

ECONOMIC RESEARCH REPORTS

WHY ARE THERE FIRMS?

by

Richard N. Langlois

D.P. 81-30

March, 1981

**C. V. STARR CENTER
FOR APPLIED ECONOMICS**



**NEW YORK UNIVERSITY
FACULTY OF ARTS AND SCIENCE
DEPARTMENT OF ECONOMICS
WASHINGTON SQUARE
NEW YORK, N.Y. 10003**

Individual copies of papers may be obtained free of charge by writing to:

*The C. V. Starr Center for Applied Economics
New York University
269 Mercer Street, 3rd Floor
New York, NY 10003-6687*

<http://www.econ.nyu.edu/working/>

WHY ARE THERE FIRMS?

Richard N. Langlois

March, 1981

Department of Economics

New York University

SUMMARY

The "firm" described in neoclassical economic theory is something very different from the business organization we observe in day-to-day life. For this and related reasons, modern economics has had a difficult time explaining why such real-life organizations exist at all.

Nevertheless, economists have proposed a number of explanations. In examining the more prominent of these, I distinguish two classes: "hard" or strict-neoclassical explanations and "soft" or extra-neoclassical explanations. My thesis is that (A) the "soft" explanations are in the end more significant; that (B) all the economic explanations surveyed reduce to a single explanation; and that (C) this explanation involves ignorance about and divergent perceptions of the categories of economic action themselves (as distinguished from simple uncertainty about events within known and agreed-upon categories). This interpretation has been obscured, I argue, by the epistemological assumptions of economic theory and by certain confusions over the nature of probability.

I close by considering the logic of explanation appropriate to answering a question like "why are there firms?" And I discuss an alternate mode of economic explanation with implications for the economics of organization and the theory of the firm.

I.

"The problem of economics," wrote William Stanley Jevons in 1871, "may, as it seems to me, be stated thus:-- Given, a certain population, with various needs and powers of production, in possession of certain lands and other sources of material: required, the mode of employing their labour which will maximise the utility of the produce."¹

Described this way, economics -- or, rather, "neoclassical" economics, for which Jevons is always accorded some paternity -- seems a very tightly circumscribed occupation, excluding, apparently, the study of such questions as how "needs" and "powers of production" come into being and how they change. Economists following after Jevons have broadened the problem somewhat, especially in recent years; but the basic framework of neoclassical economics has remained much the same: an optimization of the choice among known means to achieve given ends.

The approach to economic theory this represents is what some philosophers of science refer to as "situational determinism."² As in Jevons's problem, one considers a

¹W. S. Jevons, The Theory of Political Economy, London: MacMillan and Co., Ltd., Fourth Edition, 1911, p. 267.

²S.J. Latsis, "Situational Determinism in Economics," The British Journal for the Philosophy of Science, vol. 23, 1972, p. 207.

hypothetical situation faced by equally hypothetical economic agents who know most or all of the facts of the situation. In conjunction with the assumption of maximization, the "givens" of the situation determine the behavior of the economic agents, which in turn determines the behavior of relevant economic variables like price and quantity. Thus, the process of taking facts as given plays a crucial role in economic modeling.

Looking at economic theory in this light suggests what may be a clue to an age-old economic mystery. The mystery is this: why are there firms? Why is production in the economy so often undertaken within business organizations?

It is frequently, and in most cases unabashedly, acknowledged that the basic neoclassical formulation does not by itself provide an explanation for the existence of firms. Indeed, what is called the "theory of the firm" in neoclassical economics is not at all a theory of those business organizations that actually manifest themselves in our day-to-day life. The theory of the firm is part of neoclassical value theory; its "firms" are abstract entities making price and quantity decisions for given products as part of the overall process of allocating known and given scarce resources among known and given competing ends. This theory is not intended to explain the nature and behavior of real firms; and to equate its "firms" with those of our sensory experience, one often reads, is not only to entertain confusion but also to do injustice to an otherwise valuable theory.³

³The name most closely associated with this position is that of Fritz Machlup. See his "Theories of the Firm: Marginalist, Behavioral, Managerial," American Economic Review, v. LVII, no. 1, March 1967. As a matter of fact, though, the injunction against confusing the firm of theory with the firm of life has been honored, by economist and

In its fullest development -- general equilibrium theory -- the neoclassical theory of value sets forth distinct and uniform commodities fully known to all market participants. Each market agent consumes some of the commodities and produces others; though the distinction is unnecessary, one can think of "households" as those agents who consume goods and produce labor and "firms" as those agents who consume some goods (including the aforementioned labor) and produce other goods. All production and consumption decisions are optimization responses to the prices that manifest themselves in the market, and the final amounts produced and consumed are determined in the equilibration of that market. What goes on inside a "firm" (or a "household," for that matter) is of no concern. The transformation of commodities may be effected by a lone craftsman or by a complicated enterprise employing a large number of people and machines. The firm is a proverbial black-box, and the general equilibrium market governs only the inputs and outputs of the boxes.

Thus economists speak of "market transactions" as those which involve the inputs and outputs of the firms and which take place through the mediation of the price mechanism. The activities within the firm, by contrast, are "internal" or "non-market" transactions governed by administrative mechanisms.⁴

non-economist alike, largely in the breach. There is seldom an inclination to invoke the doctrine when questions of public policy arise; and, indeed, the neoclassical theory of the firm is frequently presented in textbooks on "industrial organization" whose other main component is an application of this theory to antitrust policy -- on precisely the assumption that its theoretical statements about "firms" are replaceable by non-theoretical statements about IBM, General Motors, and AT&T.

⁴A precise definition of the firm is a tricky and not-always-helpful undertaking. One of the most frequently given definitions, which I will be using implicitly, is any employment contract as distinguished from a market

But the interesting feature of the general-equilibrium formulation is not so much that it takes as given the mix of market and internal transactions; rather, it is that the assumptions of general-equilibrium theory themselves actually suggest that there need be no internal activity whatever. If all commodities are predetermined for all time and the techniques for producing them are given and fully known in all details, then one could easily conceive of a situation in which every separate part of the production process would be in the nature of a market transaction.

To understand what this means, consider, as a sort of gedanken-firm, a company that fabricates rifles. The way such manufacture is normally done, one suspects, is that the various parts of the rifle -- lock, stock, and barrel -- are manufactured separately in different parts of a single factory or in different factories owned by the same company; the parts are then brought together and assembled, perhaps with intervening stages of sub-assembly. This is an activity with potentially a good deal of division of labor. Each worker undertakes only a relatively small part of the gun-making process, performing his particular task in exchange for specified wages.

contract. An employment contract is one in which the contractor receives payment not for a specified good or service but for an agreement to perform various as-yet-unspecified services from within an agreed-upon set of possible tasks. Thus, I am talking about any kind of internal organization, including an unincorporated single-proprietorship, partnership, or joint-stock company; and I am not attempting to explain the existence of the the limited-liability corporation, which is a separate, though interesting question. On that see, R.B. Ekelund, Jr., and R.D. Tollison, "Mercantilist Origins of the Corporation," Bell Journal of Economics, vol. 11, no. 2, Autumn, 1980, p. 715.

Consider, now, an alternative arrangement in which, without any change in the extent of the division of labor, the assembly of the rifles might be carried out by market transactions. In this case, each worker, rather than being an employee, is, in effect, a firm unto himself; he is a separate contractor. One person might turn barrels and sell them to various other fellows who are in the business of barrel-boring. Each of these would in turn sell his product to assemblers, who would put the barrels together with other similarly acquired parts and sub-assemblies to produce a finished rifle.

In the first case, such details as the number of barrels to turn per day and who to give them to once turned were determined for the barrel-turner by some form of administrative procedure. In the second instance, the number of barrels and their subsequent disposition were determined by the process of market equilibration using prices.

Evidently, something approximating a fully price-decentralized method of assembling rifles actually did prevail for a while in England during the nineteenth century. George Stigler relates the following portrait of rifle manufacture in Birmingham, England, during the 1860's.

Of the 5800 people engaged in this manufacture within the borough's boundaries in 1861 the majority worked within a small district round St. Mary's Church. . . . The reason for the high degree of localization is not difficult to discover. The manufacture of guns, as of jewellery, was carried on by a large number of makers who specialized on particular processes, and this method of organization involved the frequent transport of parts from one workshop to another.

The master gun-maker -- the entrepreneur -- seldom possessed a factory or workshop. . . . Usually he owned merely a warehouse in the gun quarter, and his function was to acquire semi-finished parts and to give these out to specialized craftsmen, who undertook the assembly and finishing of the gun. He purchased material

from the barrel-makers, lock-makers, sight-stampers, trigger-makers, ramrod-forgers, gun-furniture makers, and, if he were engaged in the military branch, from bayonet-forgers. All of these were independent manufacturers executing the orders of several master gun-makers. . . . Once the parts had been purchased from the "material-makers," as they were called, the next task was to hand them out to a long succession of "setters-up," each of whom performed a specific operation in connection with the assembly and finishing of the gun. To name only a few, there were those who prepared the front sight and lump end of the barrels; the jiggers, who attended to the breech end; the stockers, who let in the barrel and lock and shaped the stock; the barrel-strippers, who prepared the gun for rifling and proof; the hardeners, polishers, borers and riflers, engravers, browners, and finally the lock-freers, who adjusted the working parts.⁵

To our modern ears, this sounds like an unusual way to make rifles. But there is nothing in the assumptions of general equilibrium theory to tell us why production should not be carried out in this manner.

This is not meant as a criticism of general equilibrium theory. Rather, it is meant to suggest a source of clues. For, if we can isolate differences between the "real world" and the world of general equilibrium theory, then we have potential reasons for the existence of firms. And putting the magnifying glass to the clues that emerge may in turn reveal some interesting features of the neoclassical logic of explanation.

⁵G. C. Allen, The Industrial Development of Birmingham and the Black Country, 1860-1927 (London, 1929), pp. 56-57 and 116-117, quoted in G. J. Stigler, "The Division of Labor is Limited by the Extent of the Market," Journal of Political Economy, v. LIX, no. 3, June 1951, reprinted as Chapter 12 in The Organization of Industry, Homewood Illinois: Richard D. Irwin, 1968, p. 129.

II.

One prime locus of explanation has long been the supposition that the world appears, from the point of view of the businessman, to be on the whole a good deal more uncertain than the economic models have made it out to be. Firms arise because they are a way of dealing with this uncertainty that is somehow superior to more complete decentralization. One of the earliest, and still one of the best, systematic analyses along this line was that of Frank Knight in Risk, Uncertainty, and Profit.⁶

Knight begins with the notion of uncertainty itself, of which he believes there are two kinds. The first kind, which he identifies with the idea of "objective" probability, is "measurable" uncertainty or, more simply, "risk." When this type of uncertainty is present, he maintains, the relevant probabilities of the uncertain events can be calculated objectively, either from a priori principles (as in the case of a die, whose probability of revealing any one of its numbered sides can be determined from the geometry of the die itself) or by empirical methods (as when the probability of, say, a house fire can be determined from the observed frequency of actual fires in the relevant population of houses). The second kind of uncertainty, which is to be associated with "subjective" probability, Knight calls "true" uncertainty. Here the probabilities "cannot be calculated" because there are neither relevant a priori rules nor adequate empirical bases for such a calculation. Each situation is unique; the uncertain event cannot be grouped in an actuarial way as an instance of a class of events with an underlying distribution.

⁶Chicago: University of Chicago Press, 1971 (Originally published in 1921).

Most business decisions, Knight feels, are of this last type. Each new plant expansion under consideration is sufficiently different from all previous expansions that no objective inference from past experiences can be drawn about the likely success of the project. Subjective judgment must intrude. If all uncertainty were merely "risk," Knight contends, the world could be rendered effectively certain by the mechanism of consolidation or objective grouping, the principle that makes conventional insurance possible. But if the uncertain event is idiosyncratic, such grouping is no longer possible. And this, to Knight, helps explain why firms might arise. There is, he argues, a "diversity among men in degree of confidence in their judgment and powers and in disposition to act on their opinions, to 'venture.' This fact is responsible for the most fundamental change of all in the form of organization, the system under which the confident and venturesome 'assume the risk' or 'insure' the doubtful and timid by guaranteeing to the latter a specified income in return for an assignment of the actual results."⁷

So, for Knight, it is the inherent subjectivity of business decision-making that calls forth internal organization. And, although the advent of the modern Bayesian paradigm in decision theory renders Knight's analysis of uncertainty and probability somewhat obsolete, it actually amplifies his stress on the subjective. The distinction between between "objective" and "subjective" probability in the Bayesian paradigm is very simple: there is no objective probability. All probability is subjective, "a particular individual's quantitative description of

⁷Op. Cit., p. 269.

⁸Ronald A. Howard, "An Assessment of Decision Analysis," Operations Research, v. 28, no. 1, January-February 1980,

uncertainty."⁸ Nowadays there is only one kind of uncertainty; and "risk" is taken to mean a situation in which one has something at stake, something riding on events that are uncertain. This approach permits all types of uncertain situations, from one-of-a-kind plant expansions to the colored-balls-in-the-urn lotteries of probabilistic pedagogy, to be handled consistently with the same set of conceptual and mathematical tools.

To the extent that Knight can be interpreted as distinguishing solely between what would now be called "insurable risk" and "uninsurable risk," his analysis is largely unaffected by the Bayesian paradigm-shift. But, if interpreted this way, it is also probably an inadequate explanation for the existence of firms.⁹

The subjective nature of probability does not by itself imply the need for firm-like organization of production. The "doubtful and timid" could have the burden of uncertainty lifted from their shoulders in a fully price-decentralized way through commodity futures markets. Such markets, which do in fact exist for various agricultural and other commodities, operate exactly on the principle of specialization-in-risk-bearing that Knight speaks of; those with a greater propensity for conation (as Knight might put it) speculate in the future price of the commodity, permitting the risk-averse producers and consumers of the commodity to enjoy relatively more stable prices. And such

p. 4.

⁹As a matter of fact, though, I think Knight was largely on the right track. As I will suggest below, there is a more important aspect to subjectivity that Knight seems at times to recognize but which has been obscured both by his own discussion of uncertainty and by its subsequent reinterpretation by others.

markets operate not because there is the possibility of "objective" probabilities but precisely because price expectations diverge among the speculators. Relating this to our gedanken firm, one could thus think of futures markets in locks, stocks, and gunbarrels, both bored and unbored; and our diffident gun-specialists needn't cleave to an employer for shelter from uncertainty.

Thus, as Ronald Coase suggests, the guaranteeing function of the employment relation is not a sufficient explanation for the existence of firms. Coase finds a different source of explanation. "The main reason why it is profitable to establish a firm," he argues, "would seem to be that there is a cost to using the price mechanism."¹⁰ One cost might be that attached to discovering all the relevant prices in the market; another might be the cost of making a separate contract for each transaction.¹¹

This leads Coase to consider the firm in equilibrium. Defining the size of a firm in terms of the number of transactions its organization has internalized, he suggests that a firm "becomes larger as additional transactions (which could be exchange transactions co-ordinated through the price mechanism) are organised by the entrepreneur and becomes smaller as he abandons the organisation of such

¹⁰"The Nature of the Firm," Economica, N.S. vol. 4, Nov., 1937, p. 390.

¹¹Another advantage of internal organization that Coase suggests is the existence of taxes that make internal transactions cheaper. This is subtle point in that such an advantage depends on the type of tax in question. But the point is less subtle when generalized. Government regulations of various sorts can create organizational economies of scale (a big firms has to fill out the same forms as does a small firm) that confer advantages on large size. The importance of this effect is subject to dispute.

transactions."¹² By straightforward application of standard economic reasoning, he concludes that "a firm will tend to expand until the costs of organising an extra transaction within the firm becomes equal to the costs of carrying out the same transaction by means of exchange on the open market or the costs of organising in another firm."¹³

But despite his focus on transaction costs, some notion of uncertainty still seems implicitly to be at the base of Coase's analysis. "It seems improbable that a firm would emerge without the existence of uncertainty," he admits.¹⁴ Long-term contracts among the many market transactors might be a way around some of the costs of search and renegotiation; but, because of uncertainty (in some sense of the term), there are also difficulties with long-term contracts.

Now, owing to the difficulty of forecasting, the longer the period of the contract is for the supply of the commodity or service, the less possible, and indeed the less desirable, it is for the person purchasing to specify what the other contracting party is expected to do. It may well be a matter of indifference to the person supplying the service or commodity which of several courses of action is taken, but not to the purchaser of that service or commodity. But the purchaser will not know which of these several courses he will want the supplier to take. Therefore, the service which is being provided is expressed in general terms, the exact details being left until a later date. All that is stated in the contract is the limits to what the persons supplying the commodity or service is expected to do. The details of what the supplier is expected to do is not stated in the contract but is decided later by the purchaser.¹⁵

Thus the firm does not so much replace contracting as it is itself an instance of contracting. What distinguishes a firm

¹²Op. Cit., p. 393.

¹³Op. Cit., p. 395.

¹⁴Op. Cit., p. 392.

¹⁵Op. Cit., p. 391.

from a market is the extent to which the contracts involved are open-ended. And what characterizes the employment relation, in Coase's view, is not the guaranteeing function it performs for employees but the flexibility it affords in a world of uncertainty.

In the end, though, the differences in explanation between Coase and Knight turn out to be in part illusory. Knight's catch-all conception of uncertainty, when turned to the perspective not of the employee but of the employer, produces an analysis of the employment relation identical to that of Coase.

Even when it is impossible to reduce the work itself to routine sufficiently for a machine to handle it -- due usually to lack of uniformity (i.e., uncertainty) in the material worked with -- it is possible to judge with a high degree of accuracy the capacity of a human individual to deal with the sort of irregularities to be met with in the occupation. It is the function of the operative in industry to deal with uncertainty as a matter of routine! The exact movements he shall have to perform cannot be foretold, but his ability to perform them can be, and so the uncertainty is eliminated as an element in the calculations; ignorance of the environmental situation gives place to knowledge of human judgment.¹⁶

Flexibility, far more than transactions costs or uninsurability of risk, is the underlying message in both analyses.

Alchian and Demsetz have more recently offered an explanation of the firm with a somewhat different emphasis. They work from a Coasean framework, and consider the firm as reflecting an optimal choice in equilibrium between the two competing contractual modes. What is it, they ask, that distinguishes an employment contract from a standard contract of purchase in a market? "It is a team use of

¹⁶Knight, p. 295.

inputs," they answer, "and a centralized position of some party in the contractual arrangements of all other inputs. It is a centralized contractual agent in a team productive process -- not some superior authoritarian directive or disciplinary power."¹⁷

At first, the focus of the Alchian and Demsetz analysis is on the team production aspect. "In team production," they explain, "marginal products of cooperative team members are not ... directly and separably (i.e., cheaply) observable. What a team offers to the market can be taken as the marginal product of the team but not of the team members. The cost of metering or ascertaining the marginal products of the team's members is what calls forth new organizations and procedures."¹⁸

This is a familiar problem in the neoclassical theory of market failure; formally speaking, it involves an externality in the nature of a technological indivisibility arising out of costly information. Each member of the production team has an incentive to shirk if the other members cannot easily detect relative sloth. And, since some of the cost of each member's malingering is borne by the other members, each has an incentive to shirk more than the "team optimum"; so the team produces less than it would if shirking were more easily detectable. A firm-like organization, in which an administrator monitors the work of the team, would thus have a decided advantage over, and would tend to supplant, the un-administered team.

¹⁷Armen Alchian and Harold Demsetz, "Production, Information Costs, and Economic Efficiency," American Economic Review, vol. 62, December, 1972, p. 778, emphasis original.

¹⁸Op. Cit., p. 780.

Externality problems of this sort are the mainstay of that recently popular branch of mathematical economics called the economics of information. These problems take the familiar form of an optimization by economic agents of simple and fixed objectives with respect to given means, but the assumption of "perfect information" is relaxed slightly so that there are specific facts of the situation unknown to at least some of the participants. (In the above case, for example, the participants lacked information about one another's shirking.) This sort of formulation is often quite useful, for we certainly do observe particular situations in the economic world that seem to suffer from "market failure" of this sort.¹⁹

In arguing that such team-production situations are the much-sought explanation for the firm, Alchian and Demsetz would seem to agree with Michael Spence's assertion that "the firm, in large part, consists of non-market institutions whose function is to deal with resource allocation in the presence of informational constraints that markets handle poorly or not at all."²⁰ But if "information"

¹⁹On the other hand, though, externalities involving team cooperation are somewhat overrated. Studies of such cooperative situations -- often in the context of voting, which is a good example of a team activity involving shirking -- usually find that people cooperate more than a "rational" calculation of self interest would have suggested. Thomas Cotton has postulated, in this regard, that people act on the basis of behavioral norms as well as strict calculation, and that a "norm of fairness" enhances cooperation in many team situations. (Public Man: A Model of Rational Cooperation, Ph.D. Dissertation, Department of Engineering-Economic Systems, Stanford University, 1978.) Alchian and Demsetz recognize something similar in their discussion (p. 790) of "team spirit." In the terms of the present essay, such norms are in the nature of rules -- whose origin or purpose may not be explicitly understood -- which deter shirking and encourage a sense of obligation to other team members.

²⁰"The Economics of Internal Organization: An Introduction," Bell Journal of Economics, Spring, 1975, p. 164.

is interpreted very narrowly, as it often is in the mathematical economics of information, then this assertion is almost certainly not true.

For example, the "informational constraints" imposed by technological indivisibilities in team production are clearly not a sufficient explanation for the existence of firms -- or at least of big firms. If the monitoring problem as described were the only factor involved, then all operations of production that could be performed individually would be so performed; and "firms" would operate only those aspects of the process in which one individual's work is technologically indivisible from those of others. If the assembly of a rifle requires several steps conveniently performed along an assembly line, then the workers on the line are linked technologically in the sense Alchian and Demsetz suggest. But we cannot invoke their analysis to explain why the workers who make the various parts of the rifle, to the extent such construction is not actually carried out on the assembly line itself, would need to be part of the same organization. And surely there are many aspects of even the most sophisticated industrial processes that are technically separable and thus could be -- but are not in fact -- given over to non-employee contractors.

At times, though, Alchian and Demsetz appear to back off a bit from the purely technological explanation. At the most general level, they argue, the firm "serves as a highly specialized surrogate market." Because the various "inputs" or "factors of production" are in reality quite heterogeneous, the authors argue, internal organization may have informational advantages over and above those involved

in monitoring team production. "As a consequence of the flow of information to the central party (employer), the firm takes on the characteristics of an efficient market in that information about the productive characteristics of a large set of specific inputs is now more cheaply available." The price system, they seem to be saying, cannot fully capture the important qualitative aspects of these heterogeneous inputs; thus, a firm-like organization "gives the director-employer more knowledge about the productive talents of the team's inputs, and a basis for superior decisions about efficient or profitable combinations of those heterogeneous resources. . . . Efficient production with heterogeneous resources is a result not of having better resources but in knowing more accurately the relative productive performances of those resources."²¹ Alchian and Demsetz are not clear as to whether they see the heterogeneity argument as distinct from the indivisibilities-in-team-production one; but it does seem that the two explanations are in fact separable: the firm, one would think, could embody "superior knowledge about productive combinations" even if some or all of the elements involved are technologically separable.

The most ambitious attempt to explain the respective roles of market-like and internal organizations is probably that of Oliver Williamson,²² an effort best described, in his own phrase, as eclectic. Indeed, it would be difficult to say that Williamson has any clear single explanation for internal organization. What he does have is an "organizational failures framework" consisting of four parts: (1) bounded rationality and uncertainty/complexity;

²¹Alchian and Demsetz, pp. 793 and 795, emphasis original.

²²Markets and Hierarchies: Analysis and Antitrust Implications, New York: The Free Press, 1975.

(2) opportunism and small numbers; (3) information impactedness; and (4) atmosphere. This framework enables Williamson to explain, by judicious application of one or another of these elements, virtually any instance of internal organization one might name.

The rubric of "bounded rationality and uncertainty-complexity" embodies an explanation of internal organization essentially identical to what (I have tried to argue) both Knight and Coase were ultimately getting at. Long-term contracts are a way of lowering the transaction costs involved in frequent short-term arrangements; but lower cost is achieved at the expense of that flexibility which "bounded rationality" makes necessary in a complex and uncertain world. Thus:

If, in consideration of these [rationality] limits, it is very costly or impossible to identify future contingencies and specify, ex ante, appropriate adaptations thereto, long-term contracts may be supplanted by internal organization. Recourse to the latter permits adaptations to uncertainty to be accomplished by administrative processes in a sequential fashion. Thus, rather than attempt to anticipate all possible contingencies from the outset, the future is permitted to unfold. Internal organization in this way economizes on the bounded rationality attributes of decision makers in circumstances in which ... uncertainty is substantial.²³

The "opportunism and small numbers" element of the framework is taken from the theory of strategic behavior. A small-numbers bargaining situation is a gaming situation, and thus has a number of possible pathologies that could be eliminated by internalizing the transactions involved; among the pathologies is "opportunism," by which Williamson means dissembling during the course of bargaining. The "information impactedness" aspect draws from the

²³Williamson, p. 9.

aforementioned economics of information. As we saw, there may arise situations in which specific information, as defined by a given means/ends framework, is costly and asymmetrically distributed among bargaining agents; internal organization, in such cases, might thus prove a superior arrangement. The team-production analysis of Alchian and Demsetz would seem to be an instance of such an "information impactedness" argument.

"Atmosphere" is the fourth timber of the framework. "The standard economic model," says Williamson, "assumes that individuals regard transactions in a strictly neutral, instrumental manner. However, it may be more accurate, and sometimes even essential, to regard the exchange process itself as an object of value. Concern for atmosphere tends to raise such systems issues; supplying a satisfying exchange relation is made part of the economic problem, broadly construed."²

There seem to be two arguments at work here. On one level, Williamson appears to suggest that the more atmospherically preferred mode of organization would tend to predominate over the less preferred even if the latter were equally as efficient as -- or even more efficient than -- the former in strictly instrumental terms. If rifle-makers preferred to consider themselves employees rather than independent contractors, for example, internal organization would arise even though the market mode were more effective at cranking out rifles. This may be so, but it's not an argument one should try to push too far. The other version of the "atmosphere" explanation that seems to reside in

²Op. Cit., p. 38, emphasis original.

Williamson's analysis is somewhat different -- and considerably more interesting: a mode of organization is more efficient because of its atmospheric superiority.

In saying this, we are returning very near to the Alchian and Demsetz notion of a firm as in internal market for information about heterogeneous factors of production (especially heterogeneous people). "Technological separability," says Williamson, "does not imply attitudinal separability. Reference to atmosphere is intended to make allowance for attitudinal interactions and systems consequences that are associated therewith."²⁵ Even when an operation does not qualify as team production by virtue of technological necessity, it may nonetheless be a team operation -- and be more effective as a team operation -- by virtue of less tangible "attitudinal interactions." The nature of such interactions is not entirely clear, but Williamson offers the following expansion.

Distinctions between calculative and quasimoral "involvements" are relevant. Market exchange tends predominantly to encourage calculative relations of a transaction-specific sort between the parties. Such transactions are carefully metered; unsettled obligations do not carry over from one contract, or related set of transactions, to the next. Internal organization, by contrast, is often better able to make allowances for quasimoral involvements among the parties. The sociological phenomenon of reciprocity is an example. While this can and does appear in a market context, it is much more common among members of an internal organization.²⁶

Substituting "quasimoral involvements" for "attitudinal interactions" may not be a giant step toward clarity; nevertheless, this passage does seem to suggest at least one reason (over and above the more superficial question of atmospheric pleasantness) why "quasimoral involvements"

²⁵Williamson, p. 37, emphasis deleted.

²⁶Op. Cit., p. 38.

might have advantages over market transactions: the carrying over of "unsettled obligations" from transaction to transaction has a clear sense of open-endedness to it. And this is flexibility -- once again -- in a slightly different guise.

III.

Knight, Coase, Alchian and Demsetz, and Williamson (as well as others left undiscussed) could all be said to have presented "economic" explanations of internal organization. They all implicitly or explicitly formulate a problem of choice between two alternative modes of contracting and then consider structural reasons why one mode might be more efficient or less costly than the other. Their results -- at the risk of doing more violence to their arguments than I've already accomplished -- seem to fall into two categories: "hard" or strict-neoclassical explanations and "soft" or non-neoclassical explanations.

The "hard" explanations are the "market failure" arguments. There can occur specific situations in which, for reasons arising from the structure of the situation itself, market transactions -- i.e., fully-specified exchange contracts among participants -- present difficulties that could be avoided by an internal organization of employer-employee relationships. These difficulties normally stem from a strategic advantage

possessed by some of the contract participants; specifically, costs of detection permit some of the agents to shirk or dissemble at the expense of others, a problem that could be mitigated if the transactions were under the direction of an employer-monitor.

Explanations of this sort have the advantage of relating internal organization to familiar economic concepts -- optimization, market failures -- and thus to help bring organizations into more continuous conjunction with microeconomic theory. The trouble with these formulations is that, singly or severally, they do not constitute a sufficient or fully satisfying explanation for the existence of those complicated organizations we actually observe. This is so both because (a) there are frequently ways around these informational "market failures" that do not require ditching the market-contract mode (e.g., guarantees, third-party monitors, information brokers, etc.) and, more importantly, (b) there are surely many transactions within any firm that do not involve such information problems.

The "soft" explanations -- involving "flexibility," "atmosphere," etc. -- are not so firmly within the bounds of the established economic theory. Alchian and Demsetz recognize this, in a way, when they suggest that the informational efficiency of internal organization in dealing with heterogeneous factors "is obscured in the theoretical literature by the assumption of homogeneous factors."²⁷ But the extra-neoclassical aspect of these arguments is best seen in connection with the notion of flexibility. If, as I have done rather indiscriminately throughout this essay,

²⁷Alchian and Demsetz, p. 793.

one views "neoclassical" as embracing only the adjustment of known means to given ends, then the sort of "soft" flexibility that (I am arguing) gives rise to internal organization will never appear at all in a "neoclassical" theory.

When the means/ends framework is a datum of the analysis and is assumed to be known to all economic agents involved, then the only source of uncertainty can be the variability of events within the framework. There is never any uncertainty about the means/ends framework itself. The goods to be produced are given; and both their existence and their nature is agreed upon by all producers and consumers. Even when some participants in the market have more information about some products than do other participants, they all nonetheless share the same conceptual categories about the products. In the well-known "lemons" models,²⁸ for example, sellers know better than buyers whether a particular car is a lemon or not; but all know that there are both lemons and non-lemons in the market and all agree on what it means for a car to be a lemon. These pre-existing categories, indeed, are what define the "information structure" of the problem. One has information if he knows into which preformed category he should toss the various specific instances with which he is confronted.

Uncertainty is just the flip-side of information. One is uncertain when he lacks information. Normally, a neoclassical model will define a number of mutually exclusive and collectively exhaustive "states of the world."

²⁸The exemplar of such models is G.A. Akerlof, "The Market for Lemons: Qualitative Uncertainty and the Market Mechanism," Quarterly Journal of Economics, vol. 84, August, 1970, pp. 488-500.

An economic agent may not know ex ante which state will obtain, but, once again, he and all his fellows know of and agree upon the states qua categories.²⁹

This "neoclassical" sort of uncertainty can produce a need for flexibility of a particular kind. As George Stigler long ago showed, a firm that is uncertain about how much output it will sell in any period might be inclined to use a more flexible (even if more costly per unit) production technology than would a firm subject to less variability.³⁰ But the presence of this kind of uncertainty -- or the need for this kind of flexibility -- is no explanation for the existence of firms; again, it is precisely this sort of uncertainty that contingent-claims contracts, as traded on futures markets, are best able to handle.

To put it in very broad terms, the "uncertainty" that gives rise to internal organization is uncertainty about the means/ends framework itself; firms exist, in large part, precisely because the relevant conceptual categories themselves are not objectively given, are not immutable, and are not shared among all market participants.

Consider the notion of contract flexibility discussed by Coase and Williamson. The firm exists, this argument goes,

²⁹Kenneth Arrow, pre-eminent among practitioners of the neoclassical economics of uncertainty, puts it this way. "Uncertainty means that we do not have a complete description of the world which we fully believe to be true. Instead, we consider the world to be in one or another of a range of states. Each state of the world is a description which is complete for all relevant purposes. Our uncertainty consists in not knowing which state is the true one." (The Limits of Organization, New York: W.W. Norton and Company, 1974, p. 33.)

³⁰"Production and Distribution in the Short Run," Journal of Political Economy, June, 1939, p. 305.

because one cannot anticipate and fully specify all contingencies in advance for contracting purposes. What is the nature of these future contingencies? Clearly, there must be more involved than the future prices and quantities of the inputs and outputs; this is the stuff of futures markets. The important contingencies that cannot be anticipated are the states of knowledge about inputs and outputs and their effective combination.

Consider the question of specificity. If the manufacturing process is sufficiently uncomplicated, it may be possible to specify all its elements in full detail for the purposes of a contract. For instance, it may be possible to specify in reasonably comprehensive fashion all the parameters of a nineteenth-century rifle barrel: dimensions, bore, quality of steel, and even machining tolerances. But a modern jet aircraft (say) is immensely more complex; and to specify all the details of its various subsystems would be a task of such proportions that not even the federal government would likely attempt it. At best, one could specify only certain performance criteria -- "the exact details," as Coase put it, "being left until a later date."

And there is more than complexity involved. If a large number of jets were to be built, all identical, with no change possible in the desired performance criteria, the technical specifics used to meet the criteria, or any other aspect of the planes or their method of construction, then progressively more minute details of the entity "jet aircraft" would become calcified in mind and machine. It is only the possibility of change -- the possibility that the framework of means and ends embodied in the design, production, and sale of the planes might somehow shift -- that necessitates the sort of flexibility that is at issue.

Much of the discussion of "uncertainty" and its role in internal organization is characterized by this failure to distinguish between uncertainty about (or variability in) specific events within a given framework of means and ends and uncertainty about (or variability in) the means/ends framework itself. Frank Knight helped get this confusion rolling when, despite intuitions about the causes of internal organization that seem basically sound, he equated his notion of "true" uncertainty with an inability to calculate probabilities. He recognized that the uniqueness of a decision situation is at the heart of the matter; but in highlighting the subjectivity of probability estimates, he drew attention away from the idiosyncrasy of the structure one necessarily imposes on that decision situation. It is the latter aspect of subjectivity that is important in calling forth internal organization.

Some of the conceptual apparatus of decision theory may serve here to clarify the point. This theory, which is called decision analysis in its more normative role, is a logical approach to decision-making in the face of uncertain and complex circumstances. The practice of decision analysis embodies an explicit recognition that there are two aspects to the taking of a conscious choice: first comes the imposition of a decision framework onto a complex situation, only then to be followed by considerations of probability. One usually speaks of a "deterministic phase" involving a definition of the decision, an identification of the courses of action available, and a specification of the outcomes possible as a result of each alternative. This establishes the framework of means and ends for the decision; it is a step prior to formal analysis, a matter of perception rather than calculation. "Alternative

generation," as Ronald Howard acknowledges, "is the most creative part of the decision analysis procedure."³¹

Once the framework is established, then the optimal choice, uncertainty notwithstanding, is formally determinate from the structure of the problem. Executing the remainder of the decision analysis procedure may not be trivial, but all the elements are present; it becomes a matter of what F.A. Hayek calls "the Pure Logic of Choice."³²

The deterministic phase continues with an assignment of values to each possible outcome based on the preferences of the person making the decision, and it also usually involves various degrees of modeling to elucidate more fully the connections between action and outcome. In some cases, the best alternative is immediately clear; more frequently, uncertainty will cloud the optimal choice. This necessitates a "probabilistic phase." All the relevant variables in the decision problem have been specified, but there remains the task of translating the subjective knowledge of the decision-maker (or his delegates) into quantitative statements of probability about these variables. It is a process of eliciting and "encoding" the probabilities, much akin to the process of translating subjective preferences into a valuation of the possible outcomes. For any

³¹"Decision Analysis: Applied Decision Theory," in D.B. Hertz and J. Melese, eds., Proceedings of the Fourth International Conference on Operational Research, New York: Wiley-Interscience, 1966, p. 51, reprinted in Readings in Decision Analysis, Menlo Park, California: SRI International, second edition, 1977, p. 87. The description of the decision analysis procedure given here also follows this reference.

³²"Economics and Knowledge," Economica, N. S. vol. 4, 1937, p. 33, reprinted in Individualism and Economic Order, Chicago: The University of Chicago Press, 1948 (Gateway edition, 1972), p. 35.

single individual, there is no greater difficulty (conceptually if not experientially) in "calculating the probabilities" than in setting down the means/ends framework or evaluating the alternatives within the framework. Probabilities per se, and their worrisome subjectivity, are thus a red herring.

In any particular decision-situation, it is always possible for the decision-maker subjectively to establish a means/ends framework, including a complete listing of exclusive and exhaustive states of the world. For such a one-shot decision, one's ignorance can be made explicit -- both through the probabilities one assigns and through the specification of the state variables themselves. Indeed, one can think of including a catch-all category -- a black box -- for all those unforeseeable events one knows nothing about but nonetheless knows must be possible.

What does this mean for a model of "situational determinism"? Surely it means that an economic agent who not only knows various facts of the situation but also knows that unforeseen events may happen will behave differently than will a similar agent without the latter anticipation. To the extent that he assigns a non-zero probability to the black-box outcome, the decision-making agent will choose differently than if he had not included that category. Furthermore, if he doesn't know the qualitative nature of what he doesn't know, the economic agent will behave differently than if he were merely ignorant of the quantitative values of various known variables.³³

³³Which may of course be digital variables.

This qualitative aspect of uncertainty is something, I would argue, that of necessity cannot be captured in our more rigorous and mathematical models.³ And therein lies our clue.

IV.

We must concede, however, that, at the moment, what we have is only a clue and not a complete "explanation" for the existence of firms. I have argued that a particular kind of economic ignorance or qualitative uncertainty has somehow to do with the existence of firms; making this into an explanation will require an explanatory framework -- a logic of explanation -- into which the proposed causative factor can be placed.

³That is not to say that all economists have missed this notion entirely. Knight's contemporary Joseph Schumpeter was very much concerned with the introduction into the economy -- through the agency of an economic category called the entrepreneur -- of new means/ends frameworks, of new categories of action: "the new commodity, the new technology, the new source of supply, the new type of organization..." (Capitalism, Socialism and Democracy, New York: Harper and Row, 1942 (Harper Colophon Edition, 1976), p. 84; but see especially The Theory of Economic Development, trans. Redvers Opie, Harvard, 1934 (Galaxy Edition, Oxford University Press, 1961)). More recently, Israel Kirzner has developed an intricately logical theory of entrepreneurship that generalizes the Schumpeterian entrepreneur. (Competition and Entrepreneurship, Chicago: The University of Chicago Press, 1973). Unlike Knight, both Schumpeter and Kirzner clearly distinguish risk-bearing (the function of the capitalist) from entrepreneurship (which involves perception of and action upon new economic categories).

The situational determinism approach embodied in the Jevons-like optimization models of neoclassical economics is one sort of explanatory framework. It provides a causal mechanism connecting various economic variables: a change in some variables alters the situation facing the hypothetical economic agents; these agents react according to the dictates of the maximization assumption; and this in turn produces a change in other economic variables.

The neoclassical framework has been attacked on numerous fronts for many reasons. Most often, the attacks have to do with the "realism" of the underlying assumptions, including the maximization assumption. But careful students of the philosophy of science are aware that abstraction of some form is essential to theory and that a lack of "realism" -- in the sense of detailed verisimilitude -- in one's theoretical constructs is not by itself a sufficient reason to reject a theory.

Those who have sought to defend marginalism (as the maximization approach is often called) take this point quite seriously. In his well-known article on this subject, Milton Friedman, for example, goes so far as to suggest that the verisimilitude of the underlying assumptions in marginalist theory is entirely irrelevant; all that matters is the predictive ability of the theory, measured against certain empirical standards.³⁵

³⁵"The Methodology of Positive Economics," in Essays in Positive Economics, Chicago: The University of Chicago Press, 1953, p. 3, reprinted in William Breit and Harold Hochman, eds., Readings in Microeconomics, Hinsdale, Ill.: Dryden Press, Second Edition, 1977, p. 35. All page references are from this volume. The empirical criterion involved in Karl Popper's notion of falsifiability. See "Science: Conjectures and Refutations," in Conjectures and Refutations: The Growth of Scientific Knowledge, New York: Harper and Row, 1965, p. 3.

The logical status of marginalist theories, then, should not be understood as involving direct reference to some underlying "real" process that characterizes economic phenomena. Rather, the maximization representation is a "black box," if I may use the term yet again, a mechanism that somehow generates the outcome of the underlying system without necessarily replicating its workings. Marginalism is an "as if" theory.

The implication, of course, is that there must actually exist some underlying process different from that portrayed in neoclassical theory. Friedman openly acknowledges the implication, and, indeed, suggests that the process involved is something very much akin to biological evolution.

Let the apparent immediate determinant of business behavior be anything at all -- habitual reaction, random chance, or whatnot. Whenever this determinant happens to lead to behavior consistent with rational and informed maximization of returns, the business will prosper and acquire resources with which to expand; whenever it does not, the business will tend to lose resources and can be kept in existence only by the addition of resources from outside. The process of "natural selection" thus helps to validate the hypothesis -- or, rather, given natural selection, acceptance of the hypothesis can be based largely on the judgment that it summarizes appropriately the conditions for survival.³⁶

Thus the "real" mechanism underlying the supply side of the market is natural selection; and the maximization hypothesis is not an intentionalist behavioral postulate but merely a construct useful for expressing the evolutionary model in more tractable form.

The status of the maximization of expected returns hypothesis, in Friedman's view, is not far different from the hypothesis that an expert billiard player performs "as

³⁶Friedman, p. 35.

if" he were consciously solving complex mechanical equations or the hypothesis that leaves grow fuller on the sunny side of a tree just "as if" leaves consciously adjusted their positions to maximize the sunlight received by the tree. The actual mechanisms in both cases are obviously quite different from those hypothesized. In reality, there are many billiards players, some quite skilled and some inept; but if we restrict attention to "expert" players, then we can by definition model them as conscious calculators -- since only those who act as if they could solve the physical equations deserve to be called experts. Similarly, a tree sprouts many leaves in all directions; but those with a greater access to sunlight prosper relative to those in shade. In the end, says Friedman, "the result achieved by purely passive adaptation to external circumstances is the same as the result that would be achieved by deliberate accomodation to them."³⁷ But the use of the word "adaptation" here may be misleading. As Alchian points out, in an excellent article of vintage similar to Friedman's, "the survivors may appear to be those having adapted themselves to the environment, whereas the truth may well be that the environment has adopted them."³⁸

The first thing to notice about all these examples is that the locus of explanation is not in the "as if" model. The model may be a powerful engine of calculation -- or at

³⁷Op. Cit., p. 23. Note the subtlety of the selection mechanism in the billiard player example. We "select" the experts; the "real" process underlying billiards-playing, though, has nothing to do with natural selection but is, rather, some form of tacit knowledge.

³⁸"Uncertainty, Evolution, and Economic Theory," Journal of Political Economy, v. 58, June 1950, reprinted in R. Heflebower and G. Stocking, eds., Readings in Industrial Organization and Public Policy, Howewood, Ill.: Richard D. Irwin, 1958, p. 211, emphasis original.

least a more useful heuristic device -- and, perhaps, may be better able by virtue of this to generate predictions than would an invocation of the more fundamental process itself.

But, if taken too seriously, the "assumptions don't matter" assertion removes all explanatory power from a theory. The hypothesis that firms maximize expected returns, for example, collapses, in the words of Ernest Nagel, to "a somewhat loosely expressed empirical generalization" about a firm's behavior which "specifies no determinants in explanation of that behavior."³⁹

For a theory to have explanatory power, its hypothetical statements must be understood as referring to an "ideal type" whose description is in terms of abstracted theoretical statements and limiting cases. "Accordingly," Nagel explains, "discrepancies between what is asserted for the pure case and what actually happens can be attributed to the influence of factors not mentioned in the law. Moreover, since these factors and their effects can often be ascertained, the influence of the factors can be systematically classified into general types; and in consequence, the law can be viewed as the limiting case of a set of other laws..."⁴⁰ In contrast to Friedman, Fritz Machlup agrees that hypothetical "firms" must have some behavioral content. To Machlup, explanation consists precisely in casting the elements of a theory in terms of human behavior; thus "the fundamental assumptions of economic theory are not subject to a requirement of

³⁹Ernest Nagel, "Assumptions in Economic Theory," American Economic Review, May, 1963, reprinted in Breit and Hochman, eds., Op. Cit., p. 53. Page references are again from this volume.

⁴⁰Op. Cit., p. 52.

independent empirical verification, but instead to a requirement of understandability in the sense in which man can understand the actions of fellowmen."¹

This leads Machlup to a defense of marginalism that, while not wholly unlike Friedman's, is somewhat more moderate and does insist on behavioral content "on the margin." But it is a defense that forces one, in the end, to lower his expectations about the goals and capabilities of marginal analysis. The neoclassical theory of the firm, Machlup argues, "is designed to explain and predict changes in observed prices (quoted, paid, received) as effects of particular changes in conditions (wage rates, interest rates, import duties, excise taxes, technology, etc.)." The maximization model should be used, furthermore, "to predict, not the actual reactions of any one particular firm, but the effects of hypothetical reactions of numerous anonymous 'reactors' (symbolic firms)."² It is, in other words, a theory of large numbers that does not depend on what any particular real firm does. More importantly, it is a theory of observable changes in the behavior of a population of firms arising from changes in other observable aspects of the economic environment. Such a theory thus does not rest on the overall "optimality" of the behavior of real firms, but only on the nature of their response to an isolated change in their environment. "For purposes of competitive price and allocation theory," says Machlup, "it does not make much difference whether the information which we assume

¹Fritz Machlup, "The Problem of Verification in Economics," Southern Economics Journal, vol. XXII, no. 1, July 1955, p. 17, reprinted in G. Bitros, ed., Selected Economic Writings of Fritz Machlup, New York: New York University Press, 1976, p. 73.

²Machlup, Op. Cit., pp. 8-9.

the firm to have concerning the conditions of supply, production, and demand under which it works is correct or incorrect, as long as we may safely assume that any change in these conditions is registered correctly."³

The case for marginalism is quite persuasive, and one might well wish to see its usefulness as a calculative and predictive device as outweighing any disabilities it may have in the realm of explanation -- at least for the "purposes of competitive price and allocation theory" (as Machlup put it). But for other purposes, such as explaining the existence of firms or the various other "givens" of the basic theory, a direct appeal to the underlying "real" process seems the only justifiable procedure.

And, if the underlying process is evolutionary, the logic of explanation involved will be quite different from that of marginalism. In the case of the firm, for example, the strict-neoclassical explanations start with a given set of discrete institutional alternatives (e.g., internal versus market organization) and deduce, through the "as if" intermediation of a conscious rational calculation, which of the alternatives would prevail under various circumstances of, say, cost or information asymmetry. But in an evolutionary explanation, one cannot deduce the institutional structure or organizational form from the environment. "This," as Hayek points out,

implies a sort of inversion of the relation between cause and effect in the sense that the structures possessing a kind of order will exist because the elements do what is necessary to secure the persistence of that order. The 'final cause' or 'purpose', i.e., the adaptation of the parts to the requirements of the whole, becomes a necessary part of the explanation of why structures of the kind exist: we are bound to

³Op. Cit., p. 24, emphasis original.

explain the fact that the elements behave in a certain way by the circumstance that this sort of conduct is most likely to preserve the whole...."⁴⁴

And Herbert Simon has quite recently restated the matter in strikingly similar terms.

In practice, it is very rarely that the existence or character of institutions are deduced from the functions that must be performed for system survival. In almost all cases it is the other way round; it is empirical observation of the behavior pattern that raises the question of why it persists -- what function it performs. Perhaps, in an appropriate axiomatic formulation, it would be possible to deduce that every society must have food-gathering institutions. In point of fact, such institutions can be observed in every society, and their existence is then rationalized by the argument that obtaining food is a functional requisite for all societies. This kind of argument may demonstrate the sufficiency of a particular pattern for performing the essential function, but cannot demonstrate its necessity -- cannot show that there may not be alternative, functionally equivalent, behavior patterns that would satisfy the same need."⁴⁵

The organizational structures and patterns of behavior within a firm are the organs whose existence we can "explain" in terms of the functions they serve for maintaining the firms's existence.

The last part of Simon's observation is crucial for explaining the firm. Each firm is a "persistent structure of

⁴⁴"Notes on the Evolution of Systems of Rules of Conduct," in Studies in Philosophy, Politics, and Economics, Chicago: The University of Chicago Press, 1967 (Midway Reprint 1980), p. 77.

⁴⁵"Rationality as Process and as Product of Thought," American Economic Review, vol. 68, no. 2, May, 1978, p. 4, emphasis original. In the same article, Simon criticizes the increasing tendency in economics to enlist "as if" maximization models in the service of this kind of functional analysis. Such analyses are "focused on qualitative and structural questions, typically, on the choice among a small number of discrete institutional alternatives." Nevertheless, he notes, one sees an increasing use of maximization models to "explain" these institutional choices -- despite the fact not only that such models are non-explanatory in nature but also that they do not in these cases possess any calculational or heuristic advantage and, in fact, "it is quite unclear what is gained by dressing [a functional analysis] in the garb of marginalism." (pp. 5 and 6.)

relationships" -- but not unequivocally the only such structure that could persist under the same circumstances. Each embodies a sufficient-but-not-necessary relationship between means and ends.

There are many patterns of behavior -- some as yet unimagined by anyone -- that would prove equally well adapted to any particular environment. Which pattern of structures we observe will depend not on any a priori superiority of one of these over the other (since none has such superiority) but on the specific historical sequence -- the cosmology, if you will -- that the evolutionary process followed. For "the existence of such structures may in fact depend not only on [the] environment, but also on the existence in the past of many other environments, indeed on a definite sequence of such environments. . . ." ⁶ Conceivably, then, we might compare two institutional alternatives (like the internal and market form of organization) and find neither with a claim to be more adapted than the other to the environment in question; we might even find that the arrangement we do not observe is in fact superior on some grounds to the one we do observe and wish to explain.

This puts us in a position to reinterpret the explanation offered in earlier sections for the existence of firms. My argument, in effect, was precisely that in one particular environment -- a highly static environment in which states of knowledge are not changing and in which the categories of action are known to and agreed upon by all participants -- a firm is not superior to a market contract mode of

⁶Hayek, "Notes....," p. 75.

organization. My analysis of the existence of firms does not rest on an assertion that such "neoclassical" environments are not often approximated in real life; rather, it suggests that firms exist because there must be at least some points during the evolution of technologies and markets at which the categories of action are not fixed. The economic process has to pass, as it were, through zones of ignorance, and there are certain modes of organization -- which we call firms -- that can often maintain a persistent structure while making the passage."⁷ "Organizational structure," as Simon and James March suggest, "consists simply of those aspects of the pattern of behavior in the organization that are relatively stable and that change only slowly. If behavior is 'intendedly rational,' we will expect aspects of the behavior to be relatively stable that either (a) represent adaptations to relatively stable elements in the environment, or (b) are the learning programs that govern the process of adaptation."⁸

Consider once again the gedanken-firm rifle-manufacturer introduced earlier. The fully decentralized fabrication system of 1860's Birmingham was, for a time at least, a coherent structure of behavior. While we cannot consider this arrangement a firm by the established definition, it is

⁷This interpretation is consistent with some recent theories of innovation and technical change. See, e.g., William Abernathy and James Utterback, "Patterns of Industrial Innovation," Technology Review, June/July 1978, p. 41. The view that firms must deal with an extra-neoclassical world has often been suggested by writers interested in the problem of innovation in organizations. See Brian J. Loasby, Choice, Complexity, and Ignorance, Cambridge: Cambridge University Press, 1976; Donald Schon, Technology and Change: The New Heraclitus, New York: Delacorte Press, 1967; and Burton H. Klein, Dynamic Economics, Cambridge: Harvard University Press, 1977.

⁸Organizations, New York: John Wiley and Sons, 1958, p. 170.

an "organization" in the largest sense, one that operates through market transactions rather than through employment contracts. Now, this decentralized system, I would argue, was certainly an adaptation to a particular environment, and cannot be distinguished from a firm in that regard; but, unlike a firm, the structure of relationships involved in this system was unable to persist in the face of changing circumstances.⁹ And the reason that this organization could not persist is that it possessed nothing resembling a "learning program."

Note that this is not a question of the existence of "memory." The decentralized market-contract mode as a whole "remembers" how to make guns. Like a firm, it has evolved what organization theorists sometimes describe as an "operational program" -- a way of doing things, a corpus of rules for action. But, unlike a firm, this pure market system lacks, as it were, some of the higher-level functions of memory. More to the point, it lacks not so much the ability to learn as the ability to remain an identified whole while learning -- to "co-evolve."

Part of this "co-evolutionary" ability undoubtedly arises from the centrality of purpose or direction that a firm-like organization possesses. Frank Knight made just such a suggestion as part of his attempt to explain the firm.

When uncertainty is present and the task of deciding what to do and how to do it takes the ascendancy over that of execution, the internal

⁹As Stigler remarks, "[t]he later history of the gun trade, in which American innovations in production techniques were revolutionary, suggest[s] that the organization in Birmingham was deficient in its provision for technical experimentation." ("The Division of Labor is Limited by the Extent of the Market," Journal of Political Economy, v. LIX, no. 3, June 1951, reprinted in The Organization of Industry, Homewood, Illinois: Richard D. Irwin, 1968, p. 140n.)

organization of the productive group is no longer a matter of indifference or a mechanical detail. Centralization of this deciding and controlling function is imperative, a process of "cephalization," such as has taken place in the evolution of organic life, is inevitable, as for the same reasons as in the case of biological evolution.⁵⁰

But, as Hayek notes, neither learning nor apparent purposiveness depends quite as strongly on centralized direction as one might think.

There is ... no reason why a polycentric order in which each element is guided only by rules and receives no orders from a centre should not be capable of bringing about as complex and apparently as 'purposive' an adaptation to circumstances as could be produced in a system where a part is set aside to perform such an order on an analogue or model before it is put into execution by the larger structure. In so far as the self-organizing forces of a structure as a whole lead at once to the right kind of action (or to tentative actions which can be retracted before too much harm is done) such a single-stage order need not be inferior to a hierarchic one in which the whole merely carries out what has first been tried out in a part. Such a non-hierarchic order dispenses with the necessity of first communicating all the information on which its several elements act to a common centre and conceivably may make the use of more information possible than could be transmitted to, and digested by, a centre.⁵¹

So perhaps it is "cephalization" rather than centralization that is the interesting word. For, just as psychologists are coming to believe that brain-like properties -- and especially learning functions -- are not restricted physiologically to the head, organization theorists are suggesting a more decentralized picture of organizational purposiveness and learning.

In this sense, then, Knight was on the right track all along; the process of organizational cephalization is the key to understanding the development of the firm. And here

⁵⁰Risk, Uncertainty, and Profit, Chicago: The University of Chicago Press, 1971, p. 268.

⁵¹Hayek, "Notes...", p. 74.

-- rather than in transactions costs, non-convexities, or even the subjectivity of probabilities -- is where we should look for a fundamental explanation of the firm as an economic institution.