

This is an electronic version of an Article
published in *Canadian Journal of
Economics*, 21, August 1988, 444-58

Economic theories of the firm: past, present, and future

PAUL MILGROM and JOHN ROBERTS
Stanford University

Abstract. We explore the current state of the theory of the firm, with attention to its historical origins. Answers to the crucial theoretical question of why an organization controlled by a central authority cannot always duplicate the performance of a decentralized organization are presented. Finally, we describe some important current directions in the theory of the firm and point out promising avenues for continuing research.

Les théories économiques de la firme: passé, présent, avenir. Les auteurs examinent l'état actuel de la théorie de la firme en tenant compte de ses origines. Ils proposent des réponses à la question centrale à savoir pourquoi une organisation contrôlée par une autorité centrale ne peut pas toujours avoir une performance aussi bonne que celle d'une organisation décentralisée. Les auteurs définissent aussi quelques-unes des directions importantes dans les débats en cours sur la théorie de la firme et indiquent les avenues prometteuses pour la recherche en cours.

MARKETS, FIRMS, AND WESTERN ECONOMIC HISTORY

When economists today write about the firm, they most often proceed by comparing its characteristics with those of markets. The reason is not hard to understand: As economic historians have repeatedly argued (Innis, 1938; North and Thomas, 1973; Rosenberg and Birdzell, 1985); the emergence, expansion, and eventual dominance of the market system in western economies since the Middle Ages crucially contributed to economic growth and the resulting accumulation of wealth and rising standards of living for much of the population. More recently, the spectacular successes of the market-oriented economies of the rapidly industrializing nations of Asia have shown that the strengths of the market as a basis for organizing economic activity are not limited to western societies. These considerable achievements of market economies have led some to suggest that other sorts of economic organizations

basically arise only to compensate for failures of the market (Arrow, 1974). Others, much taken with the past successes of market economies, challenge any proposed intervention with the question: Why not rely on a market solution?

There is another, quite opposite, view according to which markets are a primitive way of organizing activity – one that worked well enough in simpler times when agricultural products were traded for finished foreign goods or for the hand-made products of local craftsmen and that still works well enough for distributing consumer goods and for buying and selling standardized items (grains, financial assets), but one that has proved inferior as a way of organizing the transactions arising in complex, multistage production systems. In this view, the market's declining importance is evidenced by the emergence and growing importance of large, integrated firms, internal labour markets, joint research ventures, and the like. For if market-mediated transactions work so well, why don't firms hire more inputs – both products and services – from independent suppliers in the market? Why do they so often distribute and sell their own products? Mine their own raw materials? Provide their own personnel, accounting, computer and other services? The successes of firms that organize so many of their own activities without relying on markets has led some economists (Coase, 1937; Knight, 1971; Williamson, 1975) to pose this question: Why can't a large firm (or a centrally planned economy) always operate at least as efficiently as a chaotic, unorganized market?

In his classic 1921 study, *Risk, Uncertainty and Profit*, Frank Knight (1971) identified a number of factors that might limit the efficient size of firms. Entrepreneurial firms, he reasoned, were constrained by the limited financial resources of the founder, while large partnerships were limited by the 'free rider' problem arising from many partners sharing in a single pool of profits. However, with the rapidly developing financial markets of his day, the only deterrent Knight could identify to the unlimited growth of publicly traded corporations was the problem of properly motivating employees.¹ But even to view this as a problem presupposes that corporations cannot install market-like incentive-compensation packages.

Coase (1937) is commonly credited as the first economist to pose the question of which activities are most efficiently carried out within the firm, and which without. To answer this question, Coase posited that there are different costs in carrying out transactions in the firm and in the market. He identified the costs of negotiating agreements and determining appropriate prices as the principal costs of market-mediated transactions. Both of these since have received extensive attention. Coase had little to say, however, about what costs

1 Knight writes (p. 253): 'With reference to the first of our two points above mentioned, the extension of the scope of operations, the corporation may be said to have solved the organization problem. There appears to be hardly any limit to the magnitude of enterprise which it is possible to organize in this form, so far as mere ability to get the public to buy the securities is concerned. On the second score, however, the effective unification of interests, though the corporation has accomplished much in comparison with other forms of organization, there is still much to be desired.'

might be uniquely and necessarily associated with transactions taking place within the firm. Instead, he posited that any firm could utilize only a fixed amount of management or entrepreneurial talent, so that taking best advantage of the talent of society's entrepreneurs and managers requires an economy with many firms. This assumption obviously begs the question, but little that is more satisfactory has been proposed until recently.

RECENT THEORIES

Chief among the purported virtues of the market system are that it economizes on the information and communication needs of individuals and firms (Hayek, 1945) and that it provides proper incentives for self-interested individuals to economize on resource usage and to innovate with improved products and production techniques. The idea that a price system economizes on information and communication has been the subject of numerous theoretical studies (e.g., Hurwicz, 1977) which show that under certain conditions the announcement by buyers and sellers of quantities supplied and demanded to a hypothetical market auctioneer and by the auctioneer of prices to the buyers and sellers is the minimal amount of communication required to sustain efficient outcomes as an equilibrium of any dynamic process. These conclusions, however, are based on a paradigm in which all goods are potentially useful to all producers and consumers and all relevant information regarding resource availability can be mirrored in prices. These two conditions mean, for example, that knowledge of the wages of barbers in Ankara, Turkey is useful to a Palo Alto semi-conductor manufacturer while knowledge of the willingness of an accomplished chip-design engineer to switch jobs is extraneous or redundant. The team-theoretic models of Marschak and Radner (1972) do begin to represent the problem of communication when specialized local knowledge is necessary for effective decisions and yet some coordination of the dispersed decision-makers is vital. Research in this area is important for understanding the substitution of explicit management activity for market-guided activity: Managers spend the greater part of their time collecting informal information by telephone or in face-to-face contact for use in decision-making (Mintzberg, 1973).

Just as markets can fail as the least-cost way to communicate production information, traditional markets can fail in various ways to provide correct incentives or to provide them efficiently. First, the market traditionally provides incentives to producers by requiring them to bear the full consequences of their decisions. But bearing full responsibility may be impossible if the producer has limited financial resources, or undesirable if the producer is risk averse (Wilson, 1969). These problems are compounded when the results of production depend on the contributions of several individuals – the case of ‘team production’ (Alchian and Demsetz, 1972) – or when informational asymmetries are present, either before or after contracting.

One partial solution is to substitute monitoring of inputs for standard, output-based, incentive schemes, thus moving away from traditional market forms of organization and creating a role for other, more formal, sorts of organizations. Standard sources of market failure such as externalities or increasing returns give still other reasons for abandoning pure market solutions in favour of other organizational forms.

Another role for formal organizations is to resolve disputes or economize on bargaining costs that a market relationship would entail (Williamson, 1975). Bargaining costs presume that there is (or might be) some surplus to bargain over – some reason why the parties have something more to gain by working together rather than with some other partners. This can happen, for example, if one of the parties has made a specialized investment and the parties must later agree how to divide the returns it produces. In simple situations it may be possible to avoid bargaining costs by agreeing in advance how the assets are to be used and the benefits shared. But in complex long-term relationships, the desired contracts may be so detailed that it becomes impractical to specify and enforce them. Then, some form of governance arrangement, or *relational contract*,² in which the parties agree to *procedures* for making production-related decisions, determining how revenues are to be shared, and resolving disputes, may be more efficient than a loosely structured market relationship. Vertical integration is seen from this perspective as one extreme of relational contracting, in which the parties submit to the common authority of a chief executive (Williamson, 1985; Klein, Crawford, and Alchian, 1979).

The foregoing theories explain why there are unavoidable costs of transacting in the marketplace. But are there corresponding costs that are necessarily incurred in other methods of governing these transactions, and, if so, what are they? Our focus will be on the *inevitable* costs of non-market arrangements. If some particular tasks can be done most efficiently through the market (because, for example, prices are efficient transmitters of information), we want to allow the possibility that the organization could continue to use market arrangements for these. Still, all deviations from purely voluntary exchange involve placing some control in the hands of a central authority (e.g., the corporate head office or the state planning bureau), even if the rules prescribe that the central authority may intervene only when such intervention is efficient. The issue is why is a centralized organization using a strategy of selective intervention not able to do at least as well as the market under all circumstances, and better than the market in some circumstances?

The answer is that, for the reasons explored below, the very existence of a central authority inevitably affects how the system operates even in those situations where interventions are not helpful. First, there is the problem of opportunistic behaviour by the central authority. Its decisions might be

² See Macneil (1981) or Goetz and Scott (1981) for an account of the law and economics of relational contracting.

determined not only by considerations of efficiency but also by considerations of personal interest (Williamson, 1985; Grossman and Hart, 1986) or even by bribes and favours. This problem can sometimes be alleviated by subjecting the executive authority to monitoring by an outside judicial authority, but that procedure requires (1) specifying clearly what constitutes proper behaviour by the executive, (2) expending extra resources in monitoring and extra documentation to justify the executive's decisions, and (3) properly motivating the judicial authority. If the executive authority is opportunistic, the costs associated with (1)–(3) are likely to be quite high.

Even if the executive authority is unusually competent, public spirited, and immune to bribes – and all these qualities are commonly known – it may still be desirable to limit its discretion, for two reasons. First, in order to provide correct incentives to others in the organization, the authority must be able to make commitments to act against its own interests in the future, and these commitments are not credible unless there are some effective limits on the centre's powers. For example, the government may wish to commit to a policy of issuing patents to the discoverers of new drugs, even though, once a drug has been discovered, the public would benefit from withholding the patent and allowing unrestricted competition in the provision of the drug. The *policy* of providing patents is socially beneficial, even though the *act* of issuing of an individual patent does not serve the public interest. Similarly, the owner of a firm might want to commit to share the fruits of an innovation with the innovator, even though, once the innovation has been made, the owner's interests are best served by keeping the innovation for himself. A common form of commitment in both of these examples is the establishment of some kind of property right: the drug company has exclusive rights to market its discovery and the innovator has the right to receive royalties on sales of her innovation.³

The second reason to limit the discretion of an honest, competent decision-maker is to discourage rent-seeking behaviour by others who are affected by the centre's decisions. As we have argued elsewhere (Milgrom, 1988; Milgrom and Roberts, 1987a), the mere willingness of the centre to consider seriously a decision with large redistributive consequences will cause other economic agents to waste significant resources in attempts to influence or block it or to delay its implementation. In public decision-making, for example, enormous resources are spent in proposing legislation or regulations and in advocating or opposing these proposals, as well as in filing and manoeuvring for advantage in lawsuits. In bureaucracies – private as well as public – individuals angle for promotions and pay raises, lobby for the adoption of their programs and projects, and advocate rule changes that enhance their power or status. Selling costs in the private sector represent yet another example of costs incurred to affect decisions. As with the previous category of costs of

3 Laffont and Tirole (1985) and Baron and Besanko (1987) have explored the nature of optimal contracts when the authority is unable to commit itself.

centralized decision-making, a common and effective means of limiting the costs in this category is to establish suitable property rights, which either limit the powers of the central authority to make decisions or require full compensation for any property it seizes.

Our account so far has exaggerated somewhat the costs associated with market and centralized forms of resource allocation. Opportunistic behaviour of the kind we have described is at least partially alleviated by norms, codes of conduct, and the like. It is important to understand that these social restrictions are themselves perfectly consistent with self-interested individual behaviour. For example, suppliers may honour agreements and replace defective goods even when they aren't legally required to do so because they value their reputations for honesty, quality, or fairness (Klein and Leffler, 1981). Failure to live up to expectations may cause a seller to lose profitable business in the future. Similarly, employers may treat their employees well in order to make it easier to hire desired employees in the future.

The efficacy of the reputation mechanism depends on how much a firm gains by cheating, how quickly its cheating is likely to be detected, how widely its misbehaviour is known, how much is lost when the detected misbehaviour damages the firm's reputation, and how costly it is to rebuild a lost reputation. In particular, if cheating is detected instantly and becomes widely known throughout the relevant community, and if entry into other, more remote communities is difficult, then increasing the number and frequency of interactions within the community increases the opportunity costs of cheating and makes it possible to support rigorous codes of conduct.⁴ This suggests that a society of long-lived, formal organizations may be especially effective in using reputation mechanisms. Conversely, the reputation mechanism cannot operate effectively in fluid, impersonal, anonymous market settings – such as the Middle Eastern *suq* – where there is no expectation of long-term, repeated dealings and no market-wide reputations. Only in more developed markets, with extensive communications among traders and repeated dealings, do market reputations matter. Thus, 'the market' is seen to be not a single form of organization but a whole category,⁵ and any clear-cut distinction between markets and other organizations quickly blurs.

POSSIBLE FUTURE DIRECTIONS

All the theories reviewed above are incomplete in important ways, and our main prediction – and prescription – is that the development of these theories will – and should – continue. We shall examine some of the likely avenues of development below. A second prediction / prescription is founded on our view

4 This assertion is just one version of the Folk Theorem from the theory of repeated games – see Fudenberg and Maskin (1986) for a more complete account.

5 Douglass North (1987) emphasizes the historical and regional variations among markets in the way performance is assured.

that the incentive-based transaction costs theory has been made to carry too much of the weight of explanation in the theory of organizations. We expect competing and complementary theories to emerge – theories that are founded on economizing on bounded rationality and that pay more attention to changing technology or to evolutionary considerations. We shall have more to say about some of these possibilities, too, below.

Although widespread interest among economists in the study of organizations and, more particularly, in the firm as an actual institution is a relatively recent phenomenon, this is not the case in other disciplines. Organization behaviour, business policy, and parts of sociology have been primarily concerned with these issues for several decades. Moreover, the nature of research in these fields – especially the relatively underdeveloped state of formal theorizing – means that they have generated numerous field studies, case histories, and other empirical work on the organization, design, policies, behaviour, and performance of actual firms, and on the behaviour of individuals in these contexts. This is in striking contrast to the situation in the economics of organization, where the vast bulk of the research has been primarily deductive theorizing and where too often the questions that the latest paper seeks to answer arise not from consideration of puzzling aspects of observed practice or from present trends in business organization but from the desire to extend the analysis in an earlier paper that, in turn, may have been only tenuously connected to observation.

We would be the last to denigrate the value of specialization among researchers; it is quite likely that efficiency requires that economists first focus primarily on theoretical analyses of organizations. Moreover, it is certainly true that the research done in other disciplines, having been aimed at answering questions other than those that occur naturally to economists and, even more, having been informed by very different modes of theorizing than we employ, are not always directly relevant to our work. Still, the best work in these fields can be enormously valuable to economists (see, e.g., Baron, 1987), and it seems abundantly clear that the economics of organization could be enriched by insights and observations imported from these other fields, as well as, of course, by empirical studies by economists.

Finally, the shape of future theory will and should be influenced by the important applied issues of the day. Just as the growth of the modern firm led Knight and Coase to begin theorizing about firms, and the Russian Revolution led to new theories of socialist central planning and analyses of the market as a planning mechanism, such modern phenomena as corporate take-overs and restructuring, the increasing use of subcontractors in manufacturing industries, and the move of various finance and strategy formulation functions out of the firm to be provided by investment bankers and consulting firms, ought to (and probably will) attract the attention of economic theorists.

Here are some of the areas in which we expect progress to be made most rapidly. Topics *A–E* all fall generally under the current transaction costs

paradigm. The remaining topics look in other directions to explain important aspects of firm organization.

A. Incomplete contracts

Governance structures, or relational contracts, are now commonly seen as substitutes for detailed long-term exchange contracts (Simon, 1951; Williamson, 1986). However, relational contracts are but one alternative to detailed long-term exchange contracts. Another is a series of short-term contracts, renegotiated frequently as conditions change. Short-term contracts are often simpler and may involve far fewer contingencies than long-term contracts. Even in the presence of moral hazard, a series of ‘complete’ short-term contracts⁶ can, in a surprisingly wide range of circumstances, perform fully as well as complete long-term contracts (Fudenberg, Holmström, and Milgrom, 1987). This suggests that – when the conditions of the theorem apply⁷ – it is the costs of negotiating short-term agreements, rather than any necessary limitations of the short-term contracts themselves, that are the fundamental cost of market transactions (Milgrom and Roberts, 1987a).

Several caveats should be attached to this conclusion. Most important, even when the provisions of a long-term contract can be effectively reproduced with a sequence of short-term contracts, it is by no means true that the short-term contracts are always simpler. For example, a ‘whole life’ insurance contract in which the insured pays a fixed premium every year and his heirs collect a fixed sum upon his death is equivalent to the combination of a series of ‘one-year renewable term insurance’ contracts in which a different premium is paid each year as the insured ages and a savings plan in which deposits in each year vary with age. Contracts of both these kinds are actually marketed. We could go further, however, and imagine that the premium rates for each year were not guaranteed but instead were determined in the spot market as a function of the insured’s health as well as his age. Then, to duplicate a whole life contract, the short-term contracts would have to specify payments at the end of each year, contingent on the insured’s health, that are just sufficient to offset the increased future premiums. The long-term (‘whole life’) contract requires many fewer contingent payments, fewer and easier calculations by the insurance buyer, and fewer evaluations of the insured’s health: Surely it is simpler in these respects than a series of shorter-term contracts.

Even when costs of complexity are not at issue, short-term contracts cannot perform as well as long-term ones when either of two conditions hold: asymmetric information at recontracting dates prevents the parties from smoothly negotiating an efficient agreement; or monetary incentives are of limited effectiveness (due to either financial constraints or limited product

6 These are one-period contracts that can be made contingent on anything that is common knowledge at the end of the period.

7 Sufficient conditions are that there is common knowledge of preferences and productive opportunities at renegotiation dates and that there is no lower bound on the utility of consumption.

markets), requiring a reliance on payments in kind over time that can be delivered only in a long-term relationship. The first condition suggests that it may be enlightening to study contracting problems when one party makes an unobserved investment that affects his productivity. The second suggests a comparative study of labour markets over time and place, to see how differences in the wealth of workers and access to capital and product markets affect the nature of labour contracts. For example, to what extent was the efficient movement of workers from the feudal manor to the town delayed by the difficulty experienced by the lord in collecting a share of the resultant gain once the move occurred and by the serfs' inability to borrow against future earnings to compensate the lord in advance? In this regard, how do such factors relate to the survival of serfdom in czarist Russia long after it had disappeared in western Europe?

B. Bargaining theory

There have been several recent attempts to explain inefficiencies in bargaining – delays or total failures to agree – as resulting from asymmetric information.⁸ Most of the effort to date has focused on developing appropriate equilibrium concepts and proving existence; comparatively little of this work has proved informative in understanding the relative efficiency of bargaining in differing contexts. We still do not fully understand, for example, what determines the magnitude of efficiency losses in these models. Such results should be sought.

Given the importance of bargaining inefficiencies for transaction cost theories, it should also be useful to widen our perspective about how bargaining inefficiencies may occur. For example, suppose that two parties are bargaining over an item that is certainly more valuable to the potential buyer than to the seller. The value of the item to each party is an increasing function of some unobserved characteristic, which we call 'quality.' Information about quality can be acquired privately by either party, but only at a cost. In that case, a bargain struck quickly is no evidence of efficiency, since both parties may have wastefully acquired information to protect themselves. If quality uncertainty is great and information is moderately expensive (so that the gains from trade are less than the cost of having both parties become informed, but either party could gain by becoming unilaterally informed), then the parties may be unwilling to enter negotiations without information and yet be unable to benefit if both acquire information. Thus, they may fail to reach any agreement on an appropriate price, despite the certain gains from trade and the perfect symmetry of information at equilibrium. If the parties can anticipate that circumstances like these will arise, both could benefit by agreeing to be bound to allow the price to be set by an impartial arbitrator, that is, to substitute an 'organization' for a market.

⁸ See Sutton (1986) and Rubinstein (1987) for surveys.

C. Reputations

In long-term relationships or when a firm has many customers, reputations function to alleviate opportunism in the way that we described earlier. According to the received theory, a firm may sometimes act against its own short-term or one-customer interests in order to preserve its reputation for the longer term or among other customers. How effective is this reputation mechanism compared, say, with contractual solutions like warranties?

To illustrate the problem, imagine Customer 2 is supposed (at equilibrium) to punish a supplier by refusing to purchase from it if Customer 1 is treated unfairly. The first issue is what constitutes unfair treatment? Can Customer 2 identify it when it happens? If there is some uncertainty about that point, and the supplier does treat Customer 1 unfairly, might not Customer 2 benefit by pretending not to recognize the episode? What incentive does Customer 1 have to bring its unfair treatment to Customer 2's attention, if doing so is costly? For reputations involving punishments by third parties to operate effectively, it seems that the reputation must be about some simple, easily recognized policy, like one of refunding the purchase price to any dissatisfied customer or repairing or replacing a product free for a year from the date of purchase. But these sorts of policies are also easily written into legally binding contracts. Contracts that specify damages have the advantage, compared with reputation mechanisms, that they make it worthwhile for the damaged party to inform the third party about the violation and so enhance the information flow in the system.

Even when third-party punishments are not an issue, simplicity and ease of recognizability can still be important for the efficacy of reputations. A reputation may be most quickly and effectively built by extreme actions: one U.S. clothing retailer has built a reputation for superb service by graciously accepting 'returns' on products that could not possibly have been purchased there, including automobile tires! The role of simple stories in teaching members of an organization the sort of behaviour expected of them and in inculcating the corporation's culture has been documented in the organization behaviour literature (Martin, 1982). Generally, the circumstances in which reputation mechanisms might work well and the means by which reputations are built, used, and lost are ripe for further study.

D. Influence activities and rent-seeking

The term 'rent-seeking' was coined by Krueger (1974) to refer to the activities of private firms and individuals seeking to capture rents created by government interventions in the economy. Rent-seeking theories suggest that government involvement in the economy causes inefficiencies, and these theories have been adapted, as indicated above (Milgrom and Roberts, 1987), to explain the diseconomies that accompany centralization of authority in firms.

This line of analysis seems likely to branch in two directions. The first is to develop the basic theory of influence in organizations more fully. As a

particular example, if limiting influence activities (rent-seeking) requires that rules be established limiting the discretion of executives and managers, how are the rules to be made? Can't rent-seeking occur in the rule-making process, just as it does in legislative decisions in government? Having decisions made behind a veil of ignorance may help, at least in some matters, as may limiting opportunities for making transfers to those formulating the rules. Influence activities will in this case be practised only to the level justified by the returns to the current actors, rather than the higher level that would be generated if the benefits accruing to succeeding generations could be captured by those currently active.⁹ Still, it seems likely that an optimal arrangement will entail separating operational from rule-making decisions, and employing a more cumbersome, difficult-to-manipulate process for the latter. Rules that are too easily changed are hardly rules at all.¹⁰

The second direction is to expand the new private sector theories of rent-seeking to include rent-seeking in markets, courtrooms, and boardrooms. Are the huge salaries of arbitrageurs and securities lawyers simply returns to inefficient private sector rent-seeking? If so, then the empirical techniques of Posner (1975) suggest that all such returns are to be regarded as being a pure social loss, since competition among businessmen to earn these rents leads to their full dissipation. What of the contingency fees of lawyers, the earnings of corporate chief executives, and the greenmail paid to corporate raiders?

E. Ownership: residual returns or residual rights?

In Roman law, ownership of property meant the right to use or abuse it. Knight (1971) identifies ownership of a firm as the right to control it or to pick those who will manage it. The actual concept of ownership, at least as it applies to firms, has evolved considerably, however, from these simple, clean formulations. The connection between ownership and control is quite subtle in North American corporations; with U.S. mutual savings banks, mutual insurance companies, and most Japanese corporations, there seems to be little connection at all. Meanwhile, what relevance does the concept of ownership have to private, not-for-profit organizations, co-ops, and crown corporations? Refining our understanding of what ownership means, or whether its meaning is situation-dependent, seems essential not only for understanding a variety of phenomena from corporate take-overs in the United States to the incentives of managers in Japan, but also for formulating sensible policies, both public and private, regarding decision-making on take-overs, investments and disinvestments, compensation forms and levels, and employment practices.

9 Of course, if the interests of future generations are not represented in decisions that affect them, there is the familiar potential for an inefficiency arising from externalities.

10 See Fama and Jensen (1983) for a related view of the separation of powers and responsibilities within organizations.

Grossman and Hart (1986) have championed 'residual rights of control' as the appropriate concept of ownership of assets. These rights give the owner the power to make decisions regarding the use or disposition of an asset in all respects that are not explicitly designated or limited by a written contract. This definition is problematic in its application; for the typical large firm with many physical assets (as well as various intangible ones), it is never the case that a single party is the owner in the sense of exercising all the residual rights of control. Even if, by setting policies and monitoring adherence to them, the owners of the firm might try to control the decisions made by others, it will never be possible in a large firm to exercise total control.

A simple alternative view holds that the owner is the one who collects residual returns after all other factors have been paid. In a long-lived firm, where all the product from a single period's activities does not accrue in that period and may never be accurately measured, the way these returns are allocated may significantly affect incentives – and hence the efficiency of the organization. The full integration of residual rights with residual returns is a key problem in the newly emerging theory of ownership rights.

F. Adapting to uncertainty

Galbraith (1977) has formulated the problem of organization design as one of allowing the organization to adapt to the uncertainty in its environment. The organization can adapt either by processing more information or by reducing the need to process information. Within the second class of adaptations are environmental management (produce a different product, serve a different market), creation of slack resources (inventories, production capacity), and creation of self-contained tasks (reduced specialization). Within the first are vertical and lateral communication systems.

What determines the optimal mix of these adaptations? Galbraith's framework offers the possibility of price theoretic explanations of changes in organization form. Some initial steps in this direction are taken in Milgrom and Roberts (1987b). For example, cheaper communications, reduced costs of expanding the product line, and fall costs of flexibility in production costs and increasing levels of demand levels favour substituting vertical communications for holding inventory. Similarly, increases in the complexity of the product and the frequency of product change can be met by using more highly trained employees performing less specialized tasks.

G. Planning and budgeting

Another subject area where continued research promises new understandings is that of iterative planning. What *kinds* of information do price systems economize on in the planning process? What sorts of non-market communications are helpful in planning and controlling production? What are the

incentives for correct revelation of information in iterative planning (Roberts, 1987)? How does one properly account for the bounded rationality of the people who must make the planning process work?

There is some literature touching on aspects of these issues in the context of national economic planning but little focusing on the planning within or between autonomous firms. Yet the planning and budgeting processes seem to absorb large amounts of time and effort within formal organizations and to be important aspects of the decision-making, control, and evaluation systems of these organizations. Achieving an understanding of the nature and role of these processes and of their characteristics would seem important to understanding resource allocation in the firm.

CHANGING CONCEPTS OF THE FIRM

In the Arrow-Debreu theory of the private ownership economy, the distinction between firm and market is absolutely clear: a firm is a production set summarizing the possibilities for transforming one bundle of time-, event-, and location-differentiated commodities into another; a market is the coming together of economic agents (firms and consumers) to exchange ownership of such commodity bundles. However, as our analysis of the firm deepens, the firm-market distinction blurs; for production itself involves exchange. The boundaries of firms are fuzzy: two legally separate firms may be more closely integrated in their planning and operations than are any pair of divisions in a conglomerate; and even though there is no commonality of ownership or explicit long-term contract linking them, they may continue their close relations over indefinitely extended periods. Moreover, decentralized firms may adopt market-like solutions to their organizational problems, using, for example, arms-length negotiations to determine transfer prices and evaluating employee and divisional performance on profit criteria.

The extreme response to such observations is to deny any difference between market contracts and those made between members of a firm (Alchian and Demsetz, 1972¹¹). More reasonably, there is a multidimensional spectrum of institutional arrangements with simple, discrete markets and tightly managed hierarchies at two of the extremes. Understanding organization in its many economic dimensions – including ownership, communication, planning, and incentives – will occupy the attention of economists for many years to come.

11 They wrote: 'I can "punish" you only by withholding future business or by seeking redress in the courts for any failure to honor our exchange agreement. That is exactly all an employer can do. He can fire or sue, just as I can fire my grocer by stopping purchases from him or sue him for delivering faulty products.' This view overlooks substantial differences, perhaps most importantly that an employee typically works exclusively for his employer, which helps alleviate the problems of common agency.

REFERENCES

- Alchian, Armen and Harold Demsetz (1972) 'Production, information costs, and economic organization.' *American Economic Review* 62, 777-95
- Arrow, Kenneth (1974) *The Limits of Organization* (New York: Norton)
- Baron, James N. (1987) 'The employment relation as a social relation.' *Journal of Japanese and International Economics* (forthcoming)
- Baron, David and David Besanko (1987) 'Commitment and fairness in a dynamic regulatory relationship.' *Review of Economic Studies* (forthcoming)
- Chandler, Alfred (1977) *The Visible Hand* (Cambridge, MA: Harvard University Press)
- Coase, Ronald (1937) 'The nature of the firm.' *Economica* 4, 386-405
- Fama, Eugene and Michael Jensen (1983) 'The separation of ownership and control.' *Journal of Law and Economics* 26, 301-25
- Fudenberg, Drew, Bengt Holmström and Paul Milgrom (1987) 'Short-term contracts and long-term agency relationships.' Unpublished paper
- Fudenberg, Drew and Eric Maskin (1986) 'The folk theorem in repeated games with discounting and with incomplete information.' *Econometrica* 54, 533-54
- Galbraith, Jay R. (1977) *Organization Design* (Reading, PA: Addison-Wesley)
- Goetz, Charles and Robert Scott (1981) 'Principles of relational contracts.' *Virginia Law Review* 67, 1089-150
- Grossman, Sanford and Oliver Hart (1986) 'The costs and benefits of ownership: a theory of vertical and lateral integration.' *Journal of Political Economy* 94, 691-719
- Hayek, F.A. (1945) 'The use of knowledge in society.' *American Economic Review* 35, 519-30
- Holmström, Bengt (1982) 'Moral hazard in teams.' *Bell Journal of Economics* 13, 324-40
- Hurwicz, Leonid (1977) 'On the dimensional requirements of informationally decentralized pareto-satisfactory processes.' In K.J. Arrow and L. Hurwicz, *Studies in Resource Allocation Processes* (Cambridge: Cambridge University Press)
- Innis, Harold A. (1938) 'The penetrative powers of the price system.' *Canadian Journal of Economics and Political Science* 4, 299-319
- Klein, Benjamin, Robert Crawford, and Armen Alchian (1978) 'Vertical integration, appropriable rents, and the competitive contracting process.' *Journal of Law and Economics* 21, 297-326
- Klein, Benjamin and Keith Leffler (1981) 'The role of market forces in assuring contractual performance.' *Journal of Political Economy* 89, 615-41
- Knight, Frank H. (1971) *Risk, Uncertainty and Profit* (Chicago: University of Chicago Press)
- Laffont, Jean-Jacques and Jean Tirole (1985) 'The dynamics of incentive contracts.' MIT Economics Working Paper No. 403
- Macneil, Ian (1981) 'Economic analysis of contractual relations: its shortfalls and the need for a "rich" classificatory apparatus.' *Northwestern University Law Review* 75, 1018-63
- Marschak, Jacob and Roy Radner (1972) *Economic Theory of Teams* (New Haven: Yale University Press)
- Martin, Joanne (1982) 'Stories and scripts in organizational settings.' In A.H.

- Hastorf and A.M. Isen, eds, *Cognitive Social Psychology* (New York: Elsevier/North Holland)
- Milgrom, Paul (1988) 'Employment contracts, influence activities and efficient organizational design.' *Journal of Political Economy* 96, 42–60
- Milgrom, Paul and John Roberts (1987a) 'Bargaining and influence costs and the organization of economic activity.' Research Paper, Graduate School of Business, Stanford University
- (1987b) 'Communications and inventories as substitutes in organizing production.' *Scandinavian Journal of Economics* (forthcoming)
- (1987c) 'An economic approach to influence activities in organizations.' *American Journal of Sociology* (forthcoming)
- Mintzberg, Henry (1973) *The Nature of Managerial Work* (New York: Harper & Row)
- North, Douglass C. (1987) 'Institutions, transaction costs and economic growth.' *Economic Inquiry* 25, 419–28
- North, Douglass C. and Robert Paul Thomas (1973) *The Rise of the Western World: A New Economic History* (Cambridge: Cambridge University Press)
- Posner, Richard (1975) 'The social costs of monopoly and regulation.' *Journal of Political Economy* 83, 807–27
- Roberts, John (1987) 'Incentives in iterative planning under incomplete information.' In T. Groves, R. Radner and S. Reiter, eds, *Information, Incentives and Economic Mechanisms* (Minneapolis: University of Minnesota Press)
- Rosenberg, Nathan and L.E. Birdzell, Jr. (1986) *How the West Grew Rich* (New York: Basic Books)
- Rubinstein, Ariel (1987) 'A sequential strategic theory of bargaining.' In Truman Bewley, ed., *Advances in Economic Theory Fifth World Congress*
- Simon, Herbert (1951) 'A formal theory of the employment relation.' *Econometrica* 19, 293–305
- Sutton, John (1986) 'Non-cooperative bargaining theory: an introduction.' *Review of Economic Studies* 53, 709–24
- Williamson, Oliver (1975) *Markets and Hierarchies Analysis and Antitrust Implications* (New York: Free Press)
- (1985) *The Economic Institutions of Capitalism* (New York: Free Press)
- Wilson, Robert (1969) 'The structure of incentives for decentralization.' In *La Decision* (Paris: Centre Nationale de la Recherche Scientifique)