Incomplete Contracts and Industrial Organization: A Survey*

Patrick Legros† and Andrew Newman‡

October 14, 2011

Abstract

While most researchers in industrial organization agree that the neoclassical, cost-minimizing, unitary, view of the firm is restrictive, it is still the main object of analysis in the literature. We review the literature in industrial organization and the place occupied by a richer view of the firm both before and after [Grossman & Hart (1986)] (GH). Before GH, integration was viewed as a way to alleviate hold-up problems, agency or market power distortions; GH pointed out that integration and ownership reallocation create their own incentive problems. We show how this insight can enrich the usual analysis of firms with market power but also opens the door for a rich set of questions in more competitive environments.

*Prepared for presentation at the conference Grossman and Hart 25 on June 24-26, 2011 at the Université libre de Bruxelles. We thank our discussants Mathias Dewatripont and Kai-Uwe Kühn as well as the audience for their comments.

†Université libre de Bruxelles, ECARES, 50 avenue F.D. Roosevelt, 1050 Brussels, Belgium, e-mail: plegros@ulb.ac.be, home page: http://plegros.net

‡Boston University, Department of Economics, 270 Bay State Road, MA, 02215, e-mail: afnewman@bu.edu, home page: http://people.bu.edu/afnewman/
1 Introduction

In his iconic textbook Scherer (1980) describes industrial organization as being:

“concerned with how productive activities are brought into harmony with society’s demands for goods and services through some organizing mechanism such as a free market, and how variations and imperfections in the organizing mechanism affect the degree of success achieved by producers in satisfying society’s wants.”

This definition would probably still receive almost unanimous support from economists. Elsewhere he asserts what would receive rather less support from the practitioners of industrial organization:

“The name [industrial organization] is a curious one, distinctive mainly in its inability to communicate to outsiders what the subject is all about. It has little or nothing to say about how one organizes and directs a particular industrial enterprise, although there are industrial organization courses in business and engineering schools that deal with such matters.”

Since the goal of industrial organization is to understand how the behaviors of firms affect market performance it is puzzling how primitive a view of the firm the field is still using — despite the presence of incomplete contracting ideas and richer views of the firm that have been around since the 40s, and the explosive growth in the theory of the firm that followed Grossman and Hart’s seminal paper. It is instructive in this respect to trace the evolution of industrial organization and the place taken by other approaches to the neoclassical firm. We identify three waves in the literature: before the 70s, up to the 90s and the current period.

40s-70s: Revisiting the Assumption of Cost-Minimization

An active controversy arose in 1946-1965 about the validity of the assumption of profit maximization by firms (see Machlup (1967)), fueled by the work of Marris (1964), Coase (1937) and Williamson (1964). A ‘behavioral approach’ developed, with the goal of developing theories of the firm based on the precise observation of existing organizations. The long list of conflicting interests between managers, owners and workers that were identified then would look very familiar for a current student of the theory of the firm: managers may want a quiet life, avoid ‘resentment on the part of their colleagues and subordinate’, are more interested in their own salaries than in the profit of the firm, the flow of information through the organization may lead to misinformation about earnings, the striving for power may put aside profit motives, to name a few. The list of distortions was impressively long, suggesting that the neoclassical firm was indeed an abstraction.

Alternatives considered by the literature at the time were to equip the ‘firm’ with a different behavioral assumption: sales maximization or growth maximization being the most popular (in
particular [Baumol (1962)], see [Machlup (1967)] for a survey of the literature. However, replacing the profit maximizing assumption by another behavioral assumption is hardly convincing. Since the behavioral approach presumes that the realized behavior is the underlying objective function, this begs the question of identification if the same outcome is consistent with different behavioral assumptions and may invalidate positive or normative analysis if observed ‘behavior’ reflects in fact specific economic conditions.

The theory of industrial organization developed in parallel to the behavioral, more empirical, approach, and was mainly based on the neoclassical firm, a phenomenon reinforced by the emergence of oligopoly theory and game theory as the crucial analytical tools. This reliance on an assumption that was clearly empirically suspect was rationalized by some by referring to the view that all models are abstractions anyway and that assumptions are acceptable as long as theories developed around them have a good track record empirically (Friedman (1953)). Machlup (1967) called a “fallacy of misplaced concreteness” the goal of making the theoretical ‘firm’ close to reality and pointed out that even if there is evidence that “firms” do not minimize costs, the resulting inefficiencies are second order and transitory because competition (in a perfect of imperfectly competitive environment) has the potential to create cost-minimization discipline.

70s-90s: Information Economics and Oligopoly Theory

Following this first wave, oligopoly theory and strategic interaction between “firms” took the center stage in the field, helped by the advances in game theory and information economics, the growing importance of antitrust on the policy side and the resulting demand for models of strategic interaction in markets with market power.

The theory of industrial organization that has developed since the 70s has thus — bearing some exceptions — diverged from another early strand of the literature which was more empirical, more devoted to the understanding of the inner workings of firms and their organization (which Scherer associated with business or engineering schools). It was in this context that Grossman and Hart published their 1986 paper.

It is also in this context that the principal agent literature provided a framework for evaluating

---

1In a different vein, [Grossman & Hart (1979)] study of competitive equilibria with incomplete markets and propose a utility taking behavior of current shareholders, assuming that managers act in their best interests.

2For instance, [Williamson (1966)] shows that a growth maximizing firm maximizes profits and that output decisions under the two behavioral assumptions are identical.

3The many IO textbooks that have been written in the last thirty years typically follow the same plan. The author acknowledges that the received view of the firm is an abstraction, that it is actually nonsense to talk of the ‘behavior of a firm’: after all only individuals can have behavior, and a ‘firm’ is a place where different stakeholders have conflicting interests, where contracts are incomplete (the term seemed to have appeared in textbooks in the 70s). After emphasizing that the assumption of a unitary decision maker and profit maximization are empirically invalid, and after reviewing more modern theories of the firm and managerial behavior (this takes up to 10% of book’s pages), the remainder then abruptly ignores all of this and develops its entire exposition of the theory and empirics of industrial performance on the basis of the unitary profit maximizing firm that it has so carefully debunked.
Machlup’s claim that competition serves as a disciplining device for managers; this will be the subject of the first section of this survey. Basically, the Machlup intuition that competition disciplines managers is only partially correct (or incorrect depending on one’s prior) because it ignores the endogeneity of the incentive schemes used inside the firm and that they may either substitute for or complement the disciplining effect of the market: for instance, when demand is low, managers may work harder to avoid bankruptcy but shareholders may be less willing to provide strong incentives to managers. Both theoretical and empirical work suggest that there is a non-monotonic relationship between competition and effort provision by the manager.

This set of results suggests that it is not possible to study unitary profit maximizing firms and hope that the resulting behavior and response to shocks or changes in the environment will replicate that of managerial firms. Such a mapping is sometimes possible, for instance if marginal costs are observed and the manager can affect the distribution of marginal costs and has incentives based on profit realization, his behavior will be indistinguishable from that of a profit maximizing firm: for each cost realization the manager will choose the profit maximizing quantity; which does not mean that the effort choice is ex-ante surplus maximizing since that choice depends on the incentive scheme.

With agency problems, the total cost of production includes not only the usual variable and fixed costs of production, but also the opportunity costs to the parties from using the incentive scheme. As we will argue later in this survey, this indirect cost will also typically depend on the way the surplus is allocated between the parties and the price of output: if \( s \) is the share of profits going to the manager, and \( P \) is the price of output, the “revenue function” of a managerial firm will be as usual the product \( QP \) but the cost function will be a function of \( Q, s, P \) rather than a function of \( Q \) only as in the neoclassical case.

Therefore it is unlikely that the positive and normative analyses of industries with managerial firms are identical to that with neoclassical firms. The question, both theoretically and empirically, is then how much is lost by using the neoclassical assumption and what can be gained by using a richer view of the firm. Revealed preferences suggest that the received wisdom in industrial organization is that there is little to be gained.

90s-now: The Phase of Maturity

We are now at the end of a third phase for industrial organization, where empirical analysis has taken precedence over theoretical work, a trend amplified by advances in econometric methods and the availability of large and rich data sets. While theory has taken the back stage, it still permeates empirical analysis and is often invaluable for policy.

The GH view that ownership allocation is a key element in moderating the inefficiencies resulting from incomplete contracting has been made operative in an industrial organization context, but the
focus has mainly be on questions where integration decisions are perceived as being “first order”:

- Can ownership be used as a strategic instrument by firms?
- Do contracts facilitate or hinder entry in a market?
- Can restraints of trade necessarily lead to efficiency loses?
- Do contracts facilitate innovation?
- What are the (new) policy responses?

As we will review in section 3, the literature has indeed embraced these questions, and has dealt with bilateral contracting relationships like franchises, joint ventures for innovation, foreclosure via vertical integration in novel ways that has also influenced in significant ways – for instance – the design of the guidelines for horizontal and vertical relationships in both the US and the EU. Most of the applications of the property rights approach to industrial organization have been cast in an imperfectly competitive product market environment, suggesting that it complements the traditional oligopoly approach in industrial organization, rather than bring forth fundamentally new insights. One explanation for this may be due to the strong identification of industrial organization with market power considerations, and therefore that researchers tend to think of this when considering issues where ownership matters, like those highlighted above. It must be then acknowledged that incomplete contracting and the property rights approach has not changed in a fundamental way how researchers in industrial organization address other key questions which tend to escape simple market power explanations like:

- Do organizational choices affect the scale and scope of firms in industries?
- Why is there heterogeneity in organizational forms? Does heterogeneity correlate with industry performance?
- What explains the timing of mergers? Why do mergers tend to happen in waves?
- Do conglomerate mergers affect consumers?
- What are the (new) policy responses?

In section 3 we visit or revisit some of these questions thru thr lenses of the property rights approach. We will argue that the allocation of ownership and organizational choices create first order effects in competitive markets, suggesting that important policy questions fall within the

---

4 To cite Coase (1972) “If an economist finds something – a business practice of one sort or other – that he does not understand, he looks for a monopoly explanation. As in this field we are very ignorant, the number of ununderstandable practices tends to be rather large, and the reliance on monopoly explanation frequent” (p.67) and “Industrial organization has become the study of the pricing and output policies of firms, especially in oligopolistic situations [...]. It has not helped, of course, that there is no theory of oligopoly or, what comes to the same thing, that there are too many theories of oligopoly. But leaving this problem aside — and without intending to suggest that the questions tackled are unimportant — it is clear that modern economists writing on industrial organization have taken a very narrow view of the scope of their subject.” (p. 62). Coase contrasts this “narrow view” with the approaches in the literature in the 20s that gave him the impetus to write his famous paper.
reach of industrial organization beyond what competition policy already achieves in situations of exercice of market power.

There are many aspects of industrial structure that we do not consider in this survey, partly because they open a new set of questions and would require a survey of their own, and partly because they are covered in other surveys presented at this conference. We do not for instance provide a review of the literature on regulation and procurement, topics which are clearly important given the role that the state can play in opening markets to competition, in regulating competition or in engaging in contracting relationship with the private sector. We do not also review recent applications of the incomplete contract paradigm to the organization of markets. One such recent application is the organization of the market for intellectual property rights and the formation of technological standards in high technological sectors. In such industries, firms with patents on key technologies may hold-up adopters and discourage the development of new technologies that may infringe on these patents. Future users of a new technology as well as patent holders have therefore reasons to cooperate and put in place ex-ante rules for controlling the use and the price of these patents. Such a cooperation is neither integration nor spot-market transaction: owners of assets retain the rights to decide on the use of their assets ex-post but have accepted some constraints on this use.

2 Incentive Provision and Competition

2.1 The Ambiguous Relationship Between Competition and Incentives Provision

In the early 80s, the Machlup conjecture that competition provides discipline on managers was revisited. Nalebuff & Stiglitz (1983), Harris & Holsmtröm (1982) show that competition allows relative performance evaluation: the owner is better off since he obtains more information and incentive schemes are stronger but effort change is ambiguous because it depends on the cost of implementing incentive schemes (e.g., on the underlying distribution of signals). This type of yardstick competition requires however a lot of information about the individual performance of other firms.

Hart (1983) shows that cost correlation across firms may be sufficient to create discipline in a framework where managers have private information about their productivity and exert productive effort. He assumes that in addition to managerial firms, neoclassical firms are present in the mar-

---

5Incomplete contract setting has been applied to procurement or public-private partnerships (see Hart et al. (1997), Hart (2003), Dewatripont & Legros (2003), Jossa & Martimort (2008); to the allocation of ownerships in a regulatory setting, with public production by a private firm (Besley & Ghatak (2001)); to the provision of pure or impure public goods (Francesconi & Muthoo (2011)).

ket, and for these firms screening of managers is achieved at no additional cost; this implies that neoclassical firms produce a high output level when productivity is high and a low output level when productivity is low. By the cost correlation, if the managers’ type is high, the presence of neoclassical firms will increase the supply in the market and therefore put downward pressure on prices, making it more difficult for managers to achieve profit targets and eventually leading them to exert more effort. Hence the product market price plays its role of transmitting information about the fundamentals of the economy; this effect is absent when cost realizations are independent across firms. While Hart (1983) shows that there is indeed a positive effect of competition (measured by the mass of neoclassical firms in the market) on effort provision, Scharfstein (1988) shows that the opposite result can be achieved if we abandon the assumption of infinite risk aversion made in Hart (1983): in this case, a lower product market price may actually make the monetary transfer to the manager costlier, and lower effort provision becomes optimal for owners. While the informational requirements are weaker than in yardstick competition, the assumption of cost correlation is somewhat demanding.

Schmidt (1997) does away with yardstick competition or cost correlation. Like in the corporate finance literature, he considers a model where owners may decide to close down the firm and where bankruptcy has an opportunity cost for managers — the loss of private benefits for instance. Effort of the manager takes the form of a reduction in the probability that the cost of production is high: more effort makes it less likely that profits are low. Owners observe the cost realization and a signal about the market environment; they can therefore compute the expected profit of the firm if they decide not to produce, that is not to go into bankruptcy. A low cost will lead to positive profits but a high cost will lead to negative profits if the signal is low enough. Therefore, there is bankruptcy only if there is high cost and the signal is low. Anticipating this, the manager chooses optimally his effort trading of his cost of effort and the private loss from bankruptcy. Contrary to Hart (1983) and Scharfstein (1988), Schmidt (1997) considers risk neutral managers but subject them to limited liability. Depending on the outside option of the manager, the optimal contract may either bind or not the individual rationality condition. Suppose that more competition implies a higher probability of bankruptcy (e.g., a first order stochastic shift in the signal distribution). When the IR binds the owners are forced to provide more attractive compensation to the manager in order to meet his outside option; in turn this leads to higher effort provision. By contrast, when the individual rationality does not bind, the owners do not need to compensate the manager for the increasing risk of bankruptcy since he is already getting a rent. If the difference between the expected profits with low and high costs decreases when there is more competition, the marginal benefit of high effort decreases from the owner’s point of view. This rent reallocation effect of competition may then induce the owners to implement a lower effort for the manager.  

See Aghion et al. (1999) for an application of these ideas to growth.
Raith (2003) analyzes a free entry oligopoly model on a Salop circle where firms are composed of a risk neutral principal and a risk averse agent with CARA utility. Cost observation – the variable on which the agent effort has influence – has a normal distribution with mean a decreasing function of effort. In such a setting, a noisier environment, that is a more diffuse signal about the agent’s effort, results in weaker incentives being provided. Empirical work on this link is quite ambiguous however (e.g., Lafontaine & Slade (2007) who collect the empirical evidence on page 638; see also Prendergast (2002)). Raith reconciles the empirical results by emphasizing two types of risks that have opposite influence on effort provision. First, there is the usual agent’s risk in principal-agent models implying that noisier environment makes incentives less powerful; in the linear-normal model, this implies that risk and incentives are negatively related. However there is also a firm’s risk since the variance of profits is also a function of the shocks; and therefore it is possible that the marginal value of cost reductions may increase, implying a benefit of stronger incentive schemes even if the level of profits decreases. There are therefore potentially two opposite effects at play when a given oligopoly is subject to a shock. By focusing on the firm’s risk, one could indeed find a positive correlation between risk and incentive provision, contrary to the theoretical prediction when one’s focus is the agent’s risk.

In Raith (2003), all agents and all principals are identical and the ambiguous relationship between risk and incentive provision is channeled through the strategic interaction between firms and their responses to a decrease in their residual demands when competition increases. If however agents and principals are heterogenous, for instance principals are endowed with projects of different riskiness and agents have different levels of productivity, selection effects may be sufficient to generate positive or negative correlation between risk of the project and incentive scheme; and this can be done without having to resort to strategic interaction on the product market. This point has been made for instance by Ackerberg & Botticini (2002) in the linear-normal model. Generally, selection effects may generate dependence of the way agents and principals match, and on the types of contracts that are offered, on the distribution of characteristics as well as on the underlying technology.

In the industrial organization setting, a natural interpretation of effort is a reduction in the cost of production or innovative activity. It is well understood that competition could have positive or negative effects on innovation: on the one hand, the “monopoly replacement effect” suggests that dominant firms will tend to be less aggressive in innovating while new entrants will be more aggressive; on the other hand there is the Schumpeterian view that monopolies are better equipped to innovate since they capture all the rents from innovation. Vives (2008) offers a recent investigation of these effects under different market structures and different measures of ‘increased competition’. For instance, a larger number of firms will have a depressing effect on innovation while a greater

---

substitutability of products will have a positive effect. Schmutzler (2010) shows that endogenous entry makes the positive effect of competition more likely. Aghion et al. (1997) point out that managerial firms — subject to agency problems — are affected in an opposite way to neoclassical firms by industrial policy or competition policy. For instance, competition policy (assumed to lead to an increase in the number of firms) will tend to decrease the R&D effort of neoclassical firms – the Schumpeterian effect – but will lead to an increase in the effort of managerial firms – the Machlup effect we have discussed previously. This suggests that property right considerations may indeed change the way industrial policy should be conducted. However, ownership boundaries are exogenous in this work, which limits significantly the robustness of the policy conclusions since, as incentive schemes react to changes in the competitive environment, ownership and decentralization structures should be affected by changes in their environment, including the design of industrial or competition policies.

The agency approach and the theoretical ambiguity between competition and different measures of performance (growth, innovation, technical efficiency) has generated significant empirical research in the 90s (see the recent book Aghion & Griffith (2008) for a survey of that literature). The empirical research confirms the theoretical ambiguous relationship between competition and “performance”, in particular highlighting an inverted U relationship. While Nickell (1996) could not identify a robust positive effect of competition on corporate performance, Nickell et al. (1997) find that competition and corporate governance may be substitute. The literature on growth (e.g. Aghion et al. 2002) find a non-monotonic relationship between growth and competition. For innovation, Geroski (1990) or Blundell & Griffith (1995) find a positive relationship between competition (a lower concentration in the industry) and innovation. Finally, there is also evidence that technical efficiency starts declining when industries become more concentrated (Caves & Bailey 1992, Caves 1980, Green & Mayes 1991), consistent with an inverted U relationship between competition and technical efficiency.

### 2.2 Two Messages

A common theme in these papers is that changing market conditions, like a downward pressure on prices, may have the effect to increase effort provision assuming that the incentive scheme is kept constant but may have also the effect to discourage the owners, or whoever is responsible for defining the incentive scheme, to offer strong incentive schemes. If the latter effect is strong enough, it may overturn the benefits of competition for the provision of effort in managerial firms. It is therefore the endogeneity of incentive schemes that makes ambiguous the relationship between competition and performance. Since other organizational choices, like integration decisions for instance, also affect performance, there is no reason to assume that the relative costs of different organizational choices are not affected by changing market conditions. A corollary is that external forces shape
organizational choice and this implies in particular that one may have to look outside the firm in order to understand how it is organized. This message has not yet permeated fully the analysis of organizations and their role in market settings.

Another corollary is that we cannot simply assume that the (relative) cost of integration is exogenous, or assume from the start how this cost changes when market conditions change. Formal models of the costs and benefits of integration have therefore value since they can link these costs and benefits of integration to more fundamental economic variables and therefore facilitate empirical identification. For incomplete contract theories, the transaction cost approach inherited from Klein et al. (1978) and Williamson (1979) tends to focus on the benefit of vertical integration in reducing the hold-up problem in the presence or relation specific investments and also on the ex-post adaptability of the organization. Empirical strategies therefore correlate the level of specificity to integration patterns; for instance, changes in the environment that make “adaptability” more important should lead to more integration. The property rights approach in Grossman & Hart (1986), Hart & Moore (1990) highlights in addition to the benefit of reducing hold-up problems, the possible worsening of hold-up problems with integration and also focuses on the effects of integration on ex-ante investments. An empirical strategy should therefore focus on the changes in the relative importance of investments by different parties when the environment changes in order to correlate changes in the environment and integration patterns.

As an illustration, consider Woodruff (2002) which analyzes integration in the footwear industry and explicitly incorporates the costs and benefits of integration when both the retailer and the manufacturer have to make investments: under integration, the retailer will have lower investment incentives while under non-integration the manufacturer has lower investment incentives. TCE and PRT generate identical predictions when one correlates changes in integration with changes of heterogeneity of goods produced or the quality of inputs used in production. However TCE and PRT lead to different predictions about the correlation between integration and the rate of change of ‘fashion.’ Indeed, a higher rate of change of fashion makes ‘adaptability’ more important, hence should be positively correlated with integration if one takes a TCE approach. However, a higher rate of change of fashion also makes the investment of the retailer more important and therefore non-integration should be favored from a PRT approach. Woodruff (2002) finds support for the PRT approach in his data.

2.3 Beyond Competition in the Product Market

While most of the literature has focused on the role of product market competition on incentive provision, the disciplining role of other markets relevant for the firm, like the financial and managerial markets, have received some attention in the literature. The relationship between financial

---

9See Whinston (2001, 2003) for an argument along these lines as well as an empirical support for the property rights approach from the early literature.
market and firm performance is already reviewed in this volume and we will not comment much
on this here, but we will partly review the literature on debt and performance in the next section.

Fama (1980) poses the question of how the financial and the managerial markets can discipline
owners of capital, managers and workers. Fama was optimistic about the capacity of competition
on the managerial market to sort out the best managers. Holmström (1999) shows that discipline
may fail in a dynamic framework when output contingent contracts are not possible. In his model,
workers exert effort to signal their productivity to the market, hence to increase their future expected utility they will exert positive effort; however, once the market has learnt their type, the worker no longer has an incentive to exert effort. Hence incentives are provided by the market as long as the market’s beliefs are not degenerate.

Managerial talent is an important determinant of the performance of any type of organization,
probably a determinant of the size, performance and dynamics of firms (Lucas Jr (1978), Jovanovic
(1982)). However, most of the recent research on the role of managerial market has been confined to
the macroeconomics or labor literatures, and little is done to understand how that market interacts
with the product market. Heterogeneity of talent and sorting of managers in firms may be an
important determinant of collective reputation, an important intangible asset for competing firms.
See for instance Tadelis (1999) and Fombrun & Shanley (1990) on the meaning of ‘reputation’
on how the creation of social capital through repeated interactions affects incentive design, Tirole
(1996), Levin (2009) for the creation and sustainability of group reputations, Hart (2001) for the
role of norms, of intangible assets, for a firm’s value, and Kosfeld & von Siemens (2011) for the
role of incentives in sorting heterogenous workers in firms that value cooperation between workers
to different extents. Turnover in the labor market can also affect the ‘institutional memory’ of
organizations, hence their ability to experiment and learning from the past; for instance, Mukherjee
(2010) articulates a tradeoff between incentives to increase turnover in order to have a better ‘match
value’ between the firm and the worker and the resulting loss of history in the firm. For the same
reason, heterogeneity in talent among members of firms may shape the way firms are organized
beyond some of the “natural” forces for heterogeneity in organizational forces that we highlight in
section 5.3. There is a sense in which what happens in the labor market affects the information
available within firms and therefore their performance.

\[10\] See also Manne (1965) for the role of the managerial market for mergers, in particular on the needed “fit” of vision
between owners and managers, and the ‘evolutionary’ approach in Allen & Gale (2000) where dynamic competition
for product market competition having a winner’s take all flavor can potentially sort out ‘good’ and ‘bad’ managerial
teams.
3 Ownership as Commitment to Behavior in Markets

It has been understood for some time in industrial organization that commitment gives a competitive advantage to firms in an imperfectly competitive environment. It has been understood more recently that commitment does not come for free and that an instrument is good for commitment only if it leads to behavior consistent with sub-game perfection (or sequential rationality more generally). Ownership rights may create commitment when ownership is coupled with the right to make decisions, including the right to design organizations and incentive scheme: the owner can then commit to decisions that are sequentially rational for him or her, subject to potential renegotiations. However, ownership may also restrict the ability not to interfere with other agents’ decisions, putting constraints on the design of internal organization, in particular delegation of tasks.

3.1 Cross-Ownership

In a perfectly competitive environment, ownership of neoclassical firms has no effect on performance since prices are taken as given: changing ownership concentration for an exogenous number of firms has no effect on prices nor on the cost minimization decision. This is no longer true for imperfectly competitive environments since ‘owners’ internalize the consequences of the pricing (or the relevant strategic variable) decisions on the firms they control.

Cross-ownership of assets generates a commitment by the owner to favor high levels of profit for all firms he owns; this will therefore weaken competition between ‘firms’ if the same owners have the control of the firms, e.g., sit on their board. For this reason, cross-ownership has been a concern for antitrust from the start; the Clayton act of 1914 explicitly forbids overlapping membership in boards or cross-holdings of shares when there is a risk of lessening competition.

Beyond its facilitating role for collusion, cross-ownership may affect the relationship between usual measures of concentration, like the Herfindahl index, and welfare. For instance, Farrell & Shapiro (1990) consider Cournot oligopolies and show that more concentration in ownership may covary with consumer welfare, hence a higher Herfindahl index may not be an indication of more market power, hence lower consumer surplus. There is some evidence that cross-ownership can also affect takeovers (Allen & Gale, 2000) compare different countries and argue that differences in

11 For instance, the early models of entry in industrial organization supposed that an incumbent could commit to being aggressive in prices if there is entry tomorrow by setting a low price today.

12 For instance, lack of ownership may limit the desire of inventors to share their idea with a financier or an expert for fear of being expropriated. Biass & Perotti (2008) argue that joint ownership by the inventor and the expert serves as a commitment to share information and for the expert not to steal that information. However, this commitment value disappears for ‘really good’ ideas. Always in the realm of R&D, Aghion et al. (2008) argue that differences in control rights allocations between the private and the public sector effectively create a boundary between fundamental or early stage research and development or later stage research and serves as a commitment device for scientists to pursue their own interests for fundamental research.

13 In 2009, Google and Apple came under scrutiny by the FTC for potential violation of this act.
takeovers may not be due so much to differences in regulations but to differences in cross-holdings of shares.\textsuperscript{14}

The effect of cross-ownership is purely strategic in that literature, and the relative costs and benefits of cross-holding of shares, controlling versus non-controlling shares, is sensitive to the nature of competition,\textsuperscript{15} or to the design of the market for rights.\textsuperscript{16} But, there is no effect on the cost of production nor on the production set of individual firms, suggesting that ownership allocation will have little effect in a competitive product market environment.

That performance is independent of the ownership structure is at odds with the finding in corporate finance that the relationship between liquidity and performance is complex; more diluted ownership may also dilute incentives to monitor managers and whatever gain in coordination could be achieved through cross-ownership may be undone by managerial discretion (e.g., the survey of \textsuperscript{17}Becht et al. (2003)). \textsuperscript{18}Buccirossi & Spagnolo (2007) survey the literature on corporate governance and collusive behavior in oligopolies, and they highlight the potential role played by managerial compensation for collusive behavior, in particular the role of stock options. To date, we are not aware of empirical work on the relationship between the strength of managerial incentives and the strategic behavior on the product market.

3.2 Delegation as Strategic Commitment

Following an influential paper by \textsuperscript{19}Fershtman & Judd (1987) (see also \textsuperscript{20}Vickers (1983)), a series of papers has considered the possibility for owners to delegate strategic decisions to managers while at the same time controlling these decisions by using incentive schemes. \textsuperscript{21}Fershtman & Judd (1987) consider incentive schemes in a Cournot oligopoly where the variable part is a convex combination of the profit and the sales made by the firm; that is the manager’s compensation is proportional to $\alpha \pi_1 + (1 - \alpha)S_1$. For simplicity, consider a duopoly. After each firm’s owners choose $\alpha_1, \alpha_2$ managers then choose their individual quantities $q_1, q_2$. (The owners can choose then to observe quantities or not and yardstick competition is not allowed.) One reason for delegation could be the superior information of managers about the demand level, something they will learn after incentive contracts are agreed upon.

In this framework, as $1 - \alpha$ is positive, a manager will be more aggressive in increasing sales; by contrast if $1 - \alpha$ is negative, managers will want to reduce sales. However, if firm 1 has $1 - \alpha_1 < 0$, the owners of firm 2 will want to provide sales incentives ($\alpha_2 < 1$) to their manager in order to push

\textsuperscript{14}The free rider problem in takeovers (Grossman & Hart (1980)) suggests that the dispersed or concentrated nature of initial ownership matters for cross-ownership. There is a tradeoff between acquiring controlling shares at a high price — and being able to coordinate the strategic choices of two firms — and acquiring silent or non-controlling shares at a lower price — and being able to internalize part of the profit of the other firm.

\textsuperscript{15}Flath (1992, 1991) who shows that with an efficient financial market, since the share price reflects the post trading value of the share, acquiring shares in rival firms is not part of an equilibrium with Cournot but is with Bertrand.

\textsuperscript{16}For instance Joskow & Tirole (2000) analyze how physical or financial rights affect the strategic behavior of firms in an electrical network and how the micro-structure design influences these strategies.
further firm 1’s manager to decrease his quantities; anticipating this, firm 1’s owners should provide sales incentives to their manager for otherwise they will lose market share. Hence in equilibrium of the incentive game between owners, both $\alpha_i$ are less than one. It follows that the commitment to an incentive scheme prior to quantity choices of managers eventually leads to higher output levels, lower profit levels for the owners than when there is no delegation. The authors also consider environments with differentiated goods or uncertainty about costs; in that later case they can show that as the number of firms becomes large, equilibrium incentive schemes converge to $\alpha = 1$, that is managers behave in a profit maximizing way.\footnote{Fershtman & Judd (1987) endogenize the behavior of firms, profit maximization or sales maximization, and therefore complete the literature of the 60s that assumed rather than derived a non-profit maximizing behavior of firms. However, their analysis begs the question of commitment of owners to incentive schemes. Indeed, the explicit assumption of observability of incentive schemes by all firms and the implicit assumption that owners do not renegotiate the contract with managers are de-facto creating commitment but is often a strong assumption (see Dewatripont (1988), Caillaud et al. (1995) for the possibility of using contracts as pre-commitments when renegotiation contracts cannot be prevented).}

The issue of commitment can also clarify the role of ownership in the property right theory and how distinct it is from the organization of production – an internal decision. For instance, consider a typical hierarchy with a HQ at the top and two divisions $D_1, D_2$. One may be tempted to argue (e.g., Riordan (2003)) that ownership is distinct from integration: (i) it is not sufficient since a HQ owning two divisions may decide to organize production in such a way that each division retains control rights on decisions; (ii) it is not necessary because $D_1, D_2$ who initially own the individual assets could decide to hire a HQ and delegate to him the right to make decisions. Hence the organization of production is distinct from ownership of assets. However, in both cases, the question is how credible is the transfer of decision rights when the owner (HQ in (i), $D_1, D_2$ in (ii)) keeps ownership, legal rights on the assets. Simply assuming commitment is not satisfactory, if HQ’s role is to make decisions, why would it not interfere with $D_1, D_2$ decisions when he has ownership? We will come back to this question soon.

In an incomplete contract framework, if ex-ante contracts cannot be made contingent on future state realizations, delegation rights must be independent of the state realization (by assumption) while ownership provides more flexibility ex-post, in particular on whether $D_1, D_2$ should be delegated the right to decide in a particular state. The decision to delegate is therefore a function of the ownership structure and the “owner” may prefer to wait to delegate than to commit to do so in the initial contract (if commitment is possible). This is a point that is well understood in the appli-
cations of incomplete contracting to corporate finance (e.g., Aghion & Tirole (1997), Dewatripont & Tirole (1994)), but that is somewhat made fuzzy in simple models since issues of delegation are brought in from an internal organization perspective while issues of (ownership) integration are brought in from a boundary of the firm perspective. This link may be more apparent when there are many dimensions on which to decide. For instance, legal rights can be transferred to another party for some dimensions only or an agent may retain the ability to take non-contractible actions that affect the quality of the final product independently of ownership (see Legros & Newman (2008) for a model along these lines.)

In imperfectly competitive markets, Wickelgren (2005) shows that the owner may want to create competition between two divisions that produce differentiated products. This balances effort incentives versus internal price competition, leading to a situation similar to what would happen with non-integration and yardstick competition for managers. Alonso et al. (2008) assume commitment to decision rights allocations in a model where a unique price has to be set for two divisions producing differentiated products with linear demand functions and where managers have private information about the level of the demand. If there is delegation, the right to choose the price is given to the manager facing the flattest demand function, but delegation arises only if the slopes of the demand functions are sufficiently different.

Ownership creates commitment but is a two-way sword: with delegation, the owners cannot commit not to intervene ex-post. Commitment to delegation therefore supposes that the owners have protected the manager to whom they delegate from interfering with his decisions: this could be done by transferring the legal right to make decisions on the asset (which is very close to ownership), or could be done for incentive compatibility reasons, e.g., the belief that the agent’s decision is the ‘right one’, either because the agent has more expertise or because even if there is a conflict of interests, he has superior information and the principal does not want to overturn his decision, or invest in information acquisition because his opportunity cost of doing so is high, e.g., because the principal is overloaded with other tasks.

Aghion & Tirole (1997) articulate such a mechanism and they show that in some instances it is better to give an agent ‘real authority’ than to give him ‘legal or formal authority’: real authority means that decisions made by the agent will not be overruled by the principal, while formal or legal authority means that the principal cannot intervene in the decision made by the agent. The cost of giving formal authority to the agent is the loss of control when the principal knows that the decision taken is not the right one, the benefit is that it increases the incentives of the agent to acquire information. The decision not to overrule is endogenous rather than assumed, as should be.

Finally, outside regulation like corporate governance guidelines may constrain the internal organization of firms and the possibility to mitigate the conflict of interests between different stake-

\[\text{18}\] Or as in Tadelis (2002) when the “design” is complex and requires ex-post adjustments.
holders. There is a recent literature on the role that corporate governance may play and its relation
to competition (see for instance, Vives (2000), Allen & Gale (2000)). Some ‘natural experiments’
(Bertrand & Mullainathan (2003), Giroud & Müller (2009)) suggest that managers are indeed pur-
suing a ‘quiet life’: after regulation that prevents takeovers or sales of assets, there seems to be a
significant increase in input costs, including workers salaries, and therefore managerial discretion
seems indeed to be a source of cost inflation in firms. The results in Giroud & Müller (2009) suggest
that the effect is more pronounced in less competitive industries (as measured by the Herfindahl
index) but they do not seem to control for the potential changes in internal organization or
incentive schemes that may have followed the regulatory change.

Sometimes delegation is necessary, like in procurement contracts. There is a recent literature
emphasizing the role of incomplete contracting for the design of procurement contracts and on
the relative benefits, for instance, of auctions versus negotiation. Without incomplete contracting,
auctions tend to emerge as the optimal way to procure (Bulow & Klemperer (1996)). But when
the design of the contract is “complex” and there is a significant risk of renegotiation, negotiated
contracts, with communication between the firms and the public agency, may provide more flex-
ibility or reduce the likelihood of renegotiation. Incomplete contracting could also explain why
seemingly suboptimal contracts like cost plus contracts – that reimbursed the contractor for cost-
overruns – in a complete contracting framework (since they provide poor effort incentives) can be
adopted in an incomplete contract framework (because they reduce the need for renegotiation).

3.3 Foreclosure and Secret Contracts

Hart & Tirole (1990) provide one of the first analysis of the role of integration as a foreclosure instru-
ment. Their analysis highlights the importance of endogeneizing rather than assuming commit-
ment and entry, like it is done for instance in the contemporary paper by Ordover et al. (1990).

19 The use of the Herfindahl index for measuring the intensity of competition is borrowed from the Cournot model
but is not necessarily a good measure for other market structures, or as Farrell & Shapiro (1990) have argued when
assets can be traded.

20 See Goldberg (1977) for an early argument and Bajari & Tadelis (2001), Bajari et al. (2009) for more recent
inquiries about the relationship between complexity and the use of auctions.

21 For instance, when the upstream firms have the same marginal cost of production, Ordover et al. (1990) show that
integration is beneficial in contrast to the results in Hart & Tirole (1990) for the symmetric case. This is because, the
ability of the integrated firm to commit not to supply to the other downstream firm eliminates competition between
the upstream firms to supply to the other downstream firms, eventually benefitting both the integrated firm and the
other upstream firm. O'Brien & Shaffer (1992) consider the case of a single upstream producers and downstream
firms competition à la Bertrand (hence avoiding the usual ‘double marginalization’ motive for integration). They
assume that the upstream firm cannot commit to re-contracting and therefore that retail prices will be smaller
than under integration, even if the upstream firm can choose two-part tariffs, providing a new rationale for vertical
restraints linked to contract incompleteness. Alexander & Reiffen (1995) identify the strategic values of different price
and non-price vertical restraints, as well as the role of external enforcement mechanisms, like regulation. Schmitz
& Sliwka (2001) assumes that the supplier of a good can invest to enhance quality, which will increase the value
to the buyer and also the residual value of the asset in case bargaining fails, and in a degree of specificity, which
also enhances the value of the good to the buyer but decreases the residual value in case of failure of bargaining.
Integration depresses incentives to invest but increases specificity, while non-integration has opposite effects.
Two downstream firms $D_1, D_2$ sell a homogenous product and can produce one unit of output if they buy one unit of input. Contracts between an upstream firm and a downstream firm for the provision of input levels can be in two-part tariffs and it is not possible to prevent the upstream firm to sell to another downstream firm. Under these conditions, even if there is a unique upstream seller $U$, his profit when the two downstream firms are separate units cannot exceed that obtained under Cournot competition. Indeed, for any quantity of input $q_1$ purchased by $D_1$, $U$ maximizes his revenue from selling inputs to $D_2$ by offering that firm $q_2 = B_2(q_1)$, the best response of $D_2$ to $q_1$; anticipating this, $D_1$ will be willing to pay at most $\pi_1(q_1, B_2(q_1))$ to $U$. If $D_2$ has the same anticipations, it is optimal for $U$ to offer the Cournot quantities to the downstream firms. If however $U$ integrates with $D_1$, the merged entity will internalize the profit made by $D_1$ and therefore will prefer to sell the monopoly quantity and not to supply $D_2$: integration generates commitment not to sell to the other downstream firm.

As Hart and Tirole argue, the commitment generated by integration is superior to what could be achieved by trying to put in place exclusive contracting clauses. For instance, with other upstream suppliers — that are less efficient — an integrated firm $(U1, D1)$ will serve $D2$ since $D2$ would be served by $U2$ otherwise, something that an exclusive contract will not achieve. A ‘contingent’ contract that would limit the quantities sold to $D2$ may also be difficult to implement in practice since it requires a high degree of information.\textsuperscript{22}

In the general model,\footnote{Hart & Tirole (1990)} allow for a fixed cost of entry (or investment needed to operate) by upstream and downstream firms and profit sharing in case of integration. They show that more efficient firms have more incentives to integrate, which in turn leads to a ‘waterbed’ effect. Indeed, since the profits of the non-integrated downstream firm decrease, absent integration, this firm may exit; it is then in the interest of the other upstream firm to integrate and share the costs of investment in cost reduction.

In variants of the basic model,\footnote{Bolton & Whinston (1993)} Hart and Tirole consider situations where each upstream firm has a fixed capacity upstream (“scarce supplies”) or downstream (“scarce needs”), each $D_i$ needs only one unit of input, like a piece of equipment). Integration modifies the bargaining positions of the parties: integrated firms have a higher bargaining power since they benefit from ‘supply assurance’ and therefore the non-integrated firms will tend to have lower profits and lower incentives to invest.

\textsuperscript{22}See McAfee et al. (1989) and Marx & Shaffer (2004) for an analysis of the possibility of commitment through non-discrimination clauses. See also Segal (1999, 2003) for a general analysis of contracting with externalities.
The demand realizations are perfectly correlated: with equal probability demand realizations are \((v_H(x_1), v_L(x_2))\) or \((v_L(x_1), v_H(x_2))\). Allocating the input to the downstream firm facing the highest willingness to pay of consumers is efficient and total welfare is \(\frac{1}{2}v_H(x_1) + \frac{1}{2}v_H(x_2) - c(x_1) - c(x_2)\). Therefore the marginal incentives for investment of the firms are \(c'(x_i) = \frac{1}{2}v_H'(x_i)\).

With non-integration, each downstream firm will compete for the input à la Bertrand and therefore the highest value firm will pay the second highest value; hence conditional on having the highest value the net surplus of downstream firm \(D_1\) is \(v_H(x_1) - v_L(x_2)\) and his expected payoff is \(\frac{1}{2}(v_H(x_1) - v_L(x_2)) - c(x_1)\) implying the same marginal incentive to invest as in the social optimum. Hence investment levels with non-integration are efficient.

As Bolton & Whinston (1993) show, non-integration while leading to efficient investment levels is not always stable since the upstream firm and a downstream firm may strictly gain by integrating. Suppose that \(U, D_1\) integrate; the vertical structure has a surplus of \(v_H(x_1)\) with probability \(1/2\) (in this case \(D_2\) does not produce) and a surplus of \(v_L(x_1)\) with probability \(1/2\) (\(D_2\) purchases the input from the vertical structure at the opportunity value of \(v_L(x_1)\)). Hence the expected surplus of the vertical structure is \(\frac{1}{2}(v_H(x_1) + v_L(x_1)) - c(x_1)\): the marginal incentive to invest increases by \(v_L'(x_1)/2\) with respect to non-integration. Note that the incentives to invest of \(D_2\) are the same as with non-integration since he has the same expected surplus function. However, because the investment of \(D_1\) increases, the equilibrium value of \(D_2\) decreases (and by a larger amount than the equilibrium payoff to the vertical structure increases since investment choices are not first best).

These papers integrate two effects present in the literature: an efficiency motive for integration (see Lafontaine & Slade (2007) for a survey of the literature) and market foreclosure or raising rivals’ cost strategic considerations. Empirically, it is difficult to disentangle the two effects; for instance, highly concentrated market are prone to foreclosure effects but are also often also characterized by relation specific investments, implying therefore also an efficiency role for integration.

4 Ownership and Liquidity Constraints

4.1 Liquidity, Limited Liability and Rent Extraction

Grossman & Hart (1986) assume that the market for ownership functions without frictions, implying that the ex-ante ownership contract will be total surplus maximizing. In particular, if under non-integration the parties have payoffs \(v_i^N, i = A, B\) while under integration, and transfer of property rights to \(A\) they have \(v_i^I, i = A, B\), where \(v_A^I > v_A^N, v_B^I < v_B^N\), then the owner of the \(B\) asset is willing to relinquish ownership only if \(A\) is able to perform a transfer of \(v_B^N - v_B^I > 0\). As long as both \(A\) and \(B\) have enough cash, or can borrow from the financial market without distorting ex-post decisions (which is often a strong assumption since debt creates distortions, see for instance Jensen (1986)) it follows that \(A\) and \(B\) will choose the ownership structure that maximizes the
total surplus: that is integration arises only if $v_I^A + v_I^B > v_{IA}^A + v_{IB}^N$. However, if either $A$ or $B$ has limited cash holdings, or more generally limited access to finance that is not distortionary (as we already observed, debt financing will be distortionary since the debt repayment tends to reduce the marginal return from decisions), the surplus efficient ownership structure need not emerge. In this example if $B$ could generate more than $v_I^B$ but less than $v_N^B$ by not contracting with $A$, then if $A$ is cash constrained, non-integration will emerge.

Limited cash holdings is obviously a central theme in the corporate finance literature: it is one of the main reasons for an entrepreneur to engage into financial contracts. The literature has emphasized that some forms of financing may have better incentive effects than others (like debt versus equity in Jensen (1986)), and the ownership of assets will have to be unbundled from the residual rights on financial flows (Aghion & Bolton (1992), Dewatripont & Tirole (1994)). This literature on the applications to finance of the incomplete contract idea is surveyed by Patrick Bolton in this volume.

In agency settings, it is well understood that agent’s limited liability creates a rent extraction motive for the principal. For instance, even if the surplus maximizing (“first best”) effort is implementable, the principal may choose to implement a smaller effort in order not to give rents to the agent. When the outside option of the agent is large however, this rent extraction motive is ‘disciplined’ since the principal will need to give a sufficient share of the surplus to the agent. As we have seen earlier in the model of Schmidt (1997), whether or not the rent extraction motive is present has implications for whether product market competition covaries with incentive provision within firms. Hence, both the cash holdings and the bargaining positions (outside options) of the parties matter. While cash holdings can be observable, the bargaining positions of the stakeholders in the firm are more difficult to assess and they are often a function of what happens outside the firm, in particular what other firms do.

Beyond its effect on incentives in principal-agent models, the rent extraction motive also influences organizational choices and their efficiency. For instance, when the outside option of the agent is low, the principal may favor investing in a costly monitoring technology in order to avoid (or reduce) the moral hazard problem and paying rents to the agent. In an incomplete contract framework, as we have seen, integration decisions may be a nontrivial function of the cash holdings of the parties, and, as the moral hazard example suggests, of their bargaining positions. This

---

23Bolton & Whinston (1993) also show that if we consider more than three parties, the “efficient” ownership structure may fail to arise even if parties can make ex-ante lump-sum payments as long as payments for re-allocation of property rights cannot be made contingent on future sales of assets.

24De Meza & Lockwood (1998) show that in the Grossman & Hart (1986) model, if agents have outside options at the negotiation stage, many of the results are qualitatively different. For instance, ownership should not necessarily go to the agent whom investment is the “most important”. See also Chiu (1998) who shows that not having ownership may increase investment incentives. The difference with Grossman & Hart (1986), Hart & Moore (1990) is that they ‘bundle’ outside options and threat point in the ex-post bargaining while Chiu (1998) follows the bargaining literature tradition that views the outside option as the payoff to the agent if he quits the bargaining table while the threat point is the (discounted) payoff the bargainer obtains before negotiation is terminated. Having ownership leads to
suggests that ownership allocation may be coupled with additional constraints on the use of the assets or on the design of the renegotiation process, but also that the macroeconomy, or any shock that may change the value of outside options, will affect ownership allocation and the performance of firms.

In an application to R&D ventures, Aghion & Tirole (1994) show theoretically how cash constraints distort contracting and allocation of property rights to parties involved in joint R&D efforts: if the research unit is cash constrained and does not have a strong bargaining power, the customer will not — like for B in our example above — be willing to transfer ownership of the unit, despite the performance gain. However they focus on one relationship, assuming specific bargaining positions of the different parties. This limits the scope of positive analysis since bargaining positions are in general endogenous, reflect the organizational choices of other firms, and also reflect the distribution of cash endowment: in an equilibrium of the industry, bargaining positions and organizational choices are jointly determined. Therefore, comparative statics on organizational choices as a function of changes in the bargaining positions are not very informative, unless one can guess a-priori how bargaining positions are affected by shocks. An illustration of this point is the role that the distribution of liquidity endowments may have on contractual choices in moral-hazard settings.

A Simple Illustration of the Role of the Distribution of Liquidity

Suppose that integration is surplus maximizing in our example above: \( v_A^I + v_B^I > v_A^N + v_B^N \) but that \( v_B^N > v_B^I \). Then, in order for B to agree to integration it must obtain a lump sum transfer of at least \( t^* = v_B^N - v_B^I \). If this is the only relationship in the industry, then as the cash holding of firm A increases, integration is more likely: the size of the surplus generated in the industry is increasing in the cash of firm A. This is the usual ‘partial equilibrium’ intuition one gets from corporate governance: cash in the relationship creates a positive contracting externality.

Suppose now that there are two assets \( A_1, A_2 \) and that the owners of these assets have different cash holdings \( l_1 \leq l_2 \). Assets are identical in terms of their productive capacity: as long as an asset \( A_j \) is combined with an asset \( B \), the payoffs are given by \( v_A^k, v_B^k, k = I, N \). If an asset \( A \) is not in a relationship with an asset \( B \), his payoff is equal to zero. Because there are two assets of type \( A \), one of them will be left unmatched; since \( l_1 \leq l_2 \), the owner of \( A_1 \) is at a competitive disadvantage and will be left unmatched. Assuming that \( l_1 \leq v_A^N \), the most that this owner can offer to the owner of \( B \) is \( v_B^N + l_1 \), and this is therefore what \( B \) will want to obtain in order to be in relationship with

---

25 As in Aghion et al. (1994).
26 As far as we know, Legros & Newman (1996) is the first paper to make this point.
27 This example is based on Legros & Newman (2008).
Hence, \( v_B + l_1 \) is the outside option of \( B \) in a relationship with \( A_2 \) while the outside option of \( A_2 \) is zero.

\( A_2 \) would prefer to have integration but \( B \) will be willing to accept integration only if he receives a transfer \( t_2 \) from \( A_2 \) such that \( v_B + t_2 \geq v_B^0 + l_1 \), that is, if \( l_2 \geq t_2 \geq t^* + l_1 \), where the left inequality is the cash constraint for the owner of \( A_2 \). Clearly, as \( l_1 \) increases, the condition is less likely to be satisfied. For instance, starting from a situation where \( l_2 = t^* + l_1 \), increasing \( l_1 \) by \( 2\delta \) and increasing \( l_1 \) by \( \delta \) will lead to a violation and more cash in the industry will lead to non-integration; similarly, a redistribution of liquidity from \( l_2 \) to \( l_1 \) will also lead to non-integration. This is a new role for cash: cash in the industry may create a negative contracting externality on firms, even if they all benefit from cash injection. Note that as \( l_1 \) increases \( A_2 \) effectively loses market power in its relationship with \( B \); as a result he loses a chance to integrate, that is less likely to have power inside the firm when his market power decreases.

This example also illustrates that shocks to some firms (\( A_1 \) above) may lead to reorganization of other firms, whether or not these firms have been subject to shocks. For instance, suppose a technological shock to asset \( A_1 \) which becomes more productive: under both forms of organization the payoffs to both asset holders increase by \( \delta \). By contrast, asset \( A_2 \) is not subject to a shock. Then, if \( l_1 + \delta < l_2 \), the outside option of \( B \) is now \( v_B^0 + l_1 + \delta \). Hence, if \( l_2 \in (t^* + l_1, t^* + l_1 + \delta) \), in the relationship \((A_2, B)\) there would have been integration before the productivity shock but after the shock there will be non-integration. Hence, we would observe a change in organization for the relationship \((A_2, B)\) despite the fact that it faces the same technology: it is the change in technology of other firms that leads to the re-organization of the firm.

Since the ‘external’ effect of cash may counterbalance the ‘internal’ effect of cash, and since the distribution of cash may affect the relative strengths of the two effects, it is necessary to consider organizational choices by all firms in the market in order to understand how organizational choices by individual firms are affected by shocks.

\[ \text{Legros & Newman (1996)} \] analyze the role of the distribution of cash endowments on the use of monitoring of effort when production is joint and production requires a fixed capital investment. Two agency problems have to be solved: one internal due to a moral hazard problem in effort provision, and one external due to an asymmetry of information about the output of the firm between its members and outside lenders. Firms which cannot self-finance have to rely on a costly monitoring (of effort) or auditing (of output) to credibly commit to repay their lenders. With a severe enough financial market imperfection (when the cost of auditing is large), distribution matters in an important way, both for the allocation of surpluses in firms and for the types of firms that emerge: whether they use internally the costly monitoring technology and also the size of firms as captured by the number of members exerting effort. \[ \text{Legros & Newman (2008)} \] extend these insights

---

28 See \[ Legros & Newman (2008) \] for a more general analysis of the external effect of liquidity or technological shocks.
29 See \[ Banerjee & Newman (1993) \] for an early illustration of this point in a model where a principal has to decide whether to invest in a monitoring technology for his agent or not.
to analyze the endogenous emergence of ownership structures in a model where ownership rights can be allocated continuously. Provision of effort can be done either through monetary compensations — which affect effort provision in a continuous way — or through ownership rights — which affect effort provision in a discontinuous way since the “owner” cannot commit to internalize the disutility of effort of the other agent. When the distribution of surplus is relatively equal, it is done through changes in compensation schemes, which have second order effects, but when the distribution of surplus is relatively unequal, it is done through changes in the ownership structure, which has first order effects. Hence factors that affect the distribution of surplus in firms, e.g., a change in outside options, will have in general a discontinuous effect on ownership allocations.

As a first empirical illustration of the role of surplus distribution in determining organizational design in the presence of liquidity constraints, consider the trucking industry. In the 1980s and 1990s the trucking industry in the US experienced a shift away from drivers who owned their own trucks toward employee drivers. This organizational change has been attributed to various technological developments, such as the introduction of “on-board computers” (OBCs), which offered both better monitoring of driver actions and greater flexibility in dispatching, permitting more efficient use of trucks (Baker & Hubbard (2004)).

By the early 2000s, the prevalence of owner operators and use of OBCs had stabilized. But more recently, the industry has begun to shift some control back to drivers. Between 2004 and 2006, carriers began offering drivers such “perks” as the right to travel with spouses or to outfit their cabs with satellite televisions. Since drivers decide whether and when to exercise these rights, they constitute an increase in their control. The question is why there has been a shift of control allocations in trucking without an apparent technological shift.

A possible answer comes from the observation that an important alternative employment for truckers is construction, which experienced a boom in the early 2000s. Thinking of the drivers as A’s, and the construction-cum-trucking firms as the type B’s, since the outside options of drivers in trucking firms increases, they indeed get more “control.”

Another empirical illustration of the role of outside options and surplus division is Ciliberto (2006) which provides an empirical analysis of the integration decisions of hospitals and physicians and their effects on investment decisions by hospitals. One of the assets in the relationship between physicians and hospitals is the list of patients over which independent physicians have control rights (since they decide whether they are admitted to the hospital). While in the past patients had the freedom to choose their physician and hospital, the Managed Care Organizations (MCO) now bargains with hospitals and physicians for a set of patients. This increase in bargaining power by MCOs has been accompanied by an integration of hospitals and physicians and a transfer of control over patients from the physicians to the hospitals. Integration should therefore increase the incentives of the hospitals to invest in quality care, and Ciliberto (2006) indeed finds such a positive
effect of integration on investment in his empirical analysis.

Finally, let us note that even if there is no effect on the organization of firms, the distribution of liquidity will affect the ability of the private sector to channel funds to entrepreneurs. For instance, Holmström & Tirole (1997) analyze the effect of credit crunches when firms have different liquidity endowments. Firms with large liquidity can borrow from an (uninformed) capital market but firms with less liquidity — hence less collateral — have to borrow from intermediaries using a costly monitoring technology. Credit crunches affect the ability to borrow from low liquidity firms, increasing the cost of monitoring and the spread in interest rates paid by high and low liquidity firms. Their results are consistent with stylized facts, in particular the fact that the 1990-1991 change in bank lending is correlated with a change in the equity value of the banking sector.

One of the exciting avenues for future research would be to couple the imperfections in the financial market with the endogenous determination of the organizations of firms.

4.2 Debt as Commitment

Jensen & Meckling (1976) have highlighted the agency cost of debt: equity holders will tend to discount low profit realizations and engage into projects that have a negative present value. As we have seen in section 2, increasing the risk of bankruptcy may also serve as a disciplining device for managers in agency models and Grossman & Hart (1982) show that debt can be used as a bonding device in principal-agent models since the risk of bankruptcy increases in the case of low profit realizations. In industrial organization, in an influential paper Brander & Lewis (1986) argue that there is a nontrivial relationship between the debt-equity ratios and the performance of a duopoly: on one side, debt will induce firms to discard profits in low states that lead to bankruptcy and therefore pursue strategies that increase the profit levels in the good states but, on the other side, firms will pursue strategies that give them a competitive advantage, in particular that can drive out of the market other firms which have a high debt-equity ratio. This leads to more aggressive behavior in Cournot oligopolies and less aggressive behavior in Bertrand oligopolies, but in both cases the profit increases with the use of debt.

While the result that debt makes oligopolistic firms more aggressive has been challenged by

30Furthermore, the “more important” are MCOs – e.g., the more managed care patients belong to a MCO – the more important is the effect of integration on investment. What is less clear from this evidence is whether possible countervailing incentives of investment by the physicians has led to an overall efficiency increase.

31Holmström & Tirole (1998) consider the role that public provision of liquidity (e.g., bonds) may play to correct for the inability of the public sector to channel funds in the presence of liquidity scarcity; see also the recent book Holmström & Tirole (2011) for a survey of the literature on this point.
different authors, both theoretically and empirically, the general idea that financial structure has consequences for the strategic behavior of firms is an important message for industrial organization and receives growing attention (see the recent survey by Lyandres (2006)). It also confirms our previous claim that industry studies or analysis of market conduct should control for macroeconomic variables. For instance, Chevalier & Scharfstein (1996) provide evidence from supermarket pricing behavior that markups are counter-cyclical and that more liquidity constrained firms are more aggressive during recessions.

5 Toward an Organizational Industrial Organization?

5.1 Another Look at Perfect Competition

The previous discussion shows that there is a long history of contact between the theory of the firm and IO, but the two areas are far from integrated, so to speak. In particular, what is lacking is a set of benchmark models that could be used as a launching point for a more comprehensive organization-based industrial economics. While the previous literature has focused on managerial effort and hold-up as the main determinants of integration, another important source of conflicting interests between non-integrated asset holders could be disagreement over which of several standards should be chosen around which to coordinate production. Hart & Holmström (2010) provide a model that emphasizes this tradeoff, and we will use it to illustrate the benefit of using an incomplete contract setting in one the most basic market structure in industrial organization, perfect competition.

We note indeed that much of the literature we have discussed proceeds on the assumption of market power (with the occasional interloping search frictions). Now while a certain amount of Showalter (1995) shows, the effect of debt on strategic behavior depends not only on the substitution or complementarity of the strategic variables but also on the type of uncertainty firms face. Brander & Lewis (1986) consider demand uncertainty, and Showalter (1995) shows that with cost uncertainty firms become less aggressive. Faure-Grimaud (2000) considers endogenous financial contracts and shows that the Brander & Lewis (1986) effect of debt making firms more aggressive competitors is offset by the cost of financial instruments in a Cournot oligopoly. Povel & Raith (2004) also consider endogenous debt contracts and introduce an expected cost to liquidation that is increasing in the amount of default: this makes the equity holder liable for low profit states and reduces the agency cost of debt, which would suggest that firms with debt may be less aggressive. Jacob (1994) shows that issuing long term debt may be an instrument for collusion: since contrary to Brander & Lewis (1986) firms are engaged in a long term competition, high profit realizations make the firm less aggressive since the opportunity cost of competitive wars, following deviations from collusion, is larger. The difference of results is illustrative of the ambiguous predictions in oligopoly theory when one goes from a static analysis to a dynamic analysis: in general, ‘bad’ equilibrium outcomes in a static setting, like more aggressively in setting low prices, help in a dynamic context since they provide a credible punishment in the case of deviations.

As Chevalier (1995) uses data from supermarket and shows that an LBO announcement has a positive effect on the market value of competitors, suggesting that debt softens competition, see also Dasgupta & Titman (1998). Aghion & Bolton (1988) also emphasize the role of financial contracting as a barrier to entry. In their model a dominant firm may engage in signal jamming in order for the financial market to put a high probability on a low state of demand and restrict access to credit by financially constrained competitors.

Business cycle effects have received attention in the industrial organization literature in that they affect collusive behavior, see Rotemberg & Saloner (1986), Bagwell & Staiger (1997), and the survey by Levenstein & Suslow (2006).
of ex-post market power (including those that might arise from search frictions) is necessary for incomplete contracts to have much bite (in particular the idea that the parties are at least partly locked into the relationship), it would be wrong (not to mention unfortunately limiting) to infer that organizational design is necessarily linked to market power ex-ante. Indeed, Grossman & Hart (1986) are explicit about this: they have in mind a competitive market for partners at the contracting stage, though their assumption of non-binding liquidity may obscure this somewhat by obviating the need to consider the equilibrium distribution of surplus among the contracting parties.

There are several reasons to follow the Grossman & Hart (1986) lead and stick as much as possible to perfect competition. First, from a theoretical point of view, when trying to understand the specific contribution of contractual incompleteness and organizational design on market performance, it is desirable to consider first an environment in which no other imperfection is contributing potentially confounding distortions. This is not to say that the distortions from market power and search frictions are unimportant; indeed the interaction between them and contractual incompleteness is certainly an exciting agenda item for future research. Relatedly, looking at the perfectly competitive case is a useful tool for fighting the impression that organization is only a handmaiden to market power, with effects that are complicated to distill or identify in practice and possibly not so different from the many strategic effects that have already been identified by the IO literature.

Second, there is the question of the practical significance of market power. This has been argued empirically since the days of Harberger, whose eponymous triangles have repeatedly been shown to be small in most industries. The initial work by Harberger (1954) suggested a very low (0.1%) welfare loss but the study has been criticized, in particular, for using profit data in the estimation and for assuming linearity of demand. More recent empirical studies lead to computations of welfare loss in the range 1% – 7%. On the theory front, a symmetric Cournot model with linear demand and constant marginal costs will predict a loss of \(1/(n+1)^2\) when there are \(n\) firms, hence a loss of 1.23% for nine firms and a loss of 2.77% for five firms. And if one incorporates in the Cournot model the delegation models à la Fershtman & Judd (1987) the welfare loss is of the order of \(1/n^4\), hence a negligible welfare loss with four symmetric firms.

This does not mean obviously that imperfect product market considerations are not relevant, but it is possible that the welfare losses from the exercise of market power may be indeed small in many industries, and this makes it all the more relevant to start with competitive product markets.

\(^{36}\)See for instance Cowling & Mueller (1978) for a review of these criticisms. Another criticism is the non-inclusion by Harberger of the cost of acquiring monopoly power. When controlling for such costs, Cowling & Mueller (1978) find losses in the range of 3.9% – 7.2% while Harberger’s methodology would lead to losses in the range 0.2% – 3%.

\(^{37}\)For instance, Maudos & Deguevara (2007) find a welfare loss in the banking industry of 0.54%; Dickson & Yu (1989) in their study of Canadian manufacturing find a welfare loss bounded below by 0.51% (under Cournot) and above by 6.35% (under a collusive equilibrium) of GDP.

\(^{38}\)See Corchon (2008); Anderson & Renault (2003) show that there is a non monotonic relationship between the concavity of demand and welfare losses. This assumes lack of collusion among the oligopolists.

\(^{39}\)See Ritz (2010); obviously these computations assume that delegation creates more aggressive behavior by managers, but as we have noted in section (1.2) there is disagreement on this.
in order to understand whether organizational choices and ownership allocations matter in such settings.\footnote{And is meant to say even less that competition policy is irrelevant: insofar as actual market power is small, the reason for it could be precisely because of effective competition policy!} Somewhat surprisingly, since the initial work of \cite{Leibenstein1966}, there has been little formal empirical work estimating the welfare loss (including that of consumers rather than simply shareholders) due to managerial discretion. But there is at least some evidence that it matters, and a growing sense in public discussion that firms’ internal organization and governance do affect consumers in important ways.\footnote{Examples include a recent “Frontline” piece on the internal organization of airlines, and specifically whether outsourcing led to the crash of flight CO3407 outside Buffalo in 2009; and many indications that the organization of banks and mortgage lenders were at least contributory in generating the subprime meltdown and global financial crisis.}

In the rest of this section we will sketch what such a model might look like. We will focus on perfect competition and see what it gives us.

### 5.2 Ownership and Coordination

In the basic version of their model, \cite{Hart&Holmstrom2010} consider final goods that are produced with two complementary inputs; call them $A$ and $B$. Input suppliers are an indivisible collection of assets and workers, overseen by a manager. During the course of the relationship, non-contractible production decisions will have to be made. The important point is that these decisions need to be coordinated to enhance output: it is not so important what decision is chosen in each unit, as long as it fits what is happening in the other. The organizational problem arises from a conflict of interest over what decision is optimal. For instance, if $A$ represents product development and $B$ manufacturing, then product designs that are easy for $A$ to develop may be difficult for $B$ to manufacture, and vice versa. This could be an effect of the nature of technology, the result of past training and experience, a concern over reputation, or even a difference in beliefs about the best course of action (i.e., “vision”).\footnote{In contrast to \cite{Grossman&Hart1986}, they assume that ex-post bargaining with unlimited side transfers are not feasible since agents use contracts as reference points and would feel aggrieved if they do not get the best outcome under the initial contract.}

Thus if each retains control over his own decisions (which is interpreted as non-integration), there will tend to be poor coordination: if the two managers were to maximize profit by picking perfectly matched actions, each would want to deviate by moving (at least slightly) in his preferred direction. The coordination problem can be “solved” via integration. H-H depart from G-H in a second way here, by assuming that integration involves a sale of assets not to one of $A; B$ but rather to a third-party professional manager who then acquires decisions rights over both $A$ and $B$. The professional manager is concerned only about the enterprise’s revenue and will therefore have incentives to maximize it by perfectly coordinating decisions. The problem is that this imposes costs on the $A$ and $B$ managers in terms of private benefits.
We shall have more to say about this model soon. We will here remark on its potential significance in the light of some recent empirical work by Hortacsu & Syverson (2009). They survey intra-firm shipping data for US manufacturers. The surprising finding is that among domestic firms, very few of the inputs produced by one subsidiary are shipped to other subsidiaries of the same firm, implying that vertically integration does not comprise single supply chains (for multinational firms, the traditional picture, wherein one subsidiary produces an input that is consumed by the next firm in the chain, does seem to hold up). Rather, the picture of vertically integrated firms that emerges is that of a portfolio of input suppliers (or a university), with each subsidiary selling to its own market, in which the common owner provides certain (unmeasured) managerial services. While these findings are controversial, if true, they cast some doubt on the empirical relevance of the holdup problem as a major determinant of firm boundaries, at least among US domestic manufacturers.

The Hart and Holmström model is consistent with these findings, and indeed with Hortacsu and Syverson’s own explanation for vertical integration as a conduit for (unobserved by them) managerial services. Several suppliers that are related because they produce complementary goods but not actually transacting with each other might nevertheless benefit from services that are a single style or fit a common standard, say for brand recognition, marketing and distribution, logistics and so on. But it is harder to see how a hold-up problem could arise or spot adaptation decision would need to be made if the input produced by $A$ is not actually being shipped to $B$ in equilibrium. This does not mean, of course, that hold-up problems are irrelevant. But incomplete-contracts models articulated around a tradeoff between coordination of decisions versus private costs of these decisions may provide a reasonable and simple benchmark model for performing market analysis.

### 5.3 A Basic Model of Vertical Industry Equilibrium

The idea that it is necessary to look at the level of the industry in order to understand integration decisions of single firms has a long lineage in IO as is the view that heterogeneity is part of the equilibrium or that “the costs and benefits of vertical integration by a given firm would in general depend upon the extent of vertical integration by other firms in the industry”; see Perry (1989) for a survey of the literature on vertical industry equilibrium pre-1986. Nevertheless, there are very few attempts, both in contract theory or in industrial organization, to understand the properties of such vertical equilibria.

McLaren (2000) and Grossman & Helpman (2002) develop the first models to explain the pattern of outsourcing in industries when there is incomplete contracting. Both papers proceed in the Williamsonian tradition where integration alleviates the hold-up problem at an exogenous fixed cost. McLaren (2000) observes that when the market for specialized input thickens, there are less opportunities for hold-up and therefore a greater tendency for non-integration. This generates
strategic complementarities between organizational choices and the possibility of multiple vertical industry equilibria. This also explains why globalization, interpreted as market thickening, leads to non-integration and outsourcing. Grossman & Helpman (2002) develop the same types of tradeoffs in a monopolistic competition model with free entry, and use it to answer more industrial organization questions like the role of the matching technology, the role of demand elasticity and the possibility of heterogeneity in organizational choices.

Instead of hold-up, Legros & Newman (2009) employ a continuous-action version of the Hart & Holmström (2010) (H-H) model as the organizational “building block” for a perfectly competitive benchmark model of an industry, replacing the neoclassical “firms” in that model by “enterprises” whose ownership structures are endogenous. Specifically, they consider a market composed of a large number of H-H-style enterprises each of which takes the market price for the (single) product as given. The price is determined endogenously in the model via a market clearing condition. Here we present a simplified version of that model. As in their paper, we assume for the moment that there is a single product (so it is natural to think of H-H’s A and B as part of a supply chain).

Each enterprise (A, B) pair jointly produces the single consumption good, which fetches a price $P$ in the market. There are three non-contractible production “styles”, L, M, or R that can be adopted in each unit. If they make the same choice, the quantity produced is 1. If their choices differ by one ([L, M], [M, R], etc.) the output is $\lambda < 1$, by two (e.g., [L, R]) it’s $\lambda^2$. A’s costs of L, M, R are 0, c, C, while B’s are C, c, 0 where $2 < \frac{C}{c} < \frac{1}{\lambda+1}$. A receives a fixed payment from B, which receives the residual. The output price is $P$, which A and B take as given.

Under non-integration, A and B choose their production decisions simultaneously. The payoffs net of fixed transfers can be represented by the matrix (A is the column player, B the row)

$$
\begin{array}{ccc}
L & M & R \\
L & P - C, 0 & \lambda P - C, -c & \lambda^2 P - C, -C \\
M & \lambda P - c, 0 & P - c, -c & \lambda P - c, -C \\
R & \lambda^2 P, 0 & \lambda P, -c & P, -C \\
\end{array}
$$

A’s dominant strategy is to play L, and it is straightforward to verify that since the parametric assumptions on $\lambda$, C, and c imply $\frac{C}{\lambda+1} < \frac{C}{\lambda+1} < \frac{C}{\lambda+1}$, B plays R when $P < \frac{C}{\lambda+1} \equiv P$ and L when $P \geq \hat{P}$, yielding payoffs for B of $\lambda^2 P$ if $P < \hat{P}$ and $P - C$ if $P \geq \hat{P}$.

Under integration, a professional manager with zero opportunity cost, whose payoff is increasing in income but is indifferent about L, M, and R, purchases the assets via fixed transfers, thereby acquiring control of decisions and access to the revenue stream, which he maximizes. Since he is indifferent among (L, L), (M, M), (R, R), we assume that he picks the efficient choice, which is (M, M), as $2c < C$.

Now, B’s payoff under integration is $P - 2c$, which exceeds his payoff under non-integration when
fully conceding to $A$, which is $P - C$. However, as we have seen $B$ does not always want to concede under non-integration. If $P - 2c < \lambda^2 P$, or $P < \frac{2c}{\lambda^2} \equiv P^* < \hat{P}$, then under non-integration, the managers play $(L, R)$, and $B$ therefore prefers non-integration. But if $P \geq P^*$, then integration is preferred by $B$ to non-integration, regardless of how it is played in equilibrium.

Observe that the “transaction cost” associated with non-integration depends on the market price: the loss is $(1 - \lambda^2)P$ (the value of foregone output) when $P < \hat{P}$, and $C$ (the large private cost borne by manager $B$) when $P \geq \hat{P}$. And as a consequence, the choice of ownership structure depends on the market price.

The reason for this result is very simple: at low prices, non-integration is not well coordinated and produces little, but this is of little consequence to the managers because extra output generated by integration valued at a low price would not offset the private costs. At higher prices, the loss of revenue from inefficient production is worth the higher private cost, but since integration imposes a lower cost then non-integration at such prices, integration is preferred.

Second, in this example, the ownership structure that emerges in equilibrium is (first-best) efficient from the $A$-$B$ managers’ point of view. This is not a general conclusion, however, and is largely an artifact of the discrete action choices in the model. If choice were continuous, then (1) first best action choices would never coincide with non-integration action choices at any price, and (2) professional managers would typically “over-coordinate” the activities on the two sides because they do not take account of the private costs, so integration would not correspond to the first-best either. See Legros & Newman (2009) for an elaboration of this point.

However, as we shall see, there is another sense in which the equilibrium outcome even in this example may be inefficient.

First though, we want to do some positive analysis of a market composed of many price-taking enterprises like this one. Observe that an individual enterprise’s output is $\lambda^2$ for $P < P^*$, $1$ for $P > P^*$, and $\{\lambda^2, 1\}$ for $P = P^*$. Suppose that an industry is composed of a large number (continuum with unit measure) of enterprises just like this one. Then the industry supply will also be $\lambda^2$ for $P < P^*$ and $1$ for $P > P^*$, and will be $[\lambda^2, 1]$ for $P = P^*$.

This “organizationally augmented” supply curve embodies not only the usual price-quantity tradeoff, but also the organizational design: as we move along the supply starting at $P = 0$, we have non-integration for $P < P^*$ and integration for $P > P^*$. Moreover, at $P = P^*$, managers are indifferent between the two ownership structures. Since any mix of integrated and integrated firms would make managers happy at this price, the supply is “horizontal” there. See Figure 1 where we have also added a standard demand curve.

The competitive market equilibrium condition, quantity supplied = quantity demanded, then determines not only price, but also the ownership structures of all the enterprises. Thus in industry equilibrium, ownership structure determines supply while simultaneously demand determines ownership structure. When demand is low, the enterprises will be non-integrated. When it is high they
will be integrated. And for a “generic” set of demand curves in between the equilibrium price will be $P^*$, and there will be simultaneously some integrated firms and some non-integrated enterprises. Thus even though every enterprise is fundamentally identical, there is heterogeneity of organization and concomitant heterogeneity in performance.

Heterogeneity of performance among firms has attracted an enormous amount of attention in several literatures, including IO and organizational economics (see e.g., Gibbons (2006), Gibbons (2010), Syverson (2010)), and thus it is worthwhile investigating whether and to what extent ownership structure might account for it. On the qualitative side, the model suggests that endogenous difference in ownership structure may go part way toward explaining heterogeneity. On the quantitative side, the model also suggests that the degree of heterogeneity is clearly sensitive to demand: for low or high enough demands, there is uniformity of ownership structure, but in between most common measures of organizational or productivity variation would vary continuously with demand. Whether this or other organization IO models could account for a substantial proportion of observed performance variation remains to be seen.

The heterogeneity result is a simple consequence of market clearing where there is a discontinuity in individual supply due to the performance differences of the two ownership structures. Indeed, it is the performance difference that is responsible for the (genericity of) equilibrium heterogeneity: without it, the supply would be continuous and there’d be no reason for firms not to have identical ownership structures. Note that if potential HQs have different talents and could affect the benefit of coordination in firms, the degree of heterogeneity in organization will be also a function of the degree of heterogeneity in talent: whether the later magnifies or weakens the former is an interesting
question for future research. Extending the analysis in this direction will be also useful for welfare analysis.

Another explanation for heterogeneity in organizational forms, also based on the inherent discreteness of integration choices, is offered by Gibbons et al. (2010), which views equilibrium heterogeneity in ownership structures as the result of a particular resolution to the “Grossman-Stiglitz problem” (Grossman & Stiglitz 1980). In their (rational expectations) model, market prices aggregate information, but that information is differentially generated by firms depending on their ownership structure. Equilibrium requires that some firms invest in information gathering and some do not, which in turn entails that there is heterogeneity in ownership structure.

A second set of implications concerns the sources of organizational change. Typically, these have been attributed to changes in technology. No doubt that is important, and if we were to change the technology in this model (say $\lambda$), holding price fixed, there might be change in ownership structure. But the model also shows that for organizational change, demand matters, since it affects equilibrium price. Indeed, since all firms face the same price, a shift in demand may cause all firms to reorganize, resembling a merger (or divestiture) wave.

In fact, the importance of market forces in determining organization enriches the channel by which technical change can induce organizational change. To see this, suppose initially that demand is such that the equilibrium price just exceeds $P^*$, as in Figure 2. Now let a fraction $\phi$ of the firms experience an improvement in technology that doubles their output at every decision pair (so in equilibrium their revenue would be either $2P$ or $2\lambda^2 P$ depending on whether the are integrated or not); assume for simplicity this has no effect on the private costs. The figure shows the supply curve for the innovating firms as well as one for the firms with the old technology. The industry supply will be the $\phi$-convex combination of the two. Observe that the new equilibrium price is below $P^*$ but above $P^*/2$. Hence the innovating firms remain integrated — they don’t reorganize despite the change in technology — but the old-technology firms now re-organize by becoming non-integrated. In short, technical change within an reorganization is neither necessary nor sufficient for it to reorganize.

So far we have been discussing mainly how the market influences organization – or what IO can do for organization economics. What about the other way around? What can OE do for IO?

In addition to providing a possible explanation for productivity heterogeneity, as discussed above, consider the last example from the point of view of aggregate performance. Notice that absent reorganization, the new output level would be $2\phi + 1 - \phi$. But because of the dis-integration of the non-innovators, output is in fact only $2\phi + (1 - \phi)\lambda^2$. The difference, $(1 - \phi)(1 - \lambda^2)$, is a measure of lost innovation potential due to “re-organizational dissipation.” Legros & Newman (2009) provides an example in which this effect is so strong that there is no increase in output despite the fact that demand is elastic. This is reminiscent of the ‘missing’ productivity gains that accompanied the “information technology revolution” of the 1980s and early 1990 (Jorgenson
Figure 2: Effect of a Productivity Shock ($\phi = 1/2$)

(2001), Brynjolfsson et al. (2002)). The example suggests that organizational considerations may contribute to a better understanding of technology diffusion (Garicano & Rossi-Hansberg (2010), Legros et al. (2006) for applications to growth.)

The example also illustrates that shocks to some firms may lead to reorganizations of many other firms, effectively generating an “integration wave”. These waves could be also due to ‘defensive’ moves by firms, making integration decisions strategic complements in an imperfectly competitive market (for instance, the ‘waterbed’ effects of section 3.3. But there is also evidence indeed that they arise in more competitive industries, like for instance the cement industry (see Hortaçsu & Syverson (2007) for documentation on the vertical merger wave of the 1960s) or the US generic pharmaceutical industry (Kubo (2011)).

Another area where organizational economics and the property rights approach might contribute is the theory of firm scope, and the role of multi-product firms. The main focus in the industrial organization literature has been on the strategic consequences of the scope of (unitary, profit maximizing) firms, for instance the fact that firms may have multi-market contact, or may be better able to discriminate; on the diversification benefits or on the cost structure of these firms. In the incomplete contracting framework, the focus has been mainly on the consequences for internal organization, e.g., whether the multi-product firm should be organized along product lines or tasks (the M versus U forms); or on the role that indivisibilities play in generating excess capacity,

43 There is an active research agenda in trade on firm heterogeneity at the product level (Bernard & Redding (2006), Neary (2007), Nocke & Yeaple (2000).) This literature considers neoclassical firms and tends too assume that the marginal cost of additional products is exogenously related (often increasing) in the mass of products already offered by the firm. 31
explaining the benefits for firms to include substitute varieties in their portfolio (Penrose (1936)).

The above model can easily be modified to provide an account of whether related product lines would be produced by separate or a single firm. Indeed, firm scope is the impetus for the HH model and we feel that the property rights approach brings new perspectives for the analysis of multi-product firms or the analysis of conglomerate mergers.

Let us then relax the assumption that there is a single product, and assume instead that A and B can both produce a product that sell in competitive markets at prices $P_A$ and $P_B$. For simplicity, assume share contracts are not feasible; rather revenue division would have to be negotiated after it is realized. Follow HH in assuming that whoever owns the assets can control the disposition of any output they produce, and therefore has access to the full revenue stream generated by the owned assets. In this case, Nash bargaining over revenue will always lead to the split $P_i \times \text{realized output for supplier } i$.

Assume that output realizations in the two product markets are independent (in fact, with risk neutrality, it makes little difference if we allow for correlation), and maintain the same parametric relations among $\lambda$, $c$, and $C$. This modifies the non-integration game as follows:

\[
\begin{array}{ccc}
 L & M & R \\
 L & P_B - C, P_A & \lambda P_B - C, \lambda P_A - c & \lambda^2 P_B - C, \lambda^2 P_A - C \\
 M & \lambda P_B - c, \lambda P_A & P_B - c, P_A - c & \lambda P_B - c, \lambda P_A - C \\
 R & \lambda^2 P_B, \lambda^2 P_A & \lambda P_B, \lambda P_A - c & P_B, P_A - C \\
\end{array}
\]

It is straightforward to verify that $(R, L)$ is the unique (pure strategy) equilibrium when both $P_A$ and $P_B$ are lower than $C/(1 - \lambda^2)$, that $(M, M)$ is now an equilibrium when $P_A$ and $P_B$ both exceed $c/(1 - \lambda)$, that $(L, L)$ is an equilibrium when $P_B > C/(1 - \lambda^2)$ and that $(R, R)$ is an equilibrium when $P_A > C/(1 - \lambda^2)$. Hence $(L, L)$ is the unique equilibrium when prices are unequal ($P_A < C/(1 - \lambda^2) < P_B$). Note that the previous version of the model corresponds to a case in which $P_A \equiv 0$, so this generalizes what we did above.

Integration works as before, with the professional manager implementing $(M, M)$, after buying the assets from A and B for fixed payments. If the professional manager has a positive opportunity cost of any size, non-integration dominates integration in the “relatively equal” price region where both $P_A$ and $P_B$ exceed $c/(1 - \lambda)$. And when $P_A$ and $P_B$ are both less than $C/(1 - \lambda^2)$, neither manager wishes to integrate. For all other prices, there will be integration. If the aggregate price $P_A + P_B$ is low (lower than $2C/(1 - \lambda^2)$) or high (greater than $2c/(1 - \lambda)$) integration arises only when product prices are “unequal”.

For other aggregate prices, integration arises independently of the equality of prices. There is therefore a mean-variance effect at play for integration.

---

44 The fact that non-integration and integration perform identically in this region is again an artifact of discreteness, but does not change the conclusion that non-integration will dominate when the the surplus shares are close to each other, see Legros & Newman (2009).

45 This is qualitatively similar to Proposition 1 in Hart & Holmström (2010).
Contrary to the previous example, there is here a non-monotonic relationship between the level of prices and integration. For instance, along a ray $P_A = P_B$, there is first non-integration, then integration and then non-integration again. This illustrates the benefit of endogeneizing the scope of the firm rather than assuming that a firm’s scope is exogenously given and is invariant to changes in the market environment. For instance, the external effects of shocks can now move across sectors and a drop in demand for the $A$ good could result in reductions in firm scope (dis-integration) and concomitant output losses in sector $B$ at high prices but in increases in firm scope and output gain in sector $B$ at low prices.

Note once again that the severity of transaction costs varies with market prices. The typical TCE argument for firm scope is that contracting costs between two activities using a common resource are too high, so they must be integrated. Here, it depends on what part of the price space they are in. Surplus distribution is here realized through market prices and is again crucial in determining the ownership structure.

Finally, if professional managers (or headquarters) are liquidity constrained, some of the predictions could be altered, underscoring again the importance of integrating industrial organization and finance.

5.4 Welfare

Consumer welfare is a central concern of IO, but the bulk of attention has been on the effects of market power. But incomplete contracting and its attendant incentive problems lead to welfare efficiency losses, and it is worth asking how these might affect market performance, quite apart from market power concerns, as well as in conjunction with them. More specifically, with respect to organizational choice, we already pointed out in the previous section that organizational choice may be “inefficient” even from the point of view of the contracting parties, and one wonders whether the typical absence of consumers from the contracting table (i.e., their lack of direct involvement in organizational design) might exacerbate this situation or whether the market can discipline organizational choice enough to protect consumers.

Let us then see what the above model has to say about market performance. For simplicity, consider the one-product version, and take as the welfare measure the standard IO criterion of consumer surplus less social cost, which in this case amounts only to the managers’ private costs. From the consumer point of view, non-integration is undesirable because if produces less output than non-integration, thereby driving up prices. Of course, it does not follow that a planner ought to “force” firms to integrate unless the gains to consumers outweigh the additional costs to managers. Legros & Newman (2009) study this question for two cases, one in which the entire revenue of the enterprise accrues to the managers, and a second one in which the managers receive only a portion of the revenue, with the rest going to dispersed (therefore passive) shareholders.
Whenever non-integration is chosen in equilibrium, its private cost is 0. For integration, the cost is $2c$, which is equal to $P^*(1 - \lambda^2)$ (this follows from the definition of $P^* : P^* - 2c = \lambda^2 P^*$). This is just the light shaded rectangle in figure 3. It should be plain that if equilibrium is a price below $P^*$, there is not a social gain from forcing integration: the marginal social cost exceeds the marginal consumer willingness to pay.

Suppose though that firms are managerial, with managers accruing only $\gamma < 1$ of the revenue. The analysis of $B$’s choice of organization proceeds as in the productivity shock case: non-integration is chosen for $\gamma P < P^*$, integration for $P > P^*/\gamma$. It follows that the supply curve is now an “upward translation” of the original supply. But cost is represented as before, by the area under the supply.

Consider the case where equilibrium $P$ satisfies $P^*/\gamma > P > P^*$. If the planner forced a marginal firm to integrate, the consumer willingness to pay for the extra unit output ($P(1 - \lambda^2)$) now exceeds the social cost $P^*(1 - \lambda^2)$. It follows that for any equilibrium satisfying $P^*/\gamma > P > P^*$, there is too little integration. See figure 3 where the the dark shaded area is the “Leibenstein trapezoid” representing the deadweight loss from inefficient ownership structure.

This raises the question of policy remedies and, if any, who should impose them. It is probably unappealing to think that a government would be directly involved in forcing vertical mergers (though they do sometimes impose moves in the opposite direction, either forcing divestitures or more commonly preventing mergers). It is also worth pointing out that in this model, there can

---

46 If $\gamma$ is given the strict interpretation of a share of revenue, then for the usual rent-extraction reasons familiar from principal-agent theory, they would never pick $\gamma = 1$ unless the managers can purchase the enterprise with lump-sum payments.
never be too much integration, but that may not be true in more general models, so that policy might be asked to work in both directions.

One answer is corporate governance policy. If we interpret $\gamma$ not as a revenue share but as a measure of the degree of alignment that managers have with profit, whether due to monitoring, career concerns or professionalism, then policies that help bring about effective governance, which increase $\gamma$, will also improve consumer welfare. Thus among the things that OE brings to IO is a possible new policy instrument for the protection of consumer interests. While in the case of perfect competition it seems that “good corporate governance” helps consumers, it remains to be seen how this insight persists under imperfect competition.

6 Conclusion

Twenty years ago, Holmström & Tirole (1989) wrote an extensive survey on the theory of the firm, and covered most of the aspects present at this 2011 conference. A sign of the success of the incomplete contract literature is the fact that it is no longer possible to write a survey of the theory of the firm that would encompass all of the dimensions covered in 1989. In particular, corporate finance, internal organization, applications to industrial organization, are receiving specific treatments. In a certain way, this growth has been accompanied by more specificity, mirroring the evolution in the scope of the industrial organization literature that we highlighted in the introduction. The last twenty five years have been a period of disintegration and specialization. What theory and empirics tend to suggest however is that in order to understand the positive and normative consequences of firms boundaries, it may be important to go back to more integration.

References


Syverson, C. (2010), What Determines Productivity?


