

Organizations and Markets

Herbert A. Simon

In classical and neoclassical economic theory, markets are at the center of the stage. The actors in these markets are workers and consumers (sometimes combined into households), firms, owners of resources, governments, and perhaps others. The economic world of the neoclassical textbooks is a world of transactions, and these transactions typically involve an exchange of goods, services, and/or money that both parties to the transaction find advantageous to achieve these goals. Along with consumption, work and leisure are important components of the utility functions of households. Often, profit is assumed to be the sole objective of firms and their owners.

The description of the parties who participate in these transactions is minimal. However, as soon as firms are elaborated to become more than simple nodes in a network of transactions, to be producers—transformers of “factors” into products—difficult and important questions arise for the theory. A large part of the behavior of the system now takes place inside the skins of firms, and does not consist just of market exchanges. Counted by the head, most of the actors in a modern economy are employees, who either do not spend their days in trading, or if they do (for example, if they are salesmen or purchasing agents) are assumed to trade as agents of the firm rather than in their own interest, which might be quite different.

This raises the question of why there are firms at all. Why are not all the actors independent contractors? Why do most of them enter into employment

■ *Herbert A. Simon is Richard King Mellon University Professor of Computer Science and Psychology, Carnegie Mellon University, Pittsburgh, Pennsylvania.*

contracts, selling their labor for a wage? What determines the make-or-buy decisions of firms, hence the boundaries between them and markets? When will two domains of activity lie within a single firm, and when will they be handled by separate contracting firms?

A second set of questions asks how the employees of firms are motivated to work for the maximization of the firm's profit. What's in it for them? How are their utility functions reconciled with those of the firm? In the employee's utility function, work is usually assumed to have negative utility and leisure (including loafing and working lackadaisically) to have positive utility. Why do employees often work hard?

The simple (neoclassical) answer to the motivational question derives from the employment contract, under which workers maximize their utility by accepting the authority of the firm; that is, by agreeing to accept orders from the profit maximizers in charge. But this answer leads to the new question of how the employment contract is enforced by the employer. In particular, how are employees induced to work more than minimally, and perhaps even with initiative and enthusiasm? Why should employees attempt to maximize the profits of their firms when making the decisions that are delegated to them?

These questions about the scope of activity and operation of firms have spawned a vigorous cottage industry, a branch of which is sometimes called "the new institutional economics," which tries to explain when activities will be carried out through the market and when they will be carried out within the skins of firms, and tries to explain also how it is possible for firms to operate efficiently. In the literature of the new institutional economics, two ideas that play a major role in the explanations are "transaction costs" and "opportunism" (for example, Williamson 1975, 1985). Sometimes the explanations are couched in terms of "information asymmetry" or "incomplete information" (Ross, 1973; Stiglitz, 1974). In other writings these topics are subsumed under agency theory, which treats the employment contract as an optimal contract between principal and agents, and studies how contractual arrangements can deal with shirking and other motivational problems.

The idea behind these ideas is that a proper explanation of an economic phenomenon will reduce it to maximizing behavior of parties who are engaged in contracting, given the circumstances that surround the transaction. The terms of the contract will be influenced by the access of the parties to information, by the costs of negotiating, and by the opportunities for cheating. Access to information, negotiation costs, and opportunities for cheating are most often treated as exogenous variables that do not themselves need to be explained. It has been observed that they even introduce a sort of bounded rationality into the behavior, with the exogeneity of the limits of rationality allowing the theory to remain within the magical domains of utility and profit maximization.

A fundamental feature of the new institutional economics is that it retains the centrality of markets and exchanges. All phenomena are to be explained by translating them into (or deriving them from) market transactions based upon negotiated contracts, for example, in which employers become "principals" and

employees become “agents.” Although the new institutional economics is wholly compatible with and conservative of neoclassical theory, it does greatly multiply the number of auxiliary exogenous assumptions that are needed for the theory to work. For example, to explain the presence or absence of certain kinds of insurance contracts, moral risk is invoked; the incompleteness of contracts is assumed to derive from the fact that information is incomplete or distributed asymmetrically between the parties to the contract. Since such constructs are typically introduced into the analysis in a casual way, with no empirical support except an appeal to introspection and common sense, mechanisms of these sorts have proliferated in the literature, giving it a very *ad hoc* flavor.

In general, the new institutional economics has not drawn heavily from the empirical work in organizations and decision-making for its auxiliary assumptions. (For introductions to that literature, see March and Simon, 1958; Cyert and March, 1963; Kornai, 1971; Simon, 1979). Nevertheless, it is appropriately subversive of neoclassical theory in that it suggests a whole agenda of microeconomic empirical work that must be performed to estimate the exogenous parameters and to test the theory empirically. Until that research has been carried out (and the existing literature on organizations and decision making taken into account), the new institutional economics and related approaches are acts of faith, or perhaps of piety.

The Ubiquity of Organizations

A mythical visitor from Mars, not having been apprised of the centrality of markets and contracts, might find the new institutional economics rather astonishing. Suppose that it (the visitor—I’ll avoid the question of its sex) approaches the Earth from space, equipped with a telescope that reveals social structures. The firms reveal themselves, say, as solid green areas with faint interior contours marking out divisions and departments. Market transactions show as red lines connecting firms, forming a network in the spaces between them. Within firms (and perhaps even between them) the approaching visitor also sees pale blue lines, the lines of authority connecting bosses with various levels of workers. As our visitor looked more carefully at the scene beneath, it might see one of the green masses divide, as a firm divested itself of one of its divisions. Or it might see one green object gobble up another. At this distance, the departing golden parachutes would probably not be visible.

No matter whether our visitor approached the United States or the Soviet Union, urban China or the European Community, the greater part of the space below it would be within the green areas, for almost all of the inhabitants would be employees, hence inside the firm boundaries. Organizations would be the dominant feature of the landscape. A message sent back home, describing the scene, would speak of “large green areas interconnected by red lines.” It would not likely speak of “a network of red lines connecting green spots.”

Of course, if the vehicle hovered over central Africa, or the more rural portions of China or India, the green areas would be much smaller, and there would be large spaces inhabited by the little black dots we know as families and villages. But the red lines would be fainter and sparser in this case, too, because the black dots would be close to self-sufficiency, and only partially immersed in markets. But let us, for the present, restrict our attention to the landscape of the developed economies.

When our visitor came to know that the green masses were organizations and the red lines connecting them were market transactions, it might be surprised to hear the structure called a market economy. “Wouldn’t ‘organizational economy’ be the more appropriate term?” it might ask. The choice of name may matter a great deal. The name can affect the order in which we describe its institutions, and the order of description can affect the theory. In particular, it may strongly affect our choice of the variables that are important enough to be included in a first-order theory of the phenomena.

How does the economy look when it is viewed as an organizational economy, with market relations among organizations? I have already suggested some of the more prominent features.

First, most producers are employees of firms, not owners. Viewed from the vantage point of classical theory, they have no reason to maximize the profits of firms, except to the extent that they can be controlled by owners. Moreover, profit-making firms, nonprofit organizations, and bureaucratic organizations all have exactly the same problem of inducing their employees to work toward the organizational goals. There is no reason, a priori, why it should be easier (or harder) to produce this motivation in organizations aimed at maximizing profits than in organizations with different goals. If it is true in an organizational economy that organizations motivated by profits will be more efficient than other organizations, additional postulates will have to be introduced to account for it.

Second, the system is nearly in neutral equilibrium between the use of market transactions and authority relations to handle any particular matter: that is to say, very small changes in the situation can tip the equilibrium one way or the other. It is hard to explain degrees of integration of economic activities. In many instances, transaction cost analysis is not applicable, and even where it is, there often remains considerable latitude for different degrees of integration. For example, why are auto dealerships not a part of auto manufacturing companies, rather than having contractual relations with them?¹ Why did General Motors manage its own tool design for many years, but recently decide to contract most of it out? Under constant returns to scale and reasonably competitive markets, which characterize many manufacturing situa-

¹Williamson’s explanation—actually, Alfred P. Sloan’s explanation (see Williamson, 1985, p. 10)—that employees could not be supervised adequately in their offers for used cars, is not convincing. Dealerships are also organizations, and their salesmen are employees.

tions, make-or-buy decisions become ambiguous. The possibility of using internal division-by-division balance sheets, and internal pricing in negotiation between components of an organization further blurs the boundary between organizations and markets.

Without the introduction of very particular ad hoc assumptions, unbuttressed by empirical evidence, neoclassical theory provides no explanation for the repeated appearance of Pareto distributions of business firm sizes in virtually all situations where size distributions have been studied (Ijiri and Simon, 1977; Simon, 1979). (In a Pareto distribution, the logarithm of the number of firms above any given size decreases linearly with the logarithm of the size.) These observed distributions are difficult to reconcile with any notions that have been proposed for optimal firm size, but are easily explained by simple, plausible probabilistic mechanisms that make no appeal to optimality.

In sum, an organizational economy poses the questions of why the larger part of a modern economy's business is done by organizations, what role markets play in connecting these organizations with each other, and what role markets play in connecting organizations with consumers. Moreover, the boundary between markets and organizations varies greatly from one society to another and from one time to another. What mechanism maintains the highly fluid equilibrium between them? Until these questions are answered, it will be difficult to draw conclusions about the relative efficiencies of different forms of ownership and control of organizations, or the relative efficiency of markets versus central planning.

Motivation and Efficiency in Organizations

There are three different questions of social organization that are usually confounded, but which need to be considered separately. The first is the question of the relative efficiency of markets and organizations. The second is the question of the consequences of having a society's organizations owned by profit-making organizations, by nonprofit organizations, or by public organizations, respectively. The third is the question of the consequences of using central planning instead of markets to regulate relations among organizations. At present, our concern is only with the first question: what makes organizations work as well or badly as they do?

In particular, for whom is profit the motive? Adolf Berle and Gardiner Means posed the problem very sharply in their famous book, *The Modern Corporation and Private Property* (1933), by showing that even at the top executive levels of the modern corporation there is a great gap between ownership and control, and a correspondingly great opportunity for discrepancy between the goal of owners (profit) and the goals of managers (career status, wealth, a quiet life, and so on).

Demsetz and Lehn (1985) have contested the argument of Berle and Means on the ground that even large corporations show considerable concentration of ownership. Typically, a half dozen owners (or fewer) own 10 or 20 percent of the shares, enough to retain controlling power. Often, these owners are also the active top executives. But the objection does not hold water. If a company has an executive bonus plan, and if an executive's percentage share in bonus awards is greater than his or her percentage share of dividends, then it pays that executive to divert earnings from dividends to bonuses. Most companies have executive award systems that make this conflict of interest very real. Golden parachutes and leveraged buyouts are other significant examples of transactions where the interests of shareholders and executives may diverge strongly.²

If even top executives may be conflicted in their motives, the problem should be still greater for employees who are not owners at all, or only insignificantly. Principal-agent theory, on which the new institutional economists often rely, assumes that agents within firms will shirk unless their actions contribute directly to their own economic self-interest. It is only via monitoring combined with contracts that appeal to their self-seeking nature that such shirking may be mitigated. But the assumption that executives (and perhaps other employees) would choose to advance their own careers and wealth and consumption, rather than pursuing organizational goals like maximizing profit, is not prescribed by neoclassical theory, which leaves the specification of the utility function completely open.

Why not assume that maximizing the firm's profit is precisely what maximizes the utilities of executives and other workers? In a society of robots, an owner would not settle for less. But most of us would think this an unrealistic assumption to make for a human society. An organization theory with an unspecified utility function is not a theory at all. And one with an unrealistic utility function does not provide a basis for understanding real organizations. Instead, we should begin with empirically valid postulates about what motivates real people in real organizations. I shall argue that such postulates can be derived from four organizational phenomena whose roles are amply documented in the literature on organizations: authority, rewards, identification, and coordination.

Authority: The Employment Relation

The employment contract is an example of what is now sometimes called an "incomplete contract;" that is to say, some of its terms are unspecified.

²Demsetz and Lehn (1985) cite evidence to show that corporations where ownership is widely distributed have, on average, profits as large as those with concentrated ownership. This fact does not undermine the argument of Berle and Means for conflict of economic interest; on the contrary, it raises the question—which I will undertake to answer below—of why executives with small stakes as shareholders do appear to work for company profits.

Employees agree to do, over the life of the contract, what they are ordered to do; but the orders will not be issued until some time after the contract is negotiated (Simon, 1951; Williamson, 1975).

The usual argument (within the neoclassical framework) for the existence of incomplete contracts is that in a world of uncertainty actions will have to be taken as the situation calls for them, without time for negotiation. The employee is rewarded, in the level of the wage, for willingness to bear the brunt of this uncertainty as to what actions will be chosen, and to do, when the time comes, whatever the employer thinks the situation calls for. This argument does not imply that uncertainty is replaced by complete certainty at the time of decision. On the contrary, taking decisions under conditions of uncertainty may be one of the important skills demanded of the decision maker. The essential point is that the uncertainty for the employer is decreased by delaying the commitment to specific actions from the time employment begins until the time when action is called for.

An employment contract contains all sorts of implicit (and explicit) limitations that set boundaries to the range of actions the employee will be directed to perform. These boundaries define the “zone of acceptance” within which an employee can be expected to obey orders. The zone of acceptance is also sometimes called a “zone of indifference,” for the choice among alternative behaviors, while of major importance to the employer, may be of little or no concern to the employee. A secretary, for example, usually has little or no preference for typing a letter to one of the company’s customers rather than another, and little interest in the content of the letter. Even a factory manager will accept, within wide limits, whatever mix of products the factory is ordered to produce in a given month.

The combination of uncertainty on the part of the employer (as to what will need to be done in the future) and broad acceptance of the employee (of what he or she will be ordered to do) makes the employment contract a very attractive bargain for both parties. The new institutional economics finds that employment achieves great savings in transaction costs—the costs of negotiating separate contracts for each action.

But this theory of the employment contract must be elaborated. Authority in organizations is not used exclusively, or even mainly, to command specific actions. Most often, the command takes the form of a result to be produced (“repair this hinge”), or a principle to be applied (“all purchases must be made through the purchasing department”), or goal constraints (“manufacture as cheaply as possible consistent with quality”). Only the end goal has been supplied by the command, not the method of reaching it. The mechanic must apply all kinds of knowledge and skill to repairing the hinge. The section chief must initiate purchases of supplies needed for the work of that section; however, the company’s standard procedures must be taken as ground rules for the way the purchases are made. The factory manager must control manufacturing cost and quality.

Employees, especially but not exclusively at managerial and executive levels, are responsible not only for evaluating alternatives and choosing among them but also for recognizing the need for decisions, putting them on the agenda, and seeing to the generation of possible actions. Doing the job well is not mainly a matter of responding to commands, but is much more a matter of taking initiative to advance organizational objectives.

Commands do not usually specify concrete actions but, instead, define some of the premises that are to be used by employees in making the decisions for which they are responsible (Simon, 1947). Hence, seeing that commands are obeyed is not simply a matter of observing behavior, but of affecting the thought processes and the decision premises of employees. Further, it is usually difficult or impossible to ascertain what these decision premises have been without reviewing the whole decision—thus causing an almost complete loss of the economy that was sought in delegating it in the first place.

The command an employer might like to issue is: “Always decide in such a way as to maximize company profit!” But that would simply reintroduce the question of how the extent of obedience to the command is to be observed without losing the benefit of delegation. Even if the employees were robots, whose loyalty could be guaranteed, the problem would not be solved. For giving each robot complete discretion would surrender large efficiencies usually attainable from specialization in decision-making work. We need to delegate within guidelines, which creates the problem of monitoring the observance of guidelines without recentralizing what has just been delegated.

If authority is used to transmit premises for making decisions rather than commands for specific behaviors, then many different experts can contribute their knowledge to a single decision. Information and policy rules can flow through the organization along many channels, serving as inputs—decision premises—for many organizational behaviors.

The accounting department gathers cost data, which it supplies to the head of the blast furnace department to help make operating decisions in that department. At the same time, the blast furnace manager is receiving instructions from metallurgical specialists on the technical aspects of the operation. The faint blue lines that our visitor from Mars saw within the green areas were not just streams of orders, but flows of all kinds of decision premises (constraints and information as well as orders) from one point in the organization to another.

This explication of the employment contract and authority takes us back to the question of motivation. For the organization to work well, it is not enough for employees to accept commands literally. In fact, obeying operating rules literally is a favorite method of work slowdown during labor-management disputes, as visitors to airports when controllers are unhappy can attest. What is required is that employees take initiative and apply all their skill and knowledge to advance the achievement of the organization’s objectives.

We should not assume without evidence that organizations *do* work well. But “well” is a relative term. In most organizations, employees contribute much

more to goal achievement than the minimum that could be extracted from them by supervisory enforcement of the (vague) terms of the employment contract. Why do employees not substitute leisure for work more consistently than they do? Why do they often work so vigorously for the welfare of the organization?

Rewards as Motivations

One obvious answer to the motivational question is that employees may be motivated to accept authority by giving them material rewards, promotion, and recognition for advancing the organization's goals as defined by management. Such rewards certainly provide motivation, but they only operate satisfactorily when certain conditions are met. The most important condition is that the employee's contribution to the organization's goals must be measurable with reasonable accuracy. For example, salesmen are frequently compensated (at least partly) on a commission basis. Blue-collar employees are sometimes compensated on a piecework basis, albeit in a continually decreasing number of situations. Executives, and sometimes others, receive bonuses that are supposed to be related to their contributions to profits.

But such reward systems are effective only to the extent that success can be attributed accurately to individual behaviors. If the indices used to measure outcomes are inappropriate, either because they do not measure the right variables, or because they do not properly identify individual contributions, then reward systems can be grossly inefficient or even counterproductive. Where output quantities are measured with inadequate attention to quality, response to rewards will cause quantities to grow at the cost of lowered quality. Where compliance with company policies that constrain action is not measured, constraints will be ignored and violated. Salesmen may misrepresent the product, workmen may ignore safety rules, managers may buck difficulties to other departments.

In general, the greater the interdependence among various members of the organization, the more difficult it is to measure their separate contributions to the achievement of the organizational goals. But of course, intense interdependence is precisely what makes it advantageous to organize people instead of depending wholly on market transactions. The measurement difficulties associated with tying rewards to contributions are not superficial, but arise from the very nature and rationale of organization.

Many large U.S. corporations attempted to respond to this problem in the years after World War II by slicing their organization into components that were relatively self-contained. Then, separate balance sheets could be maintained for each division, and these balance sheets could be used to evaluate results and to compute rewards.

Of course, divisionalization can be successful only to the extent that the divisions are actually self-contained. If one division operates mainly as a

supplier of parts to other divisions, then policies have to be laid down for setting the prices for items “sold” by the one division to the others, and for determining under what conditions a division may go outside the company to purchase items at a lower price. For these and similar reasons, divisionalization can only be carried a short distance down the structure of a typical corporation, and solves the problem of attributing outcomes to individuals only at the higher levels, if at all.

Although economic rewards play an important part in securing adherence to organizational goals and management authority, they are limited in their effectiveness.³ Organizations would be far less effective systems than they actually are if such rewards were the only means, or even the principal means, of motivation available. In fact, observation of behavior in organizations reveals other powerful motivations that induce employees to accept organizational goals and authority as bases for their actions. We turn next to the most important of these mechanisms: organizational identification.

Loyalty: Identification with Organizational Goals

Pride in work and organizational loyalty are widespread phenomena in organizations (Simon, 1947). These traits are more strongly evident among skilled and managerial employees than among employees engaged in very routine work. (The latter are also more easily supervised, and can sometimes be rewarded on a piecework basis.) In part, these attitudes can be attributed to the linkage between an organization’s overall success and the personal careers and monetary rewards it can provide its employees. But this explanation ignores the problem of the commons—of benefits that are jointly gained and shared by all, non-contributors along with contributors—and the consequent possibilities for free-riding. The quality and success of an organization depends very little on the energy of any single employee (except possibly an executive at or near the very top). Why will employees work hard if they can gain almost as much by loafing?

Of course free-riding can be observed in organizations. The elimination of free riding is generally thought to be the principal reason for the success of the Chinese agricultural reforms after 1980, when responsibility and reward for agricultural production were transferred from the commune to the family. The question is not whether free riders exist—much less employees who exert something less than their maximum—but why there is anything *besides* free-riding. Why do many workers, perhaps most, exert more than minimally enforceable effort? Why do employees identify with organizational goals at all?

³Everything said here about economic rewards applies equally to privately owned, nonprofit, and government-owned organizations. The opportunity for, and limits on, the use of rewards to motivate activities toward organizational goals are precisely the same in all three kinds of organizations. For sophisticated discussions of motivation and efficiency in profit-making and nonprofit organizations, see Weisbrod (1988, 1989).

Contemporary evolutionary theory has cautioned us against postulating altruistic motives for people. In models of natural selection, nice guys generally aren't fit; they don't multiply as rapidly as their more selfish brethren. The argument from natural selection has often been used, explicitly or implicitly, to fill the utility function with selfish personal goals. But models of natural selection do not actually provide strong support for the idea that people will only pursue selfish personal economic goals. In fact, such models in no way foreclose the possibility (indeed, the probability) that people will be strongly motivated by organizational loyalty, even when they can expect no "selfish" rewards from it (Simon, 1983; 1990).

First, it should be emphasized that what natural selection increases is fitness, the number of progeny of the successful competitor. But in modern society, the attainment of wealth or other selfish rewards is not directly connected to number of progeny. In fact, first-world societies generally display a negative correlation between income level and size of family. But let us waive this point, as distracting us from the main argument, and suppose that attainment of the goals usually described as selfish (especially personal economic goals) contributes to evolutionary fitness.

We come then to the second point: each human being depends for survival on the immediate and broader surrounding society. Human beings are not the independent windowless Leibnitzian monads sometimes conjured up by libertarian theory. Society is not imposed on humans; rather, it provides the matrix in which we survive and mature and act on the environment. Families and the rest of society provide nutrition, shelter, and safety during childhood and youth, and then the knowledge and skills for adult performance. Moreover, society can react to a person's activities at every stage of life, either facilitating them or severely impeding them. Society has enormous powers, enduring though a person's lifetime, to enhance or reduce evolutionary fitness.

What kinds of traits, in addition to personal strength and intelligence, would contribute to the fitness of this socially dependent creature? One such trait, or combination of traits, might be called docility. To be docile is to be tractable, manageable, and above all, teachable. Docile people tend to adapt their behavior to norms and pressures of the society. I am not satisfied that "docile" conveys my meaning precisely, but I know of no better word.

That fitness is derivable from being docile becomes evident when we consider the opposite of docility: intractability, unmanageability, unteachability, incorrigibility. The argument is not that people are totally docile, nor that they are totally selfish, but that fitness calls for a measured but substantial responsiveness to social influence. In some contexts, this responsiveness implies motivation to learn or imitate; in other contexts, willingness to obey or conform. From an evolutionary standpoint, having a considerable measure of docility is not altruism but enlightened selfishness.

To survive as a trait, docility must contribute on average to the fitness of the individual who possesses it. Yet it may still lead to self-injurious behavior in particular cases. Thus, docile individuals may do better at earning a living, but

loyalty to the nation may lead them to sacrifice their lives in wartime. Once docility is present, society may exploit it by teaching values that are truly altruistic; that is, which contribute to the society's fitness, but not to the individual's. The only requirement is that *on balance* and on the average the docile individual must be fitter than the one who is not docile.⁴

Of course, showing that a configuration of traits or genes would contribute to fitness, if they existed, does not prove they exist. But ample empirical evidence shows that most human beings are gifted with a considerable measure of docility. The purpose of the present argument is to show that this docility and the altruism it induces is wholly consistent with the premise of selection of the fittest. In fact, the theory of natural selection strongly predicts the appearance of docility and altruism in social animals.

Docility is used to inculcate individuals with organizational pride and loyalty. These motives are based upon a discrimination between a "we" and a "they." Identification with the "we," which may be a family, a company, a city, a nation, or the local baseball team, allows individuals to experience satisfactions (to gain utility) from successes of the unit thus selected. Thus, organizational identification becomes a motivation for employees to work actively for organizational goals. Of course, identification is not an exclusive source of motivation; it exists side by side with material rewards and enforcement mechanisms that are part of the employment contract. But a realistic picture of how organizations operate must include the importance of identification in the motivations of employees.

The strength of organizational identifications will depend upon the extent to which a society uses the docility mechanism to inculcate them, and this appears to vary considerably from one society to another. For instance, it would probably be agreed by ethnographers that in Chinese society greater pressure is exerted to induce identification with the family than with employing organizations, while the reverse is true of Japanese society. Such conjectures can be tested, for example, by examining practices of nepotism, and attitudes toward it, in the two societies.

The strength of the organizational loyalties of employees is not to be attributed only to motivation induced by docility. There is also an important

⁴This is not the place to describe in detail how docility and altruism induced through the docility mechanisms can be incorporated in a formal model of evolution by natural selection. I will simply sketch the general idea. Let k be the average number of offspring of an individual in the absence of docility or altruistic behavior; $d > 0$ the gross increase in offspring due to docility; $c > 0$ the cost to a docile individual in offspring of the socially induced altruistic behavior; p the percentage of individuals in the population which are docile and hence altruistic; and b the number of offspring added to the *population* by an individual's altruistic behavior. Assume further, that the parentage of offspring contributed by altruism is distributed randomly through the population. Then it is easy to show that the difference between the net fitness of altruists and non-altruists (non-docile individuals), respectively, will equal $d - c$. Hence, provided that d is larger than c , altruists will be fitter than non-altruists. Moreover, a society will grow more rapidly the greater the fraction of altruists in it, the increase in average fitness being $(d - c + b)p$.

cognitive component. The bounded rationality of humans does not allow us to grasp the complex situations that provide the environments for our actions in their entirety. The first step in rational action is to focus attention on specific (strategic) aspects of the total situation, and to form a model of the situation in terms of those aspects that lie in that focus of attention. Rational computation takes place in the context of this model, rather than in the response to the whole external reality.

One dimension of simplification is to focus on particular goals, and one form of focus is to attend to the goals of an organization or organization unit. Having defined that unit as the “we,” actions are evaluated in terms of their contribution to the unit’s objectives. The ubiquity of this narrowing of attention is easily demonstrated. As one example, Dearborn and Simon (1958) presented a group of business executives with a description of the current situation of a large company, and asked them to identify the most serious problem facing the company. In their own companies, some of the executives were responsible for manufacturing, others for sales, others for finance. In almost every case, the “most serious problem” identified by the executive lay in the domain of his or her own department—manufacturing problems for manufacturing executives, sales problems for sales executives, and so on.

It is a commonplace of organizational life that a person’s organizational identification will shift with his or her position, although the motivational basis for the shift is perhaps more widely recognized than the cognitive basis. But a shift in organizational position exposes the employees to new “facts” and phenomena, to a new network of communications, and to new goals. A different model is inevitably formed of the decision-making situation, a model that emphasizes local components of the environment and local goals. Behavior is very much a function of position.

Because of cognitive limits, the precise form that goals take may depend on what can be measured in the situation. In business organizations, the accounting statements provide stylized measurements of profits, size, growth, market share, and so on. Even if these measurements are only rough approximations of the things they are supposed to be measuring, they are likely to replace the “real” unmeasured concepts in the decision-making process.

Willingness of employees at all levels to assume responsibility for producing results—not simply “following the rules”—is generally believed to be a major determinant of organizational success. This discussion implies that acceptance of responsibility will be affected both by the reward system and by the strengths of organizational identifications. Here again, large intercultural differences may exist. The recent establishment of a substantial number of international joint ventures, with managements and employees recruited from different cultures, provides an excellent research environment for studying these differences and their effects upon organizational efficiency.

Since the developments are quite new, little information is yet available about them. But one example where data are available is the joint venture

between Toyota and General Motors in northern California (Krafcik and Womack, 1987). Here Toyota took over a former General Motors plant, equipped it with standard state-of-the-art machinery, rehired employees mainly from the previous work force and accepted the same union. They have been able to produce automobiles with about 45 percent fewer labor hours than an entirely comparable GM plant that uses American managers and management methods, and about 30 percent fewer hours than a new GM plant having more modern “hitech” equipment, and only about 15 percent *more* labor hours than a comparable Toyota plant in Japan.

The causes for these enormous differences in efficiency have almost nothing to do with the classical physical production function. They also appear to have little to do with cultural differences at the blue-collar level.⁵ They seem to have nothing to do, either, with material reward structures, which are not significantly different in the various plants. They must be attributed in large part to differences in management practices (for example, quality control practices, and inventory policies), perhaps bolstered by differences in management attitudes and motivations.

Coordination

This examination of authority and organizational identification should help explain how organizations can be highly productive even though the relation between their goals and the material rewards received by employees, if it exists at all, is extremely indirect and tenuous. In particular, it helps explain why careful comparative studies have generally found it hard to identify systematic differences in productivity and efficiency between profit-making, nonprofit, and publicly-controlled organizations (Weisbrod, 1988, 1989). Also, it explains why Demsetz and Lehn (1985) found no difference in profits between corporations that were managed or controlled by owners and those with diffuse stock ownership.

But to understand the relative advantages of organizations and markets, and the circumstances under which one would operate more effectively than the other, one further component must be added to our description of organizations. Organizations, through the authority mechanism, provide a means for *coordinating* the activities of groups of individuals in ways that are not always easily achieved by markets.

⁵These two statements should be qualified slightly. With regard to the first, components imported by the Toyota plant from Japan may be more uniform in quality than components purchased by the other GM plants. With respect to the second, applicants interviewed for employment in the Toyota plant were screened for problem solving attitudes and skills. Note that both of these factors, whether important or not, are matters of management practice. Finally, I would not wish to claim that the factors I have mentioned were the only ones affecting the comparison.

Coordination is a rather slovenly word, often abused in organizations. An experienced executive cringes when he or she learns that someone has been appointed to “coordinate” a set of activities, since calling for coordination without specifying just what it means is simply a lazy way of passing off problems to someone else. I will try to make the concept more precise by using it to designate a specific kind of activity.

The theory of games has sharply underscored that decisions are usually indeterminate when each party in a situation is uncertain about the actions of the others. This result is quite independent of whether their goals are complementary or competitive. One simple example of this indeterminacy is that it is rational for a motorist to drive on the *same* side of the road as other drivers headed the same direction, whichever that may be. There is no question of correct behavior in relation to the environment, but only of coordinating the behaviors of all the actors. Such rules of the road, or standardization, can greatly improve the performance of systems in those (ubiquitous) situations where the correctness of an action depends on what the other actors are doing.

A more complex example of coordination is provided by a university. Conceive of a university that consisted only of some rooms, some teachers, and some students. Students and teachers would “simply” negotiate to meet at certain times and places for their classes. The resulting chaos would probably be resolved by inventing the institutions of a registrar’s office and a class schedule. While it would be extravagant to urge that class schedules provide the *raison d’être* for education by universities, rather than by contractual tutoring arrangements negotiated through markets, nevertheless, the coordinating function of schedules is not trivial.

A major use of authority in organizations is to coordinate behavior by promulgating standards and rules of the road, thus allowing actors to form more stable expectations about the behavior of the environment (including the behavior of other actors). Since organizations provide a mechanism (authority) for establishing rules of the road, which markets do not, one might even expect organizations rather than markets to be the environments in which the behavior called “rational expectations” would be most often observed.⁶

In a book on central planning during World War II, Ely Devons (1950) raised the question of why prices are supplanted by government plans, expressed as quantity goals for production and allocation, as coordinating mechanisms during wartime. The usual argument for markets, as in the well-known 1945 paper of von Hayek, is that they simplify the decision process by reducing the need of each actor to know what the other actors are doing or what situations confront them. To the extent that markets and prices perform this simplifying function, we would expect them to replace centralized decisions

⁶Of course, perfectly competitive markets do provide stable expectations of prices at least in equilibrium, and thereby permit Pareto optima to be achieved in principle. But prices are only one of many dimensions along which uncertainty of expectations may complicate rational decision making.

when a situation becomes more complex—for example, during the rapid changes that take place in shifting from a peacetime to a wartime economy. Yet, as Devons points out, it is just at such times that central planning tends to increase. Is this irrationality, or are there valid reasons for the shift?

The answer is rather obvious. Prices perform their informational function when they are known or reasonably predictable. Uncertain prices produced by unpredictable shifts in a system reduce the ability of actors to respond rationally. This point is often made by economists in arguing the costs of unexpected inflation, but its implication for the choice between organizations and markets is less often noted. Nor is it often noted that many kinds of uncertainties other than price uncertainties may make coordination through organizational procedures advantageous.

The difficulty that economics has had in giving a good account of organizations and their predominance is traceable in no small part to the fascination of economists with systems in equilibrium. Analysis under assumptions of perfect knowledge and certain expectations has little relevance, surely, for such issues of economic organization as explaining how an economy is structured between organizations and markets. Prices provide only one of the mechanisms for coordination of behavior, either between organizations or within them. Coordination by adjustment of quantities is probably a far more important mechanism from a day-to-day standpoint, and in many circumstances will do a better job of allocation than coordination by prices. For example, inventory control systems record the amounts of inputs for the organization's activities, and place orders when quantities fall below specified levels. The orders, recorded by the control systems of suppliers, initiate the scheduling of new production and are used to adjust aggregate production levels as well.

From a conceptual standpoint, it is entirely feasible to construct economies in which prices are based on costs and final demands are limited wholly by budget constraints, with demand vectors that are otherwise insensitive to prices. Quantities of goods sold and inventories, not prices, provide the information for coordinating these systems. The Leontief input-output models, with exogenous vectors of final demands, are examples, and the Hawkins-Simon theorem (1949) states the conditions under which such systems have non-negative solutions. They possess the same information-conserving virtues as price-regulated systems (von Hayek, 1945). Each actor need only know his or her own business.

Many observers of business scheduling and pricing practices have claimed that (with the possible exception of the agricultural and mining sectors) models that use quantities as signals approximate first-world national economies more closely than do models in which prices are the principal mechanisms of coordination. I don't wish to argue that point here: but simply observe that quantity adjustments play a very large role in the real world in equilibrating the operations of different organizations and different parts of organizations.

The stylized market exchanges of neoclassical economic theory generally involve only prices and quantities, which is the foundation for their parsimony

in information. But actual contracts negotiated between business firms—putting consumer products aside, for the moment—usually specify far more than prices and quantities. Contracts for construction of a building or of a product of engineering (like a generator or an airplane) specify in enormous detail the specifics of the product to be delivered. They require a massive exchange of information in both negotiation and execution. The red market traces that our Martian visitors observed from space are not narrow tracks along which only money and goods flow, but broad highways to accommodate a vast flow of detailed information as well.

Thus, the assertion that markets permit each firm to do its business with little knowledge of its partners is a fiction. In construction, in heavy industry, in manufacturing involving high technology, and in other areas, contracting partners carry on communication at a level comparable to the levels observed between departments of a firm. When products are manufactured to specifications, a great deal of information must flow among the various groups of people involved in the manufacture. But the widespread use of subcontracting in the automobile and construction industries, just to mention two, demonstrates that it is often quite feasible to transport this information across organizational boundaries, so that vertical integration is unnecessary. From this perspective, the distinction between market communications and internal communications, and the criteria for choosing between the two alternative arrangements, become correspondingly vague.

The choice between prices or quantities to coordinate the activity levels of different organizations or parts of organizations does not by itself dictate the respective roles of organizations and markets. Prices may be used to coordinate the activities of different parts of single organizations, provided that some way can be found to determine what the market prices should be, and quantity adjustments can be made between different organizations as well as within them.

There is one important difference in the operation of coordination mechanisms within and between organizations. Coordination between organizations depends almost wholly on economic motivations and rewards, and becomes seriously imperfect wherever major externalities are present that cannot be removed by enforceable contract arrangements. Within organizations, on the other hand, identification is a powerful force for combatting externalities produced by attachment to subgoals, by virtue of the loyalty it can produce to the goals of the whole system. A department will be less likely to skimp on quality to cut costs if its members identify with the final product. In particular, identification becomes an important means for removing or reducing those inefficiencies that are labeled by the terms “moral hazard” and “opportunism.”

These observations nudge us toward the conclusion that organization size and degree of integration, and the boundaries between organizations and markets, are determined by rather subtle forces. The wide range of organizational arrangements observable in the world suggests that the equilibrium between these two alternatives may often be almost neutral, with the level

highly contingent on a system's history. A traditional arrangement may be preserved until its inefficiencies become overwhelming—or even beyond. The same conclusion is suggested by the constant flux of mergers and spinoffs in the business world, many of these transformations being governed by considerations quite unrelated to productive or allocative efficiency, and many having consequences for efficiency that even those involved in them cannot evaluate.

Over a span of years, a large fraction of all economic activity has been gathered within the walls of large and steadily growing organizations. The green areas observed by our Martian have grown steadily. Ijiri and I have suggested that the growth of organizations may have only a little to do with efficiency (especially since, in most large-scale enterprises, economies and diseconomies of scale are quite small), but may be produced mainly by simple stochastic growth mechanisms (Ijiri and Simon, 1977).

But if particular coordination mechanisms do not determine exactly where the boundaries between organizations and markets will lie, the existence and effectiveness of large organizations does depend on some adequate set of powerful coordinating mechanisms being available. These means of coordination in organizations, taken in combination with the motivational mechanisms discussed earlier, create possibilities for enhancing productivity and efficiency through the division of labor and specialization.

In general, as specialization of tasks proceeds, the interdependency of the specialized parts increases. Hence a structure with effective mechanisms for coordination can carry specialization further than a structure lacking these mechanisms. It has sometimes been argued that specialization of work in modern industry proceeded quite independently of the rise of the factory system. This may have been true of the early phases of the industrial revolution, but would be hard to sustain in relation to contemporary factories. With the combination of authority relations, their motivational foundations, a repertory of coordinative mechanisms, and the division of labor, we arrive at the large hierarchical organizations that are so characteristic of modern life.

Conclusions

The economies of modern industrialized society can more appropriately be labeled organizational economies than market economies. Thus, even market-driven capitalist economies need a theory of organizations as much as they need a theory of markets. The attempts of the new institutional economics to explain organizational behavior solely in terms of agency, asymmetric information, transaction costs, opportunism, and other concepts drawn from neo-classical economics ignore key organizational mechanisms like authority, identification, and coordination, and hence are seriously incomplete.

The theory presented here is simple and coherent, resting on only a few mechanisms that are causally linked. Better yet, it agrees with empirical obser-

vations of organizational phenomena. Large organizations, especially governmental ones, are often caricatured as “bureaucracies,” but they are often highly effective systems, despite the fact that the profit motive can penetrate these vast structures only by indirect means.

This theory of organizations calls for reexamining some of the classical questions of political economy. The primacy of profit as the enforcer of organizational efficiency is replaced by organizational goals, combined with organizational identifications and with material rewards and supervision, all of which motivate employees to work toward these goals. This framework makes it necessary to reopen the question of when profit-making, nonprofit, and governmental organizations should be expected to operate well, and when market competition is needed to discipline organizations to perform efficiently.

The reopening of these questions is important for both capitalist and socialist economies. On the one side, capitalist economies are actually mixed economies, faced with a multitude of problems of regulation and deregulation, of socialization and privatization. On the other side, many socialist economies have had mediocre success in maintaining the efficiency of their organizations, and are experimenting with the reintroduction of markets, often while trying to avoid extensive privatization. Good answers to the policy questions that face all industrialized societies depend on having empirically sound theories of the behavior of large organizations. Such theories cannot be developed from the armchair. They call for fact-gathering that will carry researchers deep into the green areas, the organizations, that dominate the terrain of our economic systems.

References

Berle, A. A., and Medans, G. C., *The Modern Corporation and Private Property*. New York: Macmillan, 1933.

Cyert, R. M., and March, J. G., *A Behavioral Theory of the Firm*. Englewood Cliffs: Prentice-Hall, 1963.

Dearborn, D. C., and Simon, H. A., “Selective Perception: the Identifications of Executives,” *Sociometry*, 1958, 21. Reprinted in Simon, *Administrative Behavior*, chapter 15.

Demsetz, H., and Lehn, K., “The Structure of Corporate Ownership: Causes and Consequences,” *Journal of Political Economy*, 1985, 93, 1155–1177.

Devons, E., *Planning in Practice*. Cambridge: Cambridge University Press, 1950.

Hawkins, D., and Simon, H. A., “Note: Some Conditions of Macroeconomic Stability,” *Econometrica*, 1949, 17, 245–248.

von Hayek, F. A., “The Use of Knowledge in Society,” *American Economic Review*, 1945, 35, 519–530.

Ijiri, Y., and Simon, H. A., *Skew Distributions and the Sizes of Business Firms*. Amsterdam: North Holland, 1977.

Kornai, J., *Anti-Equilibrium*. Amsterdam: North-Holland, 1971.

Krafcik, J., and Womack, J. P., “Comparative Manufacturing Practice: Imbalances and Implications,” Appendix A., working paper, International Motor Vehicle Program, MIT, May 1987.

March, J. G., and Simon, H. A., *Organizations*. New York: Wiley, 1958.

Ross, S., "The Economic Theory of Agency: the Principal's Problem," *American Economic Review*, 1973, 63, 134-139.

Simon, H. A., *Administrative Behavior*. New York: Macmillan, 1947; 3rd edition, 1976.

Simon, H. A., "Rational Decision Making in Business Organizations," *American Economic Review*, 1979, 69, 493-513.

Simon, H. A., "A Formal Theory of the Employment Relationship," *Econometrica*, 1951, 19. Reprinted in Simon, H. A., *Models of Bounded Rationality*, Vol. II, Chapter 5.2. Cambridge: MIT Press, 1982.

Simon, H. A., *Reason in Human Affairs*. Stanford: Stanford University Press, 1983.

Simon, H. A., "A Mechanism for Social Se-

lection and Successful Altruism," *Science*, 1990, 250.

Stiglitz, J. E., "Incentives and Risk-Sharing in Sharecropping," *Review of Economic Studies*, 1974, 41, 219-255.

Weisbrod, B. A., *The Nonprofit Economy*. Cambridge: Harvard University Press, 1988.

Weisbrod, B. A., "Rewarding Performance that is Hard to Measure: The Private Non-profit Sector," *Science*, May 5, 1989, 244.

Williamson, O. E., *Markets and Hierarchies*. New York: The Free Press, 1975.

Williamson, O. E., *The Economic Institutions of Capitalism*. New York: The Free Press, 1985.

Winter, S., "Economic 'Natural Selection' and the Theory of the Firm," *Yale Economic Essays*, 1964, 4.