MECS 570-1: Organizational Economics I

Michael Powell

Boundaries of the Firm, II: Foundations; Influence Costs; Organizational IO

Foundations of Incomplete Contracts

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 justifications 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 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 been 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.

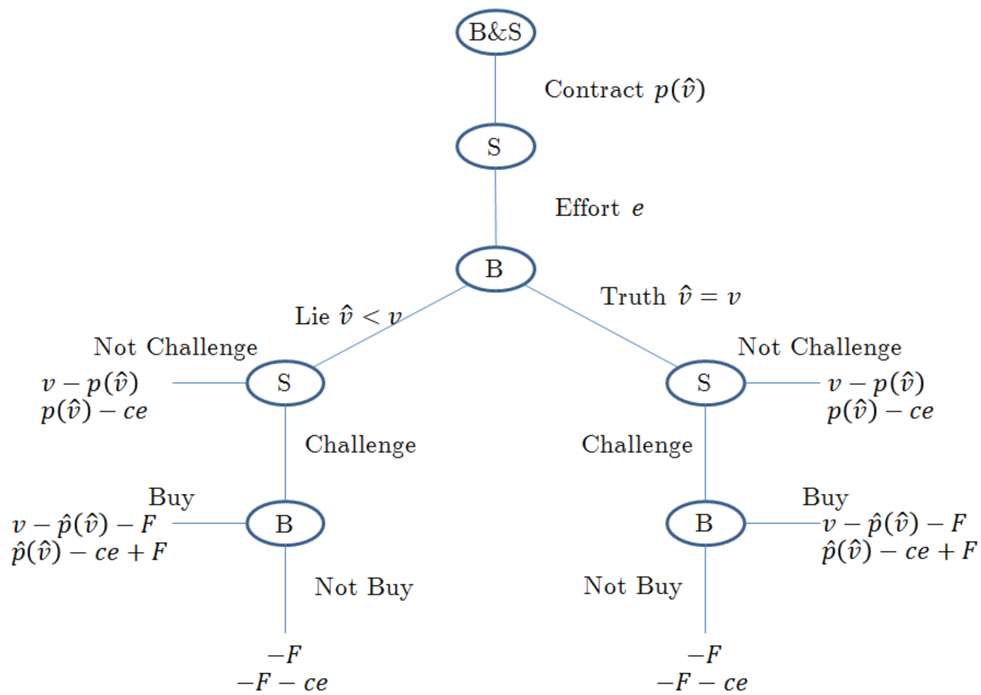
4. B announces $\hat{v} \in \{v_L, v_H\}$. $\hat{v}$ is commonly observed and verifiable by a third party.

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)

More-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 (2017) take the Moore and Repullo mechanism into the laboratory and show that indeed, when they perturb the information structure away from common knowledge of payoff-relevant variables, subjects do not make truthful announcements.

Relatedly, Fehr, Powell, and Wilkening (2018) take an example of the entire Maskin and Tirole critique into the lab and ensure that there is common knowledge of payoff-relevant variables. They show that in the game described above, there is a strong tendency for B's to reject counter offers after they have been challenged following small lies, S's are reluctant to challenge small lies, B's tend to make announcements with $\hat{v} < v$, and S's often choose low effort levels.

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

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

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

The findings of Aghion, Fehr, Holden and Wilkening and of Fehr, Powell, and Wilkening do not necessarily imply that it is impossible to find mechanisms in which in the unique equilibrium of the mechanisms, the hold-up problem can be effectively solved. What they do suggest, however, is that if subgame-perfect implementation mechanisms are to be more than a theoretical curiosity, they must incorporate relevant details of the environment in which they might be used. If people have preferences for reciprocity, then the mechanism should account for this. If people are concerned about whether their trading partner is rational, then the mechanism should account for this. If people are concerned that uncertainty about what their trading partner is going to do means that the mechanism imposes undue risk on them, then the mechanism should account for this. Framing the implementation problem in the presence of these types of "behavioral" considerations and proving possibility or impossibility results could potentially be a fruitful direction for the implementation literature to proceed.

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. In the notes, I 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 $E_s\left[s|\,\sigma_P, \hat{\lambda}(i)\right]$. 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

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

subject to R's participation constraint:

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

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, $Var\left(Y\right)=E\left[Var\left(Y|\,X\right)\right]+Var\left(E\left[Y|\,X\right]\right)$):

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

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

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

$ADAP(p)$ 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:

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

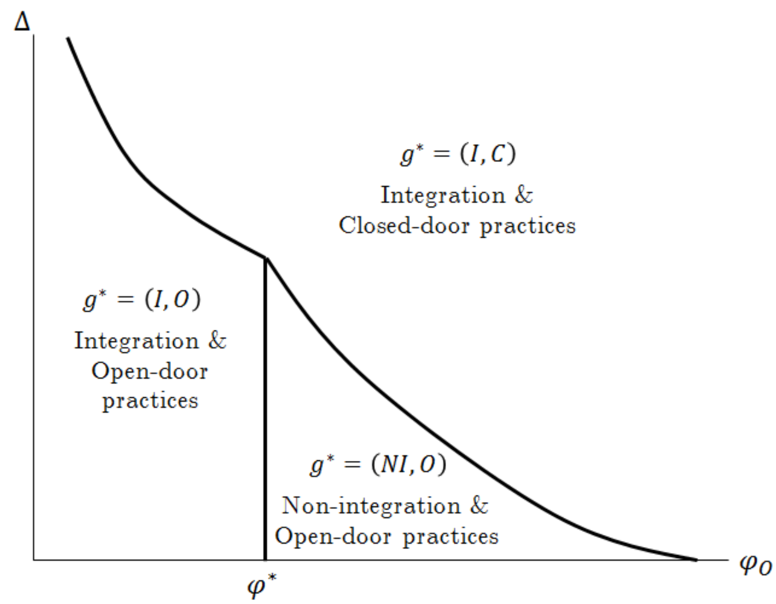
When $\alpha_L = \alpha_R$, so that $ALIGN(I) = ALIGN(NI)$, we have that $INFL(I,O) - INFL(NI,O) > 0$, so that influence costs are minimized under non-integration. When $p = C$, so that there are no influence costs, and $\alpha_L > \alpha_R$, $ALIGN(I) < ALIGN(NI)$, so that alignment costs are minimized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

Organizational Industrial Organization

The informal theory 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 revealed by 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 Holmstrom (2002/2010) model of firm boundaries into a price-theoretic framework. In the Hart and Holmstrom model, integration unifies contractible payoff rights and decision rights, thereby ensuring that decisions are 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 $d_L^{NI} = pA/(1+pA)$. 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 $\Delta^{NI} = d_R^{NI} - d_L^{NI} = 1/(1+pA)$.

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

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

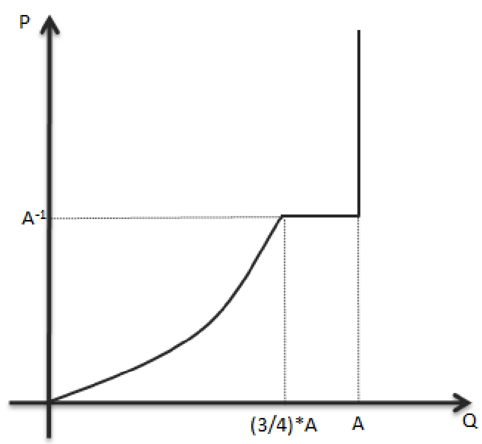
Since the coordination gap is smaller under integration than under non-integration, and expected revenues are higher under integration that under non-integration, there is a trade-off between greater coordination under integration and greater private benefits under non-integration.

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.

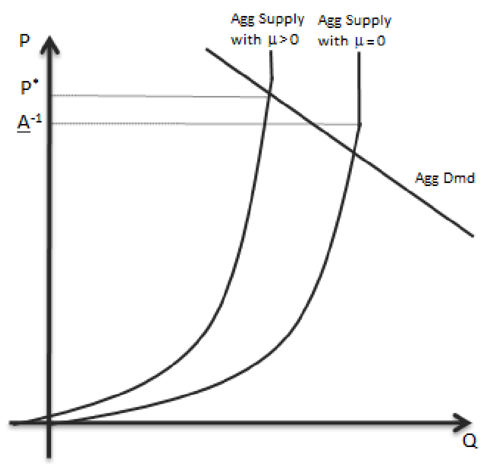
Industry Equilibrium Given a price level p, a firm of productivity A will produce expected output equal to

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

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.

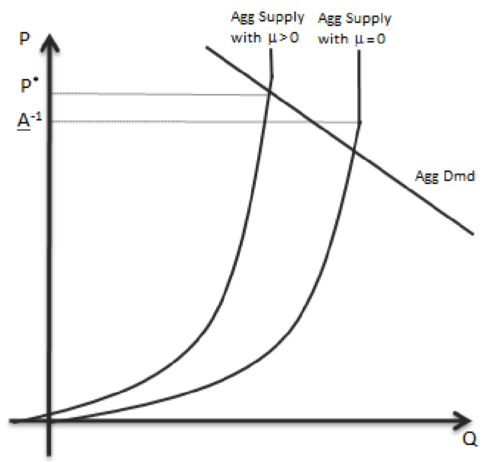


As drawn, the inverse demand curve intersects the inverse supply curve at a value of $p \in (1/\bar{A}, 1/\underline{A})$, so in equilibrium, there will be some firms that are integrated (high-productivity firms) and some that are non-integrated (low-productivity firms). If the inverse demand curve shifts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