policies in Relational Contracts. Mimeo.

[24] Battigalli, Pierpaolo and Giovanni Maggi. 2002. Rigidity, Discretion, and the Costs of Writing Contracts. *American Economic Review*, 92(4): 798-817.

[25] Bernhardt, Dan. 1995. Strategic Promotion and Compensation. *Review of Economic Studies*, 62(2): 315-339.

[26] Bernheim, Douglas and Michael Whinston. Common Agency. *Econometrica*, 54(4): 923-942.

[27] Bernheim, Douglas and Michael Whinston. Menu Auctions, Resource Allocation, and Economic Influence. *Quarterly Journal of Economics*, 101(1): 1-32.

[28] Besley, Timothy and Maitreesh Ghatak. 2001. Government Versus Private Ownership of Public Goods. *Quarterly Journal of Economics*, 116(4): 1343-1372.

[29] Board, Simon. 2011. Relational Contracts and the Value of Loyalty. *American Economic Review*, 101(7): 3349-3367.

[30] Bonatti, Alessandro and Johannes Horner. Forthcoming. Career Concerns with Exponential Learning. *Theoretical Economics*.

[31] Bond, Timothy. 2015. Internal Labor Markets in Equilibrium. Mimeo.

[32] Bose, Gautam and Kevin Lang. 2015. Monitoring for Worker Quality: Task Assignment, Job Ladders, Wages and Mobility in Nonmanagerial Internal Labor Markets. Mimeo.

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

[34] Camara, Odilon and Dan Bernhardt. 2009. The Dynamics of Promotions, Quits, and Layoffs. Mimeo.

[35] Carroll, Gabriel. 2015. Robustness and Linear Contracts. *American Economic Review*, 105(2): 536-563.

[36] Chaigneau, Pierre, Alex Edmans, and Daniel Gottlieb. 2015. The Informativeness Principle Under Limited Liability. Mimeo.

[37] Chan, William. 1996. External Recruitment versus Internal Promotion. *Journal of Labor Economics*, 14(4): 555-570.

[38] Cisternas, Gonzalo. 2016. Two-Sided Learning and Moral Hazard. Mimeo.

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

[40] Coase, Ronald. 1960. The Problem of Social Cost. *Journal of Law and Economics*, 3: 1-44.

[41] Crawford, Vincent and Joel Sobel. 1982. *Econometrica*, 50(6): 1431-1451.

[42] Cyert, Richard and James March. 1963. *A Behavioral Theory of the Firm*. Englewood Cliffs.

[43] Demougin, Dominique and Aloysius Siow. 1994. Careers in On-going Hierarchies. *American Economic Review*, 84(5): 1261-1277.

[44] Dessein, Wouter. 2002. Authority and Communication in Organizations. *Review of Economic Studies*, 69(4): 811-838.

[45] DeVaro, Jed and Hodaka Morita. 2013. Internal Promotion and External Recruitment: A Theoretical and Empirical Analysis. *Journal of Labor Economics*, 31(2): 227-269.

[46] Dewatripont, Mathias, Ian Jewitt, and Jean Tirole. 1999. The Economics of Career Concerns, Part I: Comparing Information Structures. *Review of Economic Studies*, 66(1): 183-198.

[47] Dewatripont, Mathias, Ian Jewitt, and Jean Tirole. 1999. The Economics of Career Concerns, Part II: Applications to Missions and Accountability of Government Agencies. *Review of Economic Studies*, 66(1): 199-217.

[48] Dewatripont, Mathias and Jean Tirole. 1994. A Theory of Debt and Equity: Diversity of Securities and Manager-Shareholder Congruence. *Quarterly Journal of Economics*, 109(4): 1027-1054.

[49] Diamond, Peter. 1998. Managerial Incentives: on the Near Linearity of Optimal Compensation. *Journal of Political Economy*, 106(5): 931-957.

[50] Dixit, Avinash, Gene Grossman, and Elhanan Helpman. 1997. Common Agency and Coordination: General Theory and Application to Government Policy Making. *Journal of Political Economy*, 105(4): 752-769.

[51] Doeringer, Peter and Michael Piore. 1971. *Internal Labor Markets and Manpower Analysis*. Heath Lexington Books.

[52] Dye, Ronald. Costly Contract Contingencies. *International Economic Review*, 26: 233-250.

[53] Fehr, Ernst, Michael Powell, and Tom Wilkening. 2016. Handing out Guns at a Knife Fight: Behavioral Limitations of Subgame-Perfect Implementation. Mimeo.

[54] Feltham, Gerald and Jim Xie. 1994. Performance Measure Congruity and Diversity in Multi-Task Principal/Agent Relations. *Accounting Review*, 69(3): 429-453.

[55] Foster, Lucia, John Haltiwanger, and Chad Syverson. 2008. Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability? *American Economic Review*, 98(1): 394-428.

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

[57] Fudenberg, Drew, Bengt Holmstrom, and Paul Milgrom. 1990. Short-Term Contracts and Long-Term Agency Relationships. *Journal of Economic Theory*, 51: 1-31.

[58] Fudenberg, Drew and Jean Tirole. 1990. Moral Hazard and Renegotiation in Agency Contracts. *Econometrica*, 58(6): 1279-1319.

[59] Frandsen, Brigham, Michael Powell, and James Rebizter. 2016. Persistently Inefficient? Organizational Fragmentation and Common Agency Problems in the US Health Care System. Mimeo.

[60] Frankel, Alexander. 2014. Aligned Delegation. *American Economic Review*, 104(1): 66-83.

[61] Frankel, Alexander. Forthcoming. Delegating Multiple Decisions. *American Economic Journal: Microeconomics.*

[62] Ghosh, Suman and Michael Waldman. 2010. Standard Promotion Practices versus Up-or-Out Contracts. *RAND Journal of Economics*, 41(2): 301-325.

[63] Gibbons, Robert. 2005. Four Formal(izable) Theories of the Firm? *Journal of Economic Behavior and Organization*, 58(2): 2005.

[64] Gibbons, Robert, Richard Holden, and Michael Powell. 2012. Organization and Information: Firms' Governance Choices in Rational-Expectations Equilibrium. *Quarterly Journal of Economics*, 127(4): 1813-1841.

[65] Gibbons, Robert and Kevin Murphy. 1992. Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence. *Journal of Political Economy*, 100(3): 468-505.

[66] Gibbons, Robert and Michael Waldman. 1999. A Theory of Wage and Promotion Dynamics inside Firms. *Quarterly Journal of Economics*, 114(4): 1321-1358.

[67] Grossman, Sanford and Oliver Hart. 1986. The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration. *Journal of Political Economy*, 94(4): 691-719.

[68] Halac, Marina. 2012. Relational Contracts and the Value of Relationships. *American Economic Review*, 102(2): 750-779.

[69] Halonen, Maija. 2002. Reputation and the Allocation of Ownership. *Economic Journal*, 112(481): 539-558.

[70] Hart, Oliver. 1983. The Market Mechanism as an Incentive Scheme. *Bell Journal of Economics*, 14(2): 366-382.

[71] Hart, Oliver. 1995. *Firms, Contracts, and Financial Structure*. Clarendon Press, Oxford.

[72] Hart, Oliver and John Moore. 1990. Property Rights and the Nature of the Firm. *Journal of Political Economy*, 98(6): 1119-1158.

[73] Hart, Oliver and John Moore. 1998. Default and Renegotiation: A Dynamic Model of Debt. *Quarterly Journal of Economics*, 113(1): 1-41.

[74] Hart, Oliver, Andrei Shleifer, and Robert Vishny. 1997. The Proper Scope of Government: Theory and an Application to Prisons. *Quarterly Journal of Economics*, 112(4): 1127-1161.

[75] He, Zhiguo, Bin Wei, Jianfeng Yu, and Feng Gao. 2014. Optimal Long-Term Contracting with Learning. Mimeo.

[76] Hermalin, Benjamin. 1992. The Effects of Competition on Executive Behavior. *RAND Journal of Economics*, 23(3): 350-365.

[77] Hicks, John. 1935. Annual Survey of Economic Theory: The Theory of Monopoly. *Econometrica*, 3(1): 1-20.

[78] Hsieh, Chang-Tai and Peter Klenow. 2009. Misallocation and Manufacturing TFP in China and India. *Quarterly Journal of Economics*, 124(4): 1403-1448.

[79] Holmstrom, Bengt. 1979. Moral Hazard and Observability. *Bell Journal of Economics*, 10(1): 74-91.

[80] Holmstrom, Bengt. 1984. On the Theory of Delegation. In *Bayesian Models in Economic Theory*, eds. M. Boyer and R. Kihlstrom, New York: North Holland, 115-141.

[81] Holmstrom, Bengt. 1999. Managerial Incentive Problems: A Dynamic Perspective. *Review of Economic Studies*, 66(1): 169-182.

[82] Holmstrom, Bengt. 1999. The Firm as a Subeconomy. *Journal of Law, Economics, and Organization*, 15(1): 74-102.

[83] Holmstrom, Bengt and Paul Milgrom. 1987. Aggregation and Linearity in the Provision of Intertemporal Incentives. *Econometrica*, 55(2): 303-328.

[84] Holmstrom, Bengt and Paul Milgrom. 1991. Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics, and Organization*, 7: 24-52.

[85] Holmstrom, Bengt and Paul Milgrom. 1994. *American Economic Review*, 84(4): 972-991.

[86] Holmstrom, Bengt and John Roberts. 1998. *Journal of Economic Perspectives*, 12(4): 73-94.

[87] Jewitt, Ian, Ohad Kadan and Jeroen Swinkels. 2008. Moral Hazard with Bounded Payments. *Journal of Economic Theory*, 143: 59-82.

[88] Kandori, Michihiro. 2008. Repeated Games. In *New Palgrave Dictionary of Economics*, eds. Steven Durlauf and Lawrence Blume.

[89] Ke, Rongzhu, Jin Li, and Michael Powell. 2016. Managing Careers in Organizations. Mimeo.

[90] Kerr, Steven. 1975. On the Folly of Rewarding A, While Hoping for B. *Academy of Management Journal*, 18(4): 769-783.

[91] Klein, Benjamin, Robert Crawford, and Armen Alchian. 1978. Vertical Integration, Appropriable Rents, and the Competitive Contracting Process. *Journal of Law and Economics*, 21(2): 297-326.

[92] Krakel, Matthias and Anja Schottner. 2012. Internal Labor Markets and Worker Rents. *Journal of Economic Behavior and Organization*, 84(2): 491-509.

[93] Lazear, Edward and Sherwin Rosen. 1981. Rank-Order Tournaments as Optimum Labor Contracts. *Journal of Political Economy*, 89(5): 841-864.

[94] Legros, Patrick and Andrew Newman. 2013. A Price Theory of Vertical and Lateral Integration. *Quarterly Journal of Economics*, 128(2): 725-770.

[95] Leibenstein, Harvey. 1966. Allocative Efficiency vs. "X-Efficiency." *American Economic Review*, 56(3): 392-415.

[96] Levin, Jonathan. 2003. Relational Incentive Contracts. *American Economic Review*, 93(3): 835-857.

[97] Li, Jin, Niko Matouschek, and Michael Powell. Forthcoming. Power Dynamics in Organizations. *American Economic Journal: Microeconomics*.

[98] Lipnowski, Elliot and Joao Ramos. 2015. Repeated Delegation. Mimeo.

[99] Macalay, Stewart. 1963. Non-Contractual Relations in Business: A Preliminary Study. *American Sociological Review*, 28(1): 55-67.

[100] MacLeod, Bentley and James Malcomson. 1988. Reputation and Hierarchy in Dynamic Models of Employment. *Journal of Political Economy*, 96(4): 832-854.

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

[102] MacLeod, Bentley and James Malcomson. 1998. Motivation and Markets. *American Economic Review*, 88(3): 388-411.

[103] Makowski, Louis and Joseph Ostroy. 1995. Appropriation and Efficiency: A Revision of the First Theorem of Welfare Economics. *American Economic Review*, 85(4): 808-827.

[104] Malcomson, James. 1984. Work Incentives, Hierarchy, and Internal Labor Markets. *Journal of Political Economy*, 92(3): 486-507.

[105] Malcomson, James. 2013. Relational Incentive Contracts. In *Handbook of Organizational Economics*, eds. Robert Gibbons and John Roberts, Princeton University Press.

[106] Marschak, Jacob and Roy Radner. 1972. *Economic Theory of Teams*. Yale University Press.

[107] Martimort, David and Lars Stole. 2012. Representing Equilibrium Aggregates in Aggregate Games with Applications to Common Agency. *Games and Economic Behavior*, 76(2): 753-772.

[108] Maskin, Eric. 2002. On Indescribable Contingencies and Incomplete Contracts. *European Economic Review*, 46: 725-733.

[109] Maskin, Eric and John Moore. 1999. Implementation and Renegotiation. *Review of Economic Studies*, 66(1): 39-56.

[110] Maskin, Eric and Jean Tirole. 1999. Unforeseen Contingencies and Incomplete Contracts. *Review of Economic Studies*, 66(1): 83-114.

[111] Melumad, Nahum and Toshiyuki Shibano. 1991. Communication in Settings with No Transfers. *RAND Journal of Economics*, 22(2): 173-198.

[112] Meyer, Margaret, Paul Milgrom, and John Roberts. 1992. Organizational Prospects, Influence Costs, and Ownership Changes. *Journal of Economics and Management Strategy*, 1(1): 9-35.

[113] Milgrom, Paul and John Roberts. 1988. An Economic Approach to Influence Activities in Organizations. *American Journal of Sociology*, 94: S154-S179.

[114] Mirrlees, James. 1999. The Theory of Moral Hazard and Unobservable Behaviour: Part I. *Review of Economic Studies*, 66(1): 3-21.

[115] Moore, John and Rafael Repullo. 1988. Subgame Perfect Implementation. *Econometrica*, 56(5): 1191-1220.

[116] Moroni, Sofia and Jeroen Swinkels. 2014. Existence and Non-Existence in the Moral Hazard Problem. *Journal of Economic Theory*, 150: 668-682.

[117] Nalebuff, Barry and Joseph Stiglitz. 1983. Information, Competition, and Markets. *American Economic Review*, 73(2): 278-283.

[118] Ostroy, Joseph. 1980. The No-Surplus Condition as a Characterization of Perfectly Competitive Equilibrium. *Journal of Economic Theory*, 22: 183-207.

[119] Powell, Michael. 2016. Productivity and Credibility in Industry Equilibrium. Mimeo.

[120] Powell, Michael. 2015. An Influence-Cost Model of Organizational Practices and Firm Boundaries. *Journal of Law, Economics, and Organization*, 31(suppl 1): i104-i142.

[121] Prat, Andrea and Aldo Rustichini. 2003. Games Played through Agents. *Econometrica*, 71(4): 989-1026.

[122] Prendergast, Canice. 1993. The Role of Promotion in Inducing Specific Human Capital Acquisition. *Quarterly Journal of Economics*, 108(2): 523-534.

[123] Prendergast, Canice. 2002. The Tenuous Trade-off between Risk and Incentives. *Journal of Political Economy*, 110(5): 1071-1102.

[124] Rantakari, Heikki. 2008. On the Role of Uncertainty in the Risk-Incentives Tradeoff. *B.E. Journal of Theoretical Economics*, 8(1).

[125] Rantakari, Heikki. 2008. Governing Adaptation. *Review of Economic Studies*, 75(4): 1257-1285.

[126] Raith, Michael. 2003. Competition, Risk, and Managerial Incentives. *American Economic Review*, 93(4): 1425-1436.

[127] Robertson, D.H.. 1930. *Control of Industry.* Nisbet and Company, London.

[128] Rosen, Sherwin. 1986. Prizes and Incentives in Elimination Tournaments. *American Economic Review*, 76(4): 701-715.

[129] Segal, Ilya. 1999. Contracting with Externalities. *Quarterly Journal of Economics*, 114(2): 337-388.

[130] Scharfstein, David. 1988. Product-Market Competition and Managerial Slack. *RAND Journal of Economics*, 19(1): 147-155.

[131] Schmidt, Klaus. 1997. Managerial Incentives and Product Market Competition. *Review of Economic Studies*, 64(2): 191-213.

[132] Simon, Herbert. 1991. Organizations and Markets. *Journal of Economic Perspectives*, 5(2): 25-44.

[133] Tadelis, Steven and Oliver Williamson. 2013. Transaction Cost Economics. In *Handbook of Organizational Economics*, eds. Robert Gibbons and John Roberts, Princeton University Press.

[134] Vives, Xavier. 2008. Innovation and Competitive Pressure. *Journal of Industrial Economics*, 56(3): 419-469.

[135] Waldman, Michael. 1984. Job Assignments, Signalling, and Efficiency. *RAND Journal of Economics*, 15(2): 255-267.

[136] Waldman, Michael. 2003. Ex Ante versus Ex Post Optimal Promotion Rules: The Case of Internal Promotion. *Economic Inquiry*, 41(1): 27-41.

[137] Weyl, Glen and Michal Fabinger. 2013. Pass-Through as an Economic Tool: Principles of Incidence under Imperfect Competition. *Journal of Political Economy*, 121(3): 528-583.

[138] Williamson, Oliver. 1971. The Vertical Integration of Production: Market Failure Considerations. *American Economic Review*, 63(2): 112-123.

[139] Williamson, Oliver. 1975. *Markets and Hierarchies: Analysis and Antitrust Implications*. Free Press, New York, NY.

[140] Williamson, Oliver. 1979. Transaction-Cost Economics: The Governance of Contractual Relations. *Journal of Law and Economics*, 22(2): 233-261.

[141] Williamson, Oliver. 1985. *The Economic Institutions of Capitalism*. Free Press, New York, NY.

[142] Williamson, Oliver. 2000. The New Institutional Economics: Taking Stock, Looking Ahead. *Journal of Economic Literature*, 38: 595-613.

[143] Wu, Yanhui. Forthcoming. Incentive Contracts and the Allocation of Talent. *Economic Journal*.

[144] Zabojnik, Jan and Dan Bernhardt. 2001. Corporate Tournaments, Human Capital Acquisition, and the Firm Size–Wage Relation. *Review of Economic Studies*, 68(3): 693-716.