Lectures in Organizational Economics

Daniel Barron   Michael Powell

June, 2018

Contents

Disclaimer

These lecture notes were written for a second-year Ph.D. sequence 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, 2009) Let $\mathcal{F}$ be a countable set of firms, each with a convex production-possibilities set $\mathcal{Y}^f \subset \mathbb{R}^N$. 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 $\mathcal{Y}^f(p) \subset \mathcal{Y}^f$. Then there exists a representative firm with production possibilities set $\mathcal{Y} \subset \mathbb{R}^N$ and set of profit-maximizing net supplies $\mathcal{Y}(p)$ such that for any $p \in \mathbb{R}^N_+$, $\hat{y} \in \mathcal{Y}(p)$ if and only if $\hat{y}(p) = \sum_{f \in \mathcal{F}} \hat{y}^f$ for some $\hat{y}^f \in \mathcal{Y}^f(p)$ 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 $\mathcal{Y}^f$. 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:

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

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 $\pi_i^* + \pi_j^* < \max_{y \in \mathcal{Y}^i + \mathcal{Y}^j} \pi(y)$. 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** (Holmström 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 $e^* < e^{FB}$.

1.1.1 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.

The Model There is a risk-neutral Principal (P) and a risk-averse Agent (A). The Agent chooses an **effort level** $e \in \mathcal{E} \subset \mathbb{R}_+$ and incurs a cost of $c(e)$, where $c : \mathbb{R}_+ \to \mathbb{R}_+$ is strictly increasing and strictly convex. If $\mathcal{E}$ is an interval, we will say that **effort is continuous**, and if $\mathcal{E}$ consists of a finite number of points, we will say that **effort is discrete**. We will assume $0 \in \mathcal{E}$, and $c(0) = 0$. The effort level affects the distribution over **output** $y \in \mathcal{Y}$, with y distributed according to CDF $F(\cdot|\, e)$. This output can be sold on the product market at price p, and the revenues py accrue to the Principal.

The Principal does not have any direct control over the Agent, but what she can do is write a contract that influences what the Agent will do. In particular, she can write a contract $w \in \mathcal{W} \subset \{w : \mathcal{Y} \times \mathcal{E} \to \mathbb{R}\}$, where $\mathcal{W}$ is the **contracting space**. The contract determines a transfer $w(y, e)$ that she is compelled to pay the Agent if output y is realized, and he chose effort e. If $\mathcal{W}$ does not allow for functions that depend directly on effort, we will say that **effort is noncontractible**, and abusing notation slightly, we will write the contractual payment the Principal is compelled to pay the Agent if output y is realized as $w(y,e) = w(y)$ for all $e \in \mathcal{E}$. We will be assuming throughout that effort is noncontractible, but I wanted to highlight that it is a real restriction on the contracting space, and it is one that we will impose as a primitive of the model.

The Agent can decline to work for the Principal and reject her contract, pursuing his outside option instead. This outside option provides utility $\bar{u}$ to the Agent and $\bar{\pi}$ to the Principal. If the Agent accepts the contract, the Principal's and Agent's preferences are, respectively,

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

where u is increasing and weakly concave.

We have described the players, what they can do, and what their preferences are. We still need to describe the timing of the game that the players play, as well as the solution concept. Explicitly describing the timing of the model is essential to remove any ambiguity about what players know when they make their decisions. In this model, the timing of the game is:

1. P offers A a contract $w \in \mathcal{W}$. w is commonly observed.

2. A accepts the contract $(d = 1)$ or rejects it $(d = 0)$, in which case he

receives $\bar{u}$, and the game ends. d is commonly observed.

3. If A accepts the contract, A chooses effort level e and incurs cost $c(e)$. e is privately 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 $w(y)$. The 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, much of the literature assumes that the Agent's effort level is privately observed by the Agent and therefore refers to this model as the "hidden action" model. Ultimately, though, the underlying source of the moral-hazard problem is that contracts cannot be conditioned on relevant variables, not that the relevant variables are unobserved by the Principal. Many papers assume effort is unobservable to justify it being noncontractible. While this is a compelling justification, in our framework, the contracting space itself is a primitive of the model. Later in the course, we will talk a bit about the microfoundations for different assumptions on the contracting space.

Finally, let us describe the solution concept. A **pure-strategy subgame-perfect equilibrium** is a contract $w^* \in \mathcal{W}$, an **acceptance decision**

$d^* : \mathcal{W} \to \{0,1\}$, and an **effort choice** $e^* : \mathcal{W} \times \{0,1\} \to \mathcal{E}$ 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^*.

First-Best Benchmark If we want to talk about the inefficiencies that arise in equilibrium in this model, it will be useful first to establish a benchmark against which to compare outcomes. In this model, a **feasible outcome** is a distribution over payments from the Principal to the Agent as well as an effort level $e \in \mathcal{E}$. We will say that a feasible outcome is **Pareto optimal** if there is no other feasible outcome that both players weakly prefer and one player strictly prefers. If an effort level e is part of a Pareto optimal outcome, we will say that it is a **first-best** effort level, and we will denote it by e^{FB}.

Lemma 1. The first-best effort level satisfies

$$e^{FB} \in \underset{e \in \mathcal{E}}{\operatorname{argmax}}\, E_y \left[py \middle| e \right] - c\left(e\right).$$

Proof of Lemma 1. In any Pareto-optimal outcome, payments to the agent are deterministic. Since the Agent is risk averse, given an outcome involving stochastic payments to the Agent, there is another outcome in which the Agent chooses the same effort level and receives the certainty equivalent wage instead. This outcome yields the same utility for the Agent, and since the Agent is risk averse, the certainty equivalent payment is smaller in expectation, so the Principal is strictly better off. Next, given constant deterministic wages, any Pareto-optimal outcome must solve

$$\max_{w \in \mathbb{R}, e \in \mathcal{E}} E_y \left[py \right| e] - w$$

subject to

$$u\left(w - c\left(e\right)\right) \geq \bar{u},$$

for some $\bar{u}$. In any solution to this problem, the constraint must bind, since u is increasing. Moreover, since u is increasing, it is invertible, so we can write

$$w = u^{-1}\left(\bar{u}\right) + c\left(e\right),$$

and therefore the first-best effort level must solve the problem specified in the Lemma.■

This Lemma shows that the first-best effort level maximizes expected revenues net of effort costs. If effort is fully contractible, so that the Principal could offer any contract w that depended nontrivially on e, then the first-best effort would be implemented in equilibrium. In particular, the Principal could offer a contract that pays the Agent $u^{-1}\left(\bar{u}\right) + c\left(e^{FB}\right)$ if he choose e^{FB}, and pays him a large negative amount if he chooses any $e \neq e^{FB}$. That the first-best effort level can be implemented in equilibrium if effort is contractible is an illustration of a version of the *Coase Theorem*: if the contracting space is sufficiently rich, equilibrium outcomes will be Pareto optimal.

If effort is noncontractible, and $e^{FB} > 0$, then equilibrium will not involve Pareto optimal outcomes. For an outcome to be Pareto optimal, it has to involve a deterministic wage payment to the Agent. But if the Agent's wage is independent of output, then it must also be independent of his effort level. He will therefore receive no benefit from choosing a costly effort level, and so he will choose $e = 0 < e^{FB}$. The question to which we will now turn is: what effort will be implemented in equilibrium when effort is noncontractible?

Equilibrium Effort Since the Agent's effort choice affects the Principal's payoffs, the Principal would ideally like to directly choose the Agent's effort. But, she has only indirect control: she can offer different contracts, and different contracts may get the Agent to optimally choose different effort levels. We can think of the Principal's problem as choosing an effort level e as well as a contract for which e is *incentive compatible* for the Agent to choose and for which it is *individually rational* for the Agent to accept. As a loose analogy, we can connect the Principal's problem to the social planner's problem from general equilibrium theory. We can think of e as analogous to an allocation the Principal would like to induce, and the choice of a contract as analogous to setting "prices" so as to decentralize e as an equilibrium allocation.

Formally, the Principal offers a contract $w \in \mathcal{W}$ and "proposes" an effort level e in order to solve

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

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 **incentive-compatibility constraint**:

$$e\in\operatorname*{argmax}_{\hat{e}\in\mathcal{E}}\int_{y\in Y}u\left(w\left(y\right)-c\left(\hat{e}\right)\right)dF\left(y|\,\hat{e}\right).$$

The second constraint ensures 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 **individual-rationality constraint** or **participation constraint**:

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

At this level of generality, the model is not very tractable. We will need to impose more structure on it in order to highlight some its key trade-offs and properties.

CARA-Normal Case with Affine Contracts In order to highlight one of the key trade-offs that arise in this class of models, we will make a number of strong simplifying assumptions.

Assumption A1 (CARA). 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 A2 (Normal Output). Effort shifts the mean of a normally distributed random variable. That is, $y\sim N\left(e,\sigma^{2}\right)$.

Assumption A3 (Affine Contracts). $\mathcal{W}=\left\{w:\mathcal{Y}\rightarrow\mathbb{R},w\left(y\right)=s+by\right\}$. That is, the contract space permits only affine contracts.

Assumption A4 (Continuous Effort). Effort is continuous and satisfies $\mathcal{E}=\mathbb{R}_{+}$.

In principle, we should not impose exogenous restrictions on the *functional form* of $w\left(y\right)$. There is an important class of applications, however, that restrict attention to affine contracts, $w\left(y\right)=s+by$, and a lot of the basic intuition that people have for the comparative statics of optimal contracts come from imposing this restriction.

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, 1974; 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 not 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 analogue

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 Holmström and Milgrom (1987), Diamond (1998) and, more recently, Carroll (2013) and Barron, Georgiadis, and Swinkels (2017)) This perspective will come back later in the course when we discuss the Property Rights Theory of firm boundaries.

Given Assumptions $(A1)-(A3)$, 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 $E[\exp\{tX\}]=\exp\left\{\mu t+\frac{1}{2}\sigma^2t^2\right\}$), so his preferences can be written as

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

We can take a monotonic transformation of his utility function ($f(x)=$

$-\frac{1}{r}\log(-x)$) 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

$$e \in \underset{\hat{e}}{\operatorname{argmax}}\, b\hat{e} - \frac{c}{2}\hat{e}^2$$

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. 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,

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

In other words, the Principal has to ensure that the Agent's total expected monetary compensation, $s + be\left(b\right)$, 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

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

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

$$0 = p\underbrace{e'\left(b^{*}\right)}_{1/c} - c\underbrace{e^{*}\left(b^{*}\right)}_{b^{*}/c}\underbrace{e'\left(b^{*}\right)}_{1/c} - rb^{*}\sigma^{2},$$

and therefore the optimal incentive slope satisfies

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

Moreover, 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.

Now, how does the effort that is induced in this optimal affine contract compare to the **first-best effort**? Using the result from Lemma 1, we know that first-best effort in this setting solves

$$\max_{e \in \mathbb{R}_+} pe - \frac{c}{2}e^2,$$

and therefore $e^{FB} = p/c$.

Even if effort is noncontractible, the Principal could in principle implement exactly this same level of effort by writing a contract only on output. To do so, she would choose $b = p$, since this would get the Agent to choose $e(p) = p/c$. 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.

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

Often, when an economic model can be solved in closed form, we jump right to the solution. Only when a model cannot be solved in closed form do we typically stop to think carefully about what economic properties its solution must possess. I want to spend a couple minutes *partially* characterizing this model's solution, even though we already completely characterized it above, just to highlight how this kind of reasoning can be helpful in developing intuition that might generalize beyond the present setting. In particular, many fundamental features of models can be seen as a comparison of first-order losses or gains against second-order gains or losses, so it is worth going through this first-order–second-order logic. 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

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

This first term is zero, because $b = p$ in fact maximizes $pe(b) - \frac{c}{2}e(b)^2$, since it induces the first-best level of effort. This is just an application of the envelope theorem. The second term in this expression is strictly negative. This implies that, relative to the contract that induces first-best effort, a reduction in the slope of the incentive contract yields a first-order gain to the Principal 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, and these risk costs are higher if the Agent is more risk averse and if output is noisier.

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 next week.

The First-Order Approach

Last time, we imposed a lot of structure on the Principal-Agent problem and solved for optimal affine contracts. One of the problems we identified with that approach was that there was not a particularly compelling reason for restricting attention to affine contracts. Moreover, in that particular setting, if we allowed the contracts to take more general functional forms, there in fact was no optimal contract.

Today, we will return to a slightly modified version of the more general setup of the problem and consider an alternative approach to characterizing optimal contracts without imposing any assumptions on the functional forms they might take. One change we will be making is that the Agent's preferences are now given by

$$U(w,e) = \int_{y \in \mathcal{Y}} [u(w(y)) - c(e)]\, dF(y|\, e) = E_y[u(w)|\, e] - c(e),$$

where u is strictly increasing and strictly concave, and the utility the Agent receives from money is additively separable from his effort costs.

Recall from last time that the Principal's problem is to choose an output-contingent contract $w \in \mathcal{W} \subset \{w : \mathcal{Y} \to \mathbb{R}\}$ and to "propose" an effort level e to solve:

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

subject to an incentive-compatibility constraint

$$e \in \operatorname*{argmax}_{\hat{e} \in \mathcal{E}} \int_{y \in \mathcal{Y}} u(w(y))\, dF(y|\, \hat{e}) - c(\hat{e})$$

and an individual-rationality constraint

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

One of the problems with solving this problem at this level of generality is that the incentive-compatibility constraint is quite a complicated set of conditions. The contract has to ensure that, of all the effort levels the Agent could potentially choose, he prefers to choose e. In other words, the contract has to deter the Agent from choosing any other effort level $\hat{e}$: for all $\hat{e} \in \mathcal{E}$, we must have

$$\int_{y\in\mathcal{Y}} \left[u\left(w\left(y\right)\right)-c\left(e\right)\right]dF\left(y\right|e\right) \geq \int_{y\in\mathcal{Y}} \left[u\left(w\left(y\right)\right)-c\left(\hat{e}\right)\right]dF\left(y\right|\hat{e}\right).$$

When effort is continuous, the incentive-compatibility constraint is actually a continuum of constraints of this form. It seems like it should be the case that if we impose more structure on the problem, we can safely ignore most of these constraints. This turns out to be true. If we impose some relatively stringent but somewhat sensible assumptions on the problem, then if it is the case that the Agent does not want to deviate *locally* to another $\hat{e}$, then he also does not want to deviate to an $\hat{e}$ that is farther away. When local constraints are sufficient, we will in fact be able to replace the Agent's incentive-compatibility constraint with the first-order condition to his problem.

Throughout, we will be focusing on models that satisfy the following assumptions.

Assumption A1 (Continuous Effort and Continuous Output). Effort is continuous and satisfies $\mathcal{E}=\mathbb{R}_+$. Output is continuous, with $\mathcal{Y}=\mathbb{R}$, and for each $e\in\mathcal{E}$, $F\left(\cdot\right|e\right)$ has support $\left[\underline{y},\bar{y}\right]$ and has density $f\left(\cdot\right|e\right)$, where $f\left(\cdot\right|e\right)$ is differentiable in e.

Assumption A2 (First-Order Stochastic Dominance—FOSD). The output distribution function satisfies $F_e\left(y\right|e\right)\leq 0$ for all $e\in\mathcal{E}$ and all y with strict inequality for some y for each e.

Assumption $(A2)$ roughly says that higher effort levels make lower output realizations less likely and higher output realizations more likely. This assumption provides sufficient conditions under which higher effort increases total expected surplus, ignoring effort costs.

We will first explore the implications of being able to replace the incentive-compatibility constraint with the Agent's first-order condition, and then we will provide some sufficient conditions under which doing so is without loss of generality. Under Assumption $(A1)$, if we replace the Agent's incentive-compatibility constraint with his first-order condition, the Principal's problem becomes:

$$\max_{w \in \mathcal{W}, e \in \mathcal{E}} \int_{\underline{y}}^{\bar{y}} (py - w(y)) f(y|e) \, dy$$

subject to the local incentive-compatibility constraint

$$c'(e) = \int_{\underline{y}}^{\bar{y}} u(w(y)) f_e(y|e) \, dy$$

and the individual-rationality constraint

$$\int_{\underline{y}}^{\bar{y}} u(w(y)) f(y|e) \, dy - c(e) \geq \bar{u}.$$

This problem is referred to as the **first-order approach** to characterizing second-best incentive contracts. It is now just a constrained-optimization problem with an equality constraint and an inequality constraint. We can therefore write the Lagrangian for this problem as

$$\begin{aligned}\mathcal{L} &= \int_{\underline{y}}^{\bar{y}} (py - w(y)) f(y|e)\, dy + \lambda \left(\int_{\underline{y}}^{\bar{y}} u(w(y)) f(y|e)\, dy - c(e) - \bar{u} \right) \\ &+\mu \left(\int_{\underline{y}}^{\bar{y}} u(w(y)) f_e(y|e)\, dy - c'(e) \right),\end{aligned}$$

where λ is the Lagrange multiplier on the individual-rationality constraint, and μ is the Lagrange multiplier on the local incentive-compatibility constraint. We can derive the conditions for the optimal contract $w^*(y)$ inducing optimal effort e^* by taking first-order conditions, point-by-point, with respect to $w(y)$. These conditions are:

$$\frac{1}{u'(w^*(y))} = \lambda + \mu \frac{f_e(y|e^*)}{f(y|e^*)}.$$

Contracts satisfying these conditions are referred to as Holmström-Mirrlees contracts (or (λ, μ) contracts as one of my colleagues calls them). There are several points to notice here. First, the left-hand side is increasing in $w(y)$, since u is concave. Second, if $\mu = 0$, then this condition would correspond to the conditions for an optimal risk-sharing rule between the Principal and the Agent. Under a Pareto-optimal risk allocation, the **Borch Rule** states that the ratio of the Principal's marginal utility to the Agent's marginal utility is equalized across states. In this case, the Principal's marginal utility is one. Any optimal-risk sharing rule will equalize the Agent's marginal utility of income across states and therefore give the Agent a constant wage.

Third, Holmström (1979) shows that under Assumption $(A2)$, $\mu > 0$, so that the right-hand side of this equation is increasing in $f_e(y|\, e^*)/f(y|\, e^*)$. You might remember from econometrics that this ratio is called the **score**—it tells us how an increase in e changes the log likelihood of e given output realization y. To prevent the Agent from choosing effort level e instead of e^*, the contract has to pay the Agent more for outputs that are more likely under e^* than under e. Since by assumption, we are looking at only local incentive constraints, the contract will pay the Agent more for outputs that are more likely under e^* than under effort levels arbitrarily close to e^*.

Together, these observations imply that the optimal contract $w^*(y)$ is increasing in the score. Just because an optimal contract is increasing in the score does not mean that it is increasing in output. The following assumption guarantees that the score is increasing in y, and therefore optimal contracts are increasing in output.

Assumption A3 (Monotone Likelihood Ratio Property—MLRP). Given any two effort levels $e, e' \in \mathcal{E}$ with $e > e'$, the ratio $f(y|\, e)/f(y|\, e')$ is increasing in y.

MLRP guarantees, roughly speaking, that higher levels of output are more indicative of higher effort levels.[1] Under Assumption $(A1)$, MLRP is equivalent to the condition that $f_e(y|\, e)/f(y|\, e)$ is increasing in y. We can

[1] The property can also be interpreted in terms of statistical hypothesis testing. Suppose the null hypothesis is that the Agent chose effort level e', and the alternative hypothesis is that the Agent chose effort level $e > e'$. If, given output realization y, a likelihood ratio test would reject the null hypothesis of lower effort, the same test would also reject the null hypothesis for any higher output realization.

therefore interpret the optimality condition as telling us that the optimal contract is increasing in output precisely when higher output levels are more indicative of higher effort levels. Put differently, the optimal contract "wants" to reward *informative* output, not necessarily *high* output.

The two statistical properties, FOSD and MLRP, that we have assumed come up a lot in different settings, and it is easy to lose track of what they each imply. To recap, the FOSD property tells us that higher effort makes higher output more likely, and it guarantees that there is always a benefit of higher effort levels, gross of effort costs. The MLRP property tells us that higher output is more indicative of higher effort, and it guarantees that optimal contracts are increasing in output. These two properties are related: MLRP implies FOSD, but not the reverse.

Informativeness Principle

Before we provide conditions under which the first-order approach is valid, we will go over what I view as the most important result to come out of this model. Suppose there is another contractible performance measure $m \in \mathcal{M}$, where y and m have joint density function $f(y, m|\, e)$, and the contracting space is $\mathcal{W} = \{w : \mathcal{Y} \times \mathcal{M} \to \mathbb{R}\}$. Under what conditions will an optimal contract $w(y, m)$ depend nontrivially on m? The answer is: whenever m provides additional information about e. To make this argument precise, we will introduce the following definition.

Definition 1. Given two random variables Y and M, Y is **sufficient for**

(Y, M) **with respect to** $e \in \mathcal{E}$ if and only if the joint density function $f(y, m|\, e)$ is multiplicatively separable in m and e:

$$f(y, m|\, e) = g(m|\, e)\, h(y, m).$$

We will say that M is **informative about** $e \in \mathcal{E}$ if Y is not sufficient for (Y, M) with respect to $e \in \mathcal{E}$.

We argued above that optimal contracts pay the Agent more for outputs that are more indicative of high effort. This same argument also extends to other performance measures, as long as they are informative about effort. This result is known as the *informativeness principle* and was first established by Holmström (1979) and Shavell (1979).

Theorem 1 (Informativeness Principle). Assume the first-order approach is valid. Let $w(y)$ be the optimal contract when m is noncontractible. If m is contractible, there exist a contract $w(y, m)$ that Pareto dominates $w(y)$ if and only if m is informative about $e \in \mathcal{E}$.

Proof. In both cases, the optimal contract gives the Agent $\bar{u}$, so we just need to show that the Principal can be made strictly better off if m is contractible.

If the first-order approach is valid, the optimality conditions for the Principal's problem when both y and m are contractible are given by

$$\frac{1}{u'(w^*(y, m))} = \lambda + \mu \frac{f_e(y, m|\, e^*)}{f(y, m|\, e^*)}.$$

The optimal contract $w^*(y, m)$ is independent of m if and only if y is sufficient for (y, m) with respect to e^*.

This result seems like it should be obvious: optimal contracts clearly should make use of all available information. But it is not ex ante obvious this would be the case. In particular, one could easily have imagined that optimal contracts should only depend on performance measures that are "sufficiently" informative about effort—after all, basing a contract on another performance measure could introduce additional noise as well. Or one could have imagined that optimal contracts should only depend on performance measures that are directly affected by the Agent's effort choice. The informativeness principle says that optimal contracts should depend on every performance measure that is even slightly informative.

This result has both positive and negative implications. On the positive and practical side, it says that optimal contracts should make use of benchmarks: a fund manager should be evaluated for her performance relative to a market index, CEOs should be rewarded for firm performance relative to other firms in their industry, and employees should be evaluated relative to their peers. On the negative side, the result shows that optimal contracts are highly sensitive to the fine details of the environment. This implication is, in a real sense, a weakness of the theory: it is the reason why the theory often predicts contracts that bear little resemblance to what we actually see in practice.

The informativeness principle was derived under the assumption that the first-order approach was valid. When the first-order approach is not valid, the informativeness principle does not necessarily hold. The reason for this is that when the first-order approach does not hold, there may be multiple binding incentive-compatibility constraints at the optimum, and just because an informative performance measure helps relax one of those constraints, if it does not help relax the other binding constraints, it need not strictly increase the firm's profits. Chaigneau, Edmans, and Gottlieb (2014) generalizes the informativeness principle to settings in which the first-order approach is not valid.

Validity of the First-Order Approach

Finally, we will briefly talk about some sufficient conditions ensuring the first-order approach is valid. Assumption $(A4)$, along with the following assumption, are sufficient.

Assumption A4 (Convexity of the Distribution Function Condition—CDFC). $F(\cdot|\, e)$ is twice differentiable, and $F_{ee}(\cdot|\, e) \geq 0$ for all e.

CDFC is a strong assumption. There is a fairly standard class of distributions that are often used in contract theory that satisfy it, but it is not satisfied by other well-known families of distributions. Let $F_H(y)$ and $F_L(y)$ be two distribution functions that have density functions $f_H(y)$ and $f_L(y)$ for which $f_H(y)/f_L(y)$ is increasing in y, and suppose

$$F(y|\, e) = eF_H(y) + (1-e) F_L(y).$$

Then $F(y|e)$ satisfies both MLRP and CDFC. In other words, MLRP and CDFC are satisfied if output is drawn from a mixture of a "high" and a "low" distribution, and higher effort increases the probability that output is drawn from the high distribution.

Theorem 2. Suppose $(A1)-(A4)$ are satisfied. If the local incentive-compatibility constraint is satisfied, the incentive-compatibility constraint is satisfied.

Proof sketch. The high-level idea of the proof is to show that MLRP and CDFC imply that the Agent's effort-choice problem is globally concave for any contract the Principal offers him. Using integration by parts, we can rewrite the Agent's expected utility as follows.

$$\begin{aligned}
\int_{\underline{y}}^{\bar{y}} u(w(y)) f(y|e)\,dy - c(e) &= u(w(y))\,F(y|e)\Big|_{\underline{y}}^{\bar{y}} \\
&\quad - \int_{\underline{y}}^{\bar{y}} u'(w(y))\frac{dw(y)}{dy}F(y|e)\,dy - c(e) \\
&= u(w(\bar{y})) - \int_{\underline{y}}^{\bar{y}} u'(w(y))\frac{dw(y)}{dy}F(y|e)\,dy - c(e).
\end{aligned}$$

Now, suppose $w(y)$ is increasing and differentiable. Differentiating the expression above with respect to e twice yields

$$-\int_{\underline{y}}^{\bar{y}} u'(w(y))\frac{dw(y)}{dy}F_{ee}(y|e)\,dy - c''(e) < 0$$

for every $e \in \mathcal{E}$, since $F_{ee} > 0$. Thus, the Agent's second-order condition is globally satisfied, so if the local incentive constraint is satisfied, the incentive constraint is satisfied.■

I labeled this proof as a sketch, because while it follows Mirrlees's (1976) argument, the full proof (due to Rogerson (1985)) requires showing that $w(y)$ is in fact increasing and differentiable when MLRP is satisfied. We cannot use our argument above for why MLRP implies increasing contracts, because that argument presumed the first-order approach was valid, which is exactly what we are trying to prove here. The MLRP and CDFC conditions are known as the Mirrlees-Rogerson conditions.

There are other sufficient conditions for the first-order approach to be valid that do not require such strong distributional assumptions (see, for example, Jewitt (1988)). And there are other approaches to solving the moral hazard problem that do not rely on the first-order approach. These include Grossman and Hart (1983), which decomposes the Principal's problem into two steps: the first step solves for the cost-minimizing contract that implements a given effort level, and the second step solves for the optimal effort level. We will take this approach when we think about optimal contracts under limited liability in the next section.

Further Reading Many papers restrict attention to linear contracts, even in environments in which the optimal contract (if it exists) is not linear. Holmström 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 a risk-neutral 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. Barron, Georgiadis, and Swinkels (2017) build upon this argument to show that, in general, preventing such gaming opportunities imposes a constraint that contracts induce (weakly) concave payoffs for the agent as a function of output. 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).

1.1.2 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

$$\begin{aligned}W &= \{w(y)=(1-y)\,w_0+yw_1, w_0, w_1\geq 0\}\\ &= \{w(y)=s+by, s\geq\underline{w}, b\geq 0\}.\end{aligned}$$

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.

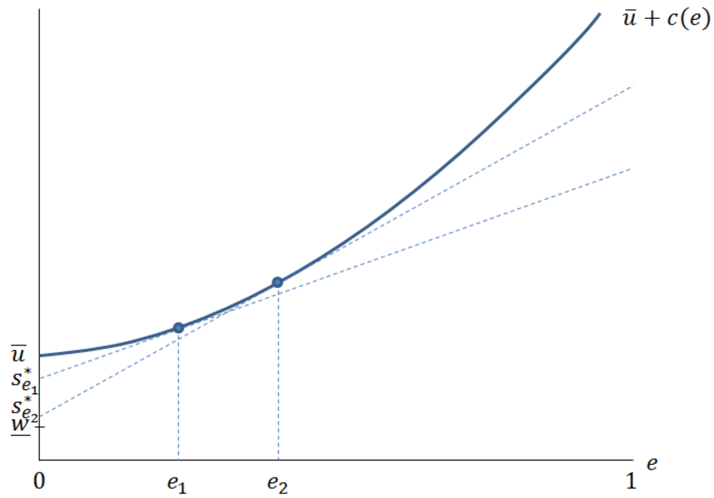


Figure 1

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** $IC(e, \bar{u}, \underline{w})$.

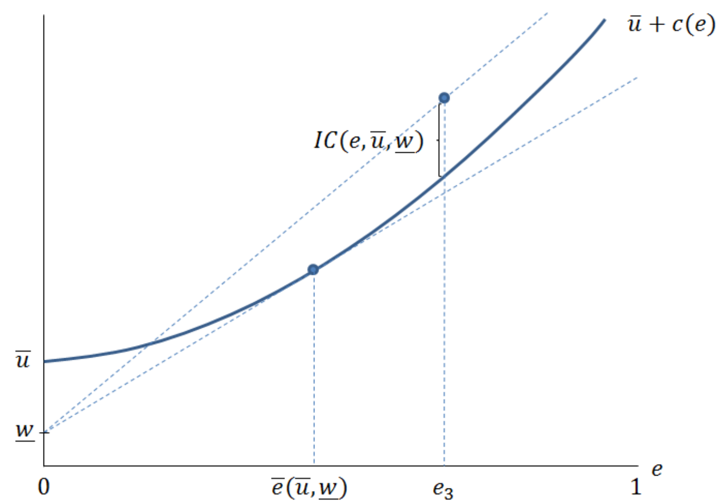


Figure 2

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 maximum 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 maximum 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 $C(e,\bar{u}-\underline{w})$.

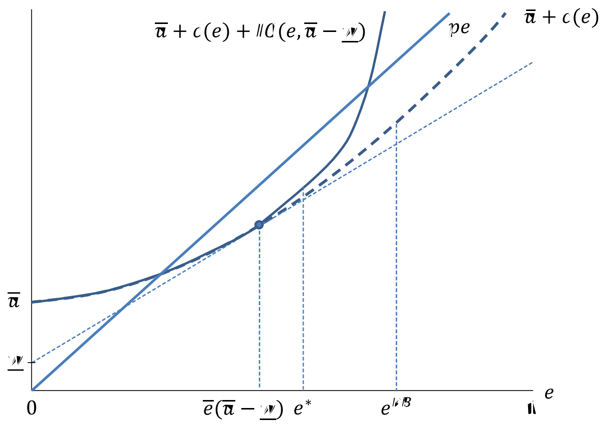


Figure 3

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. This exercise goes through a version of Diamond's (1998) and Barron, Georgiadis, and Swinkels's (2018) argument for why linear contracts are optimal when the Agent is able to "take on risk." Suppose the Principal and the Agent are both risk neutral, and let $\mathcal{Y} = [0, \bar{y}]$ and $\mathcal{E} = \mathbb{R}_+$. There is a limited-liability constraint, and the contracting space is $\mathcal{W} = \{w : \mathcal{Y} \to \mathbb{R}_+\}$. After the Agent chooses an effort level e, he can then choose any distribution function $F(y)$ over output that satisfies $e = \int_0^{\bar{y}} y dF(y)$. In other words, his effort level determines his *average* output, but he can then add mean-preserving noise to his output. Given a contract w, effort e, and distribution F, the Agent's expected utility is

$$\int_0^{\bar{y}} w(y) \, dF(y) - c(e),$$

where c is strictly increasing and strictly convex. The Principal's expected profits are $\int_0^{\bar{y}} (y - w(y)) \, dF(y)$. The Agent's outside option gives both parties a payoff of zero.

(a) Show that a linear contract of the form $w(y) = by$ maximizes the Principal's expected profits. To do so, you will want to argue that given any contract $w(y)$ that implements effort level e, there is a linear contract that also implements effort level e but at a weakly lower cost to the Principal. [Hint: instead of thinking about all the possible distribution functions the Agent can choose among, it may be useful to just look at distributions that put weight

on two levels of output, $0 \leq y_L < y_H \leq \bar{y}$ satisfying $e = (1 - q) y_L + q y_H$.]
(b) Are there other contracts that maximize the Principal's expected profits? If so, how are they related to the optimal linear contract? If not, provide an intuition for why linear contracts are uniquely optimal.

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.

1.1.3 Multiple Tasks and Misaligned Performance Measures

In the previous two models, the Principal cared about 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 previous discussion, overall firm profits are 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."

As an example, think of a retail firm that hires an employee both to make sales and to provide customer service. It can be difficult to measure the quality of customer service that a particular employee provides, but it is easy to measure that employee's sales. Writing a contract that provides the employee with high-powered incentives directly on sales will get him to put a lot of effort into sales and very little effort into customer service. And in fact, he might only be able to put a lot of effort into sales by intentionally neglecting customer service. If the firm cares equally about both dimensions, it might be optimal not to offer high-powered incentives to begin with. This is what Holmström and Milgrom (1991) refers to as the "multitask problem." We will look at a model that captures some of this intuition, although not as directly as Holmström and Milgrom's model.

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

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

$$\begin{aligned} \Pr[y = 1|\, e] &= f_1 e_1 + f_2 e_2 \\ \Pr[m = 1|\, e] &= g_1 e_1 + g_2 e_2, \end{aligned}$$

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. Output is noncontractible, but the performance measure is contractible. The Principal can write a contract $w \in \mathcal{W} \subset \{w : \mathcal{M} \to \mathbb{R}\}$ 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,

$$\begin{aligned} \Pi(w, e) &= f_1 e_1 + f_2 e_2 - s - b(g_1 e_1 + g_2 e_2) \\ U(w, e) &= s + b(g_1 e_1 + g_2 e_2) - \frac{c}{2}\left(e_1^2 + e_2^2\right). \end{aligned}$$

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

1. P offers A a contract w, 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. e is only observed by A.

4. Performance measure m and output y are drawn from the distributions described above. 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 $w^* \in \mathcal{W}$, an acceptance decision $d^* : \mathcal{W} \to \{0,1\}$, and an effort choice $e^* : \mathcal{W} \times \{0,1\} \to \mathbb{R}^2_+$ 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 and proposes an effort level e to solve

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

subject to the incentive-compatibility constraint

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

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

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

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

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

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 off 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

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

Taking first-order conditions,

$$pf_1 \underbrace{\frac{\partial e_1^*}{\partial b}}_{g_1/c} + pf_2 \underbrace{\frac{\partial e_2^*}{\partial b}}_{g_2/c} = \underbrace{ce_1^*(b^*)}_{b^* g_1/c} \underbrace{\frac{\partial e_1^*}{\partial b}}_{g_1/c} + \underbrace{ce_2^*(b^*)}_{b^* g_2/c} \underbrace{\frac{\partial e_2^*}{\partial b}}_{g_2/c},$$

or

$$b^* = p \frac{f_1 g_1 + f_2 g_2}{g_1^2 + g_2^2} = p \frac{f \cdot g}{g \cdot g} = p \frac{||f||}{||g||} \cos \theta = p \cos \theta,$$

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."**

It can be useful to view this problem geometrically. 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:

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

which has the same solution as

$$\max_e - \left(e_1 - \frac{p}{c} f_1\right)^2 - \left(e_2 - \frac{p}{c} f_2\right)^2.$$

That is, the Principal would like to choose an effort vector that is collinear with the vector f:

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

This effort vector would coincide with the first-best effort vector, since it maximizes total surplus, and the players have quasilinear preferences.

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" an effort vector that is collinear with the vector g:

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

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 $\frac{p}{c} \cdot (f_1, f_2)$?

The following figure illustrates the first-best effort vector e^{FB} and the equilibrium effort vector e^*. The concentric rings around e^{FB} are the Principal's iso-profit curves. The rings that are closer to e^{FB} represent higher profit levels. The optimal contract induces effort vector e^{*}, which also coincides with the orthogonal projection of e^{FB} onto the ray $k \cdot (g_1, g_2)$.

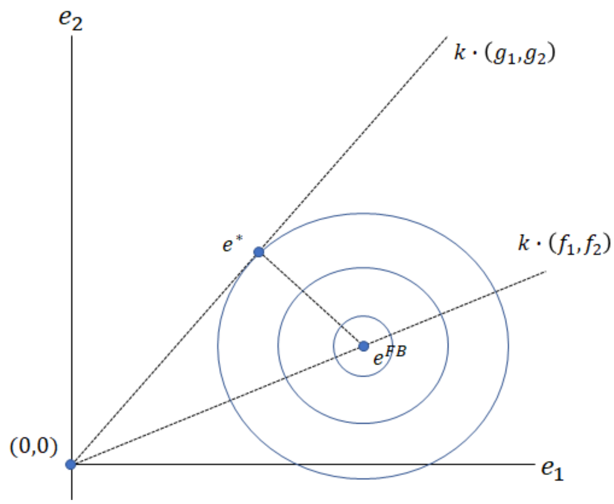


Figure 4: Optimal Effort Vector

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 simply cannot be contracted upon.

One observation that is immediate is that it may sometimes be optimal to offer incentive contracts that provide no incentives for the Agent to choose positive effort levels (i.e., $b^* = 0$). This was essentially never the case in the model in which the Agent chose only a one-dimensional effort level, yet we often see that many employees are on contracts that look like they offer no performance-based payments. As this model highlights, this may be optimal precisely when the set of available performance measures are quite bad. As an example, suppose

$$\Pr\left[y = 1 | e\right] = \alpha + f_1 e_1 + f_2 e_2,$$

where $\alpha > 0$ and $f_2 < 0$, so that higher choices of e_2 reduce the probability of high output. And suppose the performance measure is again satisfies

$$\Pr\left[m = 1 | e\right] = g_1 e_1 + g_2 e_2,$$

with $g_1, g_2 > 0$.

We can think of $y = 1$ as representing whether a particular customer buys something that he does not later return, which depends on how well he was treated when he went to the store. We can think of $m = 1$ as representing whether the Agent made a sale but not whether the item was later returned. In order to increase the probability of making a sale, the Agent can exert "earnest" sales effort e_1 and "shady" sales effort e_2. Both are good for sales, but the latter increases the probability the item is returned. If the vectors f and g are sufficiently poorly aligned (i.e., if it is really easy to make sales by being shady), it may be better for the firm to offer a contract with $b^* = 0$, as the following figure illustrates.

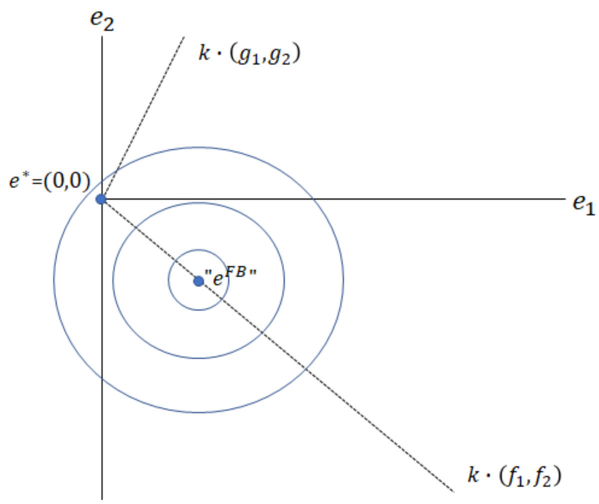


Figure 5: Sometimes Zero Effort is Optimal

This example illustrates that paying the Agent for sales can be a bad idea when what the Principal wants is *sales that are not returned*. The Kerr (1975) article is filled with many colorful examples of this problem. One such example concerns the incentives offered to the managers of orphanages. Their budgets and prestige were determined largely by the number of children they enrolled and not by whether they managed to place their children with suitable families. The claim made in the article is that the managers often denied adoption applications for inappropriate reasons: they were being rewarded for large orphanages, while the state hoped for good placements.

Limits on Activities

Firms have many instruments to help address the problems that arise in multitasking situations. We will describe two of them here in a small extension to the model. Suppose now that the Principal can put some restrictions on the types of actions the Agent is able to undertake. In particular, in addition to writing a contract on the performance measure m, she can write a contract on the dummy variables $1_{e_1>0}$ and $1_{e_2>0}$. In other words, while she cannot directly contract upon, say, e_2, she can write a contract that heavily penalizes any positive level of it. The first question we will ask here is: when does the Principal want to exclude the Agent from engaging in task 2?

We can answer this question using the graphical intuition we just developed above. The following figure illustrates this intuition. If the Principal does not exclude task 2, then she can induce the Agent to choose any effort vector of the form $k \cdot (g_1, g_2)$. If she does exclude task 2, then she can induce the Agent to choose any effort vector of the form $k \cdot (g_1, 0)$. In the former case, the equilibrium effort vector will be e^*, which corresponds to the orthogonal projection of e^{FB} onto the ray $k \cdot (g_1, g_2)$. In the latter case, the equilibrium effort will be e^{**}, which corresponds to the orthogonal projection of e^{FB} onto the ray $k \cdot (g_1, 0)$.

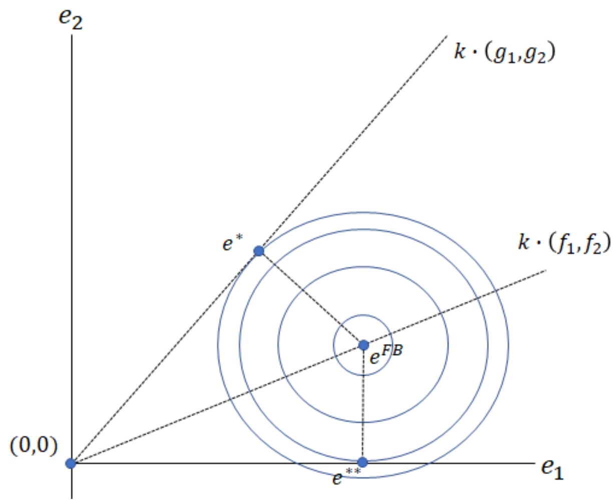


Figure 6: Excluding Task 2

This figure shows that for the particular vectors f and g it illustrates, it will be optimal for the Principal to exclude e_2: e^{**} lies on a higher iso-profit curve than e^* does. This will in fact be the case whenever the angle between vector f and g is larger than the angle between f and $(g_1, 0)$—if by excluding task 2, the performance measure m acts as if it is more closely aligned with f, then task 2 should be excluded.

Job Design

Finally, we will briefly touch upon what is referred to as job design. Suppose f and g are such that it is not optimal to exclude either task on its own. The firm may nevertheless want to hire *two* Agents who each specialize in a single task. For the first Agent, the Principal could exclude task 2, and for the second Agent, the Principal could exclude task 1. The Principal could then offer a contract that gets the first Agent to choose $\left(e_1^{FB}, 0\right)$ and the second agent to choose $\left(0, e_2^{FB}\right)$. The following figure illustrates this possibility.

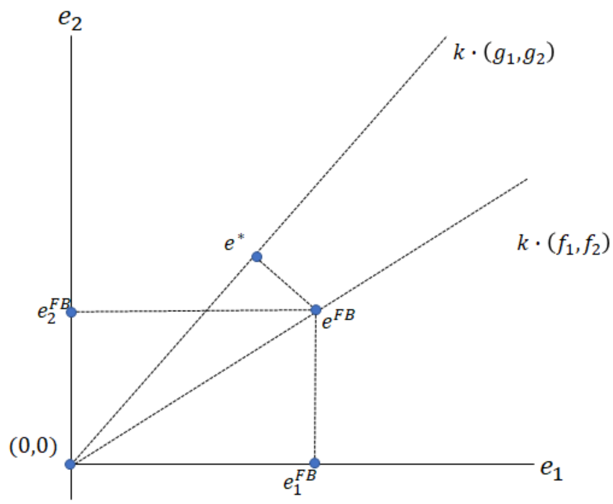


Figure 7: Job Design

When is it optimal for the firm to hire two Agents who each specialize in a single task? It depends on the Agents' opportunity cost. Total surplus under a single Agent under the optimal contract will be

$$pf \cdot e^* - \frac{c}{2} e^* \cdot e^* - \bar{u},$$

and total surplus with two specialized agents under optimal contracts will be

$$pf \cdot e^{FB} - \frac{c}{2}e^{FB} \cdot e^{FB} - 2\bar{u}.$$

Adding an additional Agent in this case is tantamount to adding an additional performance measure, which allows the Principal to choose induce any $e \in \mathbb{R}_+^2$, including the first-best effort vector. She gains from being able to do this, but to do so, she has to cover the additional Agent's opportunity cost $\bar{u}$.

Further Reading Holmström 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.

1.1.4 Indistinguishable Individual Contributions

So far, we have discussed three contracting frictions that give rise to equilibrium contracts that induce effort that is not first-best. We will now discuss a final contracting friction that arises when multiple individuals contribute to a single project, and while team output is contractible, individual contributions to the team output are not. This indistinguishability gives rise to Holmström's (1982) classic "moral hazard in teams" problem.

The Model There are $I \geq 2$ risk-neutral Agents $i \in \mathcal{I} = \{1, \ldots, I\}$ who each choose efforts $e_i \in \mathcal{E}_i = \mathbb{R}_+$ at cost $c_i(e_i)$, which is increasing, convex, differentiable, and satisfies $c_i'(0) = 0$. The vector of efforts $e = (e_1, \ldots, e_I)$ determine team output $y \in \mathcal{Y} = \mathbb{R}_+$ according to a function $y(e)$ which is increasing in each e_i, concave in e, differentiable, and satisfies $\lim_{e_i \to 0} \partial y / \partial e_i = \infty$. Note that output is not stochastic, although the model can be easily extended to allow for stochastic output. Output is contractible, and each Agent i is subject to a contract $w_i \in \mathcal{W} = \{w_i : \mathcal{Y} \to \mathbb{R}\}$. We will say that the vector of contracts $w = (w_1, \ldots, w_I)$ is a **sharing rule** if

$$\sum_{i \in \mathcal{I}} w_i(y) = y$$

for each output level y. Each Agent i's preferences are given by

$$U_i(w,e) = w_i(y(e)) - c_i(e_i).$$

Each Agent i takes the contracts as given and chooses an effort level. Output is realized and each agent receives payment $w_i(y)$. The solution concept is Nash equilibrium, and we will say that w **induces** e^* if e^* is a Nash equilibrium effort profile given the vector of contracts w.

Sharing Rules and the Impossibility of First-Best Effort Since the Agents have quasilinear preferences, any Pareto-optimal outcome under a sharing rule w will involve an effort level that maximizes total surplus, so that

$$e^{FB} \in \operatorname*{argmax}_{e \in \mathbb{R}_+^I} y(e) - \sum_{i \in \mathcal{I}} c_i(e_i).$$

Under our assumptions, there is a unique first-best effort vector, and it satisfies

$$\frac{\partial y\left(e^{FB}\right)}{\partial e_i} = c_i'\left(e_i^{FB}\right) \text{ for all } i \in \mathcal{I}.$$

First-best effort equates the social marginal benefit of each agent's effort level with its social marginal cost. We will denote the **first-best output level** $y\left(e^{FB}\right)$ by y^{FB}.

We will give an informal argument for why no sharing rule w induces e^{FB}, and then we will make that argument more precise. Suppose w is a sharing rule for which $w_i(y)$ is weakly concave and differentiable in y for all $i \in \mathcal{I}$.

For any Nash equilibrium effort vector e^*, it must be the case that

$$w_i'(y) \cdot \frac{\partial y(e^*)}{\partial e_i} = c_i'(e_i^*) \text{ for all } i \in \mathcal{I}.$$

In order for e^* to be equal to e^{FB}, it has to be the case that these equilibrium conditions coincide with the Pareto-optimality conditions. This is only possible if $w_i'(y) = 1$ for all i, but because w is a sharing rule, we must have that

$$\sum_{i \in \mathcal{I}} w_i'(y) = 1 \text{ for all } y.$$

Equilibrium effort e^* therefore cannot be first-best. This argument highlights the idea that getting each Agent to choose first-best effort requires that he be given the entire social marginal benefit of his effort, but it is not possible (at least under a sharing rule) for *all* the Agents simultaneously to receive the entire social marginal benefit of their efforts.

This argument is not a full argument for the impossibility of attaining first-best effort under sharing rules because it does not rule out the possibility of non-differentiable sharing rules inducing first-best effort. It turns out that there is no sharing rule, even a non-differentiable one, that induces first-best effort.

Theorem 3 (Moral Hazard in Teams). If w is a sharing rule, w does not induce e^{FB}.

Proof. This proof is due to Stole (2001). Take an arbitrary sharing rule w, and suppose e^* is an equilibrium effort profile under w. For any $i, j \in \mathcal{I}$, define $e_j(e_i)$ by the relation $y\left(e^*_{-j}, e_j(e_i)\right) = y\left(e^*_{-i}, e_i\right)$. Since y is continuous and increasing, a unique value of $e_j(e_i)$ exists for e_i sufficiently close to e^*_i. Take such an e_i. For e^* to be a Nash equilibrium, it must be the case that

$$w_j(y(e^*)) - c_j\left(e^*_j\right) \geq w_j\left(y\left(e^*_{-j}, e_j(e_i)\right)\right) - c_j(e_j(e_j)),$$

since this inequality has to hold for all $e_j \neq e^*_j$. Rewriting this inequality, and summing up over $j \in \mathcal{I}$, we have

$$\sum_{j\in\mathcal{I}} \left(w_j(y(e^*)) - w_j\left(y\left(e^*_{-i}, e_i\right)\right)\right) \geq \sum_{j\in\mathcal{I}} \left(c_j\left(e^*_j\right) - c_j(e_j(e_i))\right).$$

Since w is a sharing rule, the left-hand side of this expression is just $y(e^*) - y\left(e^*_{-i}, e_i\right)$, so this inequality can be written

$$y(e^*) - y\left(e^*_{-i}, e_i\right) \geq \sum_{j\in\mathcal{I}} c_j\left(e^*_j\right) - c_j(e_j(e_i)).$$

Since this must hold for all e_i close to e^*_i, we can divide by $e^*_i - e_i$ and take the limit as $e_i \to e^*_i$ to obtain

$$\frac{\partial y(e^*)}{\partial e_i} \geq \sum_{j\in\mathcal{J}} c'_j\left(e^*_j\right) \frac{\partial y(e^*)/\partial e_i}{\partial y(e^*)/\partial e_j}.$$

Now suppose that $e^* = e^{FB}$. Then $c'_j\left(e^*_j\right) = \partial y(e^*)/\partial e_j$, so this inequality becomes

$$\frac{\partial y(e^*)}{\partial e_i} \geq I\frac{\partial y(e^*)}{\partial e_i},$$

which is a contradiction because y is increasing in e_i.■

Joint Punishments and Budget Breakers Under a sharing rule, first-best effort cannot be implemented because in order to deter an Agent from choosing some $e_i < e_i^{FB}$, it is necessary to punish him. But because contracts can only be written on team output, the only way to deter each agent from choosing $e_i < e_i^{FB}$ is to simultaneously punish *all* the Agents when output is less than $y\left(e^{FB}\right)$. But punishing all the Agents simultaneously requires that they throw output away, which is impossible under a sharing rule. It turns out, though, that if we allow for contracts w that allow for **money burning**, in the sense that it allows for

$$\sum_{i \in \mathcal{I}} w_i(y) < y$$

for some output levels $y \in \mathcal{Y}$, first-best effort can in fact be implemented, and it can be implemented with a contract that does not actually burn money in equilibrium.

Proposition 1. There exist a vector of contracts w that induces e^{FB} for which $\sum_{i \in \mathcal{I}} w_i\left(y^{FB}\right) = y^{FB}$.

Proof. For all i, set $w_i(y) = 0$ for all $y \neq y^{FB}$, and let $w_i\left(y^{FB}\right) > c_i\left(e_i^{FB}\right)$ for all i so that $\sum_{i \in \mathcal{I}} w_i\left(y^{FB}\right) = y^{FB}$. Such a vector of contracts is feasible, because $y^{FB} > \sum_{i \in \mathcal{I}} c_i\left(e_i^{FB}\right)$. Finally, under w, e^{FB} is a Nash equilibrium effort profile because if all other Agents choose e_{-i}^{FB}, then if Agent i chooses

$e_i \neq e_i^{FB}$, he receives $-c_i(e_i)$, if he chooses $e_i = e_i^{FB}$, he receives $w_i\left(y^{FB}\right) - c_i\left(e_i^{FB}\right) > 0$.■

Proposition 1 shows that in order to induce first-best effort, the Agents have to subject themselves to costly joint punishments in the event that one of them deviates and chooses $e_i \neq e_i^{FB}$. A concern with such contracts is that in the event that the Agents are required by the contract to burn money, they could all be made better off by renegotiating their contract and not burning money. If we insist, therefore, that w is *renegotiation-proof*, then w must be a sharing rule and therefore cannot induce e^{FB}.

This is no longer the case if we introduce an additional party, which we will call a Principal, who does not take any actions that affect output. In particular, if we denote the Principal as Agent 0, then the following sharing rule induces e^{FB}:

$$\begin{aligned} w_i(y) &= y - k \text{ for all } i = 1, \ldots, I \\ w_0(y) &= Ik - (I-1)y, \end{aligned}$$

where k satisfies

$$k = \frac{I-1}{I} y^{FB}.$$

This vector of contracts is a sharing rule, since for all $y \in \mathcal{Y}$,

$$\sum_{i=0}^{I} w_i(y) = Iy - (I-1)y = y.$$

This vector of contracts induces e^{FB} because it satisfies $\partial w_i\left(y^{FB}\right)/\partial e_i = 1$ for all $i = 1, \ldots, I$, and if we imagine the Principal having an outside option of 0, this choice of k ensures that in equilibrium, she will in fact receive 0. In this case, the Principal's role is to serve as a **budget breaker**. Her presence allows the Agents to "break the margins budget," allowing for $\sum_{i=1}^{I} w_i'(y) = I > 1$, while still allowing for renegotiation-proof contracts.

Under these contracts, the Principal essentially "sells the firm" to *each* agent for an amount k. Then, since each Agent earns the firm's entire output at the margin, each Agent's interests are aligned with society's interest. One limitation of this approach is that while each Agent earns the entire marginal benefit of his efforts, the Principal *loses* $I-1$ times the marginal benefit of each Agent's efforts. The Principal has strong incentives to collude with one of the Agents—while the players are jointly better off if Agent i chooses e_i^{FB} than any $e_i < e_i^{FB}$, Agent i and the Principal together are jointly better off if Agent i chose $e_i = 0$.

1.1.5 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:

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

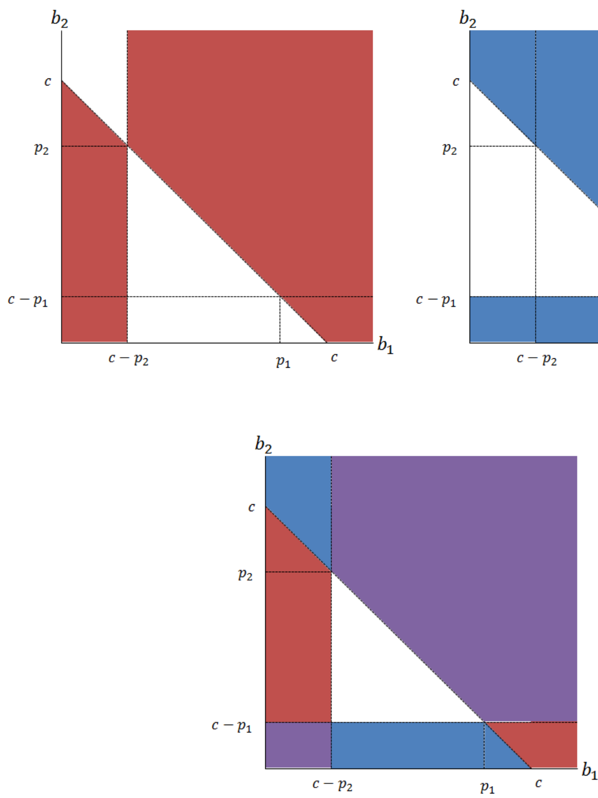
P_1's unconstrained problem is then

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

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

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

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:

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

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

1.2 Formal Contracts with Partial Commitment (TBA)

1.2.1 Renegotiation-Proofness (TBA)

1.2.2 One-Sided Commitment (TBA)

1.3 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.

1.3.1 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. Holmström (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

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

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 $w_{i1}^* : \Delta(\Theta) \to \mathbb{R}$, $w_{i2}^* : \Delta(\Theta) \to \mathbb{R}$, $d_1 : W^2 \to \{1,2\}$, $d_2 : W^2 \to \{1,2\}$, $e_1 \in \mathbb{R}_+$ 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

$$w_{12}^{*}\left(\hat{\theta}\left(y_{1}\right)\right)=w_{22}^{*}\left(\hat{\theta}\left(y_{1}\right)\right)=w_{2}^{*}\left(\hat{\theta}\left(y_{1}\right)\right)=pE\left[y_{2}\right|y_{1},\sigma^{*}]=pE\left[\theta\right|y_{1},\sigma^{*}],$$

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

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

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

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

so that he will choose

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

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

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

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

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

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 $h_\varepsilon'\left(e_t\right) > 0$. 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?*

An important source of conflicting objectives within firms is often the tension between the firm's desire to maximize profits and its workers' concerns for their careers. And importantly, as this model shows, these incentives are not chosen by the firm but rather, they are determined by the market and institutional context in which the firm operates. That is, career concerns provide *incidental*, rather than *designed*, incentives.

In this model, these incidental incentives motivate productive effort. Of course, these incentives may be excessively strong for young workers (see Landers, Rebitzer, and Taylor (1996) for evidence of this effect in law firms), and they may be especially weak for older workers (see Gibbons and Murphy (1992) for evidence of this effect among executives). More generally, however, career concerns incentives may motivate employees to make decisions that are counterproductive for the firm. If an employee is risk-averse, and he can choose between a safe project with outcomes that are independent of his ability and a risky project with outcomes that are more favorable if he is high-ability, he may opt for the safe project, even if the safe project is bad for the firm. In particular, if his expected future wage is linear in his expected ability, then since the market's beliefs about his ability are a martingale, he prefers the market's beliefs to remain constant. If a professional adviser cares about her reputation for appearing well-informed, then she may withhold valuable information when giving advice (Ottaviani and Sorensen, 2006).

If an employee cares about his reputation for being a quick learner, then an "Impetuous Youngsters and Jaded Old-Timers" dynamic can arise (Prendergast and Stole, 1996). In particular, if an employee observes private signals about payoffs of different projects, and smarter employees have more precise information, then smarter employees will put more weight on these private signals. Smarter employees' outcomes will therefore be more variable, and the market understands this, so there is an incentive for employees to "go out on a limb" by putting excessive weight on their private signals to convince the market they are smart (i.e., "youngsters may be impetuous"). Moreover, reversing a previous decision in the future signals, in part, that a worker's initial information was wrong, so older workers might inefficiently stick to prior decisions (i.e., "old-timers may be jaded").

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.

1.3.2 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 max-min 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 (Holmström, 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, 2017; Lipnowski and Ramos, 2017).

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

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

subject to

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

and

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

Functional-Form Assumptions We will assume that $s \sim U\left[-1, 1\right]$ and $y_A\left(s\right) = \beta s$, 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

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

subject to

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

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

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

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

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

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

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

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

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

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

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

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

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

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 (2017) calls "power") can be made credible.

Chapter 3

Organizational Structure

Throughout the class, we have taken as given that an organization consists of multiple individuals, but we have said little about why they interact together or why a single individual cannot do everything. For example, in our discussion of incentive conflicts, we took as given that there was an agent who was able to exert effort in production and a principal who was unable to do so. In our discussion of firm boundaries, we took as given that one party was able to make a type of specific investment that the other party was unable to do so. Garicano and Van Zandt (2013) argue that, "If the mythical unbounded rational, all-capable owner-manager-entrepreneur-manager existed, there would be no organizational processes to talk about, no need to delegate the coordination in organizations to several or many agents, and therefore no organization structure other than a center and everyone else." In other words, fundamentally, bounded rationality (of some sort) is why there are returns to specialization, it is why these returns to specialization are not unlimited, and it is why there is scope for organizations to serve, as Arrow (1974) points out, "as a means of achieving the benefits of collective action in situations in which the price system fails." (p. 33).

In this note, we will explore a couple classes of models that take aspects of bounded rationality as a key limitation that organizations have to contend with and which organizations are specifically designed to address. Before we do so, I will first describe a simplified version of the Lucas (1978) span-of-control model, which in some sense takes a form of bounded rationality as given and thinks about its aggregate implications for the economy. This model is in some ways a building block for the models that will follow.

3.1 Span-of-Control Model

If we are to understand why different firms of different size and productivity coexist in equilibrium, we need a model in which firm sizes and firms' production decisions are determined equilibrium. I will begin by describing a simplified version of the canonical Lucas (1978) "span of control" model in which people in the economy choose whether to be workers or to become entrepreneurs who employ workers. Some people are better at managing others, and these are the people who will, in equilibrium, opt to become entrepreneurs. Better managers optimally oversee more workers (the "span of control" effect), but there are diminishing marginal returns, so that it is not an equilibrium for there to only be a single firm. The underlying organizational source of diminishing marginal returns to management is unmodeled, and Lucas describes the model as providing not "serious organization theory, but perhaps some insights into why organization theory matters economically."

Description There is a unit mass of agents in the economy. Agents differ in their ability to oversee the work of others. Denote this ability by φ, and let $\Gamma(\varphi)$ denote its distribution function, which we will take to be continuous. Each agent chooses whether to be a worker and receive wage w or to become an entrepreneur and receive the profit associated with the enterprise. Output generated by an entrepreneur depends on her managerial ability (φ) and on the number of workers (n) she employs:

$$y = \varphi n^{\theta},$$

where $\theta < 1$ is a parameter that captures in a reduced-form way the organizational diseconomies of scale.

A **competitive equilibrium** is a wage w^*, a labor-demand function $n^*(\varphi)$, and an occupational choice function $d^*(\varphi) \in \{0, 1\}$ specifying which subset of agents become entrepreneurs, how many workers each firm hires, and a wage at which labor demand equals labor supply.

Suppose an agent of ability φ chooses to become an entrepreneur ($d = 1$).

Her labor demand solves

$$\max_n \varphi n^\theta - wn$$

or $n^*(\varphi, w) = (\varphi\theta/w)^{1/(1-\theta)}$, and her associated profits are

$$\pi(\varphi) = \varphi n^*(\varphi, w)^\theta - wn^*(\varphi, w) = (1-\theta)\varphi^{\frac{1}{1-\theta}}\left(\frac{\theta}{w}\right)^{\frac{\theta}{1-\theta}}.$$

An agent will therefore choose to become an entrepreneur if $\pi(\varphi) \geq w$. Since $\pi(\varphi)$ is increasing in φ, there will be some cutoff $\varphi^*(w)$ such that all agents with ability $\varphi \geq \varphi^*(w)$ will become entrepreneurs and all those with ability $\varphi < \varphi^*(w)$ will choose to be workers. Equilibrium wages, w^*, therefore solve

$$\int_{\varphi^*(w^*)}^{\infty} n^*(\varphi, w^*)\, d\Gamma(\varphi) = \Gamma(\varphi^*(w^*)),$$

where the expression on the left-hand side is aggregate labor demand at wages w^*, and the right-hand side is labor supply—the mass of agents who choose to be workers at wage w^*.

This model makes predictions about who will become an entrepreneur, and it has predictions about the distribution of wages and earning as well as firm size. The diminishing returns to the "span of control" effect are a key variable determining the model's predictions, and the model says little about what governs them. Moreover, the model has some predictions about the evolution of wage inequality that are counter to what we have observed over the post-war period in the United States. These two issues are addressed in the models that we will be examining next.

3.2 Knowledge Hierarchies

Garicano (2000) points out that knowledge is an important input into the production process that many of our existing models ignore. Knowledge, the ability to solve problems that naturally arise in the production process, is embodied in individuals who have limited time to work, and this is fundamentally why there are returns from specialization and why organization is important. As Demsetz (1988) points out, "those who are to produce on the basis of this knowledge, but not be possessed of it themselves, must have their activities directed by those who possess (more of) the knowledge. Direction substitutes for education (that is, for the transfer of knowledge)." Education is costly, so an organization will want to specialize the possession of non-routine knowledge in a few and allow production workers to ask for direction when they need help on such non-routine problems. The organizational-design question under the Garicano (2000) view is then: how should a firm organize the acquisition, use, and communication of knowledge in order to economize on the scarce time of its workers and leverage scarce knowledge?

Description A unit mass of workers interact together to produce output, and problems $z \in [0, \infty)$ arise during the production process. Workers are segmented into L distinct classes, with a fraction $\beta_i \geq 0$ in class $i \in \{1, \ldots, L\}$ so that $\sum_i \beta_i = 1$. Workers in class i spend a unit of time producing (t_i^p) or helping others (t_i^h) so that $t_i^p + t_i^h \leq 1$ and $t_i^p, t_i^h \geq 0$, and they possess knowledge set $A_i \subset [0, \infty)$, which costs the organization $c\mu(A_i)$, where $\mu(A_i)$ is the Lebesgue measure of the set A_i. The knowledge sets of workers in different classes can in principle overlap.

Each unit of time a worker spends producing generates a problem z, which is drawn according to distribution function $F(z)$ with density $f(z)$. Without loss of generality, we can order the problems so that $f'(z) < 0$. If a worker of class i encounters a problem $z \in A_i$, she solves the problem and produces one unit of output. If $z \notin A_i$, she can refer the problem to someone else in the organization. If she does not know the solution to the problem, she does not know who else in the organization might know how to solve the problem. That is, workers "don't know what they don't know."

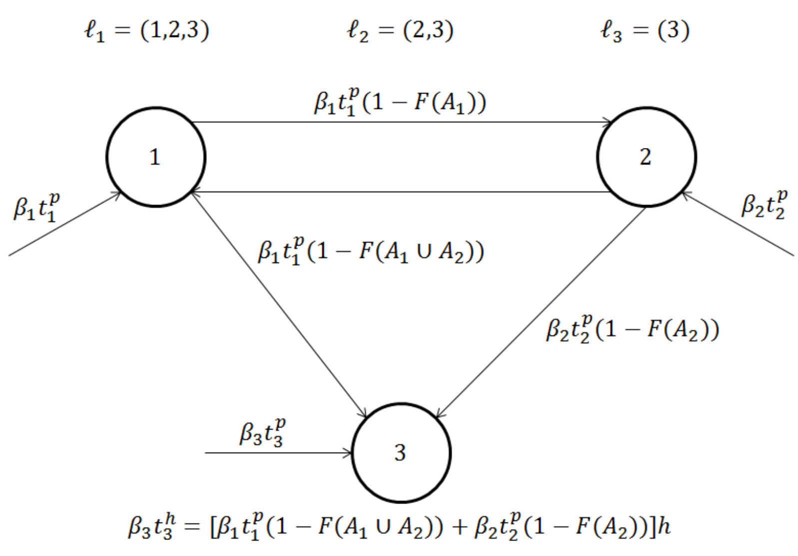
Each unit of time a worker spends helping allows her to process $1/h$ referred problems, where $h < 1$ represents the "communication costs" the helper incurs learning about and assessing the problem. She incurs these costs even when she does not know the answer.

A **referral pattern** ℓ is, for each class i, an ordered set of classes ℓ_i that i can refer to, where $\ell_i(1) = i$ for all i (so that each class can solve any problem it originates), and $\ell_i(n)$ is the nth class that i can refer to. In other words, a worker first tries to solve any problem herself, then she refers it to $\ell_i(2)$, $\ell_i(3)$, and so on. We will say that $j \prec_k i$ if class j precedes group i in the referral pattern for group k. An **organization** is a vector $g = (L, \beta, A, t, \ell)$, which specifies a **number of classes** L, a **class assignment** $\beta = (\beta_1, \ldots, \beta_L)$ with $\beta_i \geq 0$, $\sum_{i=1}^{L} \beta_i = 1$, a **knowledge assignment** $A = (A_1, \ldots, A_L)$, a **time allocation** $t = (t_1, \ldots, t_L)$, where $t_i = \left(t_i^h, t_i^p\right)$ with $t_i^h, t_i^p \geq 0$, $t_i^h + t_i^p \leq 1$, and a referral pattern ℓ.

The problem the organizational designer faces is to choose an organization to maximize the firm's net output. To determine the firm's net output, first note if class i spends enough time helping the classes that refer to it, we will have

$$\underbrace{\beta_i t_i^h}_{\text{time helping}} = \underbrace{\sum_{k:i\in\ell_k} \beta_k t_k^p \left[1 - F\left(\bigcup_{j\prec_k i}\right) A_j\right]}_{\text{problems referred to } i} \underbrace{h}_{\text{time per referral}}.$$

The following figure depicts a sample organization.



In this organization, there are three classes that each spend some time in production, so problems flow into each class. Class 1 is able to refer problems to class 2 and then to class 3, class 2 is able to refer problems to class 3, and class 3 is unable to refer any problems. Class 2 must spend $\beta_2 t_2^h = \beta_1 t_1^p (1 - F(A_1))\, h$ to process all the problems that class 1 refers to it, and class 3 must spend

$$\beta_3 t_3^h = [\beta_1 t_1^p (1 - F(A_1 \cup A_2)) + \beta_2 t_2^p (1 - F(A_2))]\, h$$

to process all the problems that classes 1 and 2 refer to it.

The **net output of class i** is equal to the mass of problems it originates times the probability that someone in its referral pattern knows the solution to those problems minus the costs of class i's knowledge. Total net output is the sum of the net output of all L classes:

$$y = \sum_{i=1}^{L} \left(\beta_i t_i^p F\left(\bigcup_{k \in \ell_i} A_k \right) - c \beta_i \mu(A_i) \right).$$

The **Coasian program** is therefore

$$\max_{L,\beta,A,t,\ell} y \text{ subject to } t_i^h + t_i^p \leq 1, \sum_{i=1}^{L} \beta_i = 1.$$

This problem is not a well-behaved convex programming problem, but several variational arguments can be used to pin down the properties of its solution. First, for any knowledge assignment A, it turns out that $t_i^p = 1$ for some class i, which we will without loss of generality set to $i = 1$, and $t_i^h = 1$

for all others. In other words, in any optimal organization, each worker uses all of her time, one class of workers specializes entirely in production, and the remaining workers specialize in solving problems the production workers refer to them. Any optimal organization has this feature, because if one class produces a higher net output than another, we can always move some of the workers from the less-productive class to the more-productive class and adjust helping times so as to maintain the high-productivity class's net output. Doing so will reduce the amount of time other classes spend producing, and their net output will fall. This perturbation is always feasible as long as multiple groups are production.

The second property of the solution is that the measure of any overlap between two knowledge sets is zero: $\mu(A_i \cap A_j) = 0$. The reason for this is that the knowledge of "problem solvers" never gets used if it is known by the producer class, and it never gets used if it is known by an earlier class in the producer class's list. So whenever $\mu(A_i \cap A_j) > 0$ for some $i \prec_1 j$, we can let $\tilde{A}_j = A_j \backslash A_i$. Under this perturbation, the same problems are solved by the organization, but at a lower cost, since the costs of the higher class's knowledge set is lower.

Next, any optimal organization will feature $A_1 = [0, z_1]$, $\ell_1 = (1, \ldots, L)$, and $A_i = [z_{i-1}, z_i]$ with $z_i > z_{i-1}$. Production workers will learn to solve the most common ("routine") problems, and problem solvers learn the exceptions. Moreover, the workers in the higher classes learn to solve more unusual problems. To see why this is true, suppose class i knows $[z, z + \varepsilon]$

and $j \prec_1 i$ knows $[z', z' + \varepsilon]$, where $f(z) > f(z')$. Then we can swap these two intervals for a small mass of each class of workers. Doing so will keep the learning costs the same. Production will be the same, since the total amount of knowledge is unchanged. But the time spent communicating problems goes down, since those earlier in the referral pattern are now less likely to confront a problem they do not know. As a result, some of the time freed up from the higher class can then be reallocated to the producer class, increasing overall output. A similar argument guarantees that there will be no gaps in knowledge between the classes (i.e., A_i and A_{i+1} overlap at exactly one point). Garicano describes this property as "management by exception" and highlights that it allows specialization in knowledge to be attained while minimizing communication costs.

Finally, any optimal organization necessarily has a pyramidal structure: if $L \geq 2$, then $\beta_1 > \beta_2 > \cdots > \beta_L$. The reason for this is that the total time spent helping by class i is $\beta_i = [1 - F(z_{i-1})] h\beta_1$, and the total time spent helping by class $i+1$ is $\beta_{i+1} = [1 - F(z_i)] h\beta_1$. Since $z_i > z_{i-1}$, $\beta_{i+1} < \beta_i$.

The Coasian program therefore becomes

$$\max_{L, z_1, \ldots, z_L, \beta} F(z_L) \beta_1 - \sum_{i=1}^{L} c\beta_i (z_i - z_{i-1}),$$

where $z_0 = 0$, subject to

$$\beta_i = [1 - F(z_{i-1})] h\beta_1 \text{ for } i > 1.$$

This problem can be solved for specific distributional assumptions. For example, it can be solved explicitly if the distribution of problems is exponential, so that $f(z) = e^{-\lambda z}$. In the solution to this problem, production workers know how to solve problems in an interval of length Z_w^*, and all problem solvers know how to solve problems in an interval of length Z_s^*. If we define the **span of control** at layer i as $s_i = \beta_i/\beta_{i+1}$, we can say something about the comparative statics of the optimal organization.

First, if communication costs fall (i.e., if h goes down) because of improvements in communication technology, then Z_s^* increases, Z_w^* falls, and the span of control increases at each layer. That is, improvements in communication technology lead to flatter organizations with less-knowledgeable production workers. If communication becomes cheaper, relying on problem solvers is "cheaper," so it is optimal for each production worker to acquire less knowledge, and each problem solver can communicate solutions to a larger team, so the span of control of problem solvers increases.

If the cost of acquiring knowledge falls (i.e., c decreases) because of improvements in information technology, then Z_s^*, Z_w^*, and s_i all increase. Improvements in information technology therefore also lead to flatter organizations with more knowledgeable helpers, but they also lead to an increase in the knowledge possessed by production workers.

3.3 Monitoring Hierarchies

Any theory of optimal firm size has to provide an answer to the *replication question*: "Why can we not simply double all the firm's inputs and double the resulting output?" Fundamentally, the answer has to be that there is some sort of fixed factor of production *at the firm level* that is not replicable, and while Penrose (1959) argues that this factor must be related in some way to what managers do, it is not obvious exactly what it is about what they do that makes it fixed in nature (i.e., if one manager's ability to coordinate activities is fixed, why not hire a second manager?), "Whether managerial diseconomies will cause long-run increasing costs [requires that] management... be treated as a 'fixed factor' and the nature of the 'fixity' must be identified with respect to the nature of the managerial task of 'coordination.' This identification has never been satisfactorily accomplished." (p. 12)

Vertical Control Loss

Williamson's (1967) answer is that even if one were to double the number of managers in a firm in order to get double their coordination efforts, someone would have to coordinate *their* activities as well, and therefore coordination activities at the highest level necessarily must be embodied within a single individual. He describes a theory in which a firm consists of a layer of workers and a hierarchy of monitors. The top-level manager supervises a layer of subordinates who each supervise a layer of subordinates, and so on until we get to the bottom layer of the firm, which consists of production workers. Production workers produce one unit of output each, but some of this output gets lost for each layer in the organization, a reduced-form way to capture communication losses and agency costs, and to capture Williamson's idea that "The larger and more authoritarian the organization, the better the chance that its top decision-makers will be operating in purely imaginary worlds." (p. 123)

Model Description In the simplest version of Williamson's model (due to Mookherjee (2013)), each manager has an exogenously specified span of control s, so if there are $N+1$ layers $i \in \{0, 1, \ldots, N\}$, there are s^i employees in layer i and therefore s^N production workers who each produce α^N units of output, where $\alpha < 1$ represents the fraction of output that gets lost for each layer of the organization. This parameter, referred to as the **vertical control loss**, is a reduced-form way to capture communication losses and agency costs.

Wages for production workers are w, and wages for employees in layer i are $\beta^{N-i}w$, where $\beta > 1$ represents the additional wages that have to be paid for employees higher in the organization. Assume $\alpha s > 1$ and $s > \beta$. An **organization** is fully characterized by a number of layers N.

The Program The firm chooses the number of layers of subordinates, N, to solve

$$\max_N \pi(N, \alpha)$$

where

$$\pi(N, \alpha) = (\alpha s)^N - w \sum_{i=0}^{N} \beta^{N-i} s^i = (\alpha s)^N - w \frac{s^{N+1} - \beta^{N+1}}{s - \beta}.$$

The objective function satisfies increasing differences in (N, α), so N^* is increasing in α (i.e., the optimal number of layers is higher if less information is lost between successive layers or if there are lower agency costs). Under some parameter restrictions, there is an interior solution as long as $\alpha < 1$ ($N^* \to \infty$ as $\alpha \to 1$), so this model pins down the optimal number of layers and hence the optimal number of workers in the firm.

The paper provides an answer to the question of why there are organizational diminishing returns to scale: activities within the firm must be coordinated, and the highest-level coordination must occur within the single individual who occupies the top position. The model's main results, however, require some pretty stringent parameter restrictions, since the firm's revenues, $(\alpha s)^N$ are convex in N (since $\alpha s > 1$). For there to be an interior solution, it has to be the case that the firm's costs, $w\left(s^{N+1} - \beta^{N+1}\right) / (s - \beta)$ are, in some sense, even more convex in N. Moreover, while the paper does answer the initial question, it is silent both on why there is vertical control loss and what determines it, as well as on why wages progress in a proportional way through the hierarchy.

Layers of Supervisors

Calvo and Wellisz (1978) put more structure on the Williamson (1967) model by explicitly introducing productive effort by production workers, monitoring effort by supervisors, and optimal wage choices by the firm.

Model Description Suppose there are $N+1$ layers in the firm, there are M_i workers in layer $i \in \{0, 1, \ldots, N\}$, and layer $i = N$ represents production workers. Workers in layer i are paid a base wage of w_i, which may be docked for non-performance, as we will see below.

Suppose each worker in the firm has a utility function $w - c(e)$, where w is the worker's wage, $c(e)$ is the cost of effort, and $e \in [0, 1]$ represents the fraction of the week worked. Each unit of time spent working by a production worker generates one unit of output for the firm, so if all production workers exert effort e_N, the firm's total output will be $M_N e_N$. Supervisors exert effort to uniformly monitor their direct subordinates. If M_i supervisors at level i work e_i units of time each, they supervise each $i+1$-level employee an amount equal to $e_i M_i / M_{i+1}$. With probability $p_{i+1} = h(e_i M_i / M_{i+1})$, the firm observes e_{i+1} and reduces worker $i+1$'s wage from w_{i+1} to $e_{i+1} w_{i+1}$. With probability $1 - p_{i+1}$, the worker receives wage w_{i+1}.

If a worker exerts effort e and is monitored with probability p, his utility will be $pwe + (1-p)w - c(e)$, and he will choose $c'(e^*(p, w)) = pw$. If his outside option yields utility $\bar{u}$, he will accept employment at the firm if $pwe^* + (1-p)\,w - c\,(e^*) \geq \bar{u}$. An **organization** is a vector $g = (M, e, w, p)$, consisting of a number of workers at each level below the top, $M = (M_1, \ldots, M_N)$, an effort vector $e = (e_1, \ldots, e_N)$, a wage vector $w = (w_1, \ldots, w_N)$, and a monitoring probability vector $p = (p_1, \ldots, p_N)$, where $p_{i+1} = h\,(e_i M_i / M_{i+1})$.

The Program Suppose the firm has $N = 1$. The firm's problem is therefore to choose the number of production workers M_1 and wage w_1 to maximize

$$\pi_1^* = \max_{M_1,p,w,e} M_1 e - [pwe + (1-p)\,w]\,M_1,$$

subject to monitoring feasibility $p = h\,(1/M_1)$, incentive-compatibility $c'\,(e) = pw$, and individual rationality: $pwe + (1-p)\,w - c\,(e) \geq \bar{u}$. Since $p = h\,(1/M_1)$ is decreasing in M_1, increasing M_1 reduces effort per worker fixing w. A larger firm must therefore either reduce worker effort or pay each worker more, which implies that there are decreasing returns to scale, holding the organizational structure fixed.

Another way the firm can expand is by increasing the number of layers. If the firm has $N+1$ layers, it chooses $g = (M, e, w, p)$ to solve

$$\pi_N^* = \max_{M,p,w,e} M_N e_N - \sum_{i=1}^{N} [p_i w_i e_i + (1-p_i)\,w_i]\,M_i$$

subject to monitoring feasibility

$$p_i = h\left(e_{i-1}M_{i-1}/M_i\right),$$

incentive-compatibility $c'\left(e_i\right) = p_i w_i$, and individual rationality $p_i w_i e_i + \left(1 - p_i\right) w_i - c\left(e_i\right) \geq \bar{u}$. Adding an additional layer allows the firm to employ more production workers while maintaining a given effort level for those production workers, since it can increase $p_N = h\left(e_{N-1}M_{N-1}/M_N\right)$ by increasing the number of direct supervisors for production workers or getting them to work harder. Doing so is costly, though, since the firm has to pay and motivate the additional layer of supervisors.

The paper's main result is that as long as it is profitable to produce with a single layer rather for the entrepreneur to do all the production herself, $\pi_1^* > 1$, then firm size is unbounded, that is, $\lim_{N\to\infty}\pi_N^* = \infty$. The paper outlines a replication argument by showing that there is a way of expanding the number of production workers indefinitely by increasing the number of layers in the firm. It does so by finding an $N = 1$ arrangement that leads to positive profits for each production worker and uses it to construct an $N > 1$ level hierarchy in which wages, effort, and the span of control are constant across levels. If it was profitable to add the second layer, it will be profitable to add the third layer, and so on, so profits are unbounded along this path.

The negative conclusion of this paper is that Williamson's proposed fixed input, once microfounded, does not lead to bounded firm size. Qian (1994) shows that Calvo and Wellisz's result is also fragile along a couple dimensions, putting Williamson's explanation on foundations that are at best highly contingent on the underlying environment.

Chapter 4

Managers and Management

4.1 Conformity of Recommendations

Prendergast (1993) provides a theory of why subordinates might be reluctant to reveal their independent assessments of decision-relevant information and instead focus on telling the boss what she already believes is true. If information acquisition is costly, then a subordinate must be motivated to acquire it. In many cases, however, the direct outcome of that information is not verifiable, so the other way to reward the agent for acquiring information is to "cross-check" it against other information sources. But then the subordinate has the incentive to report information that is likely to be successfully cross-checked, rather than offering an independent analysis. This leads to what Prendergast (1993) refers to as the "Yes Men" phenomenon as an unfortunate by-product.

Model There is a risk-neutral Principal (P) and a risk-neutral agent (A), and an unobservable decision-relevant state of the world $\theta \in \{0,1\}$ with $\Pr[\theta = 1] = 1/2$. The Agent chooses an effort level $e \in [\underline{e}, 1]$ at a private cost of $c(e)$ with $c'', c' > 0$, $\underline{e} \geq 1/2$, and $c(\underline{e}) = 0$. This effort determines the probability the agent observes a private signal $x_A \in \{0,1\}$, with $\Pr[x_A = \theta] = q$. The Principal also observes a private signal $x_P \in \{0,1\}$, with $\Pr[x_P = \theta] = q \geq 1/2$, and the Agent observes a signal of the Principal's signal. That is, he observes $z \in \{0,1\}$, with $\Pr[z = x_P] = r \geq 1/2$. Signals are noncontractible, but the Principal can write a contract that specifies a payment to the agent and can depend on messages sent by both players. The timing of the game is:

1. The state of the world $\theta \in \{0,1\}$ is realized, and it is not observed by either player.

2. The Principal offers a contract $s : \mathcal{M} \to \mathbb{R}$ to the Agent, where $\mathcal{M}$ is a contractible message space.

3. The Agent accepts or rejects. Rejection ends the game, and both players receive a payoff of 0.

4. The Principal observes her private signal x_P.

5. The Agent chooses effort e at cost $c(e)$.

6. The Agent observes private signals x_A and z.

7. The Agent and Principal simultaneously send messages m_A, m_P. Let $m \in \mathcal{M}$ denote the vector of messages sent.

8. The principal makes a decision $d \in \{0, 1\}$.

Payoffs are $\pi = v \cdot 1_{d=\theta} - s(m)$ for the principal and $u = s(m) - c(e)$ for the agent, and the solution concept is Perfect Bayesian Equilibrium.

Analysis Suppose $s(m)$ is independent of m_P. Then the agent will report whatever message gives him the highest expected payoff. Therefore, the principal can induce truthful revelation of x_A by setting $s(m)$ constant. In that case, the agent exerts effort $\underline{e}$ and reports $m_A = x_A$. If $x_A = x_P$, then the principal will choose $d = x_P$. If $x_A \neq x_P$, then the principal will choose $d = x_P$ if $q \geq \underline{e}$ and $d = x_A$ otherwise. The principal's expected payoff (gross of transfers) is therefore

$$(\underline{e}q + \max\{\underline{e}, q\}(1 - \min\{\underline{e}, q\}))\, v.$$

Now, suppose the principal tries to motivate the agent to exert effort $e > \underline{e}$. To do so, she must condition the agent's payoff on both messages; in particular, she can reward the agent if his report agrees with her information. That is, $s(m) = s_H$ if $m_A = m_P$, and $s(m) = s_L$ if $m_A \neq m_P$. Crucially, this incentive contract induces the agent to exert effort in order to better match his report to what he expects the principal to report. That is, it gives the agent the incentive to conform his report to the principal's information.

Moreover, the agent can choose to conform by reporting according to his signal z of the principal's belief, rather than his independent information x_A.

If $z = x_A$, then the agent will report $m_A = x_A$. If $z \neq x_A$, then the agent will report $m_A = x_A$ if and only if

$$s_L + (eq + (1-e)(1-q))(s_H - s_L) \geq s_L + r(s_H - s_L).$$

That is, the value from truthfully reporting x_A is larger than the value of being a "yes man," reporting z, and matching the principal's report with probability r. Since $s_H \geq s_L$ in the optimal contract, this condition can be rewritten

$$eq + (1-e)(1-q) \geq r.$$

Since $q \geq 1/2$, the left-hand side of this expression is increasing in e. Therefore, this condition requires that $q \geq r$. Intuitively, for the agent to have the incentive to acquire information, two conditions must hold. First, the principal's signal of θ must be sufficiently precise, since this signal determines the effect of a more accurate x_A on the agent's ability to match x_P. Second, the agent cannot have too good of a sense of what the principal believes, since if he does, then the opportunity cost of relying on x_A rather than simply reporting z is large.

Note that this condition does not include the initial cost of acquiring information, $c(e)$. But if the agent expects to simply report z, then he has no incentive to actually acquire information. Therefore, the agent's true incentive constraints are

$$s_L + (eq + (1-e)(1-q))(s_H - s_L) - c(e) \geq s_L + r(s_H - s_L),$$

which deters the "global" deviation to no effort and conformity, and

$$e \in \underset{\tilde{e}}{\operatorname{argmax}}\, s_L + (\tilde{e}q + (1-\tilde{e})(1-q))(s_H - s_L) - c(\tilde{e})$$

or

$$(2q-1)(s_H - s_L) = c'(e)$$

in order to deter local deviations in effort conditional on truthful revelation.

This model therefore shows that in order to motivate an Agent to acquire non-verifiable information, the Principal has to have an already-formed (although perhaps not strong) opinion on the matter because that allows her to cross-check the Agent's message. But the very act of rewarding successful cross-checking motivates the Agent to focus on confirming the Principal's opinion. Ideally, the Principal would like to have a strong but unknown opinion (i.e., q large and r close to $1/2$), but strong opinions and public opinions tend to go hand-in-hand.

4.2 Leading by Example

Hermalin (1998) considers the role of leaders in motivating their subordinates. This paper formalizes the notion of "leading by example" and considers a model in which a leader has information about the returns of her followers' efforts. For example, plant managers may have information about when demand conditions are sufficiently bad that the plant may be closed if agents continue shirking. However, the manager has a credibility problem because she always wants the agents to work as hard as possible, regardless of the demand state. In order to credibly convince her agents that their efforts are truly essential, the manager must bear a cost that would deter her from making such claims falsely.

In the simples case, this cost might be purely monetary: the manager might pay bonuses to the agents to convince them that she is telling the truth. However, this cost might also involve privately costly but socially efficient actions. In particular, the manager might "lead by example": work especially hard in order to convince her followers that working hard is particularly important for the current task. Military history is rife with examples of generals who "lead from the front," heading dangerous charges in order to motivate their soldiers to fight their hardest and not retreat. The converse implication of this argument is that a general who retreats from battle sends a strong signal to her followers that winning the battle is not worth dying and so frequently spurs a disorderly retreat.

Model A Principal offers an Agent a contract that specifies a sharing rule over output, with a fraction $s \in [0,1]$ of the output going to the agent. The Principal then privately observes the state of the world and chooses publicly observable effort, after which the agent chooses effort. The timing of the game is as follows.

1. The Principal offers the Agent a contract that specifies a split $s \in [0,1]$ of the final output.

2. The Principal privately observes $\theta \in \{0,1\}$, where $\Pr[\theta = 1] = p$.

3. The Principal chooses effort $e_P \in [0,1/2]$ at cost $c(e_P)$, which is observed by the Agent.

4. The Agent chooses $e_A \in [0,1/2]$ at cost $c(e_A)$.

5. Output $y \in \{0,H\}$ is realized with $\Pr[y = H | e_A, e_P] = \frac{1+\theta}{2}(e_P + e_A)$.

Players' effort costs are increasing and convex. The Principal's payoffs are $\pi = (1-s)y - c(e_P)$, and the Agent's payoffs are $u = sy - c(e_A)$. The solution concept is Perfect Bayesian Equilibrium, which specifies an effort strategy $e_P : \{0,1\} \to [0,1/2]$ for the principal, an effort strategy $e_A : [0,1/2] \to [0,1/2]$, and a belief process over θ that is consistent with Bayes's rule whenever possible.

Analysis Suppose that $p = 1$, so that it is common knowledge that $\theta = 1$. A nearly identical analysis can be carried out if $p = 0$. Given a split of the surplus, regardless of e_P, the Agent chooses e_A to solve

$$c'(e_A) = sH,$$

and so the Principal chooses e_P to solve

$$c'(e_P) = (1 - s) H.$$

That is, this is a standard moral-hazard-in-teams problem. Given that there is no budget breaker to make both parties residual claimant, we should expect at least one party to exert too little effort. Players have identical efforts and $c(\cdot)$ is strictly convex, so the optimal contract sets $s = 1/2$, and both players exert the same effort.

How does adding private information about θ change the analysis? Since the Principal observes θ before choosing effort, the Agent's beliefs about θ now potentially depend on e_P. Consequently, this problem becomes a signalling game, where e_P is the choice used to signal the underlying state of the world. The Principal's payoff exhibits increasing differences in (θ, e_P) as well as in (θ, e_A), so there may exist a separating equilibrium.

Define e_P^H, e_A^H, e_P^L, and e_A^L as Principal and Agent efforts if $\theta = H$ and if $\theta = L$. Then

$$c'\left(e_A^H\right) = sH$$

and

$$c'\left(e_A^L\right) = s\frac{H}{2}.$$

The Principal must have an incentive to choose e_P^H if $\theta = 1$ and e_P^L if $\theta = 0$. Therefore,

$$\left(e_P^H + e_A^H\right)(1-s)H - c\left(e_P^H\right) \geq \left(e_P^L + e_A^L\right)(1-s)H - c\left(e_P^L\right)$$

and

$$\left(e_P^L + e_A^L\right)(1-s)\frac{H}{2} - c\left(e_P^L\right) \geq \left(e_P^H + e_A^H\right)(1-s)\frac{H}{2} - c\left(e_P^H\right).$$

Relative to the symmetric-information case, the Principal derives an additional benefit from working hard: it induces the Agent to work hard as well. This additional signaling effect has two consequences in the original moral-hazard-in-teams setting. First, holding the share s fixed, the Principal might be willing to exert strictly more than her symmetric-information effort level, so that signaling partially undoes the underprovision of effort. Second, the Principal can adjust the share s so that she earns less of the total surplus, while the Agent earns more, which "spreads" the effort effect of signaling across the two players. Essentially, the signaling effect is a substitute for a monetary payoff, which implies that the Agent can be given a steeper incentive scheme. Since $c(\cdot)$ is strictly convex, equalizing efforts across the two players generates an additional gain.

4.3 Communication Channels

In addition to designing incentives and allocating control, an important problem firms face is how best to design communication channels: who should talk to whom, and who should listen to whom? Different papers in this branch of the literature focus on different aspects of the problem. Bolton and Dewatripont (1994) thinks about how best to design communication channels when communicating information and processing information are both costly to the organization. Dewatripont and Tirole (2005) highlights a complementarity that arises when players incur private costs to speak and to listen. Calvo-Armengol, de Marti, and Prat (2015) and Matouschek and Reich (2017) consider the design of communication networks in settings in which there is a fixed cost of setting up an additional communication channel between players but no marginal cost of communication.

With the exception of Dewatripont and Tirole, these papers (and others) abstract from incentive issues and focus on team-theoretic environments in which all players have the same objectives. In this note, we will focus on a recent paper by Dessein, Galeotti, and Santos (2016), which focuses on a complementarity in organizational design that arises when players have a limited "budget" to learn about decision-relevant parameters, which constrains their ability to coordinate with each other or to tailor their decisions to the underlying state of the world.

In this model, an organization consists of multiple agents, each of whom has to take actions. Each agent wants to take some actions that are tailored to their own idiosyncratic state of the world as well as some actions that are coordinated with each others' actions. If players have limited time and attention to listen to each other, then it might be optimal for players to communicate about only a few tasks and to completely ignore others. If players communicate about a task, then that task can freely adapt to the state of the world without causing miscoordination. But if the task varies with the state of the world, then it is even more important to communicate about it. So communication facilitates coordinated adaptation, which increases the returns to further communication. This complementarity leads to specialization within the organization.

Model We will consider a two-player example that highlights the complementarity that we described above. A Principal has to choose a communication network, which consists of a pair (t_P, t_A) that has to satisfy $t_A, t_P \geq 0$ and an **attention budget constraint** $t_P + t_A \leq T$. These values determine the probability with which players' actions are publicly observed. The Principal and the Agent each privately observe a **local state** $\theta_P, \theta_A \in \{0, 1\}$ with $\Pr[\theta_i = 1] = p_i$. Each player then chooses an **adaptation action** $a_i \in \{0, 1\}$, which is publicly observed with probability t_i. With the complementary probability, nothing is observed about that action. Finally, players choose a **coordinating action** c_i. The timing is:

1. The Principal chooses (t_P, t_A).

2. The Principal privately observes θ_P, and the Agent privately observes θ_A.

3. The Principal and Agent choose adaptation actions a_P and a_A.

4. A public signal $\sigma = (\sigma_P, \sigma_A)$ is realized, with $\sigma_i \in \{a_i, \emptyset\}$ $\Pr[\sigma_i = a_i] = t_i$ and $\Pr[\sigma_i = \emptyset] = 1 - t_i$.

5. The Principal and Agent choose coordinating actions c_P and c_A.

Both players have the same payoffs, which are equal to

$$\alpha_P 1_{a_P=\theta_P} + \alpha_A 1_{a_A=\theta_A} + \beta \left(1_{c_P=a_A} + 1_{c_A=a_P}\right),$$

where $\alpha_i, \beta > 0$. Assume that $p_i \geq 1/2$ for each i. The solution concept is Perfect Bayesian Equilibrium.

Analysis First, note that if T is sufficiently large (for instance, $T \geq 2$), then the Principal will optimally choose $t_P = t_A = 1$, and the attention budget constraint does not bind. In this case, both a_i are publicly observed, and so $c_P = a_A$ and $c_A = a_P$ with probability 1.

Now, suppose T is smaller, say $T = 1$. Each adaptation action a_i could either be $(i.)$ constant, in which case $a_i = 1$ because $p_i \geq 1/2$, or $(ii.)$ equal to the state, $a_i = \theta_i$. If $a_i = 1$ always, then $t_i = 0$ is optimal because the other player can already perfectly infer $a_i = 1$ and set $c_{-i} = 1$ as well. In contrast, if $a_i = \theta$, then increasing t_i increases the probability that $c_{-i} = a_i$ in equilibrium. Players' payoffs from c_{-i} then equal $\beta\left(t_i+(1-t_i)\,p_i\right)$. This is the central complementarity in the paper (highlighted here in a particularly stark way). If my choice is predictable, there is no point in communicating about it. And if nobody is communicating with me about my action, I ought to be predictable. If, however, a_i does adapt to the local state, then there is value to communicating about it, and if we are communicating about a_i, we might as well tailor it to the local state: in a richer model with more than two actions, communicating about a_i facilitates further adaptation to the state of the world, which in turn increases the return to further communication.

Conditional on adaptation, the marginal value of communication is constant. The optimal communication network, therefore, will be bang-bang in this example: $t_i = 1$ for some $i \in \{A, P\}$, and $t_i = 0$ for the other. If $t_i = 1$, then clearly $a_i = \theta$ is optimal. If $t_i = 0$, then $a_i = \theta_i$ is optimal if and only if $\alpha_i + \beta p_i \geq p_i \alpha_i + \beta$ or $\alpha_i \geq \beta$.

The $i \in \{A, P\}$ for which $t_i = 1$ solves

$$\max_{i\in\{A,P\}} \left\{\alpha_i + \beta - \max\left\{\alpha_i + \beta p_i, p_i\alpha_i + \beta\right\}\right\}.$$

That is, i maximizes the smaller of $(1-p_i)\,\beta$ or $(1-p_i)\,\alpha$: players should focus all of their communication on the task for which either $(i.)$ adaptation is especially important or $(ii.)$ the local state of the world is especially uncertain (given that $p_i \geq 1/2$, $1 - p_i$ is increasing in the risk associated with a fixed action a_i).

The model highlights several features of an optimal organization. First, there is a force toward making sure that communication channels are specialized. This complementarity is highlighted in an especially stark way in our example, but it is an essential element of the full model in Dessein, Galeotti, and Santos (2016). In particular, the paper shows that even if increasing t_i has diminishing returns on the probability of learning the state, the complementarity between adaptation decisions and communication might still push organizations to specialize in certain tasks. Second, coordination can be attained in two ways. They can adapt to their local state, and communication can be used to coordinate them. This is an optimal way to foster coordination if the state is uncertain, and adaptation is important. Alternatively, actions can be fixed in a "routine" that is not adapted to the local state but that facilitates coordination. The tasks in which an organization specializes in will be both adapted and coordinated, while other tasks will *either* be coordinated or adapted, but not both.

4.4 Judgment and Clarity

Dewan and Myatt (2008) attempts to identify the characteristics of "good leadership." A leader wants her followers to both adapt to a state of the world and to coordinate with each other. All players have identical preferences, but the leader has private information about the state and can communicate only imperfectly. The paper focuses on comparative statics for two characteristics of a leader's communication: **judgement**, which is how precisely the leader knows the state of the world, and **clarity**, which is how well the leader can communicate her information to her followers.

Both judgment and clarity help the leader provide accurate information about the state. However, a clear leader is likely to send the same (or similar) messages to everyone, which means that followers can more closely adapt to that information without sacrificing coordination. Consequently, the paper argues that clarity is more important that judgement on the margin.

The Model There is a single leader and two followers, $i \in \{1, 2\}$. The timing of the game is as follows.

1. A state of the world $\theta \in \{0, 1\}$ is realized, with $\Pr[\theta = 1] = p$.

2. The leader privately observes a signal $s \in \{\emptyset, \theta\}$, where $\Pr[s = \theta] = q_J$.

3. Each follower $i \in \{1, 2\}$ observes a message $m_i \in \{\emptyset, s\}$, where $\Pr[m_i = s] = q_C$, and m_1 and m_2 are independent conditional on the state.

4. Each follower makes a decision $d_i \in [0, 1]$.

All players have identical preferences:

$$u = \alpha \sum_{i=1}^{2} (\theta - d_i)^2 + (1 - \alpha)(d_1 - d_2)^2 .$$

That is, a follower wants to both adapt his own decision to the state of the world and coordinate with the other follower, with quadratic loss functions attached to each. The weight $\alpha \in [0, 1]$ measures the relative importance of adaptation relative to coordination.

Consider symmetric decision rules for the followers, with d_m denoting the decision if $m = m_i$. Then d_m maximizes

$$E[u| m_i = m] = \alpha \sum_{i=1}^{2} E\left[(\theta - d_m)^2 \middle| m_i = m\right] + (1-\alpha) E\left[(d_m - d_{-i})^2 \middle| m_i = m\right],$$

which has first-order condition (assuming d_m^* interior)

$$d_m^* = \alpha E[\theta| m_i = m] + (1-\alpha) E[d_{-i}| m_i = m].$$

If $m = \emptyset$, then $E[\theta| m_i = m] = p$. If $m \neq \emptyset$, then $E[\theta| m_i = m] = m \in \{0, 1\}$. The key interaction is in the coordination term $E[d_{-i}| m_i = m]$. If $m = \emptyset$, then

$$E[d_{-i}| m_i = \emptyset] = \frac{q_J}{q_J + q_C} d_\emptyset^* + \frac{q_C}{q_J + q_C}(p d_1^* + (1-p) d_0^*).$$

Similarly, if $m = \theta$, then

$$E[d_{-i}| m_i = \theta] = (1 - q_C) d_\emptyset^* + q_C d_\theta^*.$$

Therefore,

$$d_\emptyset^* = \alpha p + (1-\alpha)\left(\frac{q_J}{q_J + q_C} d_\emptyset^* + \frac{q_C}{q_J + q_C}(p d_1^* + (1-p) d_0^*)\right),$$

$$d_1^* = \alpha + (1-\alpha)\left((1-q_C)\,d_\emptyset^* + q_C d_1^*\right),$$

and

$$d_0^* = (1-\alpha)\left((1-q_C)\,d_\emptyset^* + q_C d_0^*\right).$$

Straightforward but tedious calculations yield $d_\emptyset^* = p$, which one might expect. For $\theta \in \{0,1\}$,

$$d_\theta^* = \frac{\alpha\theta + (1-\alpha)(1-q_C)\,p}{1-(1-\alpha)\,q_C}.$$

Note that d_θ^* is independent of q_J, while d_1^* and d_0^* are, respectively, increasing or decreasing in q_C. Intuitively, if the leader is clear, then each follower believes that the messages are more likely to be equal, which means that each follower can adapt to the state of the world while remaining relatively confident about coordinating.

For a fixed probability of informing a given agent, $q_J q_C$, the leader would prefer to increase q_C and decrease q_J until $q_C = 1$ (a corner solution). That is, clarity is more valuable than judgment because clarity facilitates coordination.

4.5 Differing Visions

We will now discuss a pair of papers (Che and Kartik (2009) and Van den Steen (2010)) that examine the organizational consequences of differing priors. These papers, along with much of Van den Steen's research, tackle an understudied set of questions about leadership and management. In particular, how do authority, delegation, and incentives work when parties have different *beliefs* about the correct course of action? In some respects, different beliefs are like different preferences, in that both create incentive misalignment between parties. Importantly, however, beliefs differ from preferences because they respond differently to new information. The following model (which does not do Van den Steen's work justice—Van den Steen (2009) is worth reading for a creative take on a little-used modeling ingredient) highlights this force in a stark way.

The Model There is a principal and an agent who interact once. The timing of the interaction is as follows.

1. A state of the world $\theta \in \{-1, 1\}$ is realized. Player $i \in \{A, P\}$ believes that $\Pr[\theta = 1] = q_i$, and these differing beliefs are common knowledge.

2. The agent chooses information acquisition effort $e \in [0, 1]$, which is publicly observed.

3. A signal $s \in \{0, 1\}$ is realized with probability e, with $\Pr[s = \theta] = r \geq 1/2$. If s is observed, the posteriors update to $q_A^L < q_A < q_A^H$ for the agent and $q_P^L < q_P < q_P^H$ for the principal.

4. The principal makes a decision $d \in \{0, 1\}$.

Both players have payoff $d\theta$, so the only source of incentive misalignment is the different belief p_i. We consider a Perfect Bayesian Equilibrium of this game.

Analysis Let us assume that $q_A^L > 1/2$, so the agent always wants $d = 1$, and that $q_P < 1/2 < q_P^H$, so the principal will choose $d = 1$ if and only if the high signal is realized. Then the agent's expected payoff when he chooses effort e is given by

$$e\left(q_A r - (1 - q_A)(1 - r)\right) - c(e),$$

with first-order condition

$$q_A + r - 1 = c'(e).$$

By assumption, $r + q_A \geq 1$, so this interior solution is indeed the maximum if $c'(0) = 0$ and c is strictly increasing and strictly convex.

Note that there are two forces that increase effort in this model. First, a larger r implies that the signal is more valuable because it is more accurate; a more accurate signal increases effort. In addition, the agent's effort is increasing in q_A, his prior belief that the state of the world is $\theta = 1$. Although this second effect resembles a simple bias, in the sense that the agent "likes" $d = 1$ more, it actually arises for a subtly different reason. The larger is q_A, the more convinced the agent is that her effort will yield a signal that confirms her prior. Informally, an agent who believes that $\theta = 1$ with high probability also believes that any evidence would confirm $\theta = 1$ with high probability. Such an agent is very willing to gather information, because he believes that any information is likely to support his cause and get the principal to make the "correct" decision (from his perspective).

Chapter 5

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.

5.1 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 Holmström (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.

5.1.1 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.

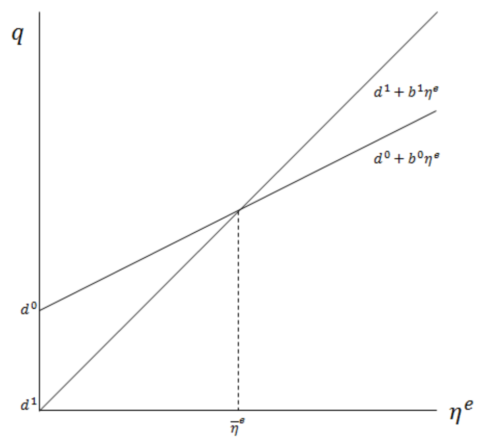


Figure 1

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 $E\left[\eta_2^e(\varphi)\right] = (1+g)\,\eta_1^e(\varphi_0) > \eta_1^e(\varphi_0)$.

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,

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

where $\hat{p}(\varphi_0)$ 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 $\hat{p}(\varphi)$ the conditional probability that $\theta = \theta^H$ given public signal φ, it must be the case that

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

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.

5.1.2 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 $E[\theta | j^1 = 1] = \left(1 + \hat{\theta}\right)/2$, and the expected ability of non-promoted workers is $E[\theta | j^1 = 0] = \hat{\theta}/2$. 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), Manove (1997), Zabojnik and Bernhardt (2001), Waldman (2003), Krakel and Schottner (2012), Auriol, Friebel, and von Bieberstein (2016), and Ke, Li, and Powell (2018) 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 6

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. Antràs 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 (Holmström 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.

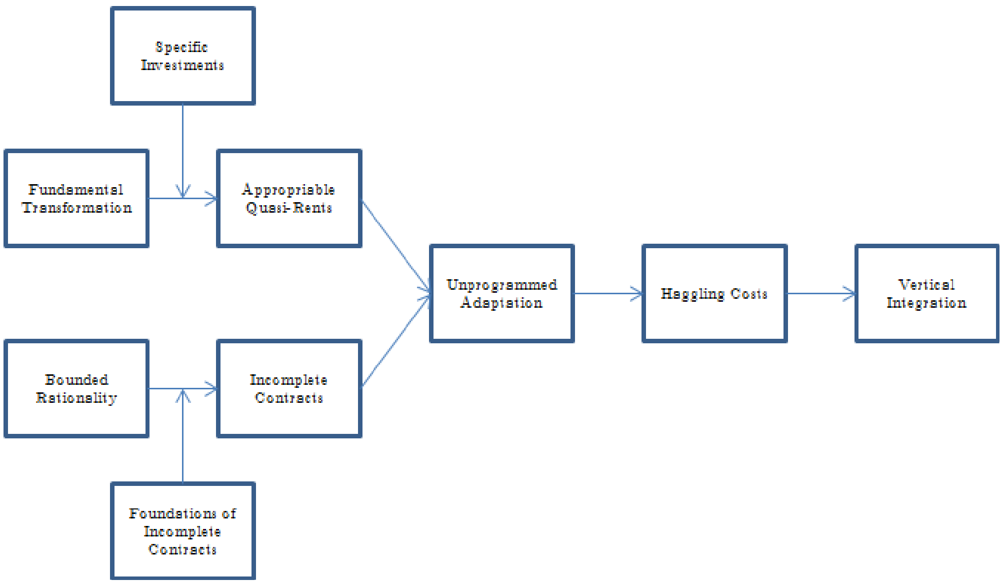
In contrast, in Transaction-Cost Economics, all the action takes place ex post, during the time in which decisions are made. Integration is chosen, precisely because it avoids inefficient haggling costs. We will look at two implications of this observation in the context of two models. The first, which 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.

6.1 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.

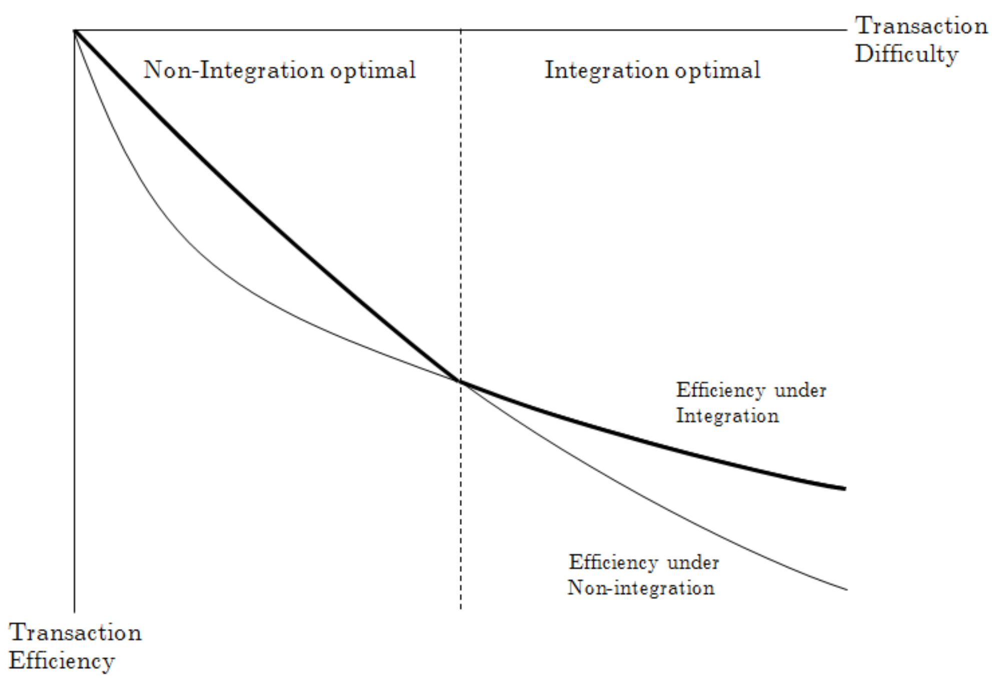
4. Vertical integration is more likely when bureaucratic costs are low.

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.

6.1.1 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

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

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.

6.2 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 $f_{ij} = h^g_{ij} = 0$ for $i \neq j$, we say that investments are **self-investments**. If $f_{ii} = h^g_{ii} = 0$, 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 $h^{UI}_{UU} = h^{NI}_{UU} = h^{RNI}_{UU}$ and $h^{DI}_{DD} = h^{NI}_{DD} = h^{RNI}_{DD}$ (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 $h^{NI}_{UU} = h^{RNI}_{UU} = h^{DI}_{UU}$ or $h^{NI}_{DD} = h^{RNI}_{DD} = h^{UI}_{DD}$ (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 $h^{DI}_{DD} = h^{UI}_{DD}$, and D's human capital is essential if $h^{UI}_{UU} = h^{DI}_{UU}$.

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 Holmström 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." Holmström 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 Holmström 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 Antràs (2003) and Acemoglu, Antràs, 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.

6.2.1 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.

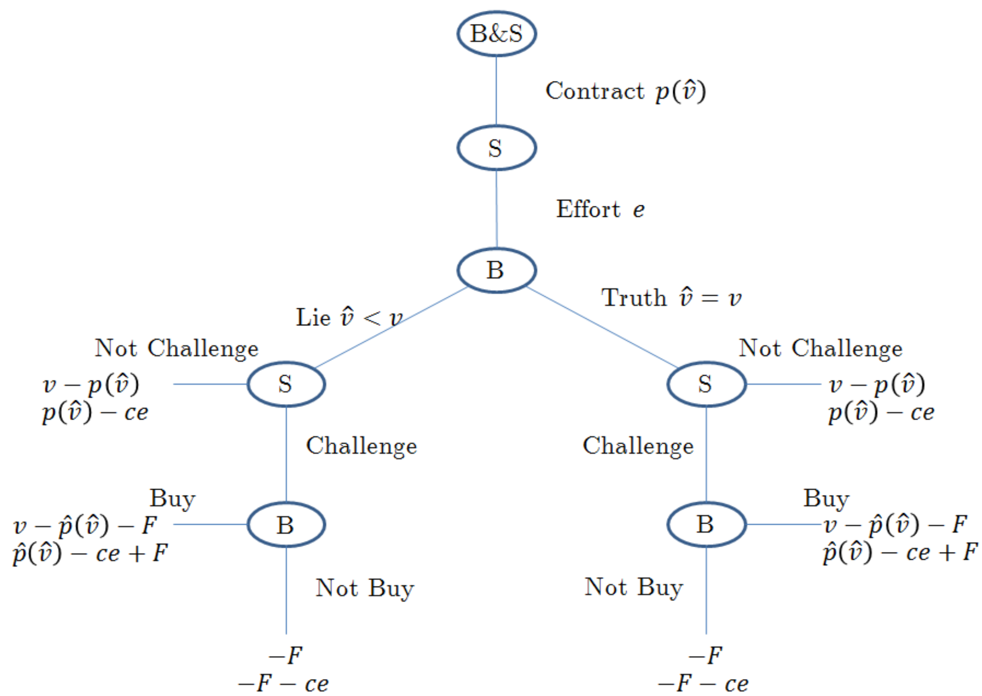
5. S can challenge B's announcement or not. The challenge decision is

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) takes 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) takes 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.

6.3 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

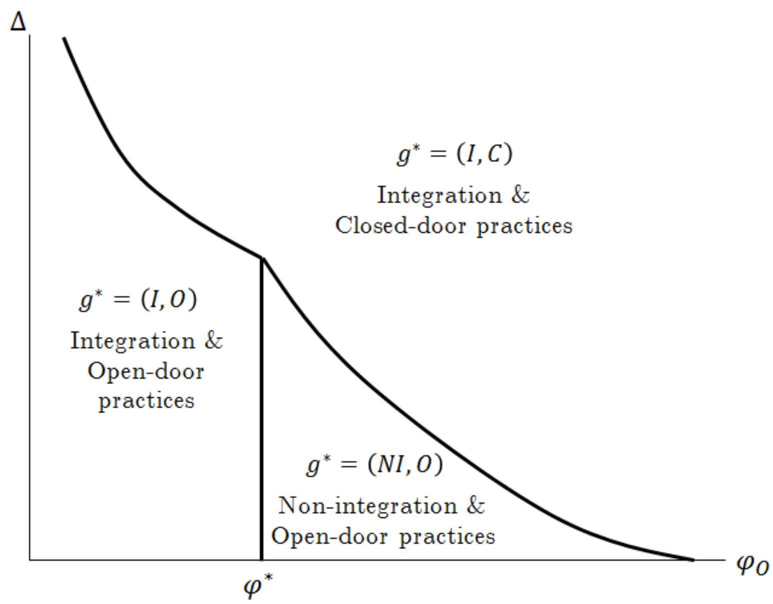
$INFL(I,O) - INFL(NI,O) > 0$, so that influence costs are minimized under non-integration. When $p = C$, so that there are no influence costs, and $\alpha_L > \alpha_R$, $ALIGN(I) < ALIGN(NI)$, so that alignment costs are minimized under integration. Unified control reduces ex post alignment costs and 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 7

Evidence on Firm Boundaries (TBA)

Part III

Dynamics

Chapter 8

Sequential Inefficiencies

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 if they do not affect the technology or options available to the firm. To borrow a phrase from Li, Matouschek, and Powell (2017), 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, Holmström, 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 (2018), "Policies in Relational Contracts."

8.1 Fudenberg, Holmström, 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 $w : \cup_{t=0}^{T} X^t \to W$ 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 $e, c : \cup_{t=0}^{T} \mathcal{H}^t \times \Theta \to \mathbb{R}$. 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, Holmström, 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.

8.1.1 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.

8.1.2 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.

8.1.3 Players have Private Information about the Future

Consider the following two-period example, which is adapted from Fudenberg, Holmström, 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 $t = 1$.

8.2 Recent Papers that Apply FHM

8.2.1 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 (2016) and Barron and Powell (2018), 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 $\pi_t = \sum_{i=}^{N} Q_{i,t} \left(v - p_t - c_{i,t}\right) Q_{i,t}$ 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**.

8.2.2 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 $q < p$.[1]

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).

[1] 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

$$E\left[b|\, e_1 = 1, y_2 = H\right] - E\left[b|\, e_1 = 1, y_2 = L\right] \geq \frac{c}{p-q}.$$

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.

8.2.3 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:

$$c\left(y^{*}\right) \leq \frac{\delta}{1-\delta}\left(S^{*}-\bar{u}-\bar{\pi}\right).$$

   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 $y_t \in \{\underline{y},\bar{y}\}$ is realized, with $\Pr[y_t = \bar{y}|\, e_t] = e_t$.

5. Payments: the fixed wage w_t is enforced by a court. The bonus b_t is

discretionary.[2]

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.

[2] 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.

Chapter 9

Dynamic Decision Making

This chapter steps back from sequential inefficiencies to consider dynamic decision-making in organizations. To that end, not all of the papers discussed in this chapter feature sequential inefficiencies or other forms of dynamics. Instead, these papers are connected by their focus on how organizations make decisions in dynamic settings. How can a principal motivate her agent to investigate new projects when she cannot commit to reward the outcome of that investigation? How do these incentives depend upon the other actions available to the principal, in particular whether she can also promise monetary bonuses? And what if the agent's private information about the correct action has an intertemporal component itself, for example because the agent wants to extend a pet project beyond its natural lifespan or end a privately costly process before it would be efficient to do so?

9.1 Authority and Commitment

When we discussed the benefits of delegation, we implicitly assumed that if the benefits of delegation outweigh the costs of delegation, then the firm will optimally choose to delegate. Formal commitment to delegating important decision rights to a subordinate is, however, not legally feasible in many settings. Instead, firms have to rely on informal promises to delegate, and delegation itself must be self-enforcing. As Baker, Gibbons, and Murphy (1999) highlight, decision rights are "loaned, not owned," and in order for it to be effective, the principal has to be able to commit not to overturn decisions that are potentially not in her favor.

The Model A principal and agent interact in periods $t = 0, 1, 2, \ldots$ and share a common discount factor $\delta < 1$. The timing is as follows.

1. The principal offers a contract $s \in \mathbb{R}$ to the agent.
2. The agent accepts or rejects. Rejecting ends the game and results in continuation payoffs of zero for both players.
3. The agent exerts effort a at cost $c(a) = \frac{c}{2}a^2$.
4. With probability a, a project with value $Y \in \{Y_L, Y_H\}$ is discovered $(x = 1)$, with probability $\Pr[Y = Y_L] = p$. $Y = 0$ if no project is discovered.
5. The agent sends a public message $m \in \{Y_L, Y_H\}$

6. The principal and agent simultaneously vote "Yes" or "No" on the project. If both vote "Yes," the project is implemented ($d = 1$); otherwise, it is not ($d = 0$).

The principal's and agent's payoffs are $\pi = xdY - s$ and $u = xdB + s$, respectively. Assume that $B > 0$ and $Y_L < 0 < Y_H$.

The principal has the formal authority in this organization; she can choose d however she likes. In a one-shot interaction, she will do so to maximize her myopic payoff, so $d = 1$ if and only if $Y = Y_H$. In the repeated game, however, the principal might informally delegate decision-making authority to the agent by choose $d = 1$ even when $Y = Y_L$. Doing so potentially increases the agent's effort, since the fruits of that effort are more likely to be implemented. However, this delegation is enforced not by a formal contract, but by an informal understanding between the two parties that is backed by the value of future interactions.

This timing intentionally leaves the observability of Y ambiguous. The agent always observes Y. The paper considers two possibilities: the principal is either (i) informed, so that she observes Y before choosing d, or (ii) uninformed, so she observes Y only after observing choosing d. In either case, the game has imperfect public monitoring, so the appropriate solution concept is Perfect Public Equilibrium (PPE).

Analysis—Informed Principal Suppose the principal observes Y before choosing d. In some circumstances, the optimal equilibrium induces her to "rubber stamp" the agent's project, even if $Y = Y_L$.

Following a deviation, let $\bar{s}$ be the total surplus produced by the punishment equilibrium. In the paper, $\bar{s}$ is assumed to be equal to the surplus produced by the repeated static Nash equilibrium, which drives some of its results. However, reversion to static Nash is not necessarily an optimal punishment, so we take a more general formulation here.

Utility is transferable, and monitoring is public, so one can show that there exists a stationary optimal PPE. Suppose that $B + Y_L > 0$, so that total surplus is maximized if $d = 1$ in each period. Let S^* be the total surplus produced in such an equilibrium. Analogous to Levin (2003), there exists an equilibrium that does so if and only if

$$-Y_L \leq \frac{\delta}{1-\delta}\left(S^* - \bar{s}\right).$$

Analysis—Uninformed Principal Now suppose that the principal observes Y after choosing d. In some circumstances, the principal might be tempted to "rubber stamp" the agent's project, even though doing so is inefficient. This is the case if $Y_L + B < 0$ and $pY_L + (1-p)\,Y_H > 0$. In that case, the agent must be deterred from voting yes on projects with $Y_L < 0$. Let S^{**} be the total surplus from such an equilibrium. An optimal stationary equilibrium with this feature exists if and only if

$$B \leq \frac{\delta}{1-\delta}\left(S^{**} - \bar{s}\right).$$

This paper presents a simple model that has a stationary optimal equilibrium, and in doing so, it highlights several important ideas that are worth keeping in mind.

First, many organizational features are best understood as informal arrangements, rather than formal contracts. The day-to-day decision-making authority within a company is frequently delegated to an employee who does not actually own the corresponding assets. In that kind of situation, it is reasonable to assume that the asset owner can revoke this delegation without formal legal repercussions.

Next, expertise and information can determine which parties are the ones at greatest risk of reneging on an agreement, and it can highlight which situations are likely to test them most severely. A deviation is most tempting if the tempted party has enough information to benefit from that decision, or if the other parties do not have enough information to immediately recognize those deviations.

Finally, formal asset ownership interacts in potentially nuanced ways with informal decision-making authority. This point is implicit if the principal owns the asset, as in the model above. It is explicitly addressed by an extension in this paper that examines divestment, which transfers formal authority from the principal to the agent. By determining who has the authority, divestment also determines who has the temptation to renege on an agreement to delegate.

9.2 Dynamic Delegation

Given our discussion of dynamics in the context of long-term formal contracts, how might the model of Baker, Gibbons, and Murphy (1999) be changed in order to investigate dynamics in informal delegation? One natural way to introduce dynamics would be to restrict transfers so that the principal could not motivate the agent using money alone. In that case, the principal may use the promise to delegate (or the threat to centralize) in the future to motivate the agent to use delegated authority wisely today. In the last few years, a few papers have explored this implication: Guo and Hörner (2017), Li, Matouschek, and Powell (2017), and Lipnowski and Ramos (2017). We will discuss as modified version of Li, Matouschek, and Powell.

Model Consider a repeated game with a principal and agent who share a common discount factor $\delta \in [0,1)$. The stage game is:

1. Each party simultaneously decides to continue or not, $a_i \in \{0,1\}$ with $i \in \{A,P\}$. Define $a \equiv a_A a_P$. If $a = 0$, the game ends and all players earn 0.

2. The principal-optimal project is either available ($\theta = 1$) or not ($\theta = 0$), with $\Pr[\theta = 1] = p$. θ is privately observed by the agent.

3. If $\theta = 1$, the agent makes a project recommendation $m \in \{P, A\}$. If $\theta = 0$, then $m = A$.

4. The principal chooses a project $d \in \{m, 0\}$, where $d = 0$ equals the default project.

The project pays off b^d and π^d to the agent and principal, respectively.

Assume that $\pi^0 = u^0 = 0$, while $\left(b^P, \pi^P\right) = \left(\pi^A, b^A\right) = (r, R)$, with $R > r > 0$, so that (i) both players prefer either project to the default, but (ii) the principal and agent disagree about the preferred project among $\{P, A\}$. The solution concept is Perfect Public Equilibrium, since the only proper subgame is the entire game because nature moves in stage 2 and its move is never observed by the principal.

One-Shot Game Consider the one-shot game. Regardless of expected m and d, there are two equilibria in stage 1: $a_A = a_P = 0$ or $a_A = a_P = 1$. In the former, payoffs equal 0 regardless of the other actions taken in equilibrium.

Suppose $a = 1$. Then $d = m$, because $r > 0$. So the agent always recommends $m = A$, since doing so yields a higher private payoff. Intuitively, it is impossible to discipline the agent to reveal $m = P$ when the principal-preferred project is available.

Repeated Game Now consider the repeated game. Might there exist an equilibrium that gives the principal an even higher payoff than repetition of the static equilibrium? Yes, so long as players are sufficiently patient. The principal can use the threat to choose $a = 0$ as a punishment following $m = A$ in order to induce the agent to recommend $m = P$ whenever it is available.

Consider the following simple equilibrium (which is not principal-optimal): in $t = 0$, the agent recommends $m = P$ if $\theta = 1$. If $m = P$, then $d = P$, and the equilibrium repeats the stage-game equilibrium with $a = 1, m = A$, and $d = A$ in every subsequent period. If $m = A$, and $d = 0$, then the equilibrium repeats the stage-game equilibrium with $a = 1, m = A$, and $d = A$ in every subsequent period. If $d = A$, then $a = 0$ in every subsequent period.

If $m = P$, then the principal will certainly choose $d = P$ because that maximizes her stage-game payoff and does not affect her continuation payoff. If $m = A$, then the principal is willing to choose $d = 0$ so long as

$$(1-\delta)\,0 + \delta r \geq (1-\delta)\,r + \delta 0$$

or $\delta \geq 1/2$. The agent is willing to choose $m = P$ whenever possible so long as

$$(1-\delta)\,r + \delta R \geq (1-\delta)\,0 + \delta R,$$

which always holds because $r > 0$. This strategy, therefore, is an equilibrium if $\delta \geq 1/2$. The principal's payoff in this equilibrium is $(1-\delta)\,pR + \delta r$, which exceeds r whenever $pR \geq r$.

Of course, this equilibrium is far from optimal. In particular, the paper constructs a more complicated set of dynamic rewards or punishments to induce the agent to recommend $m = P$ whenever possible. It turns out that the least costly way to induce truth-telling is to delay inefficient punishments for as long as possible. Therefore the principal-optimal equilibrium starts with "cooperative delegation," in which the principal chooses $d = m$, and the agent recommends $m = P$ whenever possible. Following enough periods with $m = A$, the equilibrium switches to a phase that punishes the agent by either choosing $d = 0$ or $a = 0$. After enough periods with $m = P$, the equilibrium switches to a phase that rewards the agent by inducing him to send $m = A$ in every period.

9.3 Decision Timing and Communication

This discussion presents a simplified version of the model of Grenadier, Malenko, and Malenko (2015). Consider a dynamic game between a principal and an agent.

1. The agent learns $\theta \in [0, \infty)$ according to $F(\theta)$.

2. The agent chooses a time $t \in [0, \infty)$ at which to disclose θ: $m_t = n$ or $m_t = \theta$.

3. At each time t, the principal chooses $d_t \in \{0, 1\}$. Once $d_t = 1$, the game ends.

If the game ends at time t, payoffs are

$$\begin{aligned} u &= -(t - \theta - B)^2 \\ \pi &= -(t - \theta)^2 \end{aligned}$$

for the principal and agent, respectively. Intuitively, the principal wants the game to end at θ, while the agent wants the game to end at $\theta + B$.

Suppose that $B > 0$, so that the agent wants the game to go on longer than the principal does. We construct an equilibrium in which the agent achieves his optimal stopping time, so $d_t = 1$ at $\theta + B$. Indeed, suppose that $m_t = n$ for all $t < \theta + B$. For $t < \theta + B$, the principal prefers to wait until $m_t = \theta$ as long as

$$E\left[-(t-\theta)^2 \middle| t < \theta + B\right] \leq -B^2.$$

At $B = 0$, $E\left[(t-\theta)^2 \middle| \theta > t\right] > 0$. Suppose that this expectation is uniformly bounded away from 0 for all t. Then it holds for $B > 0$ as well, under some continuity assumptions, in which case the principal strictly prefers to wait. For instance, if θ is distributed according to an exponential distribution, then this expectation is constant in t. In such circumstances, the agent's optimal decision rule is implemented.

Suppose that $B < 0$, so that the agent wants the game to end early. Then there cannot exist an equilibrium in which $t = \theta + B$ for all θ. Indeed, suppose that such an equilibrium existed. Then the principal must know θ at time $d_t = 1$. But then the principal will choose to end at $t' = \theta > \theta + B$. Indeed, if $\theta \in \{\theta_1, \ldots, \theta_N\}$ is drawn from a finite set, and $|B|$ is not too large, then we can show that the principal will stop at $t = \theta$ in equilibrium, so that the policy is the principal's optimal policy. This result follows from a version of an unraveling argument. Informally, suppose $\theta_k < \theta_{k+1}$, and assume that for $k < N$, the θ_k-type agent strictly prefers to separate herself rather than to be lumped together with θ_{k+1} (which will hold if B is small). Then the θ_1 type strictly prefers to reveal herself rather than pool with any other types; but then the θ_2 type strictly prefers to reveal herself; and so on.

Note that this result is asymmetric in the sign of the agent's bias. If the agent prefers late termination, then he can attain that goal under some conditions. But the agent's preferred policy can never be implemented if he prefers early termination. This asymmetry is due to the irreversibility of time: the principal can delay implementing a project, but she cannot go back in time to implement that project early.

The Dynamic Cheap-Talk Model Grenadier, Malenko, and Malenko (2015) considers a more complicated model than the one specified above. In particular, they assume: (*i*) cheap talk, rather than verifiable information; (*ii*) the optimal stopping time is determined by an option price that evolves according to a Markov process; and (*iii*) the type θ has a finite upper bound, $\bar{\theta}$. Consequently, the model does not feature full revelation of the state if $B > 0$ because $E\left[(t-\theta)^2 \middle| \theta > t\right] \to 0$ as $t \to \bar{\theta}$. Instead, the paper gives conditions under which the equilibrium replicates the principal's optimal mechanism that she would choose if she could commit to a stopping rule. However, the basic economics are the same as in the simple example above: the agent can wield "real authority" if $B > 0$ because she can delay revelation until $t = \theta + B$, at which point the principal cannot go back in time to implement her preferred stopping time.

9.4 Starting Small

Watson (1999) explores how players can separate potential cooperators from defectors in a repeated prisoner's dilemma. It does so by changing the stakes of the relationship, which determines both the gains from cooperation and the surplus lost by a unilateral defection.

Model Consider a continuous-time game with two players, denoted 1 and 2. At the beginning of the game, the players jointly select a level function $\alpha : \mathbb{R}_+ \rightarrow [0, 1]$ that determines the gains from cooperation at each moment $t \in \mathbb{R}_+$. How α is selected does not really matter for the analysis, which focuses on incentives to cooperate for different α. After choosing α, the players privately learn their types $\theta_i \in \{L, H\}$, with $\Pr[\theta_i = H] = p_i$.

At each $t \in \mathbb{R}_+$, agents decide to either cooperate or to take a selfish action. The game ends as soon as a player takes the selfish action. If both players cooperate, type θ earns a flow benefit $z_\theta : [0, 1] \rightarrow \mathbb{R}_+$ that depends on $\alpha(t)$. If one player defects, that player earns a lump-sum payment $x_\theta : [0, 1] \rightarrow \mathbb{R}_+$, while the other suffers a lump-sum cost of $y_\theta : [0, 1] \rightarrow \mathbb{R}_-$. If both players defect at the same time, they both earn 0. Therefore, if player

i defects at time t and the other player cooperates, then i's payoff equals

$$\int_0^t z_{\theta_i}(\alpha(s))\, e^{-rs} ds + e^{-rt} x_{\theta_i}(\alpha(t)),$$

with a similar expression (replacing x with y) if the other player defects, and i cooperates at t. Assume that:

1. All functions are "smooth enough,"

2. $z_\theta(0) = x_\theta(0) = y_\theta(0) = 0$

3. z_θ and x_θ are strictly increasing, and y_θ is strictly decreasing in α.

4. $x_L(\alpha) > z_L(\alpha)/r$ and $x_H(\alpha) < z_H(\alpha)/r$ for all $\alpha \in (0, 1]$.

Assumptions 1 and 2 are pretty innocuous. Assumption 3 implies that an increase in α increases the value from cooperating (z_θ) but also increase the temptation to defect x_θ and the cost of such a defection y_θ. Assumption 4 ensures that low types eventually want to defect, while high types are willing to cooperate with other high types. It basically says that, for fixed α, and assume that the other player cooperates forever, then low types prefer defecting to cooperating forever, while high types prefer cooperating forever.

Intuition: Level Functions Since low and high types have different optimal responses to cooperation, there is value to learning a player's type so that high types can cooperate with other high types. α provides a potential instrument to separate types, but any such attempt must overcome two challenges. First, if a high type expects her opponent to defect with high probability (e.g., because he is likely to be a low type), then her best response is to defect in order to protect herself from the penalty y_θ. Since this penalty is increasing in α, a high-type has a stronger incentive to "insure" herself in this way if α is high. Second, if $\alpha(t)$ varies over time, then low types might have the incentive to imitate high types for a while in order to maximize their betrayal gain x_L, which is increasing α. Any $\alpha(t)$ that separates the types must overcome these two obstacles.

More concretely, let us try to find a function $\alpha(t)$ such that high-type players are willing to cooperate forever, even though low-type players eventually defect. This is trivial if $\alpha(t) = 0$, so let us consider some positive stakes.

Constant level function Consider $\alpha(t) = a$ for all t and some $a \in (0, 1]$. Then low types defect immediately. High types are willing to cooperate if

$$p\frac{z_H(a)}{r} + (1-p)\, y_H(a) \geq p x_H(a) + (1-p)\, 0$$

or

$$\frac{z_H(a)}{r} - x_H(a) \geq -\frac{1-p}{p} y_H(a).$$

Since $y_H(a) < 0$ for any $a > 0$, given $a > 0$, this condition cannot hold if p is sufficiently small. That is, if the proportion of high types in the population is low, then even high types want to defect in order to save themselves the penalty $y_H(a)$.

This argument suggests that $\alpha(t)$ must be dynamic in order to induce high types to cooperate (or at least cooperate at high levels) if p is not too large.

$\alpha(t)$ jumps at $t = 0$ Let us modify the constant α slightly so that $\alpha(0) = a_L$ and $\alpha(t) = a_H$ for some $a_L \in [0,1]$ and $a_H \in (0,1]$. Assume that low types defect at $t = 0$. Then high types are willing to cooperate if

$$p\frac{z_H(a_H)}{r} + (1-p)\, y_H(a_L) \geq p x_H(a_L) + (1-p)\, 0$$

or

$$\frac{z_H(a_H)}{r} - x_H(a_L) \geq -\frac{1-p}{p} y_H(a_L).$$

Unlike the previous example, it is easy to find an $a_L < a_H$ that satisfy this expression, since $x_H(0) = y_H(0) = 0$ and $z_H(a) > 0$ for any $a > 0$. However, we must check that the agent prefers to defect at $t = 0$ rather than waiting a single instant and defecting at $t = \varepsilon$, for instance. But taking $\varepsilon \to 0$ yields the equation

$$p x_L(a_L) \geq p x_L(a_H) + (1-p)\, y_L(a_L).$$

Set $a_H = 1$, and suppose that $y_H(a) = a\bar{y}$, $y_L(a) = \underline{a}$, with similar expressions for x_θ and z_θ. Then we can write these expressions as

$$a_H \geq \frac{r}{\bar{z}} \left(\bar{x} - \frac{1-p}{p} \bar{y} \right) a_L$$

and

$$\left(1-\frac{1-p}{p}\frac{\underline{y}}{\underline{x}}\right)a_L \geq a_H.$$

For such an a_H to exist, we require that

$$\frac{\bar{z}}{r}-\bar{x} \geq \frac{1-p}{p}\left(\frac{\bar{z}\underline{y}}{r\underline{x}}-\bar{y}\right).$$

If $\bar{z}\underline{y}/\left(r\underline{x}\right)-\bar{y}>0$ or $\bar{z}/r<\underline{y}\bar{y}/\underline{x}$, then this expression cannot hold for $p\to 0$.

Together, the previous two level functions illustrate the incentive constraints that must be satisfied to sustain cooperation among high types. To simultaneously satisfy them (for p small and fixed steady-state cooperation level $a>0$) requires α to change over time.

$\alpha(t)$ jumps at $T>0$ Suppose that $\alpha(t)=a_L$ for $t<T$ and a_H thereafter, and consider a strategy profile in which high types cooperate forever and low types immediately quit. We will consider the incentive constraints required to make this strategy an equilibrium, without taking the final step of showing that such an equilibrium exists (indeed, it need not).

For high types to cooperate forever,

$$p\left(\int_0^T e^{-rt}z_H\left(a_L\right)dt+\frac{e^{-rT}}{r}z_H\left(a_H\right)\right)+\left(1-p\right)y_H\left(a_L\right)\geq px_H\left(a_L\right).$$

Low types will either defect immediately or wait until T and then defect.

They prefer to defect immediately if

$$px_L(a_L) \geq \int_0^T e^{-rt} z_L(a_L)\, dt + e^{-rT} x_L(a_H).$$

More generally, one can imagine a level function $\alpha(\cdot)$ that increases smoothly over time, such that the low types are indifferent between continuing and quitting. Then we have an additional choice variable in equilibrium: the rate at which low types defect. We can use this choice variable to control the risk associated with cooperating. The paper shows that we can find regimes $\alpha(\cdot)$ that induce equilibria of this flavor, regardless of ex ante beliefs p.

Renegotiation Proofness In the level function that jumps at $T > 0$, the low types quit immediately. However, the high types continue to cooperate at a low level a_L for a while, even though cooperating at a higher level would be possible (because only high types are left in the game, and high types are willing to cooperate at any fixed level). That is, such a renegotiation profile seems like it is not "renegotiation-proof," in the sense that both players would prefer to jump ahead to $\alpha(t) = a_H$ as soon as the low types drop out. That is, the strategy is not sequentially efficient.

The paper proposes a renegotiation protocol and analyzes how it constrains the set of level functions $\alpha(t)$ that are consistent with equilibrium. Doing so is somewhat delicate in this setting, since both players have private information about their own types. Therefore, the renegotiation protocol focuses on two potential "renegotiations" of the level function α: jumps—setting α "as if" play is at some future point, or delays—playing some "safe" level of α (at which everyone cooperates) for some period of time and then resuming according to α. Moreover, a player can unilaterally delay, but both players must agree to a jump.

Restricting attention to equilibria that are renegotiation-proof to such "jumps" and "delays," Watson (1999) shows that there exists a unique level function $\alpha(t)$ consistent with an equilibrium in which high types perpetually cooperate. This level function increases gradually over time and hits $\alpha(t) = 1$ exactly when it is common knowledge that both players have $\theta = H$.

Note that this renegotiation protocol does not quite select sequentially efficient level functions. First, this renegotiation is entire internal: the parties cannot recommend changing th level function to one that is extremely different. Note, however, that this protocol already selects a unique level function, even using this weaker criterion. So for any stronger renegotiation criterion, either no level function satisfies that stronger condition, or it selects exactly this level function.

This renegotiation protocol potentially rules out some Pareto efficient level functions, since equilibria that satisfy these conditions must prevent a single player from unilaterally forcing a delay (even if that delay is not Pareto improving). So in principle, other Pareto efficient level functions exist, and players might do better by jumping between these Pareto efficient profiles in a sequentially efficient way. This is probably not a concern if the players are symmetric, however. (In that case, it is not clear how unilateral delay and Pareto-improving delay differ, since players have symmetric payoffs in the proposed equilibrium). Of course, it is also possible that a better equilibrium exists in which high types defect with positive probability, though it is not at all clear why such behavior would lead to a Pareto improvement.

9.5 Building Routines

Chassang (2010) seeks to explain the emergence and persistence of organizational routines: systematic ways of behaving that are in some sense independent of changing economic conditions. In particular, it argues that limited transfers and learning can lead agents to coordinate on an inefficient routine, even if it is common knowledge that a better routine exists. Perhaps unsurprisingly given our previous discussions, the inefficiency of this routine depends on past performance, which leads to productivity dispersion in equilibrium among multiple firms who each play the same game. This note presents a simplified version of the paper's model and so sidesteps some of its important economic features.

Model Consider a principal and an agent who interact repeatedly, with common discount factor δ. In each period t, there is a publicly observed state variable $R_t \in \left\{R^1, \ldots, R^K\right\}$ that starts at $R_0 = R^1 = 0$ and such that $R^1 < R^2 < \cdots < R^K$. The timing is as follows:

1. The principal and agent simultaneously choose whether to exit or not. For simplicity, assume exit is permanent and yields payoff 0 for both parties.

2. If neither player exits, then the principal chooses $d_t \in \{0,1\}$, where $d_t = 0$ means "routine" and $d_t = 1$ means "experiment."

3. If $d_t = 1$, the agent chooses private effort $e_t \in \{0,1\}$.

4. R_{t+1} is realized, with $\Pr[R_{t+1} = R_t] = 1 - pd_te_t$. Otherwise, if $R_t = R^k$, then $R_{t+1} = R^{k+1}$ for $k+1 \leq K$, and otherwise $R_{t+1} = R^K$.

The value of choosing $d_t = 0$ depends on past experimentation. In particular, let $R_t = \max\{R_{t-1}, py_{t-1}\}$, with $R_{-1} \equiv 0$. Conditional on neither player exiting, payoffs for the principal and agent are

$$\begin{aligned} \pi_t &= (1-d_t) R_t \\ u_t &= (1-d_te_t) b, \end{aligned}$$

respectively, with corresponding continuation payoffs $\Pi_t = \sum_{t'=t}^{\infty} \delta^{t'-t} (1-\delta) \pi_t$ and $U_t = \sum_{t'=t}^{\infty} \delta^{t'} (1-\delta) u_t$. The solution concept is perfect public equilibrium. Let h_0^t denote a history at the start of period t and h_x^t a history following variable x in period t.

Analysis Consider the agent's incentive to exert effort, conditional on $d_t = 1$. He will do so if

$$\frac{\delta}{1-\delta}\left(E\left[U_{t+1}\middle| h_d^t, y_t > 0\right] - E\left[U_{t+1}\middle| h_d^t, y_t = 0\right]\right) \geq \frac{b}{p}.$$

Suppose that exit never occurs on the equilibrium path. Then $E\left[U_{t+1}\middle| h_d^t, y_t\right] = b$ regardless of y_t, so this incentive constraint cannot hold. Therefore, to motivate the agent to exert effort, continuation play must be inefficient following low output. This is our standard "upward-sloping payoff frontier" source of sequential inefficiencies.

The key new ingredient is that exerting effort affects the value of future routines. In effect, by exerting effort, the agent can find a "better routine" that can then be replicated in future periods without any incentive cost. Intuitively, we can think of the incentive cost as figuring out how to make a routine work, after which it is easy to implement. The paper itself microfounds and justifies routines in a much more granular and systematic way.

For simplicity, assume $R_t \in \{0, b, b+\Delta\}$ for some $\Delta > 0$, and assume that exit happens immediately following low output. If $R_t = b + \Delta$, then $d_t = 0$ in every subsequent period, and the principal and agent earn $b + \Delta$ and b, respectively. If $R_t = b$, then the principal can either experiment or not. If she experiments, her continuation payoff equals

$$p\frac{\delta}{1-\delta}(b+\Delta) + (1-p)\,\delta\,(1-q)\,\Pi^* = \Pi^*,$$

where $q > 0$ captures the probability of exit following low output. The agent's payoff from always shirking equals

$$\frac{b}{1 - \delta(1 - q)} = U_S,$$

while his payoff from working hard in each period until $R_t = b + \Delta$ equals

$$p\frac{\delta}{1 - \delta}b + (1 - p)(1 - q)\delta U^* = U^*$$

or

$$U^* = pb\frac{\delta}{(1 - \delta)(1 - (1 - p)(1 - q)\delta)}.$$

The incentive constraint therefore requires that $U_S \leq U^*$. If $q = 0$, then this inequality can never be satisfied. So this expression bounds $q > 0$ from below, and independently of Δ. Hence, the principal will experiment only if the payoff from doing so exceeds her payoff from adopting the routine:

$$\Pi^* = p\frac{\delta}{1 - \delta}\frac{b + \Delta}{1 - (1 - p)\delta(1 - q)} \geq b.$$

The larger is q, the harder this constraint is to satisfy, and hence the more likely is the principal to stick with an inefficient routine $R_t = b$ rather than experimenting to find a more efficient alternative.

This gets us to an inefficiency relative to the first best (and a sequential inefficiency due to the possibility of exit), and it gets us part of the way to productivity dispersion (because some ex ante identical firms might exit, while others find a successful routine). But among the firms that remain, all end up using the same routine. However, this result is an artifact of the simplifying assumption that R_t moves upward in a predictable way. Suppose that when $R_t = 0$, there is some probability that it jumps immediately to $R_{t+1} = b+\Delta$ after the agent works hard. Then there are parameters for which (i) the principal experiments while $R_t = 0$, but (ii) she stops experimenting for any $R_t > 0$. In other words, some principal-agent pairs will stumble upon an efficient routine $R_t = b + \Delta$ by luck, while others will find $R_t = b$ and not find it worthwhile to experiment further.

9.6 Delegating Experimentation

Consider a setting that is similar to that of Grenadier, Malenko, and Malenko (2015): a sender with private information about an optimal stopping time, but biased preferences, communicates it. But suppose that the receiver (i) learns about the true state of the world as the game progresses, and (ii) can commit to a mechanism that maps the sender's report to her chosen stopping time. In this setting, the revelation problem is effectively static by the revelation principle; in particular, it is without loss for the sender to truthfully reveal her private information at the start of the game. However, the receiver has a complex set of tools to induce truth-telling, since she must use a stopping time alone to induce truth-telling, rather than using both a decision rule and transfers.

Roughly, this is the intuition behind Guo (2016). The model presented in these notes captures only a tiny portion of the insights in the paper—in particular, Guo develops some new methodology to tackle a much more difficult problem than the simplified version presented here. But this model captures the insight that project termination can be used to induce truth-telling.

Model Consider a dynamic game between a principal and an agent with the following timing:

1. At the start of the game, a state $\theta \in \{L, H\}$ is realized and privately observed by the sender, with $\Pr[\theta = H] = p$.

2. The receiver offers a mechanism, which is a mapping $T : \{L, H\} \to \mathbb{R}_+$ that designates a stopping time $T(m)$ for each report m.

3. The sender chooses a report $m \in \{L, H\}$ and the project is stopped at $T = T(m)$.

The principal's and agent's payoffs are $\pi = -(T-\theta)^2$ and $u = -(T-\theta-B)^2$, respectively. We will assume $B > 0$, so that the agent prefers a later stopping time than the principal.

Before we analyze this model, a few notes on the differences between it and the model in the paper. In Guo's (2015) model, there is an unknown parameter that determines the arrival rate of a Poisson process. The receiver can commit to a mechanism that splits a resource between a safe and a risky asset in each moment in time. The safe asset has a fixed flow payoff, while the risky asset pays off whenever the event occurs. The sender has private information about the probability that the risky project is better than the safe project (because it has a high Poisson arrival rate), but prefers more experimentation than the sender. In this setting, the receiver has a much more complex set of tools available to induce truth-telling. For instance, she could delay experimentation (by forcing investment in the safe asset for some period of time), induce partial experimentation (by putting only part of the resource into the risky project), or require less-than-optimal investment even after the risky project has shown itself to be successful.

Analysis The sender must have the incentive to truthfully reveal θ, which requires that

$$\begin{aligned} (T(H) - H - B)^2 &\leq (T(L) - H - B)^2 \\ (T(L) - L - B)^2 &\leq (T(H) - L - B)^2 \end{aligned}$$

or

$$\begin{aligned} 0 &\leq (T(L) - T(H))\left[T(L) + T(H) - 2(H + B)\right] \\ 0 &\leq (T(H) - T(L))\left[(T(L) + T(H) - 2(L + B)\right]. \end{aligned}$$

If $H > L$, it follows immediately from these expressions that $T(L) \leq T(H)$. Therefore, either $T(L) = T(H)$, or

$$2(L + B) \leq T(L) + T(H) \leq 2(H + B).$$

The receiver's objective is to maximize her expected payoff

$$\min_{T(L),T(H)} (1 - p)(L - T(L))^2 + p(H - T(H))^2$$

subject to this constraint.

Suppose that $2(L + B) \leq L + H \leq 2(H + B)$, which is true if $H - L \geq 2B$. Then the receiver can implement her ideal stopping times for both types. Otherwise, the binding constraint will be that the low-type sender will want to report $\theta = H$: $2(L + B) = T(L) + T(H)$. Therefore, the receiver solves

$$\min_{T(H)} (1 - p)(-L - 2B + T(H))^2 + p(H - T(H))^2$$

which has first-order condition

$$2(1 - p)(-L - 2B + T(H)) - 2p(H - T(H)) = 0,$$

or

$$T(H) = (1 - p)(L + 2B) + pH > H.$$

Finally,

$$T(L) = 2(L + B) - (1 - p)(L + 2B) - pH$$

or

$$T(L) = L + p(2B - (H - L)) > L.$$

So we force both types to experiment too much, from the receiver's perspective. From the sender's perspective, the high type is required to experiment "too little," while the low type might be asked to experiment "too much" or "too little."

The model presented above has been simplified so that it can be compared more easily to the simplified version of Grenadier, Malenko, and Malenko (2015). In particular, notice that the complementary insights of the two papers highlight the economic differences between a setting with commitment and without commitment. In Grenadier, Malenko, and Malenko (2015), the revelation problem itself is dynamic because there is no commitment (and hence no Revelation Principle). This feature highlights the flow of time as a partial commitment device. In contrast, Guo (2016) assumes commitment, which means that the incentive problem is essentially static—the sender reveals her true type immediately, and the dynamics are used to make sure that she has the incentive to do so (at minimal cost to the principal). That is, unlike a static setting with transferable utility, the sender and receiver have a non-linear payoff frontier in this problem.

The paper itself does substantially more to explore the dynamic component of the problem. For example, in its model, the optimal stopping time mechanism can be implemented via delegation: the sender is allowed to choose a resource allocation for some period of time, after which the receiver takes over and invests in the safe asset. And it shows that this delegation mechanism can be made time-consistent—think "sequentially efficient given the receiver's beliefs"—by delegating without asking the sender to report her type.

9.7 Teams and Project Dynamics

This section considers what is referred to as "dynamic public goods games" in which a group of two or more agents make costly contributions over time to a joint project.[1] We will focus on particular types of settings in which (*i*) progress is gradual (i.e., each contribution only gets the project a step closer to completion) and (*ii*) the public good is discrete and generates a lump-sum payoff only once the cumulative contributions reach a threshold. We will follow Kessing's (2007) and Georgiadis's (2015) continuous-time modeling approach, since its tractability allows us to address various organizational questions.

The Model Time $t \in [0, \infty)$ is continuous. There are $N \geq 2$ identical agents who are risk-neutral, credit-constrained, and discount the future according to discount rate $r > 0$. Each agent has an outside option yielding

[1] George Georgiadis contributed greatly to writing this section.

utility 0. There is a project that begins at state $q_0 = 0$, and at every moment t, each agent i privately chooses effort $e_{i,t} \geq 0$ in order to influence the project's state according to the process

$$dq_t = \left(\sum_{i=1}^{N} e_{i,t} \right) dt,$$

where q_t denotes the state of the project at time t, which is commonly observed.[2] Each agent's flow effort costs are quadratic ($c(e) = e^2/2$). The project is completed at the first time τ such that $q_\tau = Q$, where $Q > 0$ is the **completion threshold**. If $\tau < \infty$, then each agent receives an exogenously specified reward V_N upon completion and receives no reward otherwise.

For a given set of effort profiles $\{e_{i,t}\}_{i,t}$, the discounted continuation payoff for agent i at time t satisfies

$$\Pi_{i,t} = \underbrace{e^{-r(\tau-t)} V_N 1_{\{\tau<\infty\}}}_{\text{discounted completion payoff}} - \underbrace{\int_t^\tau e^{-r(s-t)} \frac{e_{i,s}^2}{2} ds}_{\text{discounted costs}},$$

where τ denotes the **completion time** of the project. The first term captures agent i's discounted payoff from the project's completion, while the second term captures her discounted costs.

Agents' strategies are mappings from their past effort choices, the project's state, and time t to their time-t effort choice $e_{i,t}$, but we will be restricting

[2] An equivalent way of writing the project's state is, of course, $q_t = \int \left(\sum_{i=1}^{N} e_{i,t} \right) dt$, but expressing the evolution of the state will make it easier to extend the analysis to allow for noise.

the class of equilibria we will focus on. In particular, the solution concept we will be using is **Markov Perfect equilibrium** (MPE) with symmetric and differentiable strategies.

A couple comments on this choice of solution concept. First, this restriction ensures that we will be focusing on equilibria in which agents' effort strategies are public in the sense that agents do not condition their effort choices at moment t on their past effort choices, so this solution concept is in some sense a refinement of Perfect Public equilibrium (PPE). Second, agents' effort choices at moment t will depend only on the project's current state q_t. This is a real restriction—unlike in a PPE, agents' continuation payoffs cannot be used as a state variable on which they can condition continuation play. In other words, moment-t play can depend only on stage-game-payoff-relevant variables. Coupled with the restriction to strategies that are differentiable with respect to time, the solution concept rules out equilibria in which the agents can exploit the fact that the project progresses deterministically to implement non-Markov equilibria in which continuation play can be conditioned on the history in a way to ensure first-best outcomes are attainable. We will discuss this restriction further below.

Analysis We begin by characterizing the first-best outcome, which would be attained if an organizational designer could directly choose each agent's action to maximize total surplus:

$$TS = N\left[e^{-r\tau}V_N 1_{\{\tau<\infty\}} - \int_t^{\tau} e^{-r(s-t)}\frac{e_{i,s}^2}{2}ds\right],$$

where $\{e_{i,s}\}_{s\in[0,\infty)}$ is agent i's effort profile. Because each agent's costs are convex, it is without loss of generality to restrict attention to symmetric effort profiles when characterizing first-best outcomes.

This problem can be rewritten as follows:

$$\max_{\tilde{\tau}<\infty}\left\{NV_N e^{-r\tilde{\tau}} - \min_{e_t}\left\{\int_0^{\tilde{\tau}} e^{-rs}\frac{Ne_s^2}{2}ds \text{ s.t. } \int_0^{\tilde{\tau}} Ne_s ds = Q\right\}\right\},$$

and we can therefore solve the problem in two steps. First, given a completion time $\tilde{\tau}$, we can find the effort profile that minimizes total costs subject to completing the project at some fixed time $\tilde{\tau}$. Second, we can find the completion time τ^{FB} that maximizes the total discounted payoffs net of effort costs.

For the first step, fix an arbitrary completion time $\tilde{\tau} < \infty$, and solve the first-stage problem. The Lagrangian for this problem is

$$\mathcal{L}(\lambda) = \min_{e_s}\left\{N\int_0^{\tilde{\tau}}\left(e^{-rs}\frac{e_s^2}{2} - \lambda e_s\right)ds + \lambda Q\right\},$$

where λ is the Lagrange multiplier on the constraint that specifies that the cumulative effort until $\tilde{\tau}$ equals the completion threshold, Q. Since the objective is strictly convex, the solution will satisfy the first-order condition

$$e_s = \lambda e^{rs}.$$

Under the first-best outcome, each agent's discounted marginal cost $(e^{-rs} \cdot e_s)$ is constant—due to the convexity of the agents' costs, it is optimal to smooth their efforts over time.

The optimal multiplier λ^* can be pinned down by substituting $e_s = \lambda e^{rs}$ into the constraint to obtain $\lambda^* = rQ\left(e^{r\tilde{\tau}} - 1\right)^{-1}/N$, and therefore optimal effort given completion time $\tilde{\tau}$ satisfies

$$e_s = \frac{rQ}{N}\frac{e^{rs}}{e^{r\tilde{\tau}} - 1}.$$

For the second step of the problem, we can substitute the optimal effort expression into the objective, in which case the problem of choosing an optimal completion time can be written

$$\max_{\tilde{\tau}<\infty}\left\{NV_N e^{-r\tilde{\tau}} - \frac{rQ^2}{2N}\frac{1}{e^{r\tilde{\tau}} - 1}\right\},$$

and the first-best completion time satisfies $\tau^{FB} = \min\left\{\infty, \bar{\tau}^{FB}\right\}$, where $\bar{\tau}^{FB} = -\ln\left(1 - \sqrt{\frac{rQ^2}{2V_N N^2}}\right)/r$. It is efficient to complete the project if and only if $rQ^2 < 2V_N N^2$. If this condition is not satisfied, the agents are collectively better off abandoning the project and collecting their outside options.

Because the project progresses deterministically, we can write the first-best effort profile as a function of the project's progress, rather than as a function of time:

$$e^{FB}(q) = \frac{r}{N}\left(\max\left\{q - Q + \sqrt{2V_N N^2/r}, 0\right\}\right).$$

First-best effort increases as the project gets closer to completion, and the first-best effort schedule is everywhere higher for a project that is more valuable for the agents to complete.

Markov Perfect Equilibria We will now characterize the MPE. The restriction to MPE implies that effort profiles may depend only on payoff-relevant variables, which rules out agents using contingent continuation play as punishments.

Stage-game payoffs depend only on the state of the project, q, and not directly on time t. The restriction to MPE, therefore, restricts effort profiles to depend only on q (see, for example, Maskin and Tirole (2001)), and so the time subscript can be dropped.

As in discrete-time problems, we can write each agent's discounted payoffs at time t recursively. In particular, the Hamilton-Jacobi-Bellman (HJB) equation for agent i's discounted payoff is

$$r\Pi_i(q) = \max_{e_i \geq 0}\left\{-\frac{e_i^2}{2} + \left(\sum_{j=1}^{N} e_j\right)\Pi_i'(q)\right\},$$

where $e_i : [0, Q] \to \mathbb{R}_+$ is agent i's effort profile, subject to the boundary conditions

$$\Pi_i(q) \geq 0 \text{ and } \Pi_i(Q) = V_N.$$

The HJB equation can be interpreted as follows. At the optimal effort profile, each agent's optimal payoffs, $r\Pi_i(q)$, must be equal to her flow benefit of bringing the project closer to completion, $\left(\sum_{j=1}^{N} e_j\right) \Pi_i'(q)$, minus her flow costs, $e_i^2/2$. The first boundary condition requires that at every project state, each agent's discounted payoff must be nonnegative—otherwise, she can choose $e_i = 0$ in all future periods and guarantee herself a payoff of zero. The second condition reflects the fact that upon completing the project, each agent receivers her reward, and the game ends.

In a MPE, at every moment t, each agent i observes the state of the project q and chooses her effort e_i to maximize her discounted payoff, while accounting for the effort profiles of all the other agents. Assuming an interior solution to the problem, which we will verify later, agent i's first-order condition requires that $e_i(q) = \Pi_i'(q)$: at every moment, she chooses her effort so that her marginal costs equal the marginal benefit associated with bringing the project closer to completion. In this problem, the second-order conditions are satisfied, so the first-order conditions are necessary and sufficient for characterizing each agent's optimal choices. Since we are restricting attention to symmetric MPEs, it therefore follows that the discounted payoff for each agent satisfies

$$r\Pi(q) = \frac{2N-1}{2}\left[\Pi'(q)\right]^2,$$

subject to the boundary conditions.

These optimal payoffs therefore are characterized by a first-order nonlinear differentiable equation, which in this case can be solved analytically. To do so, we will use the guess-and-verify method, guessing a solution of the form $\Pi(q) = A + Bq + \Gamma q^2$. Doing so yields a solution of the form

$$\Pi(q) = \frac{r}{2(2N-1)}\left(\max\left\{q - Q + \sqrt{\frac{2V_N}{r}\frac{2N-1}{N}}, 0\right\}\right)^2,$$

so as long as $rQ^2 < 2V_N(2N-1)/N$, there exists a MPE in which the project is completed in finite time. Note that $\Pi'(q) \geq 0$ for all q, and so the solution to each agent's problem is indeed interior as we assumed above. Using the first-order condition $e(q) = \Pi'(q)$, it follows that on the equilibrium path, each agent chooses

$$e(q) = \frac{r}{2N-1}\max\left\{q - Q + \sqrt{\frac{2V_N}{r}\frac{2N-1}{N}}, 0\right\}.$$

Notice that if $rQ^2 \geq 2V_N(2N-1)/N$, then $e(0) = 0$, so no agent ever chooses $e_i > 0$, and the project is not completed in equilibrium. In general, there need not exist a unique MPE. In particular, if $rQ^2 \in (2V_N, 2V_N(2N-1)/N)$, then there exists another MPE in which no agent ever chooses $e_i > 0$, and the project is not completed. To see why such an MPE exists, suppose that all agents $j \neq i$ choose $e_j(0) = 0$. Then agent i finds it optimal to also choose $e_i(0) = 0$ because she is not willing to undertake the entire project single-handedly (since $rQ^2 > 2V_N$).

A first observation is that $e(q)$ is also increasing in q, that is, the agents choose higher effort the closer the project is to completion. This is due to the fact that agents are impatient, and costs are incurred at the time effort is chosen, while rewards only occur when the project is completed. As a result, their incentives are stronger, the closer the project is to completion. An implication of this observation, which was first made by Yildirim (2006) and Kessing (2007), is that efforts are strategic complements (across time) in this game. That is because, by raising her effort, an agent brings the project closer to completion, which induces the other agents to raise their future efforts.

Exercise. Show that in the MPE characterized above, at every project state, each agent chooses a strictly lower effort compared to the first-best outcome, and the project has a later completion date.

We have assumed that the project progresses deterministically for the sake of simplicity. One can extend the analysis to the case in which the project progresses stochastically, for example, according to $dq_t = \left(\sum_{i=1}^{N} e_{i,t}\right) dt + dB_t$, where B_t is a standard Brownian motion with $B_0 = 0$. In that case, in an MPE, each agent's discounted payoff is characterized by a second-order differential equation, which does not admit a closed-form solution, but one can show that the main results and comparative statics continue to hold. See Georgiadis (2015) for details.

This characterization of MPE lends itself to particularly straightforward comparative statics. First, $e(q)$ is increasing in V_N. If agents receive a larger reward upon completion, their incentives are stronger. Second, there is a threshold $\hat{q}$ such that $e(q)$ is increasing in r if and only if $q \geq \hat{q}$. As agents become less patient, they choose higher effort levels when the project is close to completion and lower effort levels when it is far from completion. This result follows from the fact that the marginal benefit of exerting more effort to bring the completion time forward is $-\frac{d}{d\tau}e^{-r\tau}V_N = rV_N e^{-r\tau}$, which is increasing in r if and only if τ is small (i.e., if the project is sufficiently close to completion).

Next, we can compare the effect of team size on agents' incentives. Let us compare MPEs in two settings: one in which there is a team with N agents and one in which there is a team with $M > N$ agents, holding fixed the total reward of the project, so that each agent receives $V_N = V/N$ in the first setting and $V_M = V/M$ in the second setting. Then there are two thresholds $\tilde{q}, \bar{q}$ such that:

1. $a_M(q) \geq q_N(q)$ if and only if $q \leq \tilde{q}$; and

2. $Ma_M(q) \geq Na_N(q)$ if and only if $q \leq \bar{q}$.

By increasing the size of the group, two opposing forces influence the agents' incentives. First, the standard *free-rider effect* becomes stronger.

Moreover, because the agents' effort costs are convex, this distortion intensifies as the project progresses. Second, because efforts are strategic complements, when an agent is part of a larger group, she has incentives to exert higher effort, because doing so will induce a larger number of other agents to increase their future efforts, which in turn makes her better off. This *encouragement effect* is stronger at the outset of the game, where a lot of progress remains before project completion, and it becomes weaker as the project nears completion.

The encouragement effect dominates the free-rider effect if and only if the project is sufficiently far from completion. This result has organizational implications for team size: by increasing the team size, each agent obtains stronger incentives when the project is far from completion, while their incentives become weaker near completion. Georgiadis (2015) explores how a profit-maximizing principal would design incentive contracts and dynamically adjust team size (see Georgiadis and Tang (2017) for an overview of these results). In particular, it shows that if the team size is endogenous but fixed throughout the project, optimal contracts are symmetric and can be implemented through payoffs that are made upon project completion. Optimal teams are smaller the higher the completion threshold. If $N = 2$ and the size of the team can be dynamically adjusted, the principal may put in place a contract that specifies a milestone at which one of the agents receives a lump-sum payoff and stops exerting effort, and the other agent sees the project through to completion.

Deadlines and Project Size Designing a team is about more than choosing the optimal number of members of the team. In particular, a principal who cares about joint surplus may also choose a deadline before which the project must be completed, and she may also have some latitude to choose the size or the ambition of the project itself.

To think about optimal deadlines, we can first think about exogenous deadlines and then optimize over them. In particular, suppose that the agents must complete the project by some fixed deadline $T < \infty$, so that they obtain a project payoff of 0 if the completion time satisfies $\tau > T$. In this case, both the project state q and the time remaining, $T - t$, are stage-game-payoff-relevant variables, so MPEs will specify effort profiles $e_i : [0, T] \times [0, Q] \to \mathbb{R}_+$.

In most cases, it is difficult to solve problems with more than one state variable using the techniques illustrated above because the HJB equations take the form of partial differential equations, which are less tractable than ordinary differential equations. In the appendix, we describe the optimal control approach of Georgiadis (2017) to solving this problem.

Given this characterization of MPE effort profiles, Georgiadis (2017) shows that shorter deadlines can induce agents to exert higher effort at each point in time, but regardless of the length of the deadline, effort provision is inefficient due to the agents' incentives to frontload effort because of the encouragement effect described above. Optimal deadlines are not sequentially optimal: as the deadline approaches, players would prefer to push the deadline out.

In addition to thinking about optimal deadlines, we can use this framework to think about designing the project itself. In particular, Georgiadis et al. (2014) asks how a principal should set the size of a project that is undertaken by a team, where larger projects require higher completion thresholds and result in higher payoffs conditional on reaching the threshold. Principal-optimal long-term contracts are not sequentially optimal for the principal: as the project progresses, the principal would prefer to put in place a different contract that increases the size of the project.

Further Reading Early contributions to the literature on dynamic contribution games were made by Admati and Perry (1991), Marx and Matthews (2000), and Yildirim (2006), which characterized equilibria of such games, showing that some of the static inefficiencies associated with team production are mitigated when agents can contribute over time. These papers modeled the dynamic public good games in discrete time.

In this model, players contribute to a discrete public good. Fershtman and Nitzan (1991) analyzes related games with continuous public goods. Such games exhibit different dynamics, as do related models of common-pool resource extraction problems (see, for example, Reinganum and Stokey (1985)).

The framework above abstracts from contracts by taking completion payoffs as exogenous, and it assumes agents are homogeneous. Cvitanić and Georgiadis (2016) extends the model to allow for optimal contracts and constructs a mechanism that induces efficient contributions. Bowen et al.

(forthcoming) extends the framework to allow for heterogeneous agents and collective choice over project size. When agents are heterogeneous, they disagree about the optimal size of the project: a more efficient agent will put in a larger share of the effort in equilibrium and so prefers a smaller project than a less efficient agent does. Moreover, as the project nears completion, the relative share of effort the efficient agent puts in grows, so that over time, her preferred project size shrinks.

Appendix—Deadlines To solve for MPEs with deadlines, consider an auxiliary game in which the project *must* be completed by the deadline T. Conditional on a completion time before the deadline, $\tau \leq T$, each agent minimizes her discounted costs, while anticipating that each of the other agents behaves in the same cost-minimizing manner. The MPE of this auxiliary game will be a MPE for the original game if each agent's discounted payoff at time 0 is weakly larger than her outside option of 0, and otherwise, there is no MPE in which the project is completed.

We use the maximum principle of optimal control and write strategies and payoffs as a function of time t (see, for example, Kamien and Schwartz, 2012).[3] The Hamiltonian corresponding to each agent i's objective function is

$$H_{i,t} = -e^{-rt}\frac{e_{i,t}^2}{2} + \lambda_{i,t}\left(\sum_{j=1}^{N} e_{j,t}\right),$$

[3] Note that the optimal control approach is equivalent to the HJB approach. However, it is often the case that one is more tractable than the other.

where $\lambda_{i,t} \geq 0$ is the co-state variable associated with agent i's payoff function. Note that $\lambda_{i,t}$ is equivalent to a dual multiplier in Lagrangian optimization, except that in optimal control, it is a function rather than a scalar. Moreover, it can be interpreted as agent i's marginal benefit from bringing the project closer to completion at time t.[4] Her terminal value function is $\phi_{i,\tau} = e^{-r\tau}V_N$, and the requirement that the project be completed by the deadline imposes the constraint

$$\int_0^{\tau} \sum_{i=1}^{N} e_{i,t} dt = Q, \text{ where } \tau \leq T.$$

From Pontryagin's maximum principle, we have that each agent's effort profile must maximize his Hamiltonian, so we have the optimality condition

$$\frac{dH_{i,t}}{de_{i,t}} = 0, \text{ or equivalently, } e_{i,t} = \lambda_{i,t} e^{rt},$$

and the **adjoint equation**

$$\dot{\lambda}_{i,t} = -\frac{dH_{i,t}}{dq_t},$$

which specifies the evolution path of $\lambda_{i,t}$. Finally, the **transversality condition** for each agent is

$$H_{i,\tau} + \frac{d\phi_\tau}{d\tau} \geq 0$$

[4] Note that each agent i's Hamiltonian is a function of $t, q, \{a_{j,t}\}_{j=1}^N$, and $\lambda_{i,t}$. For notational simplicity, we suppress the latter three arguments and simply write $H_{i,t}$.

with equality if $\tau < T$.

These conditions, holding for each i, are necessary conditions for a MPE of the auxiliary game. Therefore, we will proceed by characterizing a solution to this system of equations and argue that the corresponding effort profile $\{e_{i,t}\}$ constitutes a MPE for the original game if and only if each agent obtains a nonnegative ex ante discounted payoff.

First, by totally differentiating the Hamiltonian with respect to q_t, we can rewrite the adjoint equation as

$$\dot{\lambda}_{i,t} = -\sum_{j=1}^{N} \frac{dH_{i,t}}{de_{j,t}} \frac{de_{j,t}}{dt} \frac{dt}{dq_t} = -\sum_{j \neq i} \frac{dH_{i,t}}{de_{j,t}} \frac{de_{j,t}}{dt} \frac{dt}{dq_t} = -\sum_{j \neq i} \frac{\lambda_{i,t}\left(r\lambda_{j,t} + \dot{\lambda}_{j,t}\right)}{\sum_{\ell=1}^{n} \lambda_{\ell,t}},$$

where the first equality uses the optimality condition $dH_{i,t}/de_{i,t} = 0$, and the second equality uses the results that $de_{j,t}/dt = \left(r\lambda_{j,t} + \dot{\lambda}_{j,t}\right) e^{rt}$ and $dq_t/dt = \sum_{\ell=1}^{N} \lambda_{\ell,t} e^{rt}$. After rearranging terms and imposing symmetry, we have

$$\dot{\lambda}_t = -\frac{N-1}{2N-1} r\lambda_t.$$

This differential equation has a unique non-zero solution, $\lambda_t = ce^{-rt\frac{N-1}{2N-1}}$, where c is a constant to be determined. By substituting λ_t the constraint that the project will be completed at time τ, we can express c as a function of the completion time τ as follows:

$$cN \int_0^{\tau} e^{rt\left(1-\frac{N-1}{2N-1}\right)} dt = Q$$

or

$$c = \frac{rQ}{2N-1}\frac{1}{e^{\frac{rN\tau}{2N-1}}-1}.$$

The completion time τ can be pinned down by the transversality condition, which can be rewritten as

$$1-\sqrt{\frac{rQ^2}{2V_N}\frac{1}{2N-1}} \leq e^{-\frac{rN\tau}{2N-1}}$$

with equality if $\tau < T$. By noting that the right-hand side of this inequality is decreasing in τ and is less than one, it follows that the project is completed at time $\tau^{MPE} = \min\left\{T, \bar{\tau}^{MPE}\right\}$, where $\bar{\tau}^{MPE} = -\frac{2N-1}{rN}\ln\left(1-\sqrt{\frac{rQ^2}{2V_N}\frac{1}{2N-1}}\right)$. Sufficiency follows by noting that $H_{i,t}$ is strictly concave in $e_{i,t}$, and applying the Mangasarian theorem (see Seierstad and Sydsaeter, 1987).

Collecting terms, it follows that there exists a unique candidate for a symmetric, project-completing MPE, wherein each agent's effort and ex ante discounted payoff satisfies

$$e_t^{MPE} = \frac{rQ}{2N-1}\frac{e^{\frac{rNt}{2N-1}}}{e^{\frac{rN\tau^{MPE}}{2N-1}}-1},$$

and

$$\Pi_0^{MPE} = e^{-r\tau^{MPE}}V_N - \frac{rQ^2}{2\left(2N-1\right)}\frac{e^{\frac{r\tau^{MPE}}{2N-1}}-1}{\left(e^{\frac{rN\tau^{MPE}}{2N-1}}-1\right)^2},$$

respectively. This candidate is a MPE if and only if $\Pi_0 \geq 0$.

Notice that in the MPE characterized above, each agent's discounted marginal cost (i.e., $e^{rt}e_t^{MPE}$) decreases over time, which implies that equilibrium efforts are frontloaded. Notice that efforts increase with progress, and this is again a game with positive externalities. Therefore, each agent has an incentive to distort her effort profile to induce other agents to raise their future actions. This is a consequence of actions being strategic complements across time, as discussed earlier.

We conclude by discussing the restrict to MPE with differentiable strategies. Heuristically, consider an effort profile in which, at every moment t, each agent chooses the first-best effort e_t^{FB} as long as every agent has chosen the first-best effort in the past (i.e., if $q_t = \int_0^t Ne_s^{FB} ds$) and chooses the effort corresponding to the MPE with differentiable strategies otherwise. One can then construct a MPE that implements the first-best effort schedule using such an effort profile. For details, see Georgiadis (2017, p. 12). This result is an artifact of the assumption that the project progresses deterministically.

Exercise. In the auxiliary game considered above,

(*i.*) Verify that, if the project is completed strictly before the deadline in equilibrium, the effort profile in this auxiliary game coincides with the MPE effort profile in the game without deadlines.

(*ii.*) Compare the equilibrium effort profile for different deadlines to the first-best outcome. Suppose the deadline is endogenous, and the principal maximizes the sum of the agents' payoffs. Is it possible to implement the first-best outcome in a MPE? Explain.

Part IV

Assorted Topics in Organizations

Chapter 10

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.

10.1 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:

$$\rho(e) = -\frac{1}{1 - (q(p_1^*(e)) / q_1'(p_1^*(e)))'}.$$

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 Holmström'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

$$q_1(p^*(e)) = \frac{A - Bc}{2} - \frac{a_2}{2} + \int_0^e \frac{B}{2} ds.$$

If we interpret an increase in competition as a more aggressive move by firm 1's competition, then an increase in competition decreases $q_1(p^*(e))$ 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 $(A - a_2 - (B + b_2)c)$ 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 $S_2(p) = b_2(p - c)$. A higher value of b_2 is a more aggressive move by firm 2. Further we can see that

$$q_1(p^*(e)) = \frac{A - Bc}{2} + \int_0^e \frac{B + b_2}{2} ds.$$

In this case, an increase in competition increases the quantity pass-through of firm 1's residual demand curve and therefore increases $q_1(p^*(e))$ for all e, which in turn implies an increase in e^*.

10.2 Organizational Industrial Organization

The informal theory of firm boundaries and the two formal theories we have examined so far have taken a partial-equilibrium approach and explored how environmental factors such as uncertainty, the degree of contractual incompleteness, and ex post lock-in shape the firm-boundary decision. In this note, we will look at a model in which firm boundary decisions are determined in industry-equilibrium, and we will derive some predictions about how firm-level organization decisions impact the competitive environment and vice versa.

Embedding a model of firm boundaries into an industry-equilibrium framework can be difficult, so we will need to put a lot of structure both on the particular model of firm boundaries we look at as well as on the sense in which firms compete in the market. Different papers in this literature (Grossman and Helpman, 2002; Avenel, 2008; Gibbons, Holden, and Powell, 2012; Legros and Newman, 2013) focus on different models featuring different determinants of firm boundaries. Grossman and Helpman (2002) derives a trade-off between the fundamentally Neoclassical consideration of diminishing returns to scale and the search costs associated with finding alternative trading partners.

Gibbons, Holden, and Powell (2012) consider a Grossman-Hart-Moore-style model in which firms can organize either in a way that motivates a party to acquire information about product demand or in a way that motivates a different party to reduce marginal production costs. The paper embeds this model of firm boundaries into a Grossman-Stiglitz-style rational expectations equilibrium and shows that, if some firms are organized to acquire information, their information will be partially contained in the prices of intermediate goods, which in turn reduces other firms' returns to organizing to acquire information. In equilibrium, differently organized firms will coexist.

This note will focus on Legros and Newman (2013), which embeds a particularly tractable form of the Hart and Holmström (2002/2010) model of firm boundaries into a price-theoretic framework. In the Hart and Holmström model, integration unifies contractible payoff rights and decision rights, thereby ensuring that decisions made largely with respect to their effects on contractible payoffs. Under integration, different managers make decisions, and these decisions are particularly sensitive to their effects on their noncontractible private benefits. The Legros and Newman (2013) insight is that when a production chain's output price is high, the contractible payoffs become relatively more important for the chain's total surplus, and therefore integration will become relatively more desirable.

Description There are two risk-neutral managers, L and R, who each manage a division, and a risk-neutral third-party HQ. Two decisions, $d_L, d_R \in [0, 1]$ need to be made. These decisions determine the managers' noncontractible private benefits $b_L(d_L)$ and $b_R(d_R)$ as well as the probability distribution over output $y \in \{0, A\}$, where high output, A, is firm-specific and is distributed according to a continuous distribution with CDF $F(A)$ and support $[\underline{A}, \bar{A}]$. High output is more likely the more well-coordinated are the two decisions: $\Pr[y = A | d_L, d_R] = 1 - (d_L - d_R)^2$. Output is sold into the product market at price p. Demand for output is generated by an aggregate demand curve $D(p)$.

The revenue stream, $\pi = py$ is contractible and can be allocated to either manager, but each manager's private benefits are noncontractible and are given by

$$
\begin{aligned}
b_L(d_L) &= -d_L^2 \\
b_R(d_R) &= -(1 - d_R)^2,
\end{aligned}
$$

so that manager L wants $d_L = 0$ and manager R wants $d_R = 1$. The decision rights for d_L and d_R are contractible. We will consider two governance structures $g \in \{I, NI\}$. Under $g = I$, a third party receives the revenue stream and both decision rights. Under $g = NI$, manager L receives the revenue stream and the decision right for d_L, and manager R receives the decision right for d_R.

At the firm-level, the timing of the game is as follows. First, HQ chooses a governance structure $g \in \{I, NI\}$ to maximize joint surplus. Next, the manager with control of d_ℓ chooses $d_\ell \in [0, 1]$. Finally, revenues and private benefits are realized, and the revenues accrue to whomever is specified under g. Throughout, we will assume that if HQ is indifferent among decisions, it will make whatever decisions maximize the sum of the managers' private benefits. The solution concept is subgame-perfect equilibrium given an output price p. An industry equilibrium is a price level p^*, and a set of governance structures and decisions for each firm such that industry supply, $S(p)$, coincides with industry demand at price level p^*.

The Firm's Program For now, we will take the industry price level p as given. For comparison, we will first derive the first-best (joint surplus-maximizing) decisions, which solve

$$\max_{d_L, d_R} pA\left(1 - (d_L - d_R)^2\right) - d_L^2 - (1 - d_R)^2$$

or

$$d_L^{FB} = \frac{pA}{1 + 2pA}, d_R^{FB} = \frac{1 + pA}{1 + 2pA}.$$

The first-best decisions partially reflect the role that coordination plays in generating revenues as well as the role that decisions play in generating managers' private benefits. As such, decisions are not perfectly coordinated: denote the decision gap by $\Delta^{FB} = d_R^{FB} - d_L^{FB} = 1/(1 + 2pA)$.

Under non-integration, manager L receives the revenue stream, and managers L and R simultaneously choose d_L^{NI} and d_R^{NI} to solve

$$\max_{d_L} pA\left(1 - \left(d_L - d_R^{NI}\right)^2\right) - d_L^2$$

and

$$\max_{d_R} -\left(1-d_R\right)^2,$$

respectively. Clearly, manager R will choose $d_R^{NI}=1$, so manager L's problem is to

$$\max_{d_L} pA\left(1-\left(d_L-1\right)^2\right)-d_L^2,$$

and therefore she chooses $d_L^{NI}=pA/\left(1+pA\right)$. Since manager L cares both about her private benefits and about revenues, her decision will only be partially coordinated with manager R's decision: the decision gap under non-integration is $\Delta^{NI}=d_R^{NI}-d_L^{NI}=1/\left(1+pA\right)$.

Under integration, since the headquarters does not care about managers' private benefits, it perfectly coordinates decisions and chooses $d_L^I=d_R^I$, and by assumption, it sets both equal to $1/2$. The decision gap under non-integration is $\Delta^I=d_R^I-d_L^I=0$.

Denote total private benefits under governance structure g by $PB^g\equiv b_L\left(d_L^g\right)+b_R\left(d_R^g\right)$, and denote expected revenues by $REV^g=E\left[\pi|\,d_g\right]$. Total welfare is therefore

$$W\left(g\right)=\left(PB^I+REV^I\right)1_{g=I}+\left(PB^{NI}+REV^{NI}\right)1_{g=NI}.$$

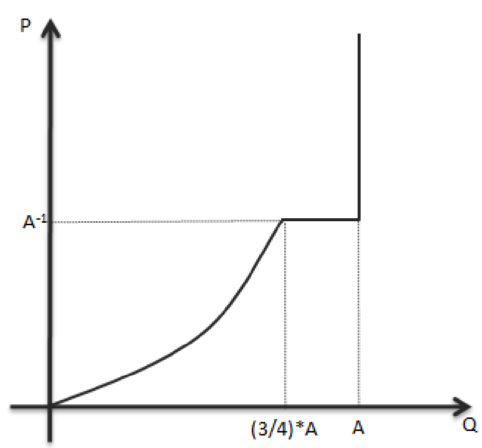
Since the coordination gap is smaller under integration than under non-integration, and expected revenues are higher under integration that under non-integration, there is a trade-off between greater coordination under integration and greater private benefits under non-integration.

Importantly the difference in expected revenues under the two governance structures, $REV^I - REV^{NI}$, is increasing in p and A, and it is increasing faster than is the difference in private benefits, $PB^{NI} - PB^I$. There will therefore be a cutoff value $p^*(A)$ such that if $p > p^*(A) = 1/A$, $g^* = I$, and if $p < p^*(A)$, $g^* = NI$. If $p = p^*(A)$, the firm is indifferent.

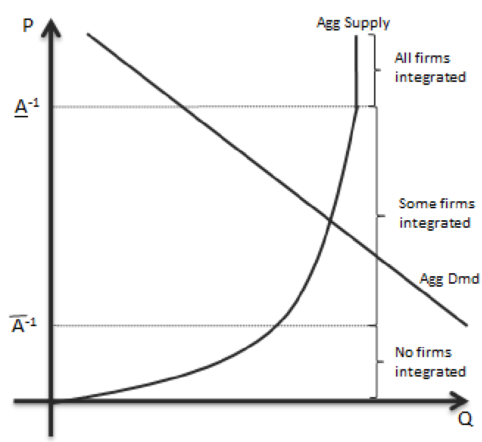
Industry Equilibrium Given a price level p, a firm of productivity A will produce expected output equal to

$$y(p; A) = \begin{cases} A\left(1 - \left(\frac{1}{1+pA}\right)^2\right) & p < 1/A \\ A & p > 1/A. \end{cases}$$

The following figure depicts the inverse expected supply curve for a firm of productivity A. When $p = 1/A$, the firm is indifferent between producing expected output $3A/4$ and expected output A.



Industry supply in this economy is therefore $Y(p) = \int y(p;A)\,dF(A)$ and is upward-sloping. For $p > 1/\underline{A}$, the inverse supply curve is vertical. If $p < 1/\bar{A}$, all firms choose to be non-integrated, and if $p > 1/\underline{A}$, all firms choose to be integrated. For $p \in \left(1/\bar{A}, 1/\underline{A}\right)$, there will be some integrated firms and some non-integrated firms. If demand shifts outward, a (weakly) larger fraction of firms will be integrated. The following figure illustrates industry supply and industry demand.

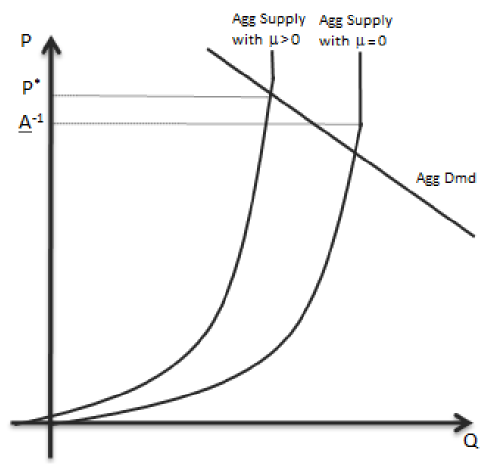


As drawn, the inverse demand curve intersects the inverse supply curve at a value of $p \in \left(1/\bar{A}, 1/\underline{A}\right)$, so in equilibrium, there will be some firms that are integrated (high-productivity firms) and some that are non-integrated (low-productivity firms). If the inverse demand curve shifts to the right, the equilibrium price will increase, and more firms will be integrated.

Output prices are a key determinant of firms' integration decisions, and one of the model's key predictions is that industries with higher output prices (or, for a given industry, during times when output prices are higher), we should expect to see more integration. This prediction is consistent with findings in Alfaro, Conconi, Fadinger, and Newman (2016), which uses industry-level variation in tariffs to proxy for output prices, and McGowan (2016), which uses an increase in product-market competition in the U.S. coal mining industry as a negative shock to output prices.

Since output prices are determined at the market level, a firm's integration decision will necessarily impact other firms' integration decisions. As an illustration, suppose a fraction $\mu \in [0, 1]$ of firms in the industry are exogenously restricted to being non-integrated, and suppose that such firms are chosen randomly and independently from their productivity.

An increase in μ from 0 to $\mu' \in (0, 1)$ will lead to a reduction in industry supply and therefore to an increase in equilibrium price. This change can lead other firms in the industry that would have otherwise chosen to be non-integrated to instead opt for integration.



The above figure illustrates the inverse supply curve under $\mu = 0$ and under $\mu' > 0$. Under $\mu = 0$, in equilibrium, there will be some firms that choose to be non-integrated. As drawn, in the $\mu' > 0$ case, output prices will be $p^* > 1/\underline{A}$, so all the firms that are not exogenously restricted to be non-integrated will in fact choose to be integrated.

10.3 Knowledge Hierarchies in Equilibrium

In this note, we continue our discussion of knowledge hierarchies to examine some of the implications of the Garicano (2000) model for the evolution of income inequality in the United States since the 1970's. Between 1974 and 1988, we saw an increase in both the 90/50 gap (the ratio of the log hourly wage for workers at the 90th percentile of the wage distribution to those at the median) and the 50/10 gap, two commonly used measures of wage inequality. But beginning in the late 1980's, the 50/10 gap began to decline, while the 90/50 gap continued to increase. This change in the late 1980's is difficult to reconcile with the two leading explanations for the increase in income inequality in the United States post-war period.

One source of wage inequality is the "superstar effect," which is present in Lucas (1978): higher-ability workers choose to become managers, and higher-ability managers manage larger firms and earn higher profits. Recall that in the Lucas (1978) model, the production technology is given by $y = \varphi n^{\theta}$, where $\theta < 1$ is a parameter that captures organizational diseconomies of scale. If we think of innovation in communication technology as reducing θ, the superstar effect becomes stronger: with a lower θ, better managers can manage larger firms, which means that for a given population of workers, a smaller share of them will become managers, and the returns to ability will increase. This reduction in θ will also have the effect of raising labor demand for a given wage level for production workers and will therefore lead to increased production-worker wages. Innovation in communication technology therefore leads to increased inequality at the top of the distribution, higher wages at the bottom, but no change in inequality at the bottom. This model can be augmented to allow production workers of ability φ to supply, say, φ efficiency units of labor. If wages are the price for efficiency units of labor, there would be inequality at the bottom of the distribution, since different workers supply different numbers of efficiency units of labor. Innovation in communication technology would lead to an increase in the price of efficiency units of labor and therefore would lead to an increase in wage inequality at the bottom of the distribution, which is consistent with the trend from the mid-1970s to the late 1980s, but not with the trend following the late 1980s.

Explanations for the evolution of inequality based on this "superstar effect" (see, for example, Gabaix and Landier (2008)) at their core predict a fanning out of the wage distribution, but this is not what we have seen in recent years, where we have seen a hollowing out of the middle of the income distribution and a rise in the top and the bottom. Garicano and Rossi-Hansberg (2006) argue that taking organizational structure seriously leads to the phenomenon they refer to as the "shadow of the superstars," and advances in communication technology have made this phenomenon more important in recent years, providing an explanation for these recent trends.

Model Description Following a simplified version of Garicano and Rossi-Hansberg (2006), assume individuals are endowed with overlapping knowledge sets $[0, z]$. We can characterize this knowledge set entirely by its upper bound z, which we will take to be the individual's type. Assume the distribution of problems is uniform on $[0, 1]$, and let $\phi(z)$ denote the distribution of skill in the population. As in Lucas (1978), each worker chooses either to work in a firm or to work as an entrepreneur, and if she chooses to be an entrepreneur, she can either be an independent entrepreneur, or she can form a hierarchy and choose how many workers to employ. Assume that hierarchies have only two layers consisting of a single manager and a mass n of production workers.

As in Garicano (2000), a production worker receives one problem and is able to solve a fraction z of them. If he works in a hierarchy, and he is unable to solve a problem on his own, he can refer the problem to his manager, who can assess $1/h$ referred problems. An **organization** then consists of a vector $g = (n, z_m, z_p)$, where n denotes the number of workers, z_m denotes the skill of its manager, and z_p denotes the skill level of its production workers. The manager of an organization will be referred $n(1 - z_p)$ problems and can solve them all as long as she has enough time to do so, or if $hn(1 - z_p) \leq 1$. A manager who hires more knowledgeable workers (i.e., hires workers with higher values of z_p) can hire more of them and still have enough time to assess the problems they refer, and a more-knowledgeable manager will be able to solve a larger fraction of these referred problems, so there is a natural complementarity between manager knowledge, worker knowledge, and the manager's span of control.

The price of hiring a worker of type z_p is $w_p(z_p)$. This production-worker wage function $w_p(z_p)$ will be endogenous to the equilibrium. Given wage function $w_p(z_p)$, a manager with skill z_m who hires n workers of skill z_p will produce output nz_m and incur a wage bill of $nw_p(z_p)$.

A **competitive equilibrium** is a **production-worker wage function** $w_p^*(z)$, which specifies the wage a firm must pay to hire a worker of knowledge z, a **labor-demand function** $n^*(z)$, which specifies the mass of production workers an entrepreneur of knowledge z hires, an **occupation-choice function** $d^*(z)$, which specifies for a worker of knowledge z, whether he becomes a production worker, an independent entrepreneur, or a manager, and an **assignment function** $m^*(z)$, which denotes the skill of the manager that a worker of knowledge z is matched to if he chooses to be a production worker.

The Program Given production-worker wage function $w_p(z_p)$, a manager with skill z_m will solve

$$w_m(z_m) \equiv \max_{z_p, n} (z_m - w_p(z_p))\, n = \max_{z_p} \frac{z_m - w_p(z_p)}{h(1 - z_p)},$$

where I substituted the manager's time constraint, holding with equality (which, under an equilibrium occupation-choice function, it will). A worker who chooses to be an independent entrepreneur will produce one unit of problems and be able to solve a fraction z of them and will therefore receive a total payoff of $w_I(z) = z$. A worker with knowledge z therefore has to choose whether to be a production worker, an independent entrepreneur, or a manager, and therefore solves

$$w(z) = \max\{w_p(z), w_I(z), w_m(z)\}.$$

The objective will be to characterize the function $w(z)$ that arises in equilibrium and to describe how it changes in response to an increase in communication technology (a reduction in h).

To do so, first note that the production-worker wage slope at z_p has to satisfy the first-order condition for the firm that employs production workers with knowledge z_p:

$$\frac{-h\left(1-z_p\right)w_p'\left(z_p\right)+\left(z_m-w_p\left(z_p\right)\right)h}{\left(h\left(1-z_p\right)\right)^2}=0$$

or $w_p'\left(z_p\right)=\left(z_m-w\left(z_p\right)\right)/\left(1-z_p\right)$. For those workers who choose to become production workers, it must be the case that $w_p\left(z_p\right)>z_p$, and therefore

$$w_p'\left(z_p\right)=\frac{z_m-w\left(z_p\right)}{1-z_p}<\frac{z_m-z_p}{1-z_p}<1.$$

For those workers who choose to become independent entrepreneurs, clearly, $w_I'(z)=1$. Finally, by the envelope theorem, for workers who choose to become managers,

$$w_m'\left(z_m\right)=\frac{1}{h\left(1-z_p\right)}>1.$$

We therefore have $w_m'(z)>w_I'(z)>w_p'(z)$, which implies that there will be two cutoffs, z^* and z^{**}, such that workers with $z\in[0,z^*]$ will choose to become production workers, workers with $z\in(z^*,z^{**}]$ will choose to become independent entrepreneurs, and workers with $z\in(z^{**},1]$ will choose to become managers. The marginal worker $z=z^*$ is indifferent between being a production worker and being an independent entrepreneur, so $w_p\left(z^*\right)=z^*$, and the marginal worker $z=z^{**}$ is indifferent between being an independent entrepreneur and being a manager, so $w_m\left(z^{**}\right)=z^{**}$, and clearly $w_m(1)=1/h$.

These conditions pin down w^* and z^*.

The equilibrium assignment function $m^*(z)$ is pinned down by the market-clearing condition. First, there will be positive sorting, so that $m^*(z)$ is increasing in z. Next, note that the labor market must clear for production workers of knowledge z for all $z \leq z^*$. Labor supply for workers of knowledge z is $\phi(z)$, and labor demand for workers of knowledge z is $n(m^*(z))\,\phi(m^*(z))$. The labor-market clearing for production workers with knowledge $z \leq z_p$ can therefore be written as

$$\int_0^{z_p} \phi(z)\,dz = \int_{z^{**}}^{m^*(z_p)} n(m^*(z_p))\,\phi(m^*(z_p))\,dz,$$

which we can differentiate with respect to z_p to get

$$m^{*\prime}(z_p) = \frac{1}{n(m^*(z_p))}\frac{\phi(z_p)}{\phi(m^*(z_p))} = h(1-z_p)\,\frac{\phi(z_p)}{\phi(m^*(z_p))}.$$

If ϕ is the uniform distribution, then this condition is simply $m^{*\prime}(z_p) = h(1-z_p)$ for all $z_p \in [0, z^*]$. This condition, along with the facts that $m^*(0) = z^{**}$ and $m^*(z^*) = 1$, pin down m^* and z^{**}.

As in Lucas (1978), this model features the superstar effect: wages for managers satisfy $w_m'(z_m) = 1/(h(1-z_p))$ so that higher-ability managers receive higher wages, and it is convex in z_m, since $m^*(z)$ is increasing. Moreover, the slope and convexity of this wage function is higher when h is lower, so innovations in communication technology can lead to increased inequality at the top of the distribution.

In addition to this superstar effect, this model features what Garicano and Rossi-Hansberg (2006) refer to as the "shadow of the superstars" effect: improvements in communication technology improve managers' ability to leverage their knowledge, so more-knowledgeable managers will manage larger teams. The threshold for being a manager therefore increases, which reduces the earnings for those that would have been managers for higher values of h. Moreover, among production workers, the slope of the matching function $m^{*\prime}(z_p) = h(1 - z_p)$ declines when there is better communication technology (since production workers will now be employed by a smaller mass of managers), which implies a reduction in the convexity of the production-worker wage function, which satisfies $w_p'(z_p) = (m(z_p) - w(z_p)) / (1 - z_p)$. Improvements in communication technology therefore raise wages at the top, reduce wages in the middle, and increase wages at the bottom.

Garicano and Rossi-Hansberg (2006) is motivated by compelling and, at a high level, puzzling facts about the evolution of income inequality in the United States. And it provides an organizationally based explanation that does a better job of matching the facts than other leading explanations. For example, one of the leading explanations for increased inequality in recent years is the skill-biased technological change hypothesis in which the decline in the cost of capital equipment has decreased the price of routine tasks that are highly substitutable with this form of capital. In contrast, analytic and manual tasks are less substitutable, so the improvement in technology leads to higher demand and higher employment of workers performing those tasks.

While this explanation can account for the observed trends, it is less obvious why this would lead to a decline in the middle class rather than a decline in the lower end of the income distribution. Explanations based on superstar effects at their core, predict a fanning out of the income distribution, which is not consistent with recent trends.

10.4 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

$$\max p(q(K,L))\, q(K,L) - wL - rK$$

subject to $L \leq \bar{L}$ and $K \leq \bar{K}$. The Lagrangian is

$$\mathcal{L} = p(q(K,L))\, q(K,L) - wL - rK + \lambda_K \left(\bar{K} - K\right) + \lambda_L \left(\bar{L} - L\right).$$

Taking first-order conditions, we get

$$\begin{aligned} MRP_K^* &= p'(q^*)\, q_K q^* + p(q^*)\, q_K = r + \lambda_K \\ MRP_L^* &= p'(q^*)\, q_L q^* + p(q^*)\, q_L = w + \lambda_L \end{aligned}$$

We can rearrange these expressions and derive

$$\begin{aligned} \left(p'(q^*)\frac{q^*}{p(q^*)} + 1\right) p(q^*)\, q_K &= r + \lambda_K \\ \left(p'(q^*)\frac{q^*}{p(q^*)} + 1\right) p(q^*)\, q_L &= w + \lambda_L \end{aligned}$$

or

$$\begin{aligned} p(q^*)\, q_K &= \frac{\left|\varepsilon_{qp}(q^*)\right|}{\left|\varepsilon_{qp}(q^*)\right| - 1}(r + \lambda_K) \\ p(q^*)\, q_L &= \frac{\left|\varepsilon_{qp}(q^*)\right|}{\left|\varepsilon_{qp}(q^*)\right| - 1}(w + \lambda_L). \end{aligned}$$

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,

$$\begin{aligned} TFPR &= \frac{p(q^*)\,q^*}{(rK^*)^{\alpha}(wL^*)^{1-\alpha}} = \frac{1}{\tilde{\alpha}^{\alpha}(1-\tilde{\alpha})^{1-\alpha}}\frac{p(q^*)\,q^*}{rK^*+wL^*} \\ &= \frac{1}{\tilde{\alpha}^{\alpha}(1-\tilde{\alpha})^{1-\alpha}}\frac{REV}{TVC} \end{aligned}$$

In particular, we see TFPR is proportional to the revenue (REV) to total variable cost (TVC) ratio. This ratio is given by

$$\frac{REV}{TVC} = \frac{|\varepsilon_{qp}(q^*)|}{|\varepsilon_{qp}(q^*)|-1}\left(\frac{\tilde{\alpha}}{\alpha}\frac{MRP_K^*}{r}\right)^{\alpha}\left(\frac{1-\tilde{\alpha}}{1-\alpha}\frac{MRP_L^*}{w}\right)^{1-\alpha}.$$

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 $p(q)$ 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

$$ALP = \frac{MRP_L^*/w}{\beta}. \tag{10.1}$$

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):

$$TFPR^* = \frac{pq^*}{(wL^*)^\beta} = p\frac{A\,(L^*)^\beta}{(wL^*)^\beta} = p\frac{A}{w^\beta}. \tag{10.2}$$

More generally, suppose the scale is assumed (by the econometrician) to be $\gamma \geq \beta$. Then

$$\frac{pq}{(wL)^\gamma} = ALP\,(wL)^{1-\gamma} \tag{10.3}$$

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 (10.3), 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 (10.2), which depends on actual TFP (A), and (10.1), 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 11

Organizations and Financial Markets

Organizations rely on financial institutions and markets to provide the capital necessary to start new initiatives and grow existing projects. But a variety of contractual frictions prevent these markets from operating efficiently, which in turn shapes the governance structure firms choose to put in place. In particular, wealth constraints or limited-liability constraints on the part of the entrepreneur, coupled with the requirement that a creditor be repaid, may imply that the entrepreneur should not be the residual claimant on the firm's proceeds. Hence, the entrepreneur might divert funds to low-profit projects that she benefits personally from, or she may otherwise act in a way that does not maximize the creditor's return.

Furthermore, asymmetric information about a firm's profits or productivity can lead to adverse selection in the financial market. These frictions can lead some positive net present value projects to go unfunded, and they may affect the productivity and profits of the projects that are funded.

This chapter explores the implications of these financial frictions for organizational performance. The first section focuses on elemental models of firms and creditors, while the second half focuses on how these frictions affect firm dynamics.

11.1 Pledgeable Income and Credit Rationing

There is a risk-neutral Entrepreneur (E) and a risk-neutral Investor (I). The Investor has capital but no project, and the Entrepreneur has a project but no capital. In order to pursue the project, the Entrepreneur needs K units of capital. Once the project has been pursued, the project yields revenues py, where $y \in \{0,1\}$ is the project's output, and p is the market price for that output. The Entrepreneur chooses an action $e \in [0,1]$ that determines the probability of a successful project, $\Pr[y=1|\, e]=e$, as well as a private benefit $b(e)$ that accrues to the Entrepreneur, where b is strictly decreasing and concave in e and satisfies $b'(0)=0$ and $\lim_{e\to 1} b'(e)=-\infty$.

The Entrepreneur can write a contract $w \in \mathcal{W}=\{w:\{0,1\}\to\mathbb{R},\ 0\leq w(y)\leq py\}$ that pays the Investor $w(y)$ if output is y and therefore shares the projects revenues with the Investor. If the Investor declines the contract, he keeps the K units of capital, and the Entrepreneur receives a payoff of 0. If the

Investor accepts the contract, the Entrepreneur's and Investor's preferences are

$$\begin{aligned} U_E(w,e) &= E[py - w(y)|\, e] + b(e) \\ U_I(w,e) &= E[w(y)|\, e]. \end{aligned}$$

There are strong parallels between this model and the limited-liability Principal-Agent model we studied earlier. We can think of the Entrepreneur as the Agent and the Investor as the Principal. There is one substantive difference and two cosmetic differences. The substantive difference is that the Entrepreneur is the one writing the contract, and while the contract must still satisfy the Entrepreneur's incentive-compatibility constraint, the individual rationality constraint it has to satisfy is the *Investor's*. The two cosmetic differences are: (1) the payments in the contract flow from the Entrepreneur to the Investor, and (2) instead of higher values of e costing the Entrepreneur $c(e)$, they reduce her private benefits $b(e)$.

Timing The timing of the game is as follows.

1. E offers I a contract $w(y)$, which is commonly observed.

2. I accepts the contract $(d = 1)$ or rejects it $(d = 0)$ and keeps K, and the game ends. This decision is commonly observed.

3. If I accepts the contract, E chooses action e and receives private benefit $b(e)$. e is only observed by E.

4. Output $y \in \{0,1\}$ is drawn, with $\Pr[y=1|\, e] = e$. y is commonly observed.

5. E pays I an amount $w(y)$. This payment is commonly observed.

Equilibrium The solution concept is the same as always. A **pure-strategy subgame-perfect equilibrium** is a contract $w^* \in \mathcal{W}$, an acceptance decision $d^* : \mathcal{W} \to \{0,1\}$, an action choice $e^* : \mathcal{W} \times \{0,1\} \to [0,1]$ such that given contract w^*, the Investor optimally chooses d^*, and the Entrepreneur optimally chooses e^*, and given d^*, the Investor optimally offers contract w^*. We will say that the optimal contract induces action e^*.

The Program The Entrepreneur offers a contract $w \in \mathcal{W}$, which specifies a payment $w(0) = 0$ and $0 \leq w(1) \leq p$ and proposes an action e to solve

$$\max_{w(1),e} (p - w(1))\, e + b(e)$$

subject to the incentive-compatibility constraint

$$e \in \underset{\hat{e}\in[0,1]}{\operatorname{argmax}} (p - w(1))\, \hat{e} + b(\hat{e}),$$

the Investor's individual-rationality (or break-even) constraint

$$w(1)\, e \geq K.$$

Analysis We can decompose the problem into two steps. First, we can ask: for a given action e, how much rents must the Entrepreneur receive in order to choose action e, and therefore, what is the maximum amount that the Investor can be promised if the Entrepreneur chooses e? Second, we can ask: given that the Investor must receive K, what action e^* maximizes the Entrepreneur's expected payoff?

The following figure illustrates the problem using a graph similar to the one we looked at when we thought about limited liability constraints. The horizontal axis is the Entrepreneur's action e, and the segment pe is the expected revenues as a function of e. The dashed line $(p - w_{e_1})\,e$ represents, for a contract that pays the Investor $w\,(1) = w_{e_1}$ if $y = 1$, the Entrepreneur's expected monetary payoff, and $-b\,(e)$ represents the Entrepreneur's cost of choosing different actions. As the figure illustrates, the contract that gets the Entrepreneur to choose action e_1 can pay the Investor at most $w_{e_1}e_1$ in expectation.

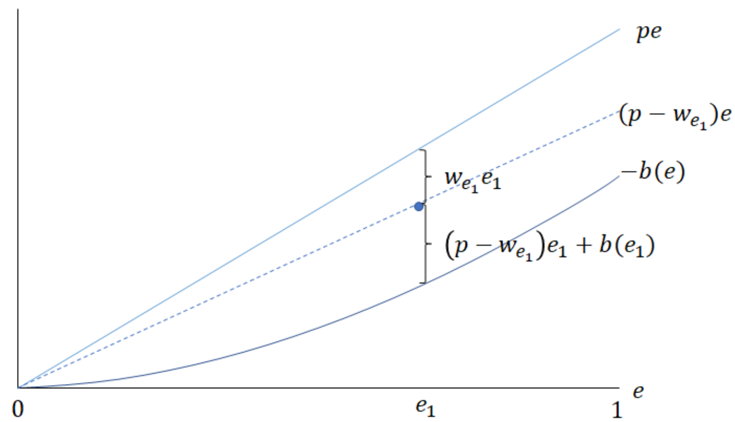


Figure 9: Entrepreneur Incentive Rents

The next figure illustrates, for different actions e, the rents $(p - w_e)\, e + b\,(e)$ that the Entrepreneur must receive for e to be incentive-compatible. Note that because $w_e \geq 0$, there is no incentive-compatible contract that gets the Entrepreneur to choose any action $e > e^{FB}$. The vertical distance between the expected revenue pe curve and the Entrepreneur rents curve is the Investor's expected payoff under the contract that gets the Entrepreneur to choose action e. For the Investor to be willing to sign such a contract, that vertical distance must be at least K, which is the amount of capital the Entrepreneur needs.

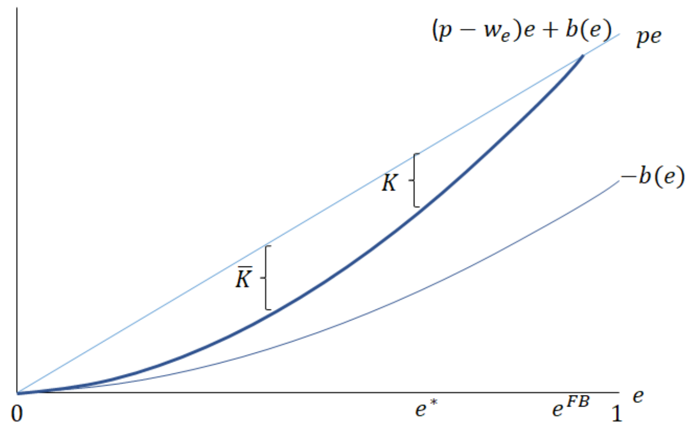


Figure 10: Equilibrium and Pledgeable Income

Two results emerge from this analysis. First, if $K > 0$, then in order to secure funding K, the Entrepreneur must share some of the project's earnings with the Investor, which means that the Entrepreneur does not receive all the returns from her actions and therefore will choose an action $e^* < e^{FB}$. Second, the value $\bar{K}$ represents the maximum expected payments the Entrepreneur can promise the Investor in any incentive-compatible contract. This value is referred to as the Entrepreneur's **pledgeable income**. If the project requires capital $K > \bar{K}$, then there is no contract the Entrepreneur can offer the Investor that the Investor will be willing to sign, even though the Entrepreneur would invest in the project if she had her own capital. When this is the case, we say that there is **credit rationing**.

As a final point about this model, with binary output, the optimal contract can be interpreted as either a debt contract or an equity contract. Under the debt contract interpretation, the Entrepreneur must reimburse w_{e^*} or else go bankrupt, and if the project is successful, she keeps the residual $p - w_{e^*}$. Under the equity contract interpretation, the Entrepreneur holds a share $(p - w_{e^*})/p$ of the project's equity, and the Investor holds a share w_{e^*}/p of the project's equity. That the optimal contract can be interpreted as either a debt contract or an equity contract highlights that if we want to actually understand the role of debt or equity contracts, we will need a richer model.

11.2 Control Rights and Financial Contracting

The previous model cannot explain the fact that equity has voting power while debt does not, except following default. Aghion and Bolton (1992) takes an incomplete contracting approach to thinking about financial contracting and brings control rights front and center. We will look at a simple version of the model that provides an explanation for debt contracts featuring *contingent* control. In this model, control rights matter because the parties disagree about important decisions that are ex ante noncontractible. The parties will renegotiate over these decisions ex post, but because the Entrepreneur is wealth-constrained, renegotiation may not fully resolve the disagreement. Investor control will therefore lead to a smaller pie ex post, but the Investor will receive a larger share of that pie. As a result, even though Investor control destroys value, it may be the only way to get the Investor to be willing to invest to begin with.

The Model As in the previous model, there is a risk-neutral Entrepreneur (E) and a risk-neutral investor (I). The Investor has capital but no project, the Entrepreneur has a project but no capital, and the project costs K. The parties enter into an agreement, which specifies who will possess the right to make a decision $d \in \mathbb{R}_+$ once that decision needs to be made. After the state $\theta \in \mathbb{R}_+$, which is drawn according to density $f(\theta)$, is realized, the decision d is made. This decision determines verifiable profits $y(d)$, which we will assume accrue to the Investor.[1] It also determines nonverifiable private benefits $b(d)$ that accrue to the Entrepreneur.

The parties can contract upon a rule that specifies who will get to make the decision d in which state of the world: let $g : \mathbb{R}_+ \to \{E, I\}$ denote the **governance structure**, where $g(\theta) \in \{E, I\}$ says who gets to make the decision d in state θ. The decision d is itself not ex ante contractible, but it is ex post contractible, so that the parties can negotiate over it ex post. In particular, we will assume that the Entrepreneur has all the bargaining power, so that she will propose a take-it-or-leave-it offer specifying a decision d as well as a transfer $w \geq 0$ from the Investor to the Entrepreneur. Note

[1] We could enrich the model to allow the parties to contract ex ante on the split of the verifiable profits that each party receives. Giving all the verifiable profits to the Investor maximizes the efficiency of the project because it maximizes the pledgeable income that he can receive without having to distort ex post decision making.

that the transfer has to be nonnegative, because the Entrepreneur is cash-constrained.

Timing

1. E proposes a governance structure g. g is commonly observed.

2. I chooses whether or not to go ahead with the investment. This decision is commonly observed.

3. The state θ is realized and is commonly observed.

4. E makes a take-it-or-leave-it offer of (d, w) to I, who either accepts or rejects it.

5. If I rejects the offer, party $g(\theta)$ chooses d.

Analysis As usual, let us start by describing the first-best decision that maximizes the sum of the profits and the private benefits:

$$d^{FB} \in \underset{d \in \mathbb{R}_+}{\operatorname{argmax}}\, y(d) + b(d).$$

Assume y and b are strictly concave and single-peaked, so that there is a unique first-best decision. Moreover, assume $y(d)$ is maximized at some decision d^I, and $b(d)$ is maximized at some other decision $d^E < d^I$. These assumptions imply that $d^E < d^{FB} < d^I$. Now, let us see what happens depending on who has control.

We will first look at what happens under Entrepreneur control. This corresponds to $g(\theta) = E$ for all θ. In this case, if the Investor rejects the Entrepreneur's offer in stage 4, the Entrepreneur will choose d to maximize her private benefit and will therefore choose d^E. Recall that the Entrepreneur does not care about the profits of the project because we have assumed that the profits accrue directly to the Investor. The decision d^E is therefore the Investor's outside option in stage 4. It will not be the decision that is actually made, however, because the Entrepreneur can offer to make a higher decision in exchange for some money. In particular, she will offer $\left(d^{FB}, w\right)$, where w is chosen to extract all the ex post surplus from the Investor:

$$y\left(d^{FB}\right) - w = y\left(d^E\right) \text{ or } w = y\left(d^{FB}\right) - y\left(d^E\right) > 0.$$

Under Entrepreneur control, the Entrepreneur's payoff will therefore be $b\left(d^{FB}\right) + y\left(d^{FB}\right) - y\left(d^E\right) > b\left(d^E\right)$, and the Investor's payoff will be $y\left(d^E\right)$, which is effectively the Entrepreneur's pledgeable income. If $y\left(d^E\right) > K$, then the Investor will make the investment, and the first-best decision will be made, but if $y\left(d^E\right) < K$, this arrangement will not get the Investor to make the investment.

Now let us look at what happens under Investor control, which corresponds to $g(\theta) = I$ for all θ. In this case, if the Investor rejects the Entrepreneur's offer at stage 4, the Investor will choose d to maximize profits and will therefore choose d^I. The decision d^I is therefore the Investor's outside option in stage 4. At stage 4, the Entrepreneur would like to get the Inventor to make a decision $d < d^I$, but in order to get him to do so, she would have to choose $w < 0$, which is not feasible. As a result, d^I will in fact be the decision that is made. Under Investor control, the Entrepreneur's payoff will be $b\left(d^I\right)$, and the Investor's payoff will be $y\left(d^I\right)$, which again is effectively the Entrepreneur's pledgeable income. Conditional on the investment being made, total surplus under Investor control is lower than under Entrepreneur control, but the benefit of Investor control is that it ensures the Investor a payoff of $y\left(d^I\right)$, which may exceed K even if $y\left(d^E\right)$ does not.

As in the Property Rights Theory, decision rights determine parties' outside options in renegotiations, which determines their incentives to make investments that are specific to the relationship. In contrast to the PRT, however, ex post renegotiation does not always lead to a surplus-maximizing outcome because the Entrepreneur is wealth-constrained. As such, in order to provide the Investor with incentives to make the relationship-specific investment of investing in the project, we may have to give the Investor ex post control, even though he will use it in a way that destroys total surplus.

If $y\left(d^I\right) > K > y\left(d^E\right)$, then Investor control is better than Entrepreneur control because it ensures the Investor will invest, but in some sense, it involves throwing away more surplus than necessary. In particular, consider a governance structure $g\left(\cdot\right)$ under which the Entrepreneur has control with probability π (i.e., $\Pr\left[g\left(\theta\right) = E\right] = \pi$), and the Investor has control with probability $1 - \pi$ (i.e., $\Pr\left[g\left(\theta\right) = I\right] = 1 - \pi$). The Entrepreneur can get the

Investor to invest if she chooses π to satisfy

$$\pi y\left(d^E\right) + (1-\pi)\, y\left(d^I\right) = K,$$

which will be optimal.

Now, stochastic control in this sense is a bit tricky to interpret, but with a slight elaboration of the model, it has a more natural interpretation. In particular, suppose that the state of the world, θ, determines how sensitive the project's profits are to the decision, so that

$$y\,(d,\theta) = \alpha\,(\theta)\, y\,(d) + \beta\,(\theta)\,,$$

where $\alpha\,(\theta) > 0$, and $\alpha'\,(\theta) < 0$. In this case, the optimal governance structure would involve a cutoff θ^* so that $g\,(\theta) = E$ if $\theta > \theta^*$ and $g\,(\theta) = I$ if $\theta \leq \theta^*$, where this cutoff is chosen so that the Investor's expected payoffs would be K.

If $\alpha'\,(\theta)\, y\,(d) + \beta'\,(\theta) > 0$ for all d, then high-θ states correspond to high-profit states, and this optimal arrangement looks somewhat like a debt contract that gives control to the creditor in bad states and gives control to the Entrepreneur in the good states. In this sense, the model captures an important aspect of debt contracts, namely that they involve contingent allocations of control. This theory of debt contracting is not entirely compelling, though, because the most basic feature of debt contracts is that the shift in control to the Investor occurs *only if the Entrepreneur does not make* *a repayment.* The last model we will look at will have this feature.

11.3 Cash Diversion and Liquidation

We will look at one final model that involves an important decision that is often specified in debt contracts: whether to liquidate an ongoing project. We will show that when the firm's cash flows are noncontractible, giving the Investor the rights to the proceeds from a liquidation event can protect him from short-run expropriation from an Entrepreneur who may want to direct the project's cash flows toward her own interests.

The Model As before, there is a risk-neutral Entrepreneur (E) and a risk-neutral investor (I). The Investor has capital but no project, the Entrepreneur has a project but no capital, and the project costs K. If the project is funded, it yields income over two periods, which accrue to the Entrepreneur. In the first period, it produces output $y_1 \in \mathcal{Y}_1 \equiv \{0, 1\}$, where $\Pr[y_1 = 1] = q$, and that output generates a cash flow of $p_1 y_1$. After y_1 is realized, the Entrepreneur can make a cash payment $0 \leq \hat{w}_1 \leq p_1 y_1$ to the Investor. The project can then be terminated, yielding a liquidation value of L, where $0 \leq L \leq K$, which accrues to the Investor. Denote the probability the project is continued by $r \in [0, 1]$. If the project is continued, in the second period, it produces output $y_2 = 1$, and that output generates cash flow of p_2. At this point, the Entrepreneur can again make a cash payment

$0 \leq \hat{w}_2 \leq p_2$ to the Investor.

The cash flows are noncontractible, so the parties are unable to write a contract that specifies output-contingent repayments from the Entrepreneur to the Investor, but they can write a contract that specifies probabilities $r : \mathbb{R}_+ \to [0,1]$ that determine the probability $r(\hat{w}_1)$ the project is continued if the Entrepreneur pays the Investor $\hat{w}_1$. The contracting space is therefore $\mathcal{W} = \{r : \mathbb{R}_+ \to [0,1]\}$. The players' payoffs, if the Investor invests K in the project are:

$$\begin{aligned} u_E(\ell, y_1, \hat{w}_1, \hat{w}_2) &= p_1 y_1 - \hat{w}_1 + r(\hat{w}_1)(p_2 - \hat{w}_2) \\ u_I(\ell, y_1, \hat{w}_1, \hat{w}_2) &= \hat{w}_1 + (1 - r(\hat{w}_1)) L + r(\hat{w}_1)\hat{w}_2. \end{aligned}$$

Throughout, we will assume that $p_2 > L$, so that liquidation strictly reduces total surplus.

Timing The timing of the game is as follows.

1. E offers I a contract $r(\hat{w}_1)$, which is commonly observed.

2. I accepts the contract $(d = 1)$ or rejects it $(d = 0)$ and keeps K, and the game ends. This decision is commonly observed.

3. If I accepts the contract, output $y_1 \in \{0, 1\}$ is realized. y_1 is commonly observed.

4. E makes a payment $0 \leq \hat{w}_1 \leq p_1 y_1$ to I. $\hat{w}_1$ is commonly observed.

5. The project is liquidated with probability $1 - r(\hat{w}_1)$. The liquidation event is commonly observed.

6. If the project has not been liquidated, output $y_2 = 1$ is realized. y_2 is commonly observed.

7. E makes a payment $0 \le \hat{w}_2 \le y_2$ to I. $\hat{w}_2$ is commonly observed.

Equilibrium The solution concept is the same as always. A **pure-strategy subgame-perfect equilibrium** is a continuation function $r^* \in \mathcal{W}$, an acceptance decision $d^* : \mathcal{W} \to \{0, 1\}$, a first-period payment rule $w_1^* : \mathcal{W} \times \{0, 1\} \to \mathbb{R}_+$, and a second-period payment rule $w_2^* : \mathcal{W} \times \{0, 1\} \times \{0, 1\} \times \mathbb{R}_+ \to \mathbb{R}_+$ such that given continuation function r^* and payment rules w_1^* and w_2^*, the Investor optimally chooses d^*, and given d^*, the Entrepreneur optimally offers continuation function r^* and chooses payment rules w_1^* and w_2^*.

The Program Models such as this one, in which the Entrepreneur's repayment decisions are not contractible, are referred to as **cash diversion** models. The Entrepreneur's problem will be to write a contract that specifies continuation probabilities and repayment amounts so that given those repayment-contingent continuation probabilities, the Entrepreneur will actually follow through with those repayments, and the Investor will at least break even. In this setting, it is clear that in any subgame-perfect equilibrium, the Entrepreneur will not make any positive payment $\hat{w}_2 > 0$, since she receives nothing in return for doing so. Moreover, it will be without loss of generality for the Entrepreneur to specify a single repayment amount $0 < w_1 \leq p_1$ to be repaid if $y_1 = 1$, and a pair of probabilities r_0 and r_1, where r_0 is the probability the project is continued (and not liquidated) if $\hat{w}_1 \neq w_1$, and r_1 is the probability the project is continued if $\hat{w}_1 = w_1$. The Entrepreneur's problem is therefore

$$\max_{r_0, r_1, w_1 \leq p_1} q\left(p_1 - w_1 + r_1 p_2\right) + (1-q)\, r_0 p_2$$

subject to the Entrepreneur's incentive-compatibility constraint

$$p_1 - w_1 + r_1 p_2 \geq p_1 + r_0 p_2$$

and the Investor's break-even constraint

$$q\left(w_1 + (1-r_1)\, L\right) + (1-q)\,(1-r_0)\, L \geq K.$$

It will be useful to rewrite the incentive-compatibility constraint as

$$(r_1 - r_0)\, p_2 \geq w_1,$$

which says that in order for repayment w_1 to be incentive-compatible, it has to be the case that by making the payment w_1 (instead of paying zero), the probability r_1 that the project is continued (and hence the Entrepreneur receives p_2) if she makes the payment is sufficiently high relative to the probability r_0 the project is continued when she does not make the payment.

Analysis In order to avoid multiple cases, we will assume that

$$p_1 > \frac{p_2}{qp_2 + (1-p)L} K,$$

which will ensure that in the optimal contract, the Entrepreneur's first-period payment will satisfy $w_1^* < p_1$.

The Entrepreneur's problem is just a constrained maximization problem with a linear objective function and linear constraints, so it can in principle be easily solved using standard linear-programming techniques. We will instead solve the problem by thinking about a few perturbations that, at the optimum, must not be profitable. Taking this approach allows us to get some intuition for why the optimal contract will take the form it does.

First, we will observe that the Investor's break-even constraint must be binding in any optimal contract. To see why, notice that if the constraint were not binding, we could reduce the payment amount w_1 by a little bit and still maintain the break-even constraint. Reducing w_1 makes the incentive-compatibility constraint easier to satisfy, and it increases the Entrepreneur's objective function. This argument tells us that the Entrepreneur will receive all of the surplus the project generates, so her problem is to maximize that surplus.

The second observation is that in any optimal contract, the project is never liquidated following repayment. To see why, suppose $r_0 < r_1 < 1$ so that the project is continued with probability less than one following repayment. Consider an alternative contract in which r_1 is increased to $r_1+\varepsilon$, for $\varepsilon > 0$ small. Since making this change alone will violate the Investor's breakeven constraint, let us also increase w_1 by εL so that

$$w_1 + \varepsilon L + (1 - r_1 - \varepsilon) L = w_1 + (1 - r_1) L.$$

Under this perturbation, the Investor's breakeven constraint is still satisfied, and the Entrepreneur's incentive-compatibility constraint is satisfied as long as

$$(r_1 + \varepsilon - r_0) p_2 \geq w_1 + \varepsilon L,$$

which is true because $(r_1 - r_0) p_1 \geq w_1$ (or else the original contract did not satisfy IC) and $\varepsilon (p_2 - L) > 0$ since continuing the project is optimal (i.e., $p_2 > L$). If the original contract satisfied IC and IR, then so does this one, but this one also increases the Entrepreneur's objective by $q(-\varepsilon L + \varepsilon p_2)$, which again is strictly positive, since $p_2 > L$. This perturbation shows that increasing the probability of continuing the project following repayment is good for two reasons: it reduces the probability of inefficient liquidation, and it increases the Entrepreneur's incentives to repay.

Finally, the last step will be to show that the incentive constraint must bind at the optimum. It clearly must be the case that $r_0 < 1$, or else the incentive constraint would be violated. Again, suppose that the incentive constraint was not binding. Then consider a perturbation in which we raise r_0 to $r_0 + \varepsilon$, and to maintain the breakeven constraint, we increase w_1 to $w_1 + \varepsilon L(1-q)/q$. If the incentive constraint was not binding, then it will still be satisfied if r_0 is raised by a little bit. Lastly, this perturbation increases the Entrepreneur's payoff by

$$-q\left[\frac{\varepsilon L(1-q)}{q}\right] + (1-q)\,\varepsilon p_2 = (1-q)(p_2 - L)\,\varepsilon > 0.$$

In other words, if the incentive constraint is not binding, it is more efficient for the Entrepreneur to pay the Investor with cash than with an increased probability of liquidation, and since the Entrepreneur captures all the surplus, she will choose to pay in this more efficient way as much as she can.

To summarize, these three perturbations show that any optimal contract in this setting has to satisfy

$$(1 - r_0^*)\,p_2 = w_1^*$$

and

$$qw_1^* + (1-q)(1-r_0^*)\,L = K.$$

This is just two equations in two unknowns, so we can solve for the probability that the project is liquidated following nonpayment:

$$1 - r_0^* = \frac{K}{qp_2 + (1-q)\,L} > 0.$$

There is a complementarity between the repayment amount and the liquidation probability: if the project requires a lot of capital (i.e., K is large), then the Investor needs to be assured a bigger payment, and in order to assure that bigger payment, the project has to be liquidated with higher probability following nonpayment. If the project has high second-period cash flows (i.e., p_2 is high), then the Entrepreneur loses a lot following nonpayment, so the project does not need to be liquidated with as high of a probability to ensure repayment. Finally, if the liquidation value of the project is high, then the Investor earns more upon liquidation, so he can break even at a lower liquidation probability.

Under the first-best outcome, the project will never be liquidated, and the project will be undertaken as long as the expected cash flows exceed the required capital, or $qp_1 + p_2 > K$. The model features two sources of inefficiencies relative to the first-best outcome. First, in order to assure repayment, the Entrepreneur commits to a contract that with some probability inefficiently liquidates the project.

Second, there is credit rationing: the maximum amount the Entrepreneur can promise the Investor is p_2 in the event that output is high in the first period and L in the event that it is not, so if

$$qp_2 + (1-q)\,L < K < qp_1 + p_2,$$

the project will be one that should be undertaken but, in equilibrium, will not be undertaken. The liquidation value of the project is related to the collateral value of the assets underlying the project, and there is a literature beginning with Kiyotaki and Moore (1997) that endogenizes the market value of those assets and shows there can be important general equilibrium spillovers across firms.

11.4 Intermediation

Holmström and Tirole (1997) studies how financial intermediaries and other investors provide capital to entrepreneurs in a financial market. In their framework, financial intermediaries are different from other investors because they can monitor an entrepreneur's behavior, which reduces the rents that an entrepreneur must be promised in order to prevent her from shirking. Since the entrepreneur earns less, both financial intermediaries and other investors can be promised a larger share of the profit, which in turn expands access to credit.

The Model There are three players: an entrepreneur (E), an intermediary (M), and an uninformed investor (I) who interact once. The game has the following timing:

1. E's assets $A \sim G(\cdot)$ are realized and publicly observed.

2. E offers a contract to M and I, which specifies amounts A_M and A_I to borrow from each, and amounts R_M and R_I to be paid to each if the

project is successful.

3. M and I simultaneously accept or reject. If either rejects, or if $A_M + A_I + A < K$, the game ends and everyone receives 0.

4. M publicly chooses to monitor $(m = 1)$ or not $(m = 0)$.

5. E chooses to work $(e = 1)$ or shirk $(e = 0)$.

6. The project succeeds $(x = 1)$ or not $(x = 0)$, where $\Pr[x = 1|\, e] = p_L + (p_H - p_L)\, e$, and $p_H > p_L$. If the project succeeds, the entrepreneur receives output Y.

Given the contract, payoffs for the three players are

$$u_E = x\,(Y - R_M - R_I) + (1 - e)\,(mb + (1 - m)\,B)\,,$$

$$u_M = xR_M - \beta A_M - cm,$$

and

$$u_I = xR_I - A_I.$$

Two notes about these payoffs. First, if the entrepreneur shirks $(e = 0)$, then the project is less likely to succeed ($x = 1$ with lower probability). However, the entrepreneur earns a private benefit from shirking. Let $B > b$, so that this private benefit is lower if the intermediary monitors. We will assume that

$$p_L Y + B < K < p_H Y,$$

so it is socially efficient to fund a project only if the entrepreneur works. Second, the investor demands a return of 1 for every dollar loaned, while the intermediary demands a return of β. We will assume $\beta \geq 1$. In the paper, β is pinned down by a market for "informed" and "uninformed" capital—that is, capital that is owned by an intermediary or by an investor.

This is a game with perfect monitoring, so our solution concept is Subgame-Perfect Equilibrium.

Analysis The entrepreneur is willing to work only if

$$Y - R_M - R_I \geq \frac{mb + (1-m)\,B}{p_H - p_L},$$

which limits the amount of money that can be promised to the intermediary and investor in equilibrium. The entrepreneur essentially has two choices when offering a contract: she can either offer a contract that induces the intermediary to monitor or not.

Suppose the contract does not induce the intermediary to monitor. Since $\beta \geq 1$, the entrepreneur will optimally borrow only from the investor. She can fund the project in this way if and only if the proceeds "left over" after giving the entrepreneur the necessary incentives are enough to cover the shortfall in the required funds:

$$p_H\left(Y - \frac{B}{p_H - p_L}\right) \geq K - A.$$

Therefore, if A is sufficiently large, then the entrepreneur can borrow funds from the uninformed investors alone.

Next, consider a contract that induces the intermediary to monitor. The intermediary is willing to do so only if

$$R_M \geq \frac{c}{p_H - p_L},$$

so that its additional return from monitoring (and thereby inducing the entrepreneur to work rather than shirk) is larger than the cost of monitoring. Therefore, the investor can be induced to contribute any A_I that satisfies

$$p_H \left(Y - \frac{b}{p_H - p_L} - \frac{c}{p_H - p_L} \right) \geq A_I.$$

The intermediary earns

$$\frac{p_H c}{p_H - p_L} - c$$

from this contract, which is strictly positive if $p_L > 0$. In other words, the intermediary is earning a rent from monitoring. The entrepreneur can therefore demand that the intermediary invest some capital in this project in exchange for these rents. That is, the intermediary is willing to contribute any loan A_M such that

$$\frac{p_H c}{p_H - p_L} - c \geq \beta A_M.$$

Since $\beta \geq 1$, the entrepreneur does not want to borrow more than this amount from the intermediary. So the project can be funded with an intermediary if and only if

$$\frac{p_L}{\beta}\frac{c}{p_H - p_L} + p_H\left(Y - \frac{b}{p_H - p_L} - \frac{c}{p_H - p_L}\right) + A \geq K.$$

By inducing an intermediary to monitor, the entrepreneur can minimize her own temptation to shirk ex post, which decreases the share of the pie that she must earn in equilibrium. If the intermediary did not need to be paid, then more money would be left over for the investor, which in turn increases the investor's willingness to loan funds. Of course, the intermediary does need "skin in the game," since otherwise, it would not monitor. Therefore, the cost of inducing monitoring is the rent that must be paid to the intermediary. However, this cost can be mitigated by demanding the intermediary contribute up-front money to the firm in exchange for its future monitoring rents. This mitigation is imperfect so long as $\beta > 1$; that is, the intermediary has a higher opportunity cost of loaning funds to the entrepreneur, relative to the investor.

Why might it be the case that $\beta > 1$? The paper considers how β is optimally set in a financial market. Demand is determined by a continuum of entrepreneurs with different initial asset levels. Intermediaries have a fixed supply of capital. If intermediary capital is scarce, then intermediaries earn some rent from loaning capital to firms in exchange for monitoring. Since the opportunity cost of loaning to one entrepreneur is loaning to another entrepreneur instead, intermediaries demand higher-than-market interest rates in exchange for their funds. Essentially, $\beta > 1$ if intermediary capital is scarce, even if total capital is not.

Holmström and Tirole (1997) apply this logic to think about how firms optimally respond to different kinds of credit crunches. For example, suppose firms suffer a collateral shock (A decreases). Firms with low capital might no longer be able to borrow enough funds. Better-capitalized firms might continue operating but be forced to start relying on intermediaries to monitor their activities. Unfortunately, the model stated above is not well-suited to asking such questions because the number of entrepreneurs who move into the "rely on intermediary zone" depends on $G(\cdot)$, and hence β might increase or decrease in response to this shock. The paper considers a model with a continuous investment choice, which is more amenable to getting clean comparative statics on interest rates and capital demand.

11.5 Repayment and Firm Dynamics

We now turn to a dynamic model of financing relationships. Clementi and Hopenhayn (2006) studies a repeated relationship between an entrepreneur and an investor, with the goal of exploring how borrowing constraints affect firm dynamics and growth. Consider an entrepreneur who must fund an initial project, then repeatedly purchase short-term working capital that determines the profitability of that project. The entrepreneur relies on an external investor to provide both the initial loan for the project and period-by-period funds for working capital. As in Hart (1995), the key contracting friction is that realized revenue is unobservable, so the entrepreneur must be induced to truthfully report it (and to make a repayment if output is high).

The Model Consider the following dynamic contracting problem between an entrepreneur and an investor. At the start of the game, the parties sign a long-term contract that specifies an up-front loan $K > 0$ and other contractible variables, discussed below. In each period,

1. The firm is liquidated ($\ell_t = 1$) or not ($\ell_t = 0$). The investor can pay the entrepreneur $Q_t \geq 0$ following liquidation, after which the game ends.

2. The investor chooses capital $k_t \in \mathbb{R}_+$.

3. Output $y_t \in \{0, R(k_t)\}$ is realized, with $\Pr[y_t = R(k_t)] = p$. Assume $R(0) = 0$ and $R(\cdot)$ is strictly increasing and strictly concave.

4. The entrepreneur chooses a message $m_t \in \{0, 1\}$.

5. The entrepreneur and investor exchange payments, with net payment to investor $\tau_t \leq y_t$.

Payoffs are $u_E = \ell_t Q_t + (1 - \ell_t)(y_t - \tau_t)$ for the entrepreneur and $u_I = \ell_t (S - Q_t) + (1 - \ell_t)(\tau_t - k_t)$ for the investor. We assume that the value of liquidation is strictly positive, $S > 0$, and that a contractible public randomization device is realized at the start of each period.

Everything except realized output y_t is contractible, and so players can commit to it as a function of the contractible history. Denote by h_0^t the history of contractible variables at the start of a period t, with h_x^t as the contractible history following the realization of a variable x in period t. Therefore, the long-term contract specifies liquidation probabilities $\alpha_t\left(h_0^t\right)$, payment to the entrepreneur upon liquidation $Q_t\left(h_0^t\right)$, working capital $k_t\left(h_\ell^t\right)$, and payment $\tau_t\left(h_m^t\right)$ in each period.

Our goal is to characterize the set of long-term contracts that maximize the investor's payoff, conditional on giving the entrepreneur some repayment guarantee. Let $B(U)$ be the maximum payoff that the investor can earn, given that the entrepreneur earns U. Then the desired contract solves:

$$B(U) = \max_{\alpha,Q,k,\tau,U_L,U_H} \alpha(S-Q) + (1-\alpha)\left(p\tau + \delta p B(U_H) + \delta(1-p)B(U_L)\right)$$

subject to the following three constraints:

1. Feasibility: $\alpha \in [0,1]$, $Q \geq 0$, $k \geq 0$, $\tau \leq R(k_t)$, $U_L, U_H \geq 0$.

2. Promise-keeping:

$$U = \alpha Q + (1-\alpha)\left(p\left(R(k) - \tau\right) + \delta p U_H + \delta(1-p)U_L\right).$$

3. Incentive-compatibility:

$$\delta\left(U_H - U_L\right) \geq \tau.$$

Here, U_H and U_L are the entrepreneur's promised continuation payoffs if she reports high or low output, respectively. It turns out that the entrepreneur's continuation payoffs, U_L and U_H, are attainable in a continuation contract if and only if they are nonnegative. This result is a consequence of two features of the problem. First, the investor can commit to earn negative continuation payoffs. Second, the entrepreneur can always guarantee a payoff of 0 by reporting no output in every period.

Analysis We can interpret the incentive-compatibility constraint as the outcome of a moral hazard problem, with the (slight) twist that the moral hazard problem arises after output is realized. Therefore, the entrepreneur can condition her action on realized output, which is ruled out by the standard formulation of private effort.[2] Given that this problem is essentially a moral hazard problem, the fact that the entrepreneur is liability-constrained suggests that the optimal contract might entail dynamics and sequential inefficiencies. Our goal is to characterize the payoff frontier.

Define k^{FB} as the capital level that maximizes total expected surplus,

$$pR'\left(k^{FB}\right) = 1.$$

[2] Such models are sometimes referred to as repeated adverse selection models.

Consider the incentive constraint $\delta\left(U_H - U_L\right) \geq \tau$. What is the cost of demanding a large repayment τ? Given the presence of a public randomization device, $B(\cdot)$ is concave. If it is strictly concave, then

$$pB\left(U_H\right) + \left(1-p\right)B\left(U_L\right) \leq B\left(pU\left(H\right) + \left(1-p\right)U_L\right).$$

Therefore, introducing dispersion in the entrepreneur's continuation payoffs U_H and U_L weakly decreases the continuation payoff that can be promised to the investor.

First, let us consider when k^{FB} can be implemented in every future period. The investor has deep pockets, so if $k_t = k^{FB}$ in every future period when the entrepreneur earns $\tilde{U}$, $k_t = k^{FB}$ in every period for any $U > \tilde{U}$. Basically, the investor can always pay the entrepreneur more without affecting efficiency. So our goal is to find the minimum $\tilde{U}$ that attains first-best working capital. Continuation play must also attain first-best, so $U_H, U_L \geq \tilde{U}$ as well. Moreover,

$$\begin{aligned}
\tilde{U} &= p\left(R\left(k^{FB}\right) - \tau\right) + \delta p U_H + \delta\left(1-p\right)U_L \\
&= pR\left(k^{FB}\right) + p\left(\delta\left(U_H - U_L\right) - \tau\right) + \delta U_L \\
&\geq pR\left(k^{FB}\right) + \delta U_L,
\end{aligned}$$

where the inequality follows from the incentive-compatibility constraint. So

$$\tilde{U} \geq \frac{pR\left(k^{FB}\right)}{1-\delta}$$

is a necessary and sufficient condition for $k_t = k^{FB}$ in every future period. Intuitively, the entrepreneur earns the entire expected output in every period and does not bear the cost of financing. Note that

$$B\left(U\right) = \frac{1}{1-\delta}\left(pR\left(k^{FB}\right) - k^{FB}\right) - \tilde{U} = -\frac{1}{1-\delta}k^{FB} < 0.$$

Therefore, at this point the entrepreneur is essentially "living off her savings": she has paid so much to the investor that he has already more than recouped the initial loan.

Initially, we want $B\left(U\right) \geq K$ so that the initial loan is repaid. Therefore, $k < k^{FB}$ in the first few periods. To understand the dynamics of the optimal contract, first note that $B\left(U\right) + U$ should be increasing in U, since the investor has deep pockets and so can efficiently increase the entrepreneur's payoff. Next, consider how to punish the entrepreneur as efficiently as possible. Given that $R\left(0\right) = 0$, the lowest payoff the entrepreneur can earn is 0. One way to attain this punishment payoff is to set $k = 0$ in every period, which results in both players receiving payoff 0. But a more efficient way to punish the entrepreneur is to liquidate the firm, which gives the investor a strictly positive payoff $S > 0$. By the same logic, for $U \approx 0$, it is efficient to liquidate the firm with some probability $\alpha \in (0, 1)$. Note that this punishment potentially entails a loss in both players' surpluses, in which case there is an upward-sloping part of the payoff frontier. It might not, however, if S is sufficiently large, even if $S < pR\left(k^{FB}\right) - k^{FB}$, since $B\left(U\right)$ is maximized for some $k < k^{FB}$.

Two further results about the dynamics. First, it is optimal for the contract to specify $\tau = R\left(k\right)$, as long as $U_H \leq \tilde{U}$, since for $\tau < R\left(k^{FB}\right)$, we could increase τ and U_H so that the incentive-compatibility constraint holds. This perturbation gives the investor a strictly higher payoff, since $B\left(U\right) + U$ is increasing in U and so $B\left(U_H\right)$ decreases by less than τ increases. Second, given that the entrepreneur earns no profit until first-best is attained, and that the incentive-compatibility constraint will bind to minimize variation in continuation payoffs, we have

$$U = \delta p U_H + \delta\left(1-p\right)U_L = p\left(\delta U_L + R\left(k\right)\right) + \delta\left(1-p\right)U_L,$$

so that

$$U_L = \frac{U - pR\left(k\right)}{\delta}.$$

Similarly,

$$U_H = \frac{U + \left(1-p\right)R\left(k\right)}{\delta}.$$

When $U < \tilde{U}$, it can be shown that $U_L < U < U_H$. Hence, total surplus and working capital increase following repayment and decrease following non-repayment. A sequence of low-output periods results in eventual liquidation, while a sequence of high-output periods results in attaining first-best.

11.6 Internal Capital Markets (TBA)

Chapter 12

Institutions

This chapter considers a set of organizations that we will call *institutions*. Institutions are "organizations of organizations" that set the "rules of the game" for interactions between other organizations and individuals. Of course, every organization has some power over the rules of the game, whether that organization is a firm that uses incentive contracts and information design to coordinate its workforce or a government that enforces formal contracts. Therefore, this definition is, by necessity, imprecise. But we should think of an institution as being "one step up" from a firm, in that it sets the rules within which firms (and workers, customers, and other stakeholders) operate. So a firm writes an optimal incentive contract given a set of contractible variables, while an institution determines *which* variables are contractible within a court system.

A wide variety of different kinds of institutions exist, from governments to business associations to informal credit-sharing networks to groups that regulate access to public or common-pool goods. These disparate groups are linked together by two common roles. First, institutions gather and disseminate information about both the state of the world and the actions of various parties. Second, institutions enforce contracts, rules, and norms that are either set by the institution itself or by parties under the shadow of that institution. These two roles can manifest in a variety of ways. A government gathers and disseminates information and enforces contracts. So does the Mafia, and indeed, the Mafia historically served as a substitute for government institutions (see, for example, Gambetta (1996)). So does a business association that levies sanctions against members who do not live up to the standards of the association (see Menard (2013)). So does a prison gang (see Skarbek (2014)). While the objective and scope of these institutions varies dramatically, their fundamental roles do not.

How does an institution enforce contracts? Fundamentally, by relying on *communal sanctions.* That is, each party is willing to follow the terms of the contract because they suffer sanctions from the rest of the community if they do not do so. In many settings, these sanctions are the purview of specialized agents of the institution, as with law enforcement and the courts in many governments. In other cases, sanctions are carried out by members of the community, as when a firm that reneges on an agreement is ostracized from a business association. Importantly, one of the "contracts" enforced by many institutions is a property rights contract, which specifies who owns a given object and what rights ownership entails.

12.1 Informal Enforcement

This section will focus on the arguments of Greif, Milgrom, and Weingast (1994), which is one of a series of papers authored by subsets Grief, Milgrom, North, and Weingast that combine historical evidence and simples theories, focusing on the dramatic expansion of long-distance trade in Europe during the 10th to the 14th centuries. During this period, trade had to overcome serious credibility issues; in particular, trading networks extended between political borders, so foreign merchants had little formal protection against rules who wanted to steal their goods. Essentially, local rules were unable to commit themselves not to appropriate goods, which deterred merchants from bringing those goods to the city at all. This problem was recognized by rules at the time; the English king Edward I noted in 1283 that "many merchants are put off from coming to this land with their merchandise to the detriment of merchants and of the whole kingdom."

The paper argues that foreign traders formed *merchant guilds* to overcome this commitment problem. Merchant guilds coordinated punishments in the form of boycotts against rules who reneged on a contract, which helped rulers commit to honor their agreements. To successfully punish a rule, guilds also had to coordinate punishments of those members who did not abide by the terms of the boycott. So guilds served an essential role in making long-distance trade feasible. Indeed, the paper provides evidence that boycotts, and punishment of those merchants who broke the boycotts, were used after a ruler reneged on a contract.

The Model A single ruler and a unit mass of traders interact repeatedly, and all players share a common discount factor of $\delta < 1$. The timing of the stage game is as follows

1. Each trader $i \in [0,1]$ chooses to send a product $(s_{i,t} = 1)$ to the ruler or not $(s_{i,t} = 0)$. This choice is publicly observed.

2. The ruler chooses an amount of money, $b_{i,t} \in \mathbb{R}_+$, to send to each trader. These transfers are publicly observed.

3. Each pair of traders, $i, j \in [0,1]$ chooses to either trade with each other or not. Denote $x_{i,j,t} = 1$ if and only if i and j both choose to trade with each other.

We will focus on what is effectively a subgame-perfect equilibrium, with the caveat that the continuum of players introduces measurability issues that we are ignoring.

Define $S_t = \int_0^1 s_{i,t} di$, $B_t = \int_0^1 b_{i,t} di$, and $X_{i,t} = \int_0^1 x_{i,j,t} dj$ be the measures of traders who send a product to the ruler, the total payment from the ruler to all traders, and the total measure of trade between i and other traders.

Then the ruler's payoff is

$$\pi_t = f(S_t) - B_t,$$

where $f(\cdot)$ is strictly increasing and strictly concave, while trader i's payoff is

$$u_{i,t} = b_{i,t} - cs_{i,t} + vX_{i,t}.$$

This game amounts to a set of trust games between the ruler and each trader i, with the trader serving as the "truster" and the ruler as the "trustee." In particular, once a trader has chosen $s_{i,t} = 1$, the ruler is tempted to choose $b_{i,t} = 0$. If the trader believes this will happen, she will not choose $s_{i,t} = 1$ in equilibrium. Repeated-game incentives can potentially induce the ruler to pay $b_{i,t} \geq c$, which will get the trader to choose $s_{i,t} = 1$.

The extra twist on this game is the value $v > 0$ that is realized if traders i and j choose to trade. This separate trade decision is a coordination game with two equilibria: one in which both choose to trade, and one in which neither chooses to trade. The coordination-game aspect of this problem is for simplicity: we could instead imagine each pair of traders playing a repeated Prisoner's Dilemma with each other. The paper argues that v should be interpreted as the monopoly rents that the guild earns in its town of origin; indeed, many guilds exercised quasi-monopoly or monopoly power during this period.

Analysis The paper considers several different equilibrium regimes that are designed to capture features of the historical context. In particular, the paper interprets a "guild" as a particular kind of equilibrium. Define the first-best level of S_t, S^{FB}, by $f'\left(S^{FB}\right) = c$.

Bilateral Agreements First, consider "bilateral agreements," defined as an equilibrium with the following restriction: for each $i \in [0,1]$, following a deviation in $b_{i,t}$, only $s_{i,t}$ can change. Then it follows immediately that $X_{i,t} = 1$ at every history both on and off the equilibrium path in any Pareto-optimal equilibrium. The main result is that first-best cannot be attained: there is no equilibrium in which $S_t = S^{FB}$ with probability 1 in every period t on the equilibrium path, regardless of δ.

To see why this result is true, note that cooperation incentives are strongest if the same set of traders choose $s_{i,t} = 1$ in every period on the equilibrium path, and if the ruler exactly compensates each of those traders: $b_{i,t} = c$. Suppose that the measure of such traders equals S^{FB} in every period on the equilibrium path. Then the principal must prefer to pay c to every trader rather than reneging on any subset of them, which requires

$$f\left(S^{FB}\right) - cS^{FB} \geq (1-\delta)\left(f\left(S^{FB}\right) - cS\right) + \delta\left(f(S) - cS\right)$$

for any $S < S^{FB}$. This condition can be rearranged to get

$$\delta\frac{f\left(S^{FB}\right) - f(S)}{S^{FB} - S} \geq c.$$

As $S \uparrow S^{FB}$, the left-hand side of this inequality converges to $\delta f'\left(S^{FB}\right)$, so a necessary condition for this inequality to hold for all $S < S^{FB}$ is that

$$\delta f'\left(S^{FB}\right) \geq c,$$

but $\delta < 1$ and $f'\left(S^{FB}\right) = c$, so this inequality cannot hold.

The intuition behind this result is that, at S^{FB}, the marginal benefit to the ruler of interacting with each trader equals his marginal cost. The ruler suffers a second-order loss in the future by reneging on the marginal trader today, but reneging yields a first-order gain of c, so the ruler has no incentive to follow through on her promised payment.

Boycott without Renegotiation Now, suppose that all traders can condition $s_{i,t}$ on the full history. Then, following any deviation by the ruler, it is without loss of generality to set $S_t = 0$ in the optimal equilibrium. Therefore, the ruler's strongest temptation to deviate comes from reneging on everyone, and hence first-best can be sustained if

$$\delta f\left(S^{FB}\right) \geq cS^{FB}.$$

This inequality holds if δ is sufficiently close to 1, since $f\left(S^{FB}\right) > cS^{FB}$ whenever trade is valuable. Notice that multilateral punishment effectively alters the principal's no-reneging from one that depends on margins to one that depends on averages.

Boycott with Renegotiation One potentially unnatural feature of the boycott without renegotiation is that the ruler could credibly promise to reward some traders to continue sending him goods. In particular, if $f'(0)$ is sufficiently large, then there exists an $0 < S^* < S^{FB}$ such that the equilibrium in bilateral agreements supports $S_t = S^*$ in each period. As a result, if S^* traders break the boycott, why doesn't the ruler offer those S^* traders a reward for doing so? Such a reward (of $b_{i,t} = c$) would be credible, even if it was supported only by bilateral punishment.

The paper argues that such renegotiation could be partially deterred if the traders sanctioned deviations from the boycott. In particular, if trader i continues to send goods to the ruler, then all $j \neq i$ can choose $x_{i,j,t} = 0$ and thereby deny trader i a continuation payoff equal to δv. Then, in order to induce such a trader to actually send a good to the ruler, the ruler's promised payment would have to satisfy $b_{i,t} \geq v + c$ in every subsequent period. In effect, the guild increases the opportunity cost of traders sending goods to a ruler under boycott by denying them access to the monopoly rents in their hometown.

Note that the ruler might still be able to support some trade in a boycott but could only do so at a premium price and with substantially less volume than even S^*. Note also that these sanctions are more effective if v is large, which gives an efficiency-enhancing rationale for why guilds pursued monopoly power in medieval Europe.

12.2 Distance and Enforcement

Geographical distance plays an important role in the functioning of community enforcement (Dixit, 2003). On the one hand, players who are located far from each other might have more to offer each other, which improves the gains and hence effectiveness of community enforcement. On the other hand, information about players' past behavior might be more difficult to obtain for more distant players, making it more difficult to sustain community enforcement over wider distances.

The Model A continuum of players, indexed by $i \in [0, 2L]$, is uniformly distributed around a circle with circumference $2L$. We will refer to the "distance" between two players as the length of the shorter of the two arcs that connect those players along the circle. Define

$$d(i,j) = \min\{i - j, 2L - (i - j)\}$$

as this distance. The players interact repeatedly, and the timing of the stage game is as follows.

1. Each player i is bilaterally matched to exactly one other player j. The distribution over matches for i is given by $p(d(i,j))$, where $\int_0^{2L} p(d(i,j))\, dj = 1$ and $p(\cdot)$ is strictly decreasing.

2. Each matched pair (i, j) plays a Prisoner's Dilemma with payoff matrix

	C	D
C	c, c	$-1, r$
D	$r, -1$	$0, 0$

   where $r > c > 0$. Let v_i be player i's realized payoff.

3. Each player i is bilaterally matched to a second player k, with distribution $p(d(i, k))$.

4. For each i, with probability $q(d(j, k))$, players i and k observe a public signal s_i equal to the player's action in the Prisoner's Dilemma.

5. Each pair (i, k) play a coordination game with payoff matrix

	0	1
0	L, L	$0, 0$
1	$0, 0$	H, H

   where $H > L \geq 0$. Let w be player i's payoff from the coordination game.

Player i's payoff is $B(d(i, j))\, v_i + w_i$, where B is strictly increasing. The solution concept, roughly speaking (again because there is a continuum of players), is perfect public equilibrium.

Analysis Consider an equilibrium of the following form: in the coordination game, players play $(1,1)$ unless they observe a public signal that indicates a deviation in the Prisoner's Dilemma, in which case they play $(0,0)$. In the Prisoner's Dilemma, players play (C,C) if $d(i,j) \leq X$ and (D,D) otherwise, for some $X \leq L$. We will solve for the optimal value of X. Note that this is not necessarily the optimal equilibrium; in particular, an alternative equilibrium might specify $(0,0)$ in the coordination game unless $s_i = s_k = C$. The paper justifies looking at this equilibrium form using a reputation game: basically, rather than playing a coordination game, players opt in or out of a game, and they strictly prefer to opt in unless their partner has revealed himself to be a "Machiavellian type" by deviating in the first period.

A player's temptation to deviate is increasing in the distance to her partner for two reasons. First, B increasing in distance implies that the gains from deviating are larger for more distant players. Second, the informative signal s_i is more likely to be observed by players close to player j. But those players are far from i if j is distant, which means that i is unlikely to be paired with them for the coordination game. Therefore, i is unlikely to be punished for choosing D in the first period.

The benefit of deviating and choosing D in the first period equals $B(d(i,j))(r-c) > 0$. The cost of doing so is that, if player k observes that deviation, then k punishes in the second period. The cost of this punishment is therefore

$$\int_0^{2L} p(d(i,k))\, q(d(j,k))\,(H-L)\,dk.$$

Therefore, player i will not deviate from a cooperation action as long as

$$r - c \leq \int_0^{2L} p\left(d\left(i,k\right)\right) \frac{q\left(d\left(j,k\right)\right)}{B\left(d\left(i,j\right)\right)} \left(H - L\right) dk.$$

We want to argue that the left-hand side is decreasing in $d\left(i,j\right)$. Here is a loose argument. For simplicity, assume that $i < j$ and that $d\left(i,j\right) = j - i$. Then increasing j increases $d\left(i,j\right)$. We know $1/B$ is increasing in $d\left(i,j\right)$, so we require:

$$\int_0^{2L} \left(p\left(d\left(i,k\right)\right) q'\left(d\left(j,k\right)\right) \frac{\partial d\left(j,k\right)}{\partial j} \right) dk < 0.$$

This integral can be broken up into two pieces:

$$\int_j^{j+L} p\left(d\left(i,k\right)\right) q'\left(d\left(j,k\right)\right) \frac{\partial d\left(j,k\right)}{\partial j} dk + \int_{j+L}^{j+2L} p\left(d\left(i,k\right)\right) q'\left(d\left(j,k\right)\right) \frac{\partial d\left(j,k\right)}{\partial j} dk,$$

where we adopt the labeling that $r + 2L$ is the same player as r. Note that $\partial d\left(j,k\right)/\partial j > 0$ for $k \in [j+L, j+2L]$, $\partial d\left(j,k\right)/\partial j < 0$ otherwise, and $\int_0^{2L} \partial d\left(j,k\right)/\partial j = 0$. Furthermore, $q'\left(d\left(j,k\right)\right)$ is symmetric for $k \in [j, j+L]$ and $k \in [j+L, j+2L]$. But $p\left(d\left(i,k\right)\right)$ is larger at every $k \in [j+L, j+2L]$ than every $k \in [j, j+L]$. So the second term in this expression is negative and larger than the first term, and hence equilibrium can sustain cooperation only if players are sufficiently close to each other. Hence, the most efficient equilibrium is characterized by an $X \in [0, L]$ that gives the *maximum distance* at which a player cooperates.

Consider varying L, the "size of the world," while "holding q constant" for a given arc length. However, we cannot hold p constant, at least if we want every player to have the chance to interact with every other player. One natural approach is to decrease p for low distances as L increases and use the "leftover mass" to assign some probability that very distant people meet. But these very distant people are unlikely to have heard of a deviation in period 1, so this approach effectively shifts mass from players who are likely to punish a deviation to players who are unlikely to punish a deviation. Hence, a player's temptation to deviate increases as we adjust L in this way, which actually decreases the maximum cooperation distance. That is, cooperation becomes strictly harder to sustain at a fixed distance d as L increases.

The model can also speak to when we might expect to see the rise of formal enforcement mechanisms. Suppose that, at the start of the game, the parties can agree to impose a formal enforcement schedule. Setting up such a schedule costs each player $k > 0$, but once implemented, it perfectly deters any deviation in period 1. How does the decision to implement formal enforcement interact with the size of the world?

If the world is small, then $X = L$, so informal enforcement performs as well as formal enforcement. As L grows, however, the argument above suggests that X shrinks, until eventually $X < L$ for large enough L. At that point, the surplus from informal enforcement is strictly decreasing in L, while the gains from formal enforcement continue to increase (because increasing L makes it possible to be paired with a partner for which B is large). Eventually, the values of the two regimes cross. Consequently, for large enough L, the society will choose to implement formal enforcement even though there are no cost-based economies of scale from doing so.

Bibliography

[1] Abreu, Dilip. 1988. On the Theory of Infinitely Repeated Games with Discounting. *Econometrica*, 56(2): 383-396.

[2] Acemoglu, Daron. 2009. *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] Admati, Anat and Motty Perry. 1991. Joint Projects without Commitment. *Review of Economic Studies*, 58(2): 259-276.

[5] Aghion, Philippe and Patrick Bolton. 1992. An Incomplete Contracts Approach to Financial Contracting. *Review of Economic Studies*, 59(3): 473-494.

[6] 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.

[7] 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.

[8] Aghion, Philippe and Jean Tirole. 1997. Formal and Real Authority in Organizations. *Journal of Political Economy*, 105(1): 1-29.

[9] Alchian, Armen and Harold Demsetz. 1972. Production, Information Costs, and Economic Organization. *American Economic Review*, 62(5): 777-795.

[10] Alfaro, Laura, Paola Conconi, Harald Fadinger, and Andrew Newman. 2016. Do Prices Determine Vertical Integration? *Review of Economic Studies*, 83(3): 855-888.

[11] Alonso, Ricardo, Wouter Dessein, and Niko Matouschek. 2008. When Does Coordination Require Centralization? *American Economic Review*, 98(1): 145-179.

[12] Alonso, Ricardo and Niko Matouschek. 2008. Optimal Delegation. *Review of Economic Studies*, 75(1): 259-293.

[13] ANDREWS, ISAIAH, AND DANIEL BARRON. 2016. The Allocation of Future Business: Dynamic Relational Contracts with Multiple Agents. *American Economic Review*, 106(9): 2742-2759.

[14] ANTRAS, POL. 2003. Firms, Contracts, and Trade Structure. *Quarterly Journal of Economics*, 118(4): 1375-1418.

[15] ARROW, KENNETH. 1974. *The Limits of Organization.* WW Norton.

[16] 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

[17] AVENEL, ERIC. 2008. Strategic Vertical Integration Without Foreclosure. *Journal of Industrial Economics*, 56(2): 247-262.

[18] BACKUS, MATTHEW. 2014. Why is Productivity Correlated with Competition? Mimeo.

[19] BAJARI, PATRICK AND STEVEN TADELIS. Incentives versus Transaction Costs: A Theory of Procurement Contracts. *RAND Journal of Economics*, 32(3): 387-407.

[20] BAKER, GEORGE. 1992. Incentive Contracts and Performance Measurement. *Journal of Political Economy*, 100(3): 598-614.

[21] Baker, George. 2002. Distortion and Risk in Optimal Incentive Contracts. *Journal of Human Resources*, 37(4): 728-751.

[22] 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.

[23] Baker, George, Michael Gibbs, and Bengt Holmstrom. 1994b. The Wage Policy of a Firm. *Quarterly Journal of Economics*, 109(4): 921-955.

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

[25] Baker, George, Robert Gibbons, and Kevin Murphy. Informal Authority in Organizations. *Journal of Law, Economics, and Organization*, 15(1): 56-73.

[26] Baker, George, Robert Gibbons, and Kevin Murphy. Relational Contracts and the Theory of the Firm. *Quarterly Journal of Economics*, 117(1): 39-84.

[27] Barron, Daniel, George Georgiadis, and Jeroen Swinkels. 2017. Optimal Contracts with a Risk-Taking Agent. Mimeo.

[28] Barron, Daniel, Robert Gibbons, Ricard Gil, and Kevin Murphy. 2015. Relational Contracts under Reel Authority. Mimeo.

[29] BARRON, DANIEL AND MICHAEL POWELL. 2018. Policies in Relational Contracts. Mimeo.

[30] BATTIGALLI, PIERPAOLO AND GIOVANNI MAGGI. 2002. Rigidity, Discretion, and the Costs of Writing Contracts. *American Economic Review*, 92(4): 798-817.

[31] BERNHARDT, DAN. 1995. Strategic Promotion and Compensation. *Review of Economic Studies*, 62(2): 315-339.

[32] BERNHEIM, DOUGLAS AND MICHAEL WHINSTON. Common Agency. *Econometrica*, 54(4): 923-942.

[33] BERNHEIM, DOUGLAS AND MICHAEL WHINSTON. Menu Auctions, Resource Allocation, and Economic Influence. *Quarterly Journal of Economics*, 101(1): 1-32.

[34] BESLEY, TIMOTHY AND MAITREESH GHATAK. 2001. Government Versus Private Ownership of Public Goods. *Quarterly Journal of Economics*, 116(4): 1343-1372.

[35] BOARD, SIMON. 2011. Relational Contracts and the Value of Loyalty. *American Economic Review*, 101(7): 3349-3367.

[36] BOLTON, PATRICK AND MATHIAS DEWATRIPONT. 1994. The Firm as a Communication Network. *Quarterly Journal of Economics*, 109(4): 809-839.

[37] Bonatti, Alessandro and Johannes Horner. Forthcoming. Career Concerns with Exponential Learning. *Theoretical Economics*.

[38] Bond, Timothy. 2015. Internal Labor Markets in Equilibrium. Mimeo.

[39] Bose, Gautam and Kevin Lang. 2015. Monitoring for Worker Quality: Task Assignment, Job Ladders, Wages and Mobility in Nonmanagerial Internal Labor Markets. Mimeo.

[40] Bowen, Renee, George Georgiadis, and Nicolas Lambert. Forthcoming. Collective Choice in Dynamic Public Good Provision. *American Economic Journal: Microeconomics*.

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

[42] Calvo, Guillermo and Stanislaw Wellisz. 1978. Supervision, Loss of Control, and the Optimum Size of the Firm. *Journal of Political Economy*, 86(5): 943-952.

[43] Calvo-Armengol, Antoni, Joan de Marti, and Andrea Prat. 2015. Communication and Influence. *Theoretical Economics*, 10(2): 649-690.

[44] Camara, Odilon and Dan Bernhardt. 2009. The Dynamics of Promotions, Quits, and Layoffs. Mimeo.

[45] Carroll, Gabriel. 2015. Robustness and Linear Contracts. *American Economic Review*, 105(2): 536-563.

[46] Chaigneau, Pierre, Alex Edmans, and Daniel Gottlieb. 2015. The Informativeness Principle Under Limited Liability. Mimeo.

[47] Chan, William. 1996. External Recruitment versus Internal Promotion. *Journal of Labor Economics*, 14(4): 555-570.

[48] Che, Yeon-Koo and Navin Kartik. 2009. Opinions as Incentives. *Journal of Political Economy*, 117(5): 815-860.

[49] Cisternas, Gonzalo. 2016. Two-Sided Learning and Moral Hazard. Mimeo.

[50] Clementi, Gian Luca and Hugo Hopenhayn. 2006. A Theory of Financing Constraints and Firm Dynamics. *Quarterly Journal of Economics*, 121(1): 229-265.

[51] Coase, Ronald. 1937. The Nature of the Firm. *Economica*, 4(16): 386-405.

[52] Coase, Ronald. 1960. The Problem of Social Cost. *Journal of Law and Economics*, 3: 1-44.

[53] Crawford, Vincent and Joel Sobel. 1982. *Econometrica*, 50(6): 1431-1451.

[54] Cvitanic, Jaksa and George Georgiadis. 2016. Achieving Efficiency in Dynamic Contribution Games. *American Economic Journal: Microeconomics*, 8(4): 309-342.

[55] Cyert, Richard and James March. 1963. *A Behavioral Theory of the Firm*. Englewood Cliffs.

[56] Demougin, Dominique and Aloysius Siow. 1994. Careers in Ongoing Hierarchies. *American Economic Review*, 84(5): 1261-1277.

[57] Demsetz, Harold. 1988. The Theory of the Firm Revisited. *Journal of Law, Economics, and Organization*. 4(1): 141-161.

[58] Dessein, Wouter. 2002. Authority and Communication in Organizations. *Review of Economic Studies*, 69(4): 811-838.

[59] Dessein, Wouter, Andrea Galeotti, and Tano Santos. 2016. Rational Inattention and Organizational Focus. *American Economic Review*, 106(6): 1522-1536.

[60] DeVaro, Jed and Hodaka Morita. 2013. Internal Promotion and External Recruitment: A Theoretical and Empirical Analysis. *Journal of Labor Economics*, 31(2): 227-269.

[61] Dewan, Torun and David Myatt. 2008. The Qualities of Leadership: Direction, Communication, and Obfuscation. *American Political Science Review*, 102(3): 351-368.

[62] 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.

[63] 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.

[64] 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.

[65] DEWATRIPONT, MATHIAS AND JEAN TIROLE. 2005. Modes of Communication. *Journal of Political Economy*, 113(6): 1217-1238.

[66] DIAMOND, PETER. 1998. Managerial Incentives: on the Near Linearity of Optimal Compensation. *Journal of Political Economy*, 106(5): 931-957.

[67] DIXIT, AVINASH. 2003. Trade Expansion and Contract Enforcement. *Journal of Political Economy*, 111(6): 1293-1317.

[68] 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.

[69] Doeringer, Peter and Michael Piore. 1971. *Internal Labor Markets and Manpower Analysis*. Heath Lexington Books.

[70] Dye, Ronald. Costly Contract Contingencies. *International Economic Review*, 26: 233-250.

[71] Fehr, Ernst, Michael Powell, and Tom Wilkening. 2016. Handing out Guns at a Knife Fight: Behavioral Limitations of Subgame-Perfect Implementation. Mimeo.

[72] Feltham, Gerald and Jim Xie. 1994. Performance Measure Congruity and Diversity in Multi-Task Principal/Agent Relations. *Accounting Review*, 69(3): 429-453.

[73] Fershtman, Chaim and Shmuel Nitzan. 1991. Dynamic Voluntary Provision of Public Goods. *European Economic Review*, 35(5): 1057-1067.

[74] 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.

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

[76] Fudenberg, Drew, Bengt Holmstrom, and Paul Milgrom. 1990. Short-Term Contracts and Long-Term Agency Relationships. *Journal of Economic Theory*, 51: 1-31.

[77] Fudenberg, Drew, David K. Levine, and Eric Maskin. 1994. The Folk Theorem with Imperfect Public Information. *Econometrica*, 62(5): 997-1039.

[78] Fudenberg, Drew and Jean Tirole. 1990. Moral Hazard and Renegotiation in Agency Contracts. *Econometrica*, 58(6): 1279-1319.

[79] Frandsen, Brigham, Michael Powell, and James Rebitzer. 2018. Sticking Points: Common-Agency Problems and Contracting in the U.S. Healthcare System. Mimeo.

[80] Frankel, Alexander. 2014. Aligned Delegation. *American Economic Review*, 104(1): 66-83.

[81] Frankel, Alexander. Forthcoming. Delegating Multiple Decisions. *American Economic Journal: Microeconomics.*

[82] Gabaix, Xavier and Augustin Landier. 2008. Why has CEO Pay Increased so Much? *Quarterly Journal of Economics*, 123(1): 49-100.

[83] Gambetta, Diego. 1996. *The Sicilian Mafia: The Business of Private Protection.* Harvard University Press: Cambridge, MA.

[84] Garicano, Luis. 2000. Hierarchies and the Organization of Knowledge in Production. *Journal of Political Economy*, 108(5): 874-904.

[85] Garicano, Luis and Esteban Rossi-Hansberg. 2006. Organization and Inequality in a Knowledge Economy. *Quarterly Journal of Economics*, 121(4): 1383-1435.

[86] Garicano, Luis and Tim Van Zandt. 2013. Hierarchies and the Division of Labor. In *Handbook of Organizational Economics*, eds. Robert Gibbons and John Roberts, Princeton University Press.

[87] Georgiadis, George. 2015. Projects and Team Dynamics. *Review of Economic Studies*, 82(1): 187-218.

[88] Georgiadis, George. 2017. Deadlines and Infrequent Monitoring in the Dynamic Provision of Public Goods. *Journal of Public Economics*, 152: 1-12.

[89] Georgiadis, George, Steven Lippman, and Christopher Tang. 2014. Project Design with Limited Commitment and Teams. *RAND Journal of Economics*, 45(3): 598-623.

[90] Georgiadis, George and Christopher Tang. 2017. Project Contracting Strategies for Managing Team Dynamics. In *Handbook of Information Exchange in Supply Chain Management*, Springer.

[91] Ghosh, Suman and Michael Waldman. 2010. Standard Promotion Practices versus Up-or-Out Contracts. *RAND Journal of Economics*, 41(2): 301-325.

[92] Gibbons, Robert. 2005. Four Formal(izable) Theories of the Firm? *Journal of Economic Behavior and Organization*, 58(2): 2005.

[93] 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.

[94] 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.

[95] Gibbons, Robert and Michael Waldman. 1999. A Theory of Wage and Promotion Dynamics inside Firms. *Quarterly Journal of Economics*, 114(4): 1321-1358.

[96] Greif, Avner, Paul Milgrom, and Barry Weingast. 1994. Coordination, Commitment, and Enforcement: The Case of the Merchant Guild. *Journal of Political Economy*, 102(4): 745-776.

[97] 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.

[98] Grossman, Gene and Elhanan Helpman. 2002. Integration Versus Outsourcing in Industry Equilibrium. *Quarterly Journal of Economics*, 117(1): 85-120.

[99] Halac, Marina. 2012. Relational Contracts and the Value of Relationships. *American Economic Review*, 102(2): 750-779.

[100] Halonen, Maija. 2002. Reputation and the Allocation of Ownership. *Economic Journal*, 112(481): 539-558.

[101] Harris, Milton and Bengt Holmstrom. A Theory of Wage Dynamics. *Review of Economic Studies*, 3(1): 315-333.

[102] Hart, Oliver. 1983. The Market Mechanism as an Incentive Scheme. *Bell Journal of Economics*, 14(2): 366-382.

[103] Hart, Oliver. 1995. *Firms, Contracts, and Financial Structure*. Clarendon Press, Oxford.

[104] Hart, Oliver and Bengt Holmstrom. 2010. A Theory of Firm Scope. *Quarterly Journal of Economics*, 75(2): 483-513.

[105] Hart, Oliver and John Moore. 1990. Property Rights and the Nature of the Firm. *Journal of Political Economy*, 98(6): 1119-1158.

[106] Hart, Oliver and John Moore. 1998. Default and Renegotiation: A Dynamic Model of Debt. *Quarterly Journal of Economics*, 113(1): 1-41.

[107] 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.

[108] He, Zhiguo, Bin Wei, Jianfeng Yu, and Feng Gao. 2014. Optimal Long-Term Contracting with Learning. Mimeo.

[109] Hermalin, Benjamin. 1992. The Effects of Competition on Executive Behavior. *RAND Journal of Economics*, 23(3): 350-365.

[110] Hermalin, Benjamin. 1998. Toward an Economic Theory of Leadership: Leading by Example. *American Economic Review*, 88(5): 1188-1206.

[111] Hicks, John. 1935. Annual Survey of Economic Theory: The Theory of Monopoly. *Econometrica*, 3(1): 1-20.

[112] Hsieh, Chang-Tai and Peter Klenow. 2009. Misallocation and Manufacturing TFP in China and India. *Quarterly Journal of Economics*, 124(4): 1403-1448.

[113] Holmstrom, Bengt. 1979. Moral Hazard and Observability. *Bell Journal of Economics*, 10(1): 74-91.

[114] Holmstrom, Bengt. 1982. Moral Hazard in Teams. *Bell Journal of Economics*, 13(2): 324-340.

[115] 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.

[116] Holmstrom, Bengt. 1999. Managerial Incentive Problems: A Dynamic Perspective. *Review of Economic Studies*, 66(1): 169-182.

[117] Holmstrom, Bengt. 1999. The Firm as a Subeconomy. *Journal of Law, Economics, and Organization*, 15(1): 74-102.

[118] Holmstrom, Bengt and Paul Milgrom. 1987. Aggregation and Linearity in the Provision of Intertemporal Incentives. *Econometrica*, 55(2): 303-328.

[119] 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.

[120] Holmstrom, Bengt and Paul Milgrom. 1994. *American Economic Review*, 84(4): 972-991.

[121] Holmstrom, Bengt and John Roberts. 1998. *Journal of Economic Perspectives*, 12(4): 73-94.

[122] Holmstrom, Bengt and Jean Tirole. 1997. Financial Intermediation, Loanable Funds, and the Real Sector. *Quarterly Journal of Economics*, 112(3): 663-691.

[123] Jewitt, Ian, Ohad Kadan and Jeroen Swinkels. 2008. Moral Hazard with Bounded Payments. *Journal of Economic Theory*, 143: 59-82.

[124] KAMIEN, MORTON AND NANCY SCHWARTZ. 2012. *Dynamic Optimization: The Calculus of Variations and Optimal Control in Economics and Management*. Dover Publications.

[125] KANDORI, MICHIHIRO. 2008. Repeated Games. In *New Palgrave Dictionary of Economics*, eds. Steven Durlauf and Lawrence Blume.

[126] KE, RONGZHU, JIN LI, AND MICHAEL POWELL. 2018. Managing Careers in Organizations. *Journal of Labor Economics*, 36(1): 197-252.

[127] KERR, STEVEN. 1975. On the Folly of Rewarding A, While Hoping for B. *Academy of Management Journal*, 18(4): 769-783.

[128] KESSING, SEBASTIAN. 2007. Strategic Complementarity in the Dynamic Private Provision of a Discrete Public Good. *Journal of Public Economic Theory*, 9(4): 699-710.

[129] 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.

[130] KRAKEL, MATTHIAS AND ANJA SCHOTTNER. 2012. Internal Labor Markets and Worker Rents. *Journal of Economic Behavior and Organization*, 84(2): 491-509.

[131] LANDERS, RENEE, JAMES B. REBITZER, AND LOWELL TAYLOR. Rat Race Redux: Adverse Selection in the Determination of Work Hours in Law Firms. *American Economic Review*, 86(3): 329-348.

[132] Lazear, Edward and Sherwin Rosen. 1981. Rank-Order Tournaments as Optimum Labor Contracts. *Journal of Political Economy*, 89(5): 841-864.

[133] Legros, Patrick and Andrew Newman. 2013. A Price Theory of Vertical and Lateral Integration. *Quarterly Journal of Economics*, 128(2): 725-770.

[134] Leibenstein, Harvey. 1966. Allocative Efficiency vs. "X-Efficiency." *American Economic Review*, 56(3): 392-415.

[135] Levin, Jonathan. 2003. Relational Incentive Contracts. *American Economic Review*, 93(3): 835-857.

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

[137] Lipnowski, Elliot and Joao Ramos. 2015. Repeated Delegation. Mimeo.

[138] Lucas, Robert. 1978. On the Size Distribution of Business Firms. *Bell Journal of Economics*, 9(2): 508-523.

[139] Macaulay, Stewart. 1963. Non-Contractual Relations in Business: A Preliminary Study. *American Sociological Review*, 28(1): 55-67.

[140] MacLeod, Bentley and James Malcomson. 1988. Reputation and Hierarchy in Dynamic Models of Employment. *Journal of Political Economy*, 96(4): 832-854.

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

[142] MacLeod, Bentley and James Malcomson. 1998. Motivation and Markets. *American Economic Review*, 88(3): 388-411.

[143] 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.

[144] Malcomson, James. 1984. Work Incentives, Hierarchy, and Internal Labor Markets. *Journal of Political Economy*, 92(3): 486-507.

[145] Malcomson, James. 2013. Relational Incentive Contracts. In *Handbook of Organizational Economics*, eds. Robert Gibbons and John Roberts, Princeton University Press.

[146] Manove, Michael. 1997. Job Responsibility, Pay and Promotion. *Economic Journal*, 107: 85-103.

[147] Marschak, Jacob and Roy Radner. 1972. *Economic Theory of Teams*. Yale University Press.

[148] 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.

[149] Marx, Leslie and Steven Matthews. 2000. Dynamic Voluntary Contributions to a Public Project. *Review of Economic Studies*, 67(2): 327-358.

[150] Maskin, Eric. 2002. On Indescribable Contingencies and Incomplete Contracts. *European Economic Review*, 46: 725-733.

[151] Maskin, Eric and John Moore. 1999. Implementation and Renegotiation. *Review of Economic Studies*, 66(1): 39-56.

[152] Maskin, Eric and Jean Tirole. 1999. Unforeseen Contingencies and Incomplete Contracts. *Review of Economic Studies*, 66(1): 83-114.

[153] Maskin, Eric and Jean Tirole. 2001. Markov Perfect Equilibrium. *Journal of Economic Theory*, 100(2): 191-219

[154] McGowan, Daniel. 2016. Digging Deep to Compete: Vertical Integration, Product Market Competition, and Prices. Mimeo.

[155] Melumad, Nahum and Toshiyuki Shibano. 1991. Communication in Settings with No Transfers. *RAND Journal of Economics*, 22(2): 173-198.

[156] Menard, Claude. 2013. Hybrid Modes of Organization. In *Handbook of Organizational Economics*, eds. Robert Gibbons and John Roberts, Princeton University Press.

[157] 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.

[158] Milgrom, Paul and John Roberts. 1988. An Economic Approach to Influence Activities in Organizations. *American Journal of Sociology*, 94: S154-S179.

[159] Mirrlees, James. 1999. The Theory of Moral Hazard and Unobservable Behaviour: Part I. *Review of Economic Studies*, 66(1): 3-21.

[160] Mookherjee, Dilip. 2013. Incentives in Hierarchies. In *Handbook of Organizational Economics*, eds. Robert Gibbons and John Roberts, Princeton University Press.

[161] Moore, John and Rafael Repullo. 1988. Subgame Perfect Implementation. *Econometrica*, 56(5): 1191-1220.

[162] Moroni, Sofia and Jeroen Swinkels. 2014. Existence and Non-Existence in the Moral Hazard Problem. *Journal of Economic Theory*, 150: 668-682.

[163] Nalebuff, Barry and Joseph Stiglitz. 1983. Information, Competition, and Markets. *American Economic Review*, 73(2): 278-283.

[164] Ostroy, Joseph. 1980. The No-Surplus Condition as a Characterization of Perfectly Competitive Equilibrium. *Journal of Economic Theory*, 22: 183-207.

[165] Ottaviani, Marco and Peter Sorensen. 2006. *Journal of Economic Theory*, 126(1): 120-142.

[166] Penrose, Edith. 1959. *The Theory of the Growth of the Firm.* Basil Blackwell, Oxford, U.K..

[167] Powell, Michael. 2016. Productivity and Credibility in Industry Equilibrium. Mimeo.

[168] Powell, Michael. 2015. An Influence-Cost Model of Organizational Practices and Firm Boundaries. *Journal of Law, Economics, and Organization*, 31(suppl 1): i104-i142.

[169] Prat, Andrea and Aldo Rustichini. 2003. Games Played through Agents. *Econometrica*, 71(4): 989-1026.

[170] Prendergast, Canice. 1993. A Theory of "Yes Men". *American Economic Review*, 83(4): 757-770.

[171] Prendergast, Canice. 1993. The Role of Promotion in Inducing Specific Human Capital Acquisition. *Quarterly Journal of Economics*, 108(2): 523-534.

[172] Prendergast, Canice. 2002. The Tenuous Trade-off between Risk and Incentives. *Journal of Political Economy*, 110(5): 1071-1102.

[173] Prendergast, Canice and Lars Stole. 1996. Impetuous Youngsters and Jaded Old-Timers: Acquiring a Reputation for Learning. *Journal of Political Economy*, 104(6): 1105-1134.

[174] Qian, Yingyi. 1994. Incentives and Loss of Control in an Optimal Hierarchy. *Review of Economic Studies*, 61(3): 527-544.

[175] Rantakari, Heikki. 2008. On the Role of Uncertainty in the Risk-Incentives Tradeoff. *B.E. Journal of Theoretical Economics*, 8(1).

[176] Rantakari, Heikki. 2008. Governing Adaptation. *Review of Economic Studies*, 75(4): 1257-1285.

[177] Raith, Michael. 2003. Competition, Risk, and Managerial Incentives. *American Economic Review*, 93(4): 1425-1436.

[178] Reinganum, Jennifer and Nancy Stokey. 1985. Oligopoly Extraction of a Common Property Natural Resource: The Importance of the Period of Commitment in Dynamic Games. *International Economic Review*, 26(1): 161-173.

[179] Robertson, D.H.. 1930. *Control of Industry.* Nisbet and Company, London.

[180] Rosen, Sherwin. 1986. Prizes and Incentives in Elimination Tournaments. *American Economic Review*, 76(4): 701-715.

[181] Segal, Ilya. 1999. Contracting with Externalities. *Quarterly Journal of Economics*, 114(2): 337-388.

[182] Scharfstein, David. 1988. Product-Market Competition and Managerial Slack. *RAND Journal of Economics*, 19(1): 147-155.

[183] Schmidt, Klaus. 1997. Managerial Incentives and Product Market Competition. *Review of Economic Studies*, 64(2): 191-213.

[184] Seierstad, Atle and Knut Sydsaeter. 1987. *Optimal Control with Economic Applications*. North Holland.

[185] Simon, Herbert. 1991. Organizations and Markets. *Journal of Economic Perspectives*, 5(2): 25-44.

[186] Skarbek, David. 2014. *The Social Order of the Underworld: How Prison Gangs Govern the American Penal System*. Oxford University Press: Oxford, U.K..

[187] Tadelis, Steven and Oliver Williamson. 2013. Transaction Cost Economics. In *Handbook of Organizational Economics*, eds. Robert Gibbons and John Roberts, Princeton University Press.

[188] Van den Steen, Eric. 2009. Interpersonal Authority in a Theory of the Firm. *American Economic Review*, 100(1): 466-490.

[189] Van den Steen, Eric. 2010. Culture Clash: The Costs and Benefits of Homogeneity. *Management Science*, 56(10): 1718-1738.

[190] Vives, Xavier. 2008. Innovation and Competitive Pressure. *Journal of Industrial Economics*, 56(3): 419-469.

[191] Waldman, Michael. 1984. Job Assignments, Signalling, and Efficiency. *RAND Journal of Economics*, 15(2): 255-267.

[192] Waldman, Michael. 2003. Ex Ante versus Ex Post Optimal Promotion Rules: The Case of Internal Promotion. *Economic Inquiry*, 41(1): 27-41.

[193] 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.

[194] Williamson, Oliver. 1967. Hierarchical Control and Optimum Firm Size. *Journal of Political Economy*, 75(2): 123-138.

[195] Williamson, Oliver. 1971. The Vertical Integration of Production: Market Failure Considerations. *American Economic Review*, 63(2): 112-123.

[196] Williamson, Oliver. 1975. *Markets and Hierarchies: Analysis and Antitrust Implications*. Free Press, New York, NY.

[197] WILLIAMSON, OLIVER. 1979. Transaction-Cost Economics: The Governance of Contractual Relations. *Journal of Law and Economics*, 22(2): 233-261.

[198] WILLIAMSON, OLIVER. 1985. *The Economic Institutions of Capitalism*. Free Press, New York, NY.

[199] WILLIAMSON, OLIVER. 2000. The New Institutional Economics: Taking Stock, Looking Ahead. *Journal of Economic Literature*, 38: 595-613.

[200] WU, YANHUI. Forthcoming. Incentive Contracts and the Allocation of Talent. *Economic Journal*.

[201] YILDIRIM, HUSEYIN. 2006. Getting the Ball Rolling: Voluntary Contributions to a Large Scale Public Project. *Journal of Public Economic Theory*, 8(4): 503-528.

[202] 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.