4

Skills

THE PRECEDING CHAPTER explored the triad of ideas that underlie orthodox explanations of why firms do what they do—objectives, choice sets, behavior as maximizing choice. This chapter begins the task of developing the basic postulates about behavior in evolutionhry theory. Although our theory is concerned with the behavior of business firms and other organizations, we find it useful to begin the analysis with a discussion of some aspects of individual behavior. An obvious reason for doing so is that the behavior of an organization is, in a limited but important sense, reducible to the behavior of the individuals who are members of that organization. Regularities of individual behavior must therefore be expected to have consequences, if not counterparts, at the organizational level. More directly relevant to our development here is the value of individual behavior as a *metaphor* for organizational behavior: the idea that "individuals are complex organizations too" has considerable power. And the indirect approach to organizational behavior, by way of this metaphor, has the advantage that the discussion can be based to a large extent on the empirical data of everyday observation and introspection.

Because our real concern is with organizations, we make no attempt to be balanced and comprehensive in our discussion of individual behavior. Rather, we highlight those aspects of the subject that provide, in our view, the most helpful introduction and truest guide to phenomena at the organizational level. Even in the pursuit of that objective, we depart somewhat from a balanced appraisal in

the direction of attempting to compensat'e for the biases of the orthodox treatment of the subject. Our attention is drawn to example situations that tend to reveal the inadequacies of orthodox conceptual categories at the same time as they illustrate the relevance of the categories we propose. We neglect the areas where the orthodox view is informative and fruitful; were we to consider those areas in detail, we would argue that the evolutionary scheme subsumes the orthodox one and delineates its proper uses.

Specifically, the focus of this chapter is on the skilled behavior of individuals. We propose that individual skills are the analogue of organizational routines, and that an understanding of the role that routinization plays in organizational functioning is therefore obtainable by considering the role of skills in individual functioning. We do not, of course, suggest that the concept of skill is the ***unique*** key to individual behavior, but it is a very important key. Routinization is relatively more important as a feature of organizational behavior than skill is as a feature of individual behavior, but it is still less than the whole story. In both realms, close examination of the nature of skillful/routinized behavior brings to light the shortcomings of optimization notions as an approach to understanding the basis of the effective functioning of an individual/organization in an environment.

By a "skill" we mean a capability for a smooth sequence of coordinated behavior that is ordinarily effective relative to its objectives, given the context in which it normally occurs. Thus, the ability to serve a tennis ball well is a skill, as is the ability to engage in competent carpentry, drive a car, operate a computer, set up and solve a linear programming model, or judge which job candidate to hire. The first few of these skills might be regarded by orthodox theory as capabilities in a choice set; the last few are intimately involved with the act of choosing. We emphasize that these skills have many characteristics in common, regardless of whether we think of them as capabilities or choice behavior.

In the first place skills are programmatic, in that they involve a sequence of steps with each successive step triggered by and following closely on the completion of the preceding one. Second, the knowledge that underlies a skillful performanceis in large measure tacit knowledge, in the sense that the performer is not fully aware of the details of the performance and finds it difficult or impossible to articulate a full account of those details. Third, the exercise of a skill often involves the making of numerous "choices"—but to a considerable extent the options are selected automatically and without awareness that a choice is being made.

These three aspects of skilled behavior are closely interrelated. If,

for example, it were *not* the case that behavior options are selected "automatically" in the course of the exercise of a skill, then the performance as a whole would not have the quality of being a connected, unitary "program." And the difficulty of articulating the basis for such automatic choices forms an important part of the total problem of explaining how the performance is accomplished. Nevertheless, the three aspects are conceptually distinguishable, have been emphasized in different degrees by different authors in the past, and play somewhat different roles in our own account of individual and organizational behavior. We therefore discuss them separately.

1. SKILLS AS PROGRAMS

A variety of terms have been used in the literature of social science to denote a smooth sequence of behavior that functions, in some sense, as an effective unit. "Skill" is obviously one such; there is, in particular, a substantial psychological literature relating to skills and skill learning. The terms "plan," "script," "habit," "routine," and "program'' have also been used to name either the same concept or a very closely related one. But there are obvious differences in connotation among these terms, and exploration of these various connotations can be informative.

To think of skills as programs is to evoke the image of a computer program. Clearly, the development of the modern electronic computer and its associated software has had an important and widely diffused influence on theoretical thinking about the phenomena that concern us here.[1] Computer programs that simulate complex, patterned behaviors have been developed over a wide range of human and organizational activity. These efforts have shown, above all, that the logical processes of a digital computer can mimic very "skillful" and "intelligent" behaviors, at least in the sense of providing a sufficient account of numerous observable aspects of such behavior. Here, however, we will not review specific examples of this sort of research, but will consider only the broad parallels between skills and (computer) programs.

The following features of computer programs are analogous to, and instructive regarding, corresponding features of human skills. First, a program functions as a unit, and its execution is ordinarily a

1. For discussions of the influence of cybernetic theory and computer modeling on psychology, see Miller, Galanter, and Pribam (1960, ch. **3)** and Newell and Simon (1972, historical addendum, esp. pp. 878–882).

highly complex performance relative to the actions required to initiate the performance. Second, although loops and "go to" statements and conditional branching statements complicate the picture, the basic organization scheme of a program is serial. There is a beginning and an end (or at least there is supposed to be an end). Also, resumption following an unplanned interruption of program execution is often problematic, and it is easier to start over from the beginning than it is to complete the partial performance. Third, considering that it is performed by an automaton, it is clear that the execution of a computer program is literally "automatic." Finally, the speed and accuracy with which an appropriately programmed computer accomplishes its task are often considered impressive. One standard of "impressiveness" may be human performance on the same task, but perhaps a more useful standard from the point of view of the informativeness of the analogy would be the performance that could be achieved using the computer but not the program—that is, by directly commanding each individual step.

The points about skills implied in the above statements about programs are largely self-evident, but some brief elaboration may be useful. As regards "functioning as a unit," it may be noted that, for both programs and skills, there are recognizable "units" at various levels of organization. Larger units are organized complexes of smaller ones, in which the latter may nevertheless retain some individuality. Thus, for even a moderately proficient touch-typist, the typing of words like "the," "and," "here," "in," and "as" is executable at a stroke, while "sincerely yours" is both a unit and a two-unit complex. Probably very few typists have fingers for which "antidisestablishmentarianism" is a familiar rhythm; nevertheless, a skilled typist will break that word into familiar units and thereby execute it much more quickly than a novice can. Typing skill also serves to illustrate the point about serial organization—essentially, that the order in which component units of a skill are executed is a significant fact about the structure of the skill itself. A typist who can rattle off "through" without a thought is likely to have to slow down and pay attention to type "hguorht," or even "ughthro."

Skilled human performance is automatic in the sense that most of the details are executed without conscious volition. Indeed, a welcome precursor of success in an effort to acquire a new skill is the diminishing need to attend to the details. And it is a familiar fact that attempting to attend to the details often has a disruptive effect: in many competitive situations in athletics, the arts, and other spheres, success depends importantly on the ability of the performer to "stay loose" and "not clutch''—that is, to resist the pressures that might cause destructive attention to intrude into the details of the per-

formance.[2] It is not uncommon for a performer who is particularly noted for this ability to be compared, approvingly, to a computer or other machine.

Although "impressiveness" is obviously a matter of degree and relative to expectation, only the most phlegmatic can escape being impressed, at some point, by a skillful performance. Indeed, "world class" performances in a variety of intellectual, artistic, and athletic pursuits often fall in the range of the "awesome" rather than that of the merely impressive. In such cases, of course, one is led to speculate about the role that the basic mental and physical equipment of the performer plays in high skill. For this reason, it is perhaps more relevant to our concerns to consider the reaction of the novice to the moderately skilled tennis player, skier, pianist, or solver of differential equations. At least for an observer unjaded by exposure to superstars, performances made possible by a few years of lessons and regular practice are often highly impressive—and depressing, because illustrative of a goal that seems unattainable. This gap between a skilled performer and a novice with the same "basic equipment" is the analogue of the difference between having the computer and also the right program for the task, and having the computer only.

2. SKILLS AND TACIT KNOWING

The late scientist-philosopher Michael Polanyi wrote extensively of the central place in the general scheme of human knowledge occupied by knowledge that cannot be articulated—tacit knowledge. On the simple observation "We know more than we can tell," Polanyi built an entire philosophical system (Polanyi, 1967, p. 4). Though the full import of "tacit knowing" in Polanyi's philosophy can only be hinted at by examples of what would ordinarily be called "skills," such examples do provide familiar and compelling illustrations of phenomena of broad significance. In fact, in Polanyi's *Personal Knowledge* (1962), the discussion of skills (ch. 4) plays a role analogous to our own discussion here. It provides a useful perspective on other realms of knowledge—in his case, that of scientific knowledge; in ours, that of organizational capability.

To be able to do something, and at the same time be unable to explain how it is done, is more than a logical possibility—it is a common situation. Polanyi offers a good example early in his discus-

2. Of course, the skilled performer must also avoid the opposite error of being too relaxed and "losing his concentration." But the concentration required is on the objective of the performance at each moment, not on the details of the procedure.

sion of skills: "I shall take as my clue for this investigation the well-known fact *that the aim of a skillful performance is achieved by the observance of a set of rules which are not known as such to the person following them*. For example, the decisive factor by which the swimmer keeps himself afloat is the manner by which he regulates his respiration; he keeps his bouyancy at an increased level by refraining from emptying his lungs when breathing out and by inflating them more than usual when breathing in; yet this is not generally known to swimmers" (Polanyi, 1962, p. 49).

The difficulty of explaining the basis of a skilled performance comes to the fore in the teaching or learning of skills. Polanyi's swimming example suggests that in some cases the difficulty may arise from the fact that the "instructor" is quite unaware of the key principles, and that he actually serves less to instruct than to detect and reward randomly occurring improvements in performance. In other cases, the instructor may be able, or at least be subjectively confident that he is able, to explain the matter in detail. But the detailed instruction offered typically consists of a list of subskills to be executed in sequence, and the instructions neither convey the ability to perform the subskills with requisite efficiency nor assure the smooth integration of those subskills into the main skill. This point is emphasized by Miller, Galanter, and Pribam, commenting on a description of how to land an airplane: "When skillfully elaborated and executed it will serve to get pilot and craft safely back to earth. It is a short paragraph and could be memorized in a few minutes, but it is doubtful whether the person who memorized it could land a plane, even under ideal weather conditions. In fact, it seems likely that someone could learn all the individual acts that are required in order to execute the Plan, and still be unable to land successfully. The separate motions, the separate parts of the Plan, must be fused together to form a skilled performance. Given the description of what he is supposed to do, the student still faces the major task of learning how to do it" (Miller, Galanter, and Pribam, 1960, pp. 82-83).

Instruction in a skill typically consists in large part of the imposition of a discipline of practice, a portion of which is supervised by the instructor. Verbal instruction is included, but is predominantly in the form of critique of practice. Illustration by the instructor and (attempted) imitation by the learner is often employed as an alternative mode to verbal instruction and critique. As Miller et al. indicate, verbal instruction by itself—the information in the "how-to-do-it" book—provides only a starting point at best for the acquisition of the skill. Possession of such a book—the articulable portion of the knowledge involved—may be indicative of ambition to learn, but it certainly does not certify possession of the skill.

The limitations of verbal instruction are even more apparent when the learner is attempting to reacquire a skill that has become rusty. Only in extreme cases does the how-to-do-it book prove useful in the reacquisition of a rusty skill. The remnant of the skill itself, lying latent in the brain, is typically more helpful as a restarting point than any collection of more words could be. What is needed is renewed practice and constructive criticism, not the beginner's handbook.

These propositions do not relate only to psychomotor skills. With minor modification, they extend to the realm of specific cognitive skills such as facility in mathematical manipulation of a particular type, the ability to solve the theoretical exercises characteristic of a certain area and method of scientific inquiry, or the ability to generate good solutions to complex production scheduling problems. The manipulation of equations in elementary algebra will serve as an example. Clearly, the axioms of the real number system together with a relatively short list of problem-solving heuristics (like "isolate the unknown") do constitute, in a sense, an articulated account of the skill involved. Equally clearly, the skilled manipulator in action has little or no conscious awareness of this articulated characterization of his activity. He does not think "distributive law—rearrange terms—factor out X" and so on, but simply "perceives" productive transformations of the expression and carries them out, often making several transformations at once in the course of rewriting the expression. There is, in Polanyi's terms, only "subsidiary awareness" of the rules being employed, whereas there is "focal awareness" of the expression manipulated.

It seems clear that the "tacitness" of a skill, or rather of the knowledge underlying a skill, is a matter of degree. Words are probably a more effective vehicle for communicating the skills of elementary algebra than for those of carpentry, and more effective for carpentry than for gymnastic stunts. Also, a trait that distinguishes a good instructor is the ability to discover introspectively, and then articulate for the student, much of the knowledge that ordinarily remains tacit. The same knowledge, apparently, is more tacit for some people than for others. Incentives, too, clearly matter: when circumstances place a great premium on effective articulation, remarkable things can sometimes be accomplished. For example, it has been established in occasional emergency situations that it is not impossible to convey by radioed verbal commands enough information on how to fly a small plane so that a person who lacks a pilot's skills can bring the plane in for a landing.[3]

3. This observation runs somewhat contrary to the statement of Miller, Galanter, and Pribam quoted above. But it is clear that a pilot who entirely lacks tacit knowledge of how to land is a pilot with whom one would prefer not to ride.

As we observed previously, a variety of terms have been used in the social science literature to refer to concepts closely related to "skill." It is interesting and somewhat curious that the array of terms employed in this connection includes several whose connotations are to a degree *adverse* to tacitness. The above passage from Miller, Galanter, and Pribam is indicative of the fact that their notion of a "Plan" is intimately related to the usual idea of a skill, and also to the idea that words may not suffice to communicate a plan. Yet the word itself, in ordinary usage, usually refers to something that is articulable and capable of being represented symbolically. A similar observation holds for "program," a term favored by March and Simon, among others.

Schank and Abelson employ the term "script" to refer to "a structure that describes appropriate sequences of events in a particular context, . . . a predetermined, stereotyped sequence of actions that defines a well-known situation" (Schank and Abelson, 1977, p. 41). As with "plan," the connotations of "script" clearly favor the notion that the knowledge involved can be articulated. Nevertheless, scripts turn out to resemble skills rather closely, as the reference to "stereotyped sequence" suggests. To the extent that there is a distinction, the key to it lies in the fact that Schank and Abelson are concerned above all with the process by which natural language is understood. This concern entails a focus upon the successful use of language: the inquiry relates to how this is accomplished in human beings and how it might be accomplished by a computer. A vast realm of tacit knowledge is nevertheless implied by the computer programs that Schank and Abelson devise to represent the processes of understanding. They are well aware of, but do not focus upon, the fact that these programs imply a great deal of information processing that is not part of the conscious activity of a human being who is trying to understand. Indeed, were it *not* the case that the inferential processes they attempt to model are imperfectly accessible to conscious thought, the modeling task would be trivial and unworthy of the attention they bestow upon it. Thus, it seems that their approach to understanding of language does parallel Polanyi's characterization of skill as involving "the observance of a set of rules which are not known as such to the person following them."

In an important sense, the researcher whois attempting to build a computer model of human psychological processes is in a position analogous to that of a student attempting to learn a skill from an instructor. Both are betting that language can serve to communicate useful guidance to the underlying structure and details of a complex performance: the student seeks such guidance from his instructor and the researcher seeks it from his subject or, introspectively, from himself. Both would like to know how the thing is really done, the

student for the sake of being able to do it and the researcher for the sake of being able to explain how it is done. Both are aware that, to the extent they experience difficulty in achieving their goals, language is an imperfect tool for conveying the information they need. Language can communicate a framework, but a great deal of filling-in remains to be done after the resources of language are exhausted; much of the filling-in involves laborious trial-and-error search. Perhaps both the student and the researcher tend to suffer from ambivalence regarding the limitations of language. Both hope that words will smooth their individual paths to achievement; both know that there is no distinction in the achievement if the path is too smooth.

For many reasons, it is important to try to identify the determinants of the "degree of tacitness"—that is, the considerations that make tacit knowledge a more important part of the picture in some cases than in others. As a preliminary step in this direction, we will consider here the sources from which the limits on the articulation of knowledge derive. Such limits seem to arise in three distinguishable ways.

There is, first of all, a limit imposed by the feasible time rate of information transfer through symbolic communication, which may be well below the rate necessary or appropriate in the actual performance. In the case of serving a tennis ball or performing a gymnastic stunt, the law of gravity imposes a tight constraint on the rate at which critical portions of the maneuver are performed. Thus, although step-by-step description is possible, and pretrial instruction and posttrial criticism are both helpful, it is not realistic to offer detailed instruction *during* an attempt. And although the learner can attempt to store pretrial instruction in memory and consciously retrieve it as the action is performed, the effectiveness of this tactic is severely limited by the speed and simultaneity of the information processing required. Ultimately, therefore, the learner has to work out the details of the coordination problem for himself. His knowledge of those details remains tacit, is recollected without conscious awareness, and is probably no more susceptible to articulation than his instructor's corresponding knowledge was.

Time-rate considerations also figure, though in a somewhat different way, in learning touch typing or piano playing. In these cases, it is at least possible to enhance the role of articulation and of conscious awareness by slowing the time rate of the performance, and this fact is commonly exploited in learning. Nevertheless, the details of an *accomplished* performance are tacit: it is not the case that one can learn to perform the task on the "slow" setting and then simply push the "fast" button to produce an expert performance.

A second consideration that limits the articulation of the knowl-

edge underlying a skill is the limited *causal depth* of the knowledge. Polanyi's swimming example illustrates the point that possession of a skill does not require theoretical understanding of the basis of the skill. In fact, it seems quite clear for all psychomotor skills that the actual mode of storage of the knowledge in the nervous system makes no use of the terms in which physicists, physiologists, and psychologists would describe the skilled performance. Yet this does not imply that an attempt to *articulate* the basis of the skill would not benefit from the availability of this terminology. Perhaps some novice swimmers would be helped by Polanyi's brief explanation of the body's buoyancy. More generally, we may note that a skilled performance takes place in a context defined by the values of a wide range of variables relevant to the performance; these may include aspects of the performer's physical state, as well as conditions of air pressure and lighting, gravitational forces, and so forth. The performer need not be aware of the existence of all of these variables, let alone of their relevance to the performance. This means that the performer simply relies upon these variables being in acceptable ranges, and is in no position to describe what it is that he relies upon. Should the values of some of the variables change so that the constraints are violated, the limited causal depth of the knowledge involved will impede or prevent effective adjustment to the change.

The third aspect of the limitation of articulation is the *coherence* aspect—that of the whole versus the parts. Efforts to articulate "complete" knowledge of something by exhaustive attention to details and thorough discussion of preconditions succeed only in producing an incoherent message. This difficulty is probably rooted to a substantial extent in the related facts of the linear character of language-based communication, the serial character of the "central processor" of the human brain, and the relatively limited capacity of human short-term memory. Given these facts, the possibilities of articulating both the details and the coherent patterns they form—the relationships among the details—are necessarily limited. At a given point in a text, a passage is encountered in a context established by nearby passages; to convey the fact that it is also meaningfully connected to other parts of the text requires more words, and places demands on the reader's memory. Similarly, it is difficult to form coherent three-dimensional mental images from exposures to a number of two-dimensional cross-sections of an object. To cope with these limitations of human powers of articulation and symbolic information processing, a variety of aids are employed that present information about patterns and structures directly to the eyes—aids such as photographs, diagrams, graphs, flowcharts, and holograms. There is a rapidly advancing technology of such aids.

In short, much operational knowledge remains tacit because it

cannot be articulated fast enough, because it is impossible to articulate all that is necessary to a successful performance, and because language cannot simultaneously serve to describe relationships and characterize the things related. This observation provides us with at least a starting point for assessing the relative significance of tacit knowledge in different situations. The knowledge contained in the how-to-do-it book and its various supplements and analogues tends to be more adequate when the pace of the required performance is slow and pace variations are tolerable, where a standardized, controlled context for the performance is somehow assured, and where the performance as a whole is truly reducible to a set of simple parts that relate to one another only in very simple ways. To the extent that these conditions do not hold, the role of tacit knowledge in the performance may be expected to be large.

Finally, it should be emphasized that costs matter. Whether a particular bit of knowledge is *in principle* articulable or necessarily tacit is not the relevant question in most behavioral situations. Rather, the question is whether the costs associated with the obstacles to articulation are sufficiently high so that the knowledge *in fact* remains tacit.

3. SKILLS AND CHOICES

While the exercise of a skill involves the selection of behavior options, the selection process is highly automatic. This raises the question of whether it is at all appropriate to discuss this process in terms of "choice." In the terminology of the previous chapter, the sort of choice that takes place in the process of exercising a skill is choice without deliberation. To the extent that the conceptual baggage carried by the term "choice" includes a lot of things that are associated with deliberation, it may be quite misleading when applied to the automatic choices involved in skills. As we noted, orthodox theoretical discussion is inconsistent and ambiguous on whether choice involves deliberation, but it is quite clear in maintaining that there is a sharp distinction between capability and choice behavior. The two issues are obviously related: the choice among behavior options that takes place in the exercise of a skill typically involves no deliberation *and* it is a constituent of the capability that the skill represents. These issues are deep and important ones.

From one point of view, *all* of the coordinated sequential behavior involved in the exercise of a skill is chosen behavior. A large range of available alternative behaviors is continually being rejected in favor of the behavior sequence called for in the program. When a driver

makes the small adjustments of the steering wheel required to keep his car on an approximately straight path down the road, he "chooses" not to let the car drift off the road, and also "chooses" not to turn the wheel abruptly and throw the car into a skid. When he decelerates as he catches up to a car in front of him, he "chooses" not to maintain his speed and crash into the rear of that car.

However, any experienced driver can attest on the basis of introspection that these and many other micro-units of driving skill are normally selected and performed entirely without attention or awareness. The conscious mind may be devoted to looking for a street sign, planning the day's activity, or carrying on a conversation while these "choices" are being made. That this phenomenon of programmed choice is of the essence of driving skill becomes apparent when the contrasting case of the student driver is considered: it is the novice who really *chooses* not to drive off the edge of the road—if "really choosing" means "paying attention to what is desired and deliberately acting to accomplish what is desired." The skilled driver does not (deliberately) choose to keep the vehicle on the road, but merely accomplishes this result incidental to a choice to exercise his driving skill for the purpose of getting from one place to another.

In general, choice plays a larger role in the selection of large units of behavior than of small ones. The action of directing the car onto the northbound on-ramp of a freeway is more likely to involve choice than the multitude of shallow turns involved in negotiating a straight stretch of road. But this generalization must be qualified very significantly by reference to the frequency with which the unit of behavior occurs. For example, if the turn onto the northbound on-ramp is part of the regular commuting trip to work, it may have a degree of automaticity approaching that involved in the microskills of control of the car. Such automaticity reflects, of course, the fact that the turn onto the ramp is but a component in the macroskill "driving to work"; it is accomplished in a "programmed" way in its normal place in that larger sequence of behavior.

The picture is further complicated by the fact that particular units of behavior, of whatever scale, are not assigned permanently and uniquely to the categories "chosen" and "automatic." Rather, circumstances affecting the immediate goals and attention allocation of the performer are an important determinant of whether a particular unit is run off automatically, or as a result of deliberate choice. A driver's selection of the speed of his vehicle may be a choice made in response to posted limits, with conscious reflection on the probabilities of speed traps and on the costs and benefits of alternative times of arrival at his destination. But speed is also subject to automatic adjustment in response to traffic density, driving conditions, and other

influences. The driver may choose to pay attention to his speed—that is, he may choose to choose his speed—but he may also let speed selection occur automatically, just as he keeps the car on the road automatically. An important possibility, especially for a driver who has recently had a speeding ticket, is that he may choose to try to choose his speed *and fail:* his automatic responses may take over in spite of his intentions. Similarly, to revert to our previous example, a driver may find himself going up the on-ramp "on the way to work" when it is actually Saturday morning and he had intended to go to the hardware store.

There are corresponding points to be made about the relation of a skilled performance to its preconditions. We noted above that such a performance takes place in a context set by the values of a large number of variables; the effectiveness of the performance depends on those variables being in appropriate ranges. The performer typically relies, without conscious thought, on the constraints being satisfied. In some cases, and certainly when the existence of the constraints is unknown to the performer, there may be no practical alternative to such unconsidered reliance. In other cases, the performer may have occasion to worry about possible difficulties and perhaps be led to consider adjustments in the performance, or to forgo it altogether. For example, a driver normally relies on the effective functioning of the braking system, but worries about brake failure may sometimes receive conscious attention and there may then be a choice between normal reliance and doing something about the possible problem. As in the case of selection of behavior options, contingencies of intention and attention will determine where, in the enormous range of preconditions that might conceivably fail, occasional worries rise to consciousness.

We may now take stock of the relations of skills and choice. The picture is complex, but in general it seems to contrast sharply with the emphasis that orthodoxy gives to choice in the explanation of behavior, and also with its insistence on a strict conceptual distinction between capability and choice. Skills are deep channels in which behavior normally runs smoothly and effectively. It is far from the case that behavior must take a unique course, but the reconciliation of smoothness and effectiveness with the availability of numerous options is accomplished by making option selection largely automatic. Skillful acts of selection from the available options are constituents of the main skill itself: they are "choices" embedded in a capability.[4] Deliberate choice plays a narrowly circumscribed role,

4. March and Simon (1958, pp. 26, 141–142) and Schank and Abelson (1977, pp. 42–57) are explicit on the point that the entities they respectively call "programs" and

limited under normal circumstances to the selection of the large-scale behavior sequence to be initiated. This suppression of choice is certainly associated with, and is probably a *condition* for, the smoothness and effectiveness that skilled behavior confers. On the other hand, it is possible for choice to intrude into the skilled performance. Option selections that are normally automatic may be made deliberately, or behavior may be diverted entirely from the deep channels of skill. The modification of skilled performance by deliberate choice greatly expands the potential diversity, flexibility, and adaptability of behavior — but always at an opportunity cost in terms of forgone uses of conscious attention, and usually at the cost of introducing some hesitation and awkwardness into an otherwise smooth flow of behavior.

Thus, there is in a sense a tradeoff between capability and deliberate choice, a tradeoff imposed ultimately by the fact that rationality is bounded. The advantages of skill are attained by suppressing deliberate choice, confining behavior to well-defined channels, and reducing option selection to just another part of the program. There are attendant risks that the thing done well may be the wrong thing, or that unnoticed contextual abnormalities may be rendering the performance ineffective or irrelevant. There are, on the other hand, advantages to being open-minded, deliberate, and wary in the choice of actions at all levels of detail — but there are attendant risks of being tardy, poorly coordinated, and unskillful in action itself.

4. THE USES OF SKILL NAMES

Skills, like computer programs, govern performances that are complex relative to the actions that are required to initiate them. The manifold coordinated details of the performance seem to take care of themselves once the decision to exercise the skill is made and a few initial steps are taken. This differential in complexity between initiation and the full performance is mirrored in the use of language to describe and discuss skills. It is, as we have emphasized, difficult or impossible to use language to characterize the "inner workings" of a skill, but words serve quite well in thinking and communicating about skills considered as units of purposive behavior. We make effective use of skill names and skill-related verbs in planning and

"scripts" do not determine ***unique*** sequences of behavior, but rather are complex entities involving numerous options, dependencies *on* environmental cues, and embedded "choices."

problem solving, and rarely reflect on the extreme complexity of the actual behaviors that these symbols represent.

If we are planning a trip from New Haven to Boston, and going by car is one of the transportation options, we consider that option with very little regard to the "overwhelming" magnitude of the information-processing task involved in driving the car—ordinarily, it suffices to assure ourselves that at least one of the potential occupants of the vehicle knows how to drive. If we are remodeling the kitchen, we may plan to hire the services of a plumber, a carpenter, and an electrician, and we care that we hire "good ones" and do not pay too much—but we do not concern ourselves with the detailed structure of these complex skills and their relationship to the particular problems posed by the kitchen plan. If we are bothered by a vision problem it is helpful to know the meaning of "ophthalmologist" and "optician," but the relevant meaning is the "what for" meaning, not the "how to" meaning that is known to the possessors of these skills.

Of course, planning and problem solving are skills in their own right. There are detailed behavioral programs for planning specific sorts of activities, and more loosely defined problem-solving skills of broader applicability. In the exercise of these cognitive skills, an important role is played by language and, in particular, by the names of other skills that may or may not be possessed by the planner or problem solver. This observation leads to an important distinction regarding the scope of the capabilities possessed by an individual —namely, the distinction between "knowing how to do *X*" and "knowing how to get X accomplished." Given an appropriate environment, and the resources and skills required for implementation of plans in that environment, an effective planner can get a lot of things accomplished that he does not personally know how to do. One does not need to be an ophthalmologist or an optician to get new glasses prescribed and made. However, even in this simple case the problem of getting the desired result accomplished may be quite difficult for a planner who does not have command of the relevant vocabulary of skill names.[5] In cases where the required vocabulary is larger and more esoteric, the planning difficulties associated with the lack of that vocabulary are correspondingly greater.

Thus, the planning vocabulary of an individual is an important determinant of the range of things that the individual can get accom-

5. At least in the opinion of many ophthalmologists, public understanding of the distinction between "ophthalmologist" and "optometrist" is sufficiently shaky so that many individuals with vision problems (as opposed to "needs for new glasses") do not receive appropriate care.

plished. That there exist people in the economy who could perform a task that one cannot perform oneself is of little help unless one knows how to locate such a person for the purpose of arranging a transaction, and such a quest is difficult to pursue effectively unless one knows or can discover the name of the skill or capability one is seeking. But vocabulary is clearly only one variable among many that affect the ability to get things accomplished, and the vocabulary variable interacts subtly with the others. We have noted that all skills are context-dependent in various ways, but the effectiveness of planning and implementation skills is particularly dependent upon detailed features of the *social* context.

For one thing, the "right" vocabulary is itself socially defined. The word that it is really important to know may be the heading under which the required capability is listed in the Yellow Pages. Or the key feature of the social context may be an organization of which the individual is a member, and the vocabulary the individual needs to command may be the specialized planning vocabulary of that organization. In a great many situations — such as getting a car repaired — the effectiveness of planning and implementation by an individual who will not ultimately do the thing himself is considerably enhanced by possession of some level of the required skill, as a complement to knowledge of the skill name. The extent to which this is the case depends on social arrangements affecting such things as the degree of standardization of services performed, the costs of verifying performance, certification arrangements, .interpersonal trust, and the definition and enforcement of contractual obligations. If the service performed is of a standardized type, if the requisite quality of performance is sharply defined and easily verified, and if the performer is clearly and effectively liable for the consequences of deficiencies in his performance, a simple market purchase of the service is likely to be a satisfactory means of implementation for a planner who knows only the name of the service he needs to buy. Where these conditions are absent and the planner is not protected by certification and trust from the possible incompetence or opportunism of the performer, he may have to concern himself with the details of the performance in an effort to assure that he gets what he needs at a reasonable price. To be useful, such concern needs to be guided by normative standards for the details — by knowledge of *how* the thing should be done.

Obviously, it would be nice if social arrangements involving standardization, certification, and so forth could be further elaborated so as to sharpen and assure the meanings of skill names. This would promote efficiency through the division of labor, by relieving planner-purchasers of the need to concern themselves with the de-

tails of the skilled performances they obtain from others. Unfortunately, skills really are complex, and there are intrinsic limits to the extent to which effective planning can be conducted by manipulating a limited vocabulary of symbols representing these complex entities, limits that are particularly stringent when the planning relates to novel circumstances. We now turn to an examination of the sources of these intrinsic limits.

5. AMBIGUITY OF SCOPE

Performance of a complex skill involves, we have remarked, the integration of a number of more elementary units of action. Often, these more elementary units constitute subskills that are optional components of the main skill, selected in response to cues in the performer's environment. Thus, the integration required is not just a matter of the relation of the subskills to one another, but also of their relation to information arising from the environment. Further, the same observations apply to the subskills: they involve integration of still more elementary units, or "subsubskills," and the integration may again involve relations with the environment as well as within the units. Continuing this descent through the hierarchical structure of the main skill, one comes ultimately to a domain of neurological and physiological considerations for which the "subskill" terminology is not really appropriate—but reducibility to still more elementary units of action remains possible.

Because skills are such complex, structured entities, and also because of the considerations that limit the articulation of the knowledge applied in a skillful performance, there is inevitably some ambiguity regarding the scope of a skill. This ambiguity has two aspects. There is, first of all, what may be termed *operational* ambiguity. It involves predictive uncertainty as to what a particular individual who possesses "the skill" can actually accomplish in an attempt to exercise that skill under particular circumstances. The second aspect is the *semantic* ambiguity of the skill name, the uncertainty regarding the denotation of the term. Operational ambiguity is obviously one source of semantic ambiguity: to be uncertain about whether a particular electrician, functioning as an electrician, will be able to bring about a desired result under particular circumstances is to be a bit vague about what it means to be an electrician. What is more important, semantic ambiguity arises in discussions that abstract from the particular possessor of the skill and the particular circumstances of its exercise. Uncertainty about what an electrician is arises in large part from the diversity of electricians and the diversity

of tasks and circumstances involved in the exercise of the skills of an electrician.

Both sorts of ambiguity are subject to reduction by deliberate effort to that end. By considering the past performances of a particular possessor of the skill, and the characteristics of the particular circumstances in question in relation to those that surrounded the past performances, it may be possible to sharpen predictions concerning the specific instance. By extending the discussion to subskills, particular tasks, and quality differentials among possessors of the skill, some of the ambiguity that surrounds the generic skill name can be eliminated. However, neither of these sorts of clarification is costless, and neither can be totally effective. Both require detailed knowledge of the skill in terms of the mix of subskills involved, the preconditions of effective performance, and so forth. To the extent that this sort of knowledge is tacit, only a person who possesses the skill himself is likely to be in a position to reduce ambiguity by the methods described. To the extent that there are preconditions for effective performance that are simply unknown, or that the tacit knowledge underlying actual performance cannot be brought to bear on the more abstract tasks of assessment and prediction, some part of the ambiguity is simply irremediable.

To amplify these points somewhat, consider again the example of the ability to drive a car. This skill is not just the ability to make the vehicle follow a desired course with acceptable accuracy, but also the ability to use a wide range of cues in the environment—other vehicles, traffic signs and lights, and so on—as the basis for determining the details of the course itself. The integration and coordination involved in the skilled performance as a whole is not merely of the sort represented in taking a curve smoothly through the coordination of pressure on the accelerator and turning of the wheel, but also the relatively automatic use of a large store of information as the basis for interpretive intermediation of sensory input and muscular response. In ordinary discussion about driving, we have little occasion to attend to the complexity of the skill and the implications of that complexity for the variability of specific driving performances across individuals and across situations. We treat the ability to drive as a dichotomous variable, assuming that the skill is possessed in satisfactory degree or not at all, and regard driver training as a process that transfers individuals from the "unskilled" category to the "skilled." This way of talking and thinking about driving skill is typically adequate and we have no need to belabor the complexities and distinctions of the matter. Occasionally, though, distinctions are confronted. If a teenage son or daughter is planning a trip with friends, we may concern ourselves with experience levels, atti-

tudes toward taking chances, and specific experience with passing on two-lane roads. We may need someone to run an errand and have only a stick-shift car available, and confront the question of whether the assembled "drivers" include anyone who can shift gears. In such cases, we drop our habitual, implicit homogenization of driving skill and—with the aid of a good many additional words—articulate the distinctions that concern us regarding subskill mixes and so forth.

Sometimes, however, highly relevant distinctions escape conscious consideration or effective articulation. Adverse effects on performance may arise from causes that do not announce themselves. The ability to control a skid on an icy road will not come in for timely consideration if it is not expected that the roads will be icy. What is not identified cannot be considered, and what is not anticipated cannot be considered in advance. But even fully identifiable and anticipated causes of performance change can resist *effective* consideration because of the tacit basis of skill. Consider the American driver who, after the overnight flight to London, confronts for the first time the problem of driving on the left, in an unfamiliar vehicle with the steering wheel on the right. It may be clear enough, in advance of the trip, that the combination of jet lag, fatigue, and unfamiliar task environment is potentially capable of producing a degradation of driving performance. It may also be clear that "being careful''—which in this case means deliberately attempting to rely less on tacitly known skill—is likely to be at least partially effective as a compensating factor. But the problem of assessing the weight of these considerations, for the purpose of deciding whether the plan is acceptable or not, is intractable because of the tacit basis of driving skill. A full conscious override of habitual response is not possible, and if it were it would mean the abandonment, not the effective adaptation, of driving skill. The planner might reflect that the problem is surely not that serious; the muscular coordination aspect of controlling the vehicle will not require much attention. On the other hand, those muscular responses are tightly linked to visual cues, and the cues do not have their accustomed import. Habitual responses will be modified and the American driver will "get the hang of it" after a while, but it is hard to say how much experience will be needed or what risk levels might be involved in acquiring it. There is thus a significant degree of ambiguity about whether an American driver, driving for the first time "on the wrong side of the road," knows how to drive or not. The ambiguity is partly a matter of uncertainty concerning the fate of the individual driver, and partly a reflection of the fact that the phrase "knows how to drive" papers over many significant distinctions.

Of course, if the American driver never goes to England, he may

never directly confront this particular illustration of the ambiguous scope of skills. If he goes with sufficient frequency, he may develop a driving-on-the-left subskill that is as much an integrated, tacitly known part of his overall driving skill as the ability to adjust to dense city traffic after coming off the ramp of a relatively uncrowded freeway. It is the differences between the environment in which a skill (and associated terminology) is developed and a relatively novel environment in which it is exercised that highlight its operational (and semantic) ambiguities. A fully static world would never pose the problem of using relatively concise language to consider the matching of complex skills with novel, complex task environments. The matches would all have been made, and could be counted on to work precisely as well in the future as they had in the past. But the real world is not static.

6. THE SKILLS OF THE BUSINESSMAN

Our primary purpose in this examination of individual skills has been to establish a useful starting point for the appraisal of the corresponding issues in the case of the large, complex organization. Much of the discussion of "theory and realism" in the economic theory of the firm has, however, been implicitly or explicitly concerned with the case of the single proprietorship. The question of whether or in what sense the business firm can be said to maximize profits has for the most part been treated in the literature as equivalent to a question about the decision-making skills of the proprietor.

The contributions of Machlup to the marginalist controversy of the forties, Friedman's methodological essay (Friedman, 1953), and Machlup's review of the issues in his presidential address of 1967 are the major papers that set forth the defense of the orthodox theory of the firm against critics who complained of its lack of realism. Although the scope and technical sophistication of orthodox theory have vastly increased during the more than three decades since the marginalist controversy, and although a number of contributions have been made to the discussions of the broader methodological issues involved, the main arguments in defense of doing economic theory in the orthodox style remain approximately where Friedman and Machlup left them. Or perhaps, indeed, there has been a retrogression — some contemporary theorists seem to operate on the basis of a methodological creed that is little more than a caricature of Friedman's sophisticated and carefully hedged position. We therefore confine our review to the classic statements.

In the course of making their methodological points about why it

is not worthwhile for economists to think concretely and in detail about a realistic account of the sources of business behavior, Friedman and Machlup managed to say or imply a great deal about what such a realistic account would be like. Much of what they said can easily be translated into and summarized in the language that we have employed in this chapter. The following attempt at such a translation reveals a high degree of convergence between their perspective and ours.

An experienced businessman acting in the pursuit of pecuniary gain is an individual exercising a complex skill. As with any such skill, the pursuit of gain is based on tacit knowledge of relevant conditions and involves at most subsidiary awareness of many of the details of the procedures being followed. The economic theorist's abstract account of business decision making is not to be confused with the businessman's skills; it serves different purposes and those purposes place a high premium on articulation. Clear articulation of his methods may be valueless, or even counterproductive, for the businessman. It is therefore quite illegitimate to seek to appraise the validity of the theoretical account of business decisions by asking businessmen whether their procedures match the theoretical constructs. Such a method founders first on the general observation that the possibilities for articulating the basis of high skill are limited; second, even if this fact were somehow of minimal importance in the specific context of business decision, there would be no reason to expect that the language chosen by the businessman to articulate his skill would be the language of economic theory. There is, after all, no reason to expect a bicyclist to be able to explain in the language of physics how he remains upright, but this does not imply that he usually falls over.

That the foregoing is a plausible encapsulation of many of the Friedman–Machlup points may be corroborated by the following specific references. In the context of his famous analogy between the businessman and the expert billiard player, Friedman remarked as follows: "The billiard player, if asked how he decides where to hit the ball, may say that he 'just figures it out' but then also rubs a rabbit's foot just to make sure; and the businessman may well say that he prices at average cost, with of course some minor deviations when the market makes it necessary. The one statement is about as helpful as the other, and neither is a relevant test of the associated hypothesis" (Friedman, 1953, p. 22**).** Even more explicitly, Machlup wrote in 1946: "Businessmen do not always 'calculate' before they make decisions, and they do not always 'decide' before they act. For they think that they know their business well enough without having to make repeated calculations; and their actions are frequently

routine. But routine is based on principles which were once considered and decided upon and have then been frequently applied with decreasing need for conscious choices. The feeling that calculations are not always necessary is usually based upon an ability to size up a situation without reducing its dimensions to definite numerical values" (Machlup, 1946, pp. 524–525). Since driving an automobile has been prominent among our own examples of the exercise of individual skill, we acknowledge Machlup's earlier treatment of the topic by quoting at some length from his well-known analogy between the theory of the maximizing firm and the "theory of overtaking":

> What sort of considerations are behind the routine decision of the driver of an automobile to overtake a truck proceeding ahead of him at slower speed? What factors influence his decision? Assume that he is faced with the alternative of either slowing down and staying behind the truck or of passing it before a car which is approaching from the opposite direction will have reached the spot. As an experienced driver he somehow takes into account (a) the speed at which the truck is going, (b) the remaining distance between himself and the truck, (c) the speed at which he is proceeding, (d) the possible acceleration of his speed, (e) the distance between him and the car approaching from the opposite direction, (f) the speed at which that car is approaching, and probably also the condition of the road (concrete or dirt, wet or dry, straight or winding, level or uphill), the degree of visibility (light or, dark, clear or foggy), and the condition of the tires and brakes of his car, and—let us hope—his own condition (fresh or tired, sober or alcoholized) permitting him to judge the enumerated factors. Clearly, the driver of the automobile will not "measure" the variables; he will not "calculate" the time needed for the vehicles to cover the estimated distances at the estimated rates of speed; and, of course, none of the "estimates" will be expressed in numerical values. Even *so*, without measurements, numerical estimates or calculations, he will in a routine way do the indicated "sizing-up" of the total situation. He will not break it down into its elements. Yet a "theory of overtaking" would have to include all these elements (and perhaps others besides) and would have to state how changes in any of the factors were likely to affect the decisions or actions of the driver. The "extreme difficulty of calculating," the fact that "it would be utterly impractical" to attempt to work out and ascertain the exact magnitudes of the variables which the theorist alleges to be significant, show merely that the *explanation* of an action must often include steps of reasoning which the acting individual himself does not consciously perform (because the action has become routine) and which perhaps he would never be *able* to perform in scientific exactness (because such exactness is not necessary in everyday life).
>
> The businessman who equates marginal net revenue productivity and marginal factor cost when he decides how many to employ need not engage in higher mathematics, geometry, or clairvoyance. Ordinarily he would not even consult with his accountant or efficiency expert in order to arrive at his decision; he would not make any tests or formal calculations; he would sim-

ply rely on his sense or his "feel" of the situation. There is nothing very exact about this sort of estimate. On the basis of hundreds of previous experiences of a similar nature the businessman would "just know," in a vague and rough way, whether or not it would pay him to hire more men. (Machlup, 1946, pp. 534–535)

It appears that it might be difficult for a disinterested judge to distinguish between the Friedman–Machlup perspective on the realities of business decision making and our own. Some greater divergence will appear as we develop our own argument further, but the paradox that has arisen here will by no means be resolved by that development alone. On the same stylized fact—"business decision making is the exercise of a skill comparable to other skills, such as driving a car or playing billiards''—Friedman and Machlup built a defense for orthodox theory and we propose to build an alternative to that theory.

What is one to make of this? At a superficial level, the paradox is easily dealt with. The disagreement is not, indeed, about the stylized fact; it is about the arguments that link the fact to conclusions about the relative merits of its interpretation in orthodox or evolutionary theory. A full analysis and comparison of these linking arguments, as between orthodoxy and evolutionary theory, would be a major task. Much of this book is concerned with it, directly or indirectly. However, merely noting that the central problem is how to model skilled behavior opens the way for a substantial clarification of the issues. Orthodoxy treats the skillful behavior of the businessman as maximizing *choice,* and "choice" carries connotations of "deliberation." We, on the other hand, emphasize the *automaticity* of skillful behavior and the suppression of choice that this involves. In skillful behavior, behavioral options are selected, but they are not deliberately chosen. This observation directs attention to the processes by which skills are learned, the preconditions for the effective exercise of skill, and the possibilities for gross error through automatic selection of the wrong option.

To identify skillful behavior with *maximizing* choice is an even larger step from the realities of skill. Skills are attributed to individuals largely on the basis of comparisons with other individuals who are less skilled or unskilled. Formal orthodox theory, on the other hand, does not rate solutions as maximizing because they are better than some other observed solutions, but because they are the best feasible solutions. It thus premises a standard of performance that is independent of the characteristics of performers; the attribution "skilled driver" involves no such premise. This observation points us toward the deeper problems involving the definition of the fea-

sible set. What are the possibilities that a skillful performance makes the best of? Are the tacit skills of the driver overtaking a truck such as to make no-passing zones unnecessary or counterproductive? Do they warrant the practice of giving American drivers licenses to drive in England without a driving test? Are we entitled to doubt — as Machlup seems momentarily to doubt — that the typical driver adequately assesses possible impairments of his own capacities?

Such questions have their parallels in the numerous policy issues that involve, in one way or another, the scope and quality of business decision making. To assess business decision making as (merely) skillful is to recognize the potential significance of a number of questions that orthodoxy tends to ignore. Are market conditions the same as they have been? Is the range of technological options the same? If conditions have changed, are businessmen aware of that? Even if conditions have not changed, have businessmen experimented enough with the available options? If the answers to such questions are in the negative, the observation that business decision making involves the exercise of skill is not entirely reassuring as *to* its likely quality. One may legitimately be concerned about problems analogous to the possibility that the American driver in England will seek to avoid the oncoming traffic by steering his car to the right.

5

Organizational Capabilities and Behavior

THIS CHAPTER PRESENTS an alternative to orthodoxy's view of organizational behavior as optimal choice from a sharply defined set of capabilities. Our view of organizational behavior has been molded by the contributions of a number of organization theorists and economists—March and Simon, Allison, Gouldner, Perrow, Doeringer and Piore, Williamson, Schumpeter, and others. What is distinctive about our treatment of organizations derives first of all from its place in our broader evolutionary framework; this accounts in particular for the attention we devote to the nature and sources of continuity in the behavioral patterns of an individual organization. Second, the analysis here builds upon that of the previous chapter and exploits the parallels between individual skills and organizational routines. Relatedly, the influence of Michael Polanyi (not usually counted as an organization theorist) is strong in this chapter, though less explicit than in the previous one.

Scope. There are a great many different sorts of organizations, and it is implausible that a given collection of concepts and propositions would apply uniformly, or even usefully, to all of them. The sorts of organizations we have in mind are, first of all, organizations that are engaged in the provision of goods and services for some outside clientele, and have at least vague criteria for doing well or poorly. The salient examples are business firms concerned with survival and profits, but much of our analysis is relevant, perhaps with minor modification, to other sorts of organizations.

Second, since "routine" is a key concept in our theoretical frame-

work, the framework applies most naturally to organizations that are engaged in the provision of goods and services that are visibly "the same" over extended periods—manufacturing hand tools, teaching second graders, and so forth—and for which well-defined routines structure a large part of organizational functioning at any particular time. As we shall argue later on in this chapter the notion of routine can usefully be stretched to relate to a number of activities that would not ordinarily be described by that term. Nevertheless, organizations that are involved in the production or management of economic change as their *principal* function—organizations such as R&D laboratories and consulting firms—do not fit neatly into the routine operation mold.

Third, the discussion relates primarily to organizations that are "large and complex." The role of this restriction is simply to maintain the focus on phenomena that are distinctively organizational. The organizations we envisage are ones that face a substantial coordination problem, typically because they have many members, performing many distinct roles, who make complementary contributions to the production of a relatively small range of goods and services. In such organizations, most of the working interactions of a large number of the members are primarily with other members rather than with the organization's environment. Also, while the organizations we describe are of the sort that have a top management that is concerned with the general direction of the organization, the scale and complexity of the organization are presumed to make it impossible for that top management to direct or observe many of the details of the organization's functioning.[1]

Terminology. The importance of the concept of organizational routine in our discussion and the parallel with individual skill have already been noted. We use "routine" in a highly flexible way, much as "program" (or, indeed, "routine") is used in discussion of computer programming. It may refer to a repetitive pattern of activity in an entire organization, to an individual skill, or, as an adjective, to the smooth uneventful effectiveness of such an organizational or individual performance. The term "organization member" is also

1 Some parts of the discussion that follows are of clearest relevance at the "establishment" level—that is, at the level of an organizational unit that has a particular geographic location. Our analysis suggests that the memory of an organization that comprises many widely separated establishments may exist mainly in the establishments, or if not it is of quite a different sort than it is in a single establishment. Significant questions relating to economic policy are involved here—for example, the question of how much difference it is likely to make to the operations of a particular plant if it is transferred as a functioning unit from one very large corporation to another. We have not pursued these questions.

used flexibly: although in most cases we use it to mean an individual, it is sometimes convenient to think of an organizational subunit as a "member" of the larger organization. Such a perspective is called for, in particular, when the information exchanges by which coordination is achieved within the subunit are quite rapid and predominantly nonsymbolic, so that the coordinating processes resist articulation in a way that parallels the case of individual skills.

In our conceptualization, an organization member is **by** definition a unit that can accomplish something on its own. A production worker, for example, may be able to put together subassembly H without interacting with other members, provided that the necessary parts are at hand, the lights are on in the work area, and so forth. He might also be able to put together subassembly K, provided likewise that the parts are at hand, and the lights are on. A typical organization member has certain skills or routines. The set of skills or routines that a particular member could perform in some appropriate environment will be called the *repertoire* of that member. Although the activities of other working members affect the local working environment of a particular member, and thereby his feasible behaviors, it is to be understood that strictly concurrent action by other members is not a precondition for his performance. Thus, in the example of the assembly operation, the state of the parts bins mediates the relationship between the member doing the assembly and the member or members who keep the bins full, but there is no requirement for concurrent action or very short-term interaction.[2]

Plan. The method and structure of our discussion parallels that followed by Schumpeter in *The Theory* of *Economic Development* (1934). We begin by considering the analogue of Schumpeter's "circular flow" at the level of the individual organization. The situation portrayed is unchanging or cyclically repetitive; it is an unrealistically quiet and static condition. We then gradually introduce into the picture more of the processes of change, displaying some of the connections between planned change and unplanned change, and examine finally the role of routine in innovation.

The first section below considers routine as organizational memory; we provide here an answer to the question raised earlier as to where organizational capabilities reside. Section *2* discusses routine

2. For the purposes of a detailed analysis of organizational coordination, it might be helpful to admit to the roster of "organization members" any feature of the total situation that constitutes an identifiable unit with a distinctive role in the total performance—including machines, parts bins, and even tables or particular areas of the floor. A complex machine, for example, may embody what amounts to tacit knowledge: the machine gets the job done, but nobody can explain how it does it.

as truce; here we recognize the divergence of interests among organization members and provide the basis for a rationale that, nonetheless, organizations can be modeled without explicit attention to the fact that many participants are involved. (We do not intend to deny here that for some purposes it is important, for some essential, to recognize the conflict of interest contained in and reflected by organizational behavior.) In Section 3, we consider routine operation as the target of efforts directed to organizational control, to replication of existing routines, and to imitation of routines employed by other organizations. We pause in Section 4 to take explicit note of some of the parallels between organizational routines and individual skills. Section 5 examines the relationship of our concept of routine operation to orthodoxy's claim that firms optimize — and to the fact that at least some firms employ explicit optimization methods to make some sorts of decisions some of the time. The penultimate section explores the connections between routinized behavior and innovative behavior — and finds much less opposition between these two ideas than is commonly thought. The concluding section summarizes the message to be carried forward to the modeling efforts of Part III

1. ROUTINE AS ORGANIZATIONAL MEMORY

It is easy enough to suggest that a plausible answer to the question "Where does the knowledge reside?" is "In the organization's memory." But where and what is the memory of an organization? We propose that the routinization of activity in an organization constitutes the most important form of storage of the organization's specific operational knowledge. Basically, we claim that organizations *remember* by doing — although there are some important qualifications and elaborations.

The idea that organizations "remember" a routine largely by exercising it is much like the idea than an individual remembers skills by exercising them. The point that remembering is achieved largely through exercise, and could not be assured totally through written records or other formal filing devices, does not deny that firms keep formal memories and that these formal memories play an important role. But there must be much more to organizational memory than formal records. Further, cost considerations make "doing" the dominant mode of information storage even in many cases where formal records could in principle be kept.

To see how exercise of a routine serves as parsimonious organizational memory, consider an organization in fully routine operation and ask what really needs to be remembered, given that such a state

has been achieved. Under such a regime, the situations of individual members and of the organization as a whole contain no significant novelties: the situations confronted replicate ones that were confronted the previous day (or week, month, or year) and are handled in the same way. The scope of the activity that actually takes place in such a static condition and the operational knowledge involved are extremely restricted. Members perform only a minute fraction of the routines they have in repertoire. Thelathe operator and the lathe turn out a few specific parts; there is an indeterminately larger number that they could (after appropriate setup and learning) produce. The operator's skills as truck driver and short-order cook are never drawn upon, and perhaps are unknown to other organization members. Routine operation of the organization as a whole certainly does not require that the lathe operator maintain his skill in cooking bacon and eggs, or in the machining of parts for products that were discontinued three years previously; neither does it require that other members remember that the lathe operator possesses or once possessed these skills. If the same state of routine operation is expected to continue indefinitely, there is no economic benefit to be anticipated from holding this sort of information in the organization's memory. (As an obvious corollary, if there is a positive cost to storing information, this sort of "irrelevant" information will tend *not* to be held in memory under the "equilibrium" condition of continuing routine operation.)

What is required for the organization to continue in routine operation is simply that all members continue to "know their jobs" as those jobs are defined by the routine. This means, first of all, that they retain in their repertoires all routines actually invoked in the given state of routine operation of the organization.

There is, however, much more to "knowing one's job" in an organization than merely having the appropriate routines in repertoire. There is also the matter of knowing what routines to perform and when to perform them. For the individual member, this entails the ability to receive and interpret a stream of incoming messages from other members and from the environment. Having received and interpreted a message, the member uses the information contained therein in the selection and performance of an appropriate routine from his own repertoire. (This may, of course, be merely a "relay message" routine, or even a "file and forget" routine.)

The class of things that count as "messages" in this characterization is large and diverse. There are, first of all, the obvious examples of written and oral communications that take overtly the form of directives to do this or that. Such directives involve the exercise of formal authority, a phenomenon that has been the focus of a great

deal of organizational literature. Then there are the written and oral communications that do not take this form but that are responded to in much the same way. For example, descriptions of what is "needed," when directed to the member whose job it is to meet that need, often function as directives. Even a simple description of the situation, without explicit reference to a need, may function this way. Then there are all the hand signals, gestures, glances, whistles, bell ringing, and so on that can serve in lieu of oral and written communication for these same purposes. Another broad subclass of examples follows a pattern wherein the performance of a routine by one member produces an alteration in the local working environment of another, and the alteration simultaneously makes the performance of a particular routine feasible and carries the message that it should be performed. An assembly line is one example: the arrival of the partly assembled product at a particular station (as a consequence of the performances of other members) both makes possible the performance of the operation done at that station and indicates that the performance is now called for. The arrival of a draft of a letter or document on a secretary's desk makes possible its typing, and may also indicate that its typing is now called for. In still another large subclass, there are messages to which an individual member responds that do not, in any immediate sense, come from other human members. They may come from clocks and calendars—the start of the working day is an obvious example. They may come from meters, gauges, and display boards that convey information on the current state of machines or of other aspects of the working environment and the progress of activity. Or they may come from outside the organization, as when an order or invoice or application form arrives in the mail.[3]

The ability to receive these various sorts of messages involves the possession of certain sensory capacities, plus, let us say, an ordinary ability to understand the natural language of written and oral communication in the wider society of which the organization is a part. These are abilities that usually characterize an organization member quite apart from his role in the organization—that is, they are the sorts of things a new member typically brings to the organization.

3. The fact that there are such diverse sources and media for the messages to which organization members respond in carrying out their duties is suggestive of the problems of defining "authority" in a useful way. To confine attention to directives from superior to subordinate, or even to communications of all sorts from superior to subordinate, is to ignore most of the details of the coordinating information flow. On the other hand, it is hard to deny that the relations of superior and subordinate often have a lot to do with how the subordinate responds to, for example, messages from the clock.

What about the ability to interpret the messages—to make the link between a message and the performance that it calls for? It is just as necessary as knowing the job, but much more specific to the organization and the job. It is one thing to know how to tell time; it is another to know when to arrive at work, and what it is that you do at about 10 A.M. on the last working day of the month. It is one thing to see a partly assembled automobile in front of you on the line and another to see it as a call for the particular steps that are yours to perform. Even directives that appear to be in "plain English" often require interpretation in a manner that is quite specific to the organizational context. For example, they often omit reference to the typical locations of objects or individuals named in the directives; only someone who has been around the place long enough can easily supply the interpretation. But, in addition, the internal language of communication in an organization is never plain English: it is a dialect full of locally understood nouns standing for particular products, parts, customers, plant locations, and individuals and involving very localized meanings for "promptly," "slower," "too hot," and so on.[4]

The activity of formulating and sending appropriate messages we regard as the performance of a routine by the organization member concerned. This view seems convenient because, as we have noted, there is an important range of cases in which message origination occurs incidentally in the performance of a routine that nominally is directed to other ends. For example, no distinct problem of message formulation arises if the message is conveyed by the partly finished product, passed along to the member who should deal with it next. The burden of the communication process in this case and many similar ones falls upon the receiver who (to know his job) must be able to discern the implications for his own action that are implicit in the changes in his immediate environment—changes that others, by merely doing their jobs, have produced. But there are, of course, many organizational roles whose performance does involve message formulation in a conventional sense. For organization members in such roles, there are additional requisites of knowing the job that parallel the ones involved in receiving and interpreting such messages. These include, again, the abilities to speak and write the natural language of the society to which the organization belongs, but also the important additional requirement of command of the organizational dialect. Such command is certainly not to be taken for

4. Kenneth Arrow, among others, has given particular emphasis to the internal dialect or "code" of an organization as a key source of the economies that formal organization provides and as an important cause of persistent differences among organizations. See Arrow (1974, pp. 53-59).

granted in a new organization member, but is imputed by assumption to members in an organization in a state of routine operation.

The overall picture of an organization in routine operation can now be drawn. A flow of messages comes into the organization from the external environment and from clocks and calendars. The organization members receiving these messages interpret them as calling for the performance of routines from their repertoires. These performances include ones that would be thought of as directly productive—such as unloading the truck that has arrived at the loading dock—and others of a clerical or information-processing nature—such as routing a customer's inquiry or order to the appropriate point in the organization. Either as an incidental consequence of other sorts of action or as deliberate acts of communication, the performance of routines by each organization member generates a stream of messages to others. These messages in turn are interpreted as calling for particular performances by their recipients, which generate other performances, messages, interpretations, and so on. At any given time, organization members are responding to messages originating from other members as well as from the environment; the above description of the process as starting with information input from external sources or timekeeping devices is merely an expositional convenience. There is, indeed, an internal equilibrium "circular flow" of information in an organization in routine operation, but it is a flow that is continuously primed by external message sources and timekeeping devices.

For such a system to accomplish something productive, such as building computers or carrying passengers between airports or teaching children to read and write, some highly specific conditions must be satisfied, different in each particular case. The specific features that account for the ability of a particular organization to accomplish particular things are reflected, first of all, in the character of the collection of individual members' repertoires. Airlines are the sorts of organizations that have pilots as members, while schools have teachers. The capabilities of a particular sort of organization are similarly associated with the possession of particular collections of specialized plant and equipment, and the repertoires of organization members include the ability to operate that plant and equipment. Finally, of course, the actual exercise of productive capability requires that there be something upon which to exercise it—some computer components to assemble, or passengers to carry, or children to teach. These are the considerations recognized in the "list of ingredients" level of discussion of productive capability, which is standard in economic analysis. There is also a "recipe" level of discussion, at which "technologies" are described in terms of the prin-

ciples that underlie them and the character and sequencing of the subtasks that must be performed to get the desired result. This is the province of engineers and other technologists, and to some extent of designers and production managers.

But just as an individual member does not come to know his job merely by mastering the required routines in the repertoire, so an organization does not become capable of an actual productive performance merely by acquiring all the "ingredients," even if it also has the "recipe." What is central to a productive organizational performance is coordination; what is central to coordination is that individual members, knowing their jobs, correctly interpret and respond to the messages they receive. The interpretations that members give to messages are the mechanism that picks out, from a vast array of possibilities consistent with the roster of member repertoires, a collection of individual member performances that actually constitute a productive performance for the organization as a whole.[5] To the extent that the description above is valid, skills, organization, and "technology" are intimately intertwined in a functioning routine, and it is difficult to say exactly where one aspect ends and another begins. This is another way of arguing that "blueprints" are only a small part of what needs to be in an organizational memory in order that production proceed effectively. Furthermore, once the set of routines is in memory by virtue of use, blueprints may not be necessary save, perhaps, as a checkpoint to assess what might be wrong when the routine breaks down.

Given this picture, it is easy to see the relationship between routine operation and organizational memory—or, alternatively, to identify the routinization of activity as the "locus" of operational knowledge in an organization. Information is actually stored primarily in the memories of the members of the organization, in which reside all the knowledge, articulable and tacit, that constitutes their individual skills and routines, the generalized language competence and the specific command of the organizational dialect, and, above all, the associations that link the incoming messages to the specific performances that they call for. In the sense that the memories of individual members do store so much of the information required for the performance of organizational routines, there is substantial truth in the proposition that the knowledge an organization possesses is reducible to the knowledge of its individual members. This is the

---

5. We have passed over here the problem of what makes the organization member *willing* to respond appropriately to a message he receives and correctly interprets. This issue is addressed in the following section.

perspective that one is led to emphasize if one is committed to the view that "knowing" is something that only humans can do.

But the knowledge stored in human memories is meaningful and effective only in some context, and for knowledge exercised in an organizational role that context is an organizational context. It typically includes, first, a variety of forms of external memory—files, message boards, manuals, computer memories, magnetic tapes—that complement and support individual memories but that are maintained in large part as a routine organizational function. One might, therefore, want to say that they are part of organizational memory rather than an information storage activity of individual members. Second, the context includes the physical state of equipment and of the work environment generally. Performance of an organizational memory function is in part implicit in the simple fact that equipment and structures are relatively durable: they and the general state of the work environment do not undergo radical and discontinuous change. A fire or severe storm may break the continuity. The destruction caused by such an event is informational as well as physical, for there is a disruption of the accustomed interpretive context for the information possessed by human members. One might therefore be tempted to say that an organization "remembers" in part by keeping—and to the extent that it succeeds in keeping—its equipment, structures, and work environment in some degree of order and repair. Finally, and most important, the context of the information possessed by an individual member is established by the information possessed by all other members. Without the crane operator's ability to interpret the hand signal for "down a little more" and to lower the hook accordingly, the abilities to perceive the need for the signal and to generate it are meaningless. To view organizational memory as reducible to individual member memories is to overlook, or undervalue, the linking of those individual memories by shared experiences in the past, experiences that have established the extremely detailed and specific communication system that underlies routine performance.

What requires emphasis in the foregoing account is the power of the supposition that "the organization is in a state of routine operation" to limit the scope of the organizational memory function that needs to beperformed. While each organization member must know his job, there is no need for anyone to know anyone *else's* job. Neither is there any need for anyone to be able to articulate or conceptualize the procedures employed by the organization as a whole. Some fraction of the necessary coordinating information may be communicated among members in explicit, articulated form, but

there is heavy reliance on the communication implicit in performances that nominally serve other, directly productive purposes. There is no need for an exhaustive symbolic account of the organization's methods; in any case, because much of the knowledge involved is tacit knowledge held by individual members, such an account cannot exist. Yet the amount of information storage implicit in the successful continuation of the routinized performance of the organization as a whole may dwarf the capacity of an individual human memory. The complexity and scale of the productive process may far surpass what any "chief engineer," however skilled, could conceivably guide.[6]

It is by no means the case, however, that routinization entirely frees organizational memory and organizational performance from constraints imposed by human memory limitations. It is important here to distinguish between the memory requirements of a complex coordinated performance taking place at a given time and the requirements of a flexible performance in which the organization as a whole does quite different things at different times. The complexity of performance at a given time can be greater in a larger organization. With a larger number of members and thus a larger number of human memories among which the organizational memory function can be divided, greater complexity can be consistent with constant or declining demands on the memories of individual members. All members can, simultaneously, remember their jobs by doing them. The situation is quite different with respect to flexibility of organizational performance over time. Flexibility involves variation of the organizational performance in response to variation in the environment.[7] For the organization to respond routinely with a wide variety of specialized routine performances, each "customized" for a particular configuration of the environment, members must be able to retain in repertoire the specialized individual routines involved, and to recall the meaning of a set of messages sufficiently rich to differentiate all the required performances from one another. They must do so in spite of the long time intervals elapsing between the performances of at least some specialized routines and the receipts of some particular messages. (That there are such intervals is of course

6. We have already noted in Chapter 3 the limitations of the "chief engineer" and "book of blueprints" parables that occur in orthodox accounts of productive knowledge.

7. It might also involve response to variations in directives from top management, but presumably those variations reflect changes in the environment. In any case, the story would be much the same for arbitrary changes in directives.

implied by the supposition that the list of performances or messages to be distinguished is long.) Especially in the case of the tacit components of high skill, the phenomenon of memory loss or increasing rustiness over time is important. A skill that is only exercised briefly every year or two cannot be expressed with the smoothness and reliability of one consistently exercised five days a week. And unexpected lapses by individual members tend to have amplified disruptive effects on organizational performance, since by themselves they create further novelties in the organization's state—novelties with which existing routines and communication systems may be unprepared to deal.

These are the considerations that link routine operation with remembering by doing. It is not just that routinization reflects the achievement of coordination and the establishment of an organizational memory that sustains such coordination. It is that coordination is preserved, and organizational memory refreshed, by exercise—just as, and partly because, individual skills are maintained by being exercised. It may be possible to achieve flexibility by scheduling drills for the specific purpose of maintaining infrequently exercised capabilities, or even by having standby units that do nothing but drill for particular contingencies. But these are obviously costly ways of maintaining organizational memory, at least as compared with genuine "doing" that is directly productive. And, as is well known, the quality of the practice afforded by a drill is inevitably degraded by the fact that it is merely a drill.

2. ROUTINE AS TRUCE

Our discussion to this point has been concerned with the cognitive aspects of the performances of organization members—with the question of whether they know what to do and how to do it. We have ignored the motivational aspect—the question of whether they would actually choose to do what is "required" of them in the routine operation of the organization as a whole. Relatedly, the image of coordination that we have presented involves no mention of authority figures, backed by a system of incentives and sanctions, who cajole or coerce the required performances from other members. It is not, however, part of our intention to ignore the divergence of interests among organization members, or to assume implicitly that members are somehow fully committed to the smooth functioning of the organization. Here we fill in the part of the picture of routine

operation that involves motivational considerations and intraorganizational conflict.'

First of all, our concept of routine operation should not be confused with performance according to the nominal standards of the organization. Neither should the proposition that members correctly interpret and appropriately respond to messages they receive be taken to imply that members do what they are told. Nominally, the workday in a particular organization may run from 9.00 to 5:00, but it may be the case (routinely) that very little activity that is productive from the organization's point of view gets done before 9:30 or after 4:45. Similarly, days or weeks may pass between the nominal deadlines for the completion of particular tasks and the typical dates at which they are actually completed. Repeated follow-up requests or orders may, quite routinely, be part of the system of messages that ultimately results in "timely" performance by other organization members. The priority system used by a particular member in allocating effort among tasks may make use, routinely, of the information contained in the overtones of panic or fury in the incoming messages. In short, routine operation is consistent with routinely occurring laxity, slippage, rule-breaking, defiance, and even sabotage. Such behaviors typically violate nominal standards and expectations in an organization, but they do not necessarily violate empirically based expectations or have consequences for output that are inconsistent with results being statistically stable and within the expected range. They may be expected, adapted to, and allowed for—even to the point where a sudden reversion to nominal standards by some organization members would be disruptive of the achieved state of coordination.

Although nominal standards of performance are not necessarily relevant, it is nevertheless true that some sort of stable accommodation between the requirements of organizational functioning and the motivations of all organization members is a necessary concomitant of routine operation. What signals the existence of an accommodation is not the conformity of behavior to standards of performance laid down by supervisors or codified in job descriptions, but that members are rarely surprised at each other's behavior and also that involuntary separations of members from the organization do not occur.

The usual mechanisms of internal control are, of course, a part of

8. In regard to the context of this section, we acknowledge a diffuse intellectual indebtedness to a large number of authors: Coase (1937), Simon (1951), March and Simon (1958), Doeringer and Piore (1971), Ross (1973), Williamson (1975, ch. 4), and Leibenstein (1976).

the context that helps define the *de facto* contracts that individual members make with the organization. Some of the clerks in the retail store might simply ignore the customers if the manager did not check up occasionally—but the manager does, routinely, check up occasionally, and this keeps the problem within limits. Some fraction of workers may in fact take every opportunity to shirk. This means that the "contracts" of these workers call for them to deliver an amount of work that is defined by the level of managerial supervision; a change in that level would mean a change in the *de facto* contract, but no such change occurs in the context of routine operation. Again, if banks did not have elaborate routinized systems of financial control, it is likely that more bank employees would exploit their positions to their own financial advantage, whether by dipping directly into the till or by approving doubtful loans to undertakings in which they have an interest. As it is, the operation of the control system is a major component of the routine tasks of many bank employees: every job is partially defined by the system's existence and illicit appropriation of bank funds is not (routinely) an important form of compensation.

The examples just given illustrate the way in which control of organization members is effected through mechanisms operating routinely as part of the jobs of other organization members, and serving primarily to threaten sanctions, including dismissal, for behavior that deviates from organizational requirements in specified prohibited directions and in excessive degree. Such rule-enforcement mechanisms play a crucial but limited role in making routine operation possible. On the one hand, they largely prevent or deter individual members from pursuing their own interests along lines that are so strongly antithetical to organizational requirements as to threaten the feasibility of any coordinated performance at all. In this sense, they are crucial in keeping the underlying conflicts among organization members from being expressed in highly disruptive forms.

Ordinarily, however, control systems of this type leave individual members with substantial areas of behavioral discretion, areas that embrace performances of widely differing appropriateness or value from the organizational perspective. Except for tasks involving very low levels of skill, performed under conditions favorable to close observation of several workers by a single supervisor, it is not practical to monitor and control behavior so closely that only organizationally appropriate behaviors are permitted. Within the substantial zone of discretion that exists in most cases, the conformity of individual member behavior to organizational requirements is motivated by considerations other than the routinized organizational mechanisms

that "enforce the rules." A variety of other motivating considerations exist. In some cases it is possible to measure individual member "output" reasonably well; reward (or freedom from sanction) can then be conditioned on achievement of a satisfactory output level. In others, organizationally appropriate behavior may be as attractive to the individual member as any other behavior in the zone of discretion left by the rule-enforcement system. Or members may regard themselves as being in a long-term exchange relationship with the organization and may expect future rewards for effective behavior in the present. The importance and efficacy of these motivators and of others not mentioned may be expected to vary among tasks, among rule enforcement, output monitoring and promotion systems, and also, importantly, across member cultures and subcultures that inculcate differing attitudes toward the responsibilities and rewards of organizational membership.[9]

In routine operation, the combined effect of the rule-enforcement mechanism and other motivators is such as to leave members content to play their roles in the organizational routine—but "content" only in the sense that they are willing to continue to perform up to their usual standard, to the accompaniment of the usual amount of griping and squabbling. Conflict, both manifest and latent, persists, but manifest conflict follows largely predictable paths and stays within predictable bounds that are consistent with the ongoing routine. In short, routine operation involves a comprehensive truce in intraorganizational conflict. There is a truce between the supervisor and those supervised at every level in the organizational hierarchy: the usual amount of work gets done, reprimands and compliments are delivered with the usual frequency, and no demands are presented for major modifications in the terms of the relationship. There is similarly a truce in the struggle for advancement, power, and perquisites among high-level executives. Nobody is trying to steer the organizational ship into a sharp turn in the hope of throwing a rival overboard—or if someone is trying, he correctly expects to be thwarted.

When one considers routine operation as the basis of organizational memory, one is led to expect to find routines patterned in ways that reflect characteristics of the information storage problem that they solve. When one considers routine operation as involving a truce in intraorganizational conflict, one is led to expect routines to

9. The considerations just mentioned are among those involved in discussion of "internal" labor markets and the "dual labor market" theory. See Doeringer and Piore (1971) and Williamson (1975, ch. 4).

be patterned in ways that reflect features of the underlying problem of diverging individual member interests. The obvious example of such patterning is the existence of rule-enforcement mechanisms as an ongoing feature of organizational routine, even when serious breaches of the rules are infrequent and most of the sanctions that are nominally available are not applied.

But more subtle manifestations, specific to a particular organizational context, frequently exist. Like a truce among nations, the truce among organization members tends to give rise to a peculiar symbolic culture shared by the parties. A renewal of overt hostilities would be costly and would also involve a sharp rise in uncertainty about the future positions of the parties. Accordingly, the state of truce is ordinarily considered valuable, and a breach of its terms is not to be undertaken lightly. But the terms of a truce can never be fully explicit, and in the case of the intraorganizational truce are often not explicit at all. The terms become increasingly defined by a shared tradition arising out of the specific contingencies confronted and the responses of the parties to those contingencies. In the interpretive context of such a tradition, actions by individual members have connotations related to the terms of the truce. In particular, a contemplated action otherwise sensible both for the organization and for the member taking it may have to be rejected if it is likely to be interpreted as "provocative"—that is, as signaling a lessened commitment to the preservation of the truce and a corresponding willingness to risk overt conflict for the sake of modifying the routine in a manner favored by the member who initiates the change. On the defensive side, each member strives to protect his interests by standing prepared to deliver a firm rebuff not only to actions by others that clearly threaten those interests, but also to actions that might be quite innocuous were it not for their possible interpretation as probes of his alertness or determination to defend his rights under the truce.

The apparent fragility of the prevailing truce and the implied need for caution in undertaking anything that looks like a new initiative is thus reinforced by the defensive alertness (or alert defensiveness) of organization members seeking to assure that their interests continue to be recognized and preserved. The result may be that the routines of the organization as a whole are confined to extremely narrow channels by the dikes of vested interest. Adaptations that appear "obvious" and "easy" to an external observer may be foreclosed because they involve a perceived threat to internal political equilibrium.

Of course, organizations vary in the extent to which these mecha-

nisms operate, as they do in other respects. But it seems safe to say that fear of breaking the truce is, in general, a powerful force tending to hold organizations on the path of relatively inflexible routine.

3. Routine as Target: Control, Replication, and Imitation

So far, we have emphasized that a state of routine operation in an organization is in many ways self-sustaining. Judging by the preceding sections, an organization might be expected to encounter difficulty in departing from its prevailing routines, but it should have no trouble in conforming to them. Although this generalization is more than half of the story and is a basic assumption of our evolutionary models, it is subject to important qualification. Just keeping an existing routine running smoothly can be difficult. When this is the case, the routine (in its smoothly functioning version) takes on the quality of a norm or target, and managers concern themselves with trying to deal with actual or threatened disruptions of the routine. That is, they try to keep the routine under control.

The preceding sections do suggest that there is typically going to be some difficulty encountered in deliberately creating a complex new routine where none existed before. Organization members have to learn the system of coordinating messages. They may have to add new skills to their individual repertoires, and they need to achieve a first reconciliation of their expectations regarding the distribution of costs and benefits in the situation. In such a context—for example, the initial operation of a new plant—the eventual achievement of a state of routine operation also serves as a target for managerial effort, much as it does in the context of control of an existing routine. Because there are important parallels between these "routine as target" situations, we discuss them together here. But there are also important differences, relating to the definiteness of the target presented and the adequacy of the available information as to how it may be attained. With regard to these dimensions of difference, there is a continuum of situations ranging from the edge of full routine—"getting this production line working well, like it was yesterday"—to the edge of major innovation—"opening a plant to build small computers similar to those just introduced by our rival, only better and cheaper." In the formal models of the following chapters, this continuum gets represented by distinct categories and sharp discontinuities. Here we admit that everything is a matter of degree—and examine some of the variables that distinguish the "degrees" of different cases.

Control

An organization is not a perpetual motion machine; it is an open system that survives through some form of exchange with its environment. Even its most durable machines and oldest hands undergo change with the passage of time and through the organizational process itself, and ultimately are replaced. On a much shorter time scale, current inputs of various kinds flow in, and outputs flow out. The organization's routine, considered as an abstract "way of doing things," is an order that can persist only if it is imposed on a continually changing set of specific resources. Some part of this task of imposing the routine's order on new resources is itself handled routinely; another part is dealt with by *ad hoc* problem-solving efforts. Either the routinized or the *ad hoc* part of the task may fail to be accomplished if the environment does not cooperate—for example, if it fails to yield, on the usual terms, the resources that are required.

A major part of the control problem is related, directly or indirectly, to the fact that productive inputs are heterogeneous. The firm itself creates distinctions among inputs in the course of "imposing the routine's order" upon them; it buys a standard type of machine in the market and bolts it to the floor in a particular location in the shop, and it hires a machinist and familiarizes him with the particular capabilities and layout of its equipment and the tasks that are typically performed. Further differentiation occurs incidental to the input's cumulative experience with the idiosyncratic environment of the firm; the machine suffers particular wear patterns and the machinist particular patterns of frustration with his supervisor. But of course the firm also confronts the fact that different units of the "same" input may have distinctive characteristics when they are offered to the firm for purchase, and that the entire distribution of characteristics displayed by different units offered concurrently may itself be changing over time. This prepurchase heterogeneity in the market complicates the problem of postpurchase modification, since the same treatment applied to different units will not necessarily produce the same result. Finally, because machines and workers may pass through the market again after a stay in a firm, the modifications resulting from experience in firms contribute to heterogeneity in the market.

The problem posed for the firm is somehow to acquire inputs with the particular characteristics required for the smooth functioning of its routines, in the face of the fact that such inputs may not be available on the market at all, or, if available, may not be readily distinguishable from other inputs whose characteristics make them less effective or positively dangerous. Since this problem cannot be solved

totally and consistently, a corollary task is to limit the damage associated with imperfections in the solution to the primary problem.

The general tactics applied in dealing with these matters are much the same regardless of the class of inputs considered. A basic tactic is to *select* from the alternatives available from the supply side of the input market those particular inputs that are compatible with the routine. This process is complicated and imperfect if input characteristics are difficult and costly to ascertain, and is further complicated by tension with the cost-control problem, arising from the fact that the range of alternatives available is affected by the price offered. There is then an effort to modify acquired inputs so that they meet the requirements of the routine—to dilute, grind, trim, or sort the raw material to a uniform standard, to teach the clerk the filing system and the portion of the organizational dialect relevant to its use, to bolt down and adjust the new machine, or to instruct the new executive in the rudiments of the technology he is now managing. Of course, if too big a mistake has been made at the selecting stage, adequate modification may be impossible. The central damage-limiting tactic is to monitor the organizational process to detect the shirking or slow worker, the embezzler, the purchased component that fails too often, the paint that does not adhere, and so forth—and, having detected them, to reinvoke the "modify" tactic or to "select" anew from the market. Some of these problems are of course difficult to detect, particularly the ones that actively seek to avoid detection. As a last resort it may be possible to adapt the routine itself so that it either is more tolerant of heterogeneity or so that it can respond routinely to information on varying input characteristics with compensatory adjustments elsewhere. The latter presumes, of course, that available information permits a sorting of inputs into categories of adequate homogeneity.

The first three of these tactics are routinely pursued by various functional subunits within virtually all large organizations. The "selecting" function described is what purchasing and personnel departments do. Some "modifying" is also done by the personnel department and by trainers, supervisors, and co-workers, or, for nonhuman inputs, by engineers or production workers. "Monitoring" is done by line supervisors, but is also an aspect of financial control and of quality control. However, the fact that such routinized arrangements exist does not assure that they are comprehensive or fully efficacious. Some input selection problems arise too infrequently to be dealt with routinely: major purchases of durable equipment and recruitment of high-level executives cannot be entirely routine matters themselves and may be the occasion of major discontinuities in the functioning of the organization as a whole. And if the arrays of alter-

natives that input markets present to the firm change rapidly enough in adverse directions, existing routines for dealing with input heterogeneity are likely to be overwhelmed. Then the organization will either have to adapt its routines or see them go seriously out of control. Finally, the less that is known about what input characteristics are relevant and the more difficult it is to detect the relevant characteristics, the more likely it is that the only symptoms of adverse change in input characteristics will be inexplicable difficulties in carrying out the routine.

As the examples above indicate, the consequences of control lapses are diverse and variable. The plant may have to shut down for a few hours or days while the mess is straightened out. **A** bad batch may have to be thrown away. Perhaps the customers will get an inferior product; with luck they won't even notice, but there is the possibility of getting hit with a big product liability suit. Or perhaps the stockholders collectively will just be a bit poorer, to the tune of whatever the embezzler got away with.

The sorts of consequences that are of particular interest here are those that relate to organizational memory and the long-run continuity of routine. Control lapses may be the cause or effect of memory lapses. We have, for example, emphasized that the memories of individual organization members are a primary repository of the operational knowledge of the organization. Some part of the information thus stored may be readily replaced if the particular member storing it leaves the firm; the former employee may have been the only one who knew how to run a particular machine, but it may be easy to hire a replacement who knows how to run it. Or it may be that the knowledge of the employee who has departed is fully subsumed in the knowledge of his supervisor, who remains. But in some cases the memory of a single organization member may be the sole storage point of knowledge that is both idiosyncratic and of great importance to the organization. The knowledge may be tacit—say, an intuitive grasp of the priority structure of the competing demands on the employee's time that are signaled by incoming messages. It may be articulable but not written down—the first names, marital status, and preferred recreations of the important customers in the region, or the action that is called for when a particular machine starts to vibrate too much.

The loss of an employee with such important idiosyncratic knowledge poses a major threat to the continuity of routine—indeed, if the departure is unanticipated, continuity is necessarily broken. The new person hired to fill the role may eventually restore a semblance of the old routine, but only by picking up the knowledge more or less from scratch, guided by whatever clues his predecessor left lying

about and by the indications provided by those in adjacent roles, within or outside the organization. However, those in adjacent positions may be taking the opportunity to attempt to redefine his organizational role in their own interest, so their advice is not fully trustworthy. For this reason, and because the new role occupant may himself be different in significant and durable ways from his predecessor, and also as the result of other contingencies affecting the role-learning process, it is highly unlikely that a near replica of the predecessor's role performance will result. In short, the organizational routine will mutate.

Mutations, of course, are not always deleterious. To put it another way, maintenance of prevailing routine is often an operational target, but it is not an ultimate objective. Modifications of routine that involve improvements in role performance are presumably welcome. However, in functioning complex systems with many highly differentiated and tightly interdependent parts, it is highly unlikely that undirected change in a single part will have beneficial effects on the system; this, of course, is the basis for the biological proposition that mutations tend to be deleterious on the average. An organization member trying to do a better job can presumably accomplish something more than "undirected change," but changes that seem like obvious improvements viewed from a particular role can easily have adverse effects elsewhere in the system. With the aid of a comprehensive understanding of the system as a whole, beneficial directed change in a part might reliably be accomplished. But since nobody in a complex organization actually has that sort of comprehensive understan'ding, it is clear *a fortiori* that a new employee does not have it.

It is not surprising, therefore, that the control processes of (surviving) organizations tend to resist mutations, even ones that present themselves as desirable innovations. For the particular mutagenic event of loss of a member with a unique knowledge store, the form of the resistance obviously depends on whether the departure is anticipated or not. On the assumption that it is not, control efforts will focus on the selection of a suitably malleable successor who will at least try to respond to the routinized demands placed on the role. The efforts of the veterans to instruct the recruit in the requirements of his role will be colored by their concern to achieve a new truce at least as favorable as the old one; as a result, those efforts will tend to disabuse that successor of "naive" aspirations toward innovative change. When the departure is anticipated, on the other hand, the incumbent is likely to be enlisted in an effort to train one or more possible successors. How well this goes depends on, among other things, the degree to which the knowledge involved is tacit, the de-

gree to which experience during the training period is representative of the full job, and — importantly — whether the incumbent really wants to succeed in imparting the knowledge to his successor.

Although the question of whether the organization can maintain continuity of routine is posed particularly clearly by the example of turnover in a key role, all organizational problems of "keeping things under control" pose that question in some degree. Time and environmental changes buffet the organization with potentially mutagenic events, against which its control systems struggle. In the long run, the most important threats to the maintenance of a successful routine may be the insidious ones, the changes that either escape the control system's notice entirely or else are susceptible to "symptomatic relief" that leaves adverse underlying trends uncorrected. If, for example, the organization fails to maintain an adequate general level of pay relative to alternatives in the market, it may happen that the quality and motivation of its personnel gradually decline, perhaps with adverse consequences for the quality of its product or service that develop a little too slowly to be detected and linked to the pay problem. Against the simpler and more visible problems, on the other hand, the routinized control system may be deployed so massively that it has the collateral effect of impeding adaptation when adaptation is actually necessary. The fact that organizations need to have routinized forms of resistance to unwanted change in routines thus becomes yet another reason why organizational behavior is so strongly channeled by prevailing routine.

Replication

The axiom of additivity is fundamental in orthodox production theory. It implies, among other things, that any feasible pattern of productive activity can be faultlessly replicated: an exact doubling of output per unit time is accomplished by an exact doubling of input. In concrete terms, the claim advanced in this proposition is captured by the image of a plant on a particular site producing a particular output mix in a particular way; on an identical site elsewhere, an identical plant is constructed and produces the identical output mix in the identical way. Or, as F. H. Hahn put it, "If two identical entrepreneurs set up two identical plants, with an identical labor force, to produce an identical commodity x, then the two together will produce twice the amount of x that could be produced by one alone" (Hahn, 1949, p. 135).

So stated, the proposition seems to have the compelling quality of the answer to a very elementary arithmetic problem. Presumably, the posit of identical entrepreneurs is supposed to entail an identity of

productive technique, and the identical plants are not just identical in themselves, but situated in identical environments. After suitable amplification of this sort, the claim may be regarded as a simple tautology or perhaps as an assertion of the universal validity of physical law.

The question is whether the proposition says anything that is helpful in interpreting economic reality. For it to do so, the terms "identical entrepreneurs," "identical plants," and "identical labor force" must have empirical counterparts at least in the sense that they describe limiting cases that are often approached in real situations. In the context of orthodox thought, the idea that these connections to reality exist is supported by: (1) a habit of taking the idea of homogeneous input categories seriously, so that the "identical labor force" assumption is not blatantly contrafactual; (2) a propensity to think of individual entrepreneurs as the repositories of productive knowledge, so that positing "identical entrepreneurs" assumes identity of productive knowledge; and (3) a tendency to regard productive knowledge as articulable and free of idiosyncratic elements, so that the supposition of "identical entrepreneurs" does not relate to an exceedingly remote happenstance.

In our evolutionary models, we make the same assumption that perfect replication is possible, with a similar image in mind of a second plant identical to the first and employing identical routines.10 However, our interpretation of the assumption is quite different from the orthodox one, and our commitment to it considerably less deep. A basic conceptual distinction is that we think of replication as being a costly, time-consuming process of copying an *existing* pattern of productive activity. Though in our modeling we abstract from the costs and make the simplest assumption about the time required, this is still a very different concept from the orthodox one, which is concerned entirely with the structure of *ex ante* possibilities. To put it another way, our assumption relates to what can be accomplished starting from the status quo of a functioning routine, whereas the long-run orthodox theory to which the additivity axiom relates has no notion of a status quo at all. Further, we regard the feasibility of close (let alone perfect) replication as being quite problematic — more problematic than the feasibility of continuation through time of the

10. We will limit our discussion of replication to the simple case of establishing the same routine in a plant identical to the original. Some of the same issues arise in almost any case of capacity expansion; a typical situation is that capacity is increased by a partial replication that relaxes the constraint imposed by a particular class of input services. However, partial replications involve some additional complications that we do not treat here.

existing routine, which is itself no foregone conclusion, as the above discussion points out. As an initial perspective on the problem, we would not recommend the Hahn tautology, but the following account from Polanyi: "The attempt to analyze scientifically the established industrial arts has everywhere led to similar results. Indeed, even in modern industries the indefinable knowledge is still an essential part of technology. I have myself watched in Hungary a new, imported machine for blowing electric lamp bulbs, the exact counterpart of which was operating successfully in Germany, failing for a whole year to produce a single flawless bulb" (Polanyi, 1964, p. 52).

The point emphasized by evolutionary theory is that a firm with an established routine possesses resources on which it can draw very helpfully in the difficult task of attempting to apply that routine on a larger scale. Because the creation of productive organizations is *not* a matter of implementing fully explicit blueprints by purchasing homogeneous inputs on anonymous markets, a firm that is already successful in a given activity is a particularly good candidate for being successful with new capacity of the same sort. The replication assumption in evolutionary models is intended primarily to reflect the advantages that favor the going concern attempting to do more of the same, as contrasted with the difficulties that it would encounter in doing something else or that others would encounter in trying to copy its success.

To understand the nature of these advantages, it is helpful first of all to consider the similarities between replication and control, and the deeper connections to the problem of organizational memory. In replicating an existing routine, the firm seeks to impose that routine's order on an entire new set of specific inputs. That task is a magnified version of one for which the firm already possesses routinized arrangements. For example, its existing personnel and training operations have the capability to "select and modify" the sorts of employees the routine requires. By diverting these existing capabilities at least in part to the tasks associated with the new facility, it can avoid difficulties that would be very likely to arise if the manning of that new facility were accomplished by an equally new and inexperienced personnel operation. The new plant will ultimately need its own personnel department (at least if "replication" is taken literally), but the new production system does not have to be hampered by the early mistakes of a new personnel department that may be learning to operate in a novel labor market environment. And a functioning production system that is effective enough to detect mistakes by the new personnel department can then help that department to learn its job.

More generally, the existing routine serves as a template for the

new one. The use of the template makes possible a relatively precise copying of a functioning system that is far too large and complex to be comprehended by a single person. It is not necessary for there to be a central file that contains an articulate account of how the whole thing is done. Rather, for each organizational role that is a unique storage point for important and idiosyncratic organizational knowledge, it is necessary that the individual who will occupy that role in the new plant acquire the knowledgerequired for its performance. This may be accomplished by having that individual observe or be actively trained by the incumbent of that role in the old system, or by transferring the incumbent to the new system and leaving his trained successor in the old one. The collection of new role occupants thus created will make a coordinated, routinely functioning productive organization of the new facility, because the roles were coordinated in the old one—provided that the copying of the individual roles is accurate enough.

Of course, the process described will in general impose some costs in terms of the functioning of the old plant. It is unlikely that there will be enough slack resources available for training new personnel or for actually performing, temporarily, some functions in the new plant. For the replication story to make economic sense, the benefits obtained must exceed or be expected to exceed these costs. This issue is basically one of investment analysis. If the old plant is enjoying a *temporary* period of high prosperity, to be followed by normal or low profits, the opportunity costs of replication may indeed be excessive." The knowledge transfer must make it possible to capture a flow of rents in the new plant that lasts long enough to compensate, in present value terms, the loss of rents in the old plant. The likelihood of this sort of pattern is obviously enhanced to the extent that a large knowledge transfer can be carried out with only small sacrifices in the old plant. Here it is relevant that the costs of a small number of anticipated departures or absences from key positions in the old plant are likely to be small, since such isolated gaps pose just the sort of problem that the control system routinely handles. On the other hand, the value of only a few people who know what they are doing may be enormous in providing the basic matrix of the routine in the new plant. That is, there are likely to be diminishing returns to experienced personnel, in terms of learning costs saved, in both plants. The transfer of a small number of experienced personnel from the

11. When long-run prospects are favorable but current profits are also high, it can happen that constructing a new plant *de novo* is preferable to replication involving current opportunity costs, even though replication is absolutely profitable and would be the preferred mode of expansion under less favorable conditions.

old, predominantly experienced plant to the new, predominantly inexperienced one saves a lot of learning costs in the latter and incurs only small ones in the former. Finally, because of imbalances arising from indivisibilities or for other reasons, there may be some resources in the old plant that are actually idle and can be costlessly applied to the replication effort or transferred to the new plant.

There are some potential obstacles to replication that may be difficult to overcome even at very high cost. Some employees at the old plant may be exercising complex skills with large tacit components, acquired through years of experience in the firm. Others may have skills of lesser complexity and tacitness, but be very poor at teaching those skills to someone else—doing and teaching are, after all, different. Some members may for various reasons be unwilling to cooperate in the process of transferring their segment of the memory contents to someone else; they may, for example, be unwilling to disclose how easy their job really is, or the extent of the shortcuts they take in doing it.[12] Finally, personal relationships may be an important factor, particularly in the structure and stability of the truce that the existing routine represents. The personnel department is not likely to be up to the challenge of locating a suitably matched *set* of new role occupants who can be relied upon to maintain the same sort of truce. For these reasons and more, the template provided by the existing routine may not yield a good copy. There will be some mutation of the routine as it is transferred to the new plant.

Of course, perfect replication is no more of an ultimate objective than perfect control. What matters is not that the plant be the same, but that it work with overall efficiency comparable to the old one.

Contraction

If an existing routine is a success, replication of that success is likely to be desired. In particular, in the models to follow, the organization in question is a business firm for which success is roughly measured by profits, and replication of productive routines is motivated by a desire to replicate the profit flows that those routines make possible. There are symmetric questions to be addressed if the existing routine is a failure—that is, unprofitable. But while the questions are at least roughly symmetric, the answers are not. Because of their obvious importance to our models of economic selection, we digress briefly to consider them.

12. The question of the incentives of organization members to disclose idiosyncratic information of importance to the organization's functioning is addressed by Williamson under the rubric "information impactedness" (Williamson, 1975, ch. 4).

One important asymmetry between replication and contraction is that while the former is typically an *optional* response to success, the latter is typically a *mandatory* response to failure. As usual, the situation is clearest in the case of business firms, though there are analogous problems in other sorts of organizations. If the revenues derived from the sale of the routine's outputs fail to cover the costs of the routine's inputs, then — barring governmental bail-outs, philanthropically inclined investors and similarly unlikely contingencies — it will ultimately become impossible to acquire the inputs to continue the routine on the existing scale and something will have to happen.

Under this pressure, a business firm may be expected to initiate some sort of search for a new routine that would be viable in the prevailing environment. The analysis of this sort of search runs roughly parallel to the analysis of imitation and innovation that will concern us later in this chapter, with the proviso that the initiation of the search under conditions of adversity has implications for the quantity and quality of the resources that may be devoted to it. But if the search is successful in the limited sense that the firm begins to attempt to carry out a new routine, then the old routine is no longer the target and has fallen victim to the condition of adversity. The firm itself may live on, at least temporarily.

Although some sort of search response to adversity is probably typical, it may happen that the organization remains firmly committed to its existing ways of doing things — a course of action that can be rationalized as an attempt to last out a period of adversity that is perceived or hoped to be temporary. In this case, the only "search" that goes on is for the resources to continue to finance the existing routine. A likely occasion for such attempts to fall short is when it comes time to replace a large, indivisible item of durable equipment. If unable to carry out such a replacement, the firm may simply shrink and carry on roughly as before, but on a diminished scale. This sort of response (in addition to the search response) is envisaged in the formal models that follow. After a series of scale reductions of this sort, the firm and its routine may ultimately disappear entirely.

In reality, a great many factors are involved in determining the consequences of sustained adversity — for example, the degree of owner versus management control, merger opportunities, tax and bankruptcy law considerations, the liquidity or illiquidity of the firm's assets, and the state of the firm's balance sheet when adversity began. It is beyond the scope of our present discussion to sort out these factors and relate them to the likely persistence or change of routines. One point perhaps is worth noting here: a firm without a viable routine is a firm without a viable truce in intraorganizational

conflict. That consideration, by itself, affords abundant reason to doubt that firms behave in adversity "as if" they were under the rational control of a single actor.[13]

Imitation

As a final example of a routine serving as a target, let us consider the case in which the target is a routine of some other firm. The interest in this sort of situation arises, of course, because it often happens that a firm observes that some other firm is doing things that it would like to be able to do—specifically, making more money by producing a better product or producing a standard product more cheaply. The envious firm then attempts to duplicate this imperfectly observed success. We will consider here only the case in which the imitatee is not cooperating with the imitation effort, and will assume that non-cooperation implies, at a minimum, that the imitator's personnel cannot directly observe what goes on in the imitatee's plant.[14]

What distinguishes this situation from replication is the fact that the target routine is not in any substantial sense available as a template. When problems arise in the copy, it is not possible to resolve them by closer scrutiny of the original. This implies that the copy is, at best, likely to constitute a substantial mutation of the original, embodying different responses to a large number of the specific challenges posed by the overall production problem. However, the imitator is not directly concerned with creating a good likeness, but with achieving an economic success—preferrably, an economic success at least equal to that of the original. Differences of detail that are economically of no great consequence are perfectly acceptable.

By this economically relevant criterion, the prospects for successful imitation vary dramatically from one situation to another. At one extreme, the production in question may be a novel combination of highly standardized technological elements. If so, close scrutiny of the product itself—"reverse engineering"—may permit the identification of those elements and the nature of their combination, and

---

13. Philip Nelson's book (1981) provides fascinating details on the behavior of a number of business firms operating under adversity. One point that stands out is that, under severe adversity the divergence of member interests contributes importantly to the inability of the organization as a whole to cope effectively with its problems.

14. There are cases intermediate between the categories of "replication" and "imitation"—cases of attempted near-replication in environments very different from the original one, or of imitation with the active support of the firm being imitated. These are usually addressed under the heading of "transfer of technology." Our own thinking in this general area has benefited particularly from the work on technology transfer of Hall and Johnson (1967) and Teece (1977).

this may suffice for an economically successful imitation. Indeed, even vague rumors about the nature of the product may suffice, perhaps permitting the copy to hit the market almost as soon as the original. At the other extreme, the target routine may involve so much idiosyncratic and "impacted" tacit knowledge that even successful replication is highly problematic, let alone imitation from a distance.

In the wide range of intermediate cases, the imitator's basic tactic is to follow the example of a replicator wherever possible (and not too expensive), and to fill in the remaining gaps by independent effort. One important application of this tactic is to try to hire away from the imitatee those employees that the imitatee would reasonably want to transfer to a new plant in an attempt to replicate the existing one. Another is to obtain, by whatever means may be available, indirect clues to the nature of the target routine.

An imitator working with an extremely sparse set of clues about the details of the imitatee's performance might as well adopt the more prestigious title of "innovator," since most of the problem is really being solved independently. However, the knowledge that a problem *has* a solution does provide an incentive for persistence in efforts that might otherwise be abandoned.

4. ROUTINES AND SKILLS: PARALLELS

As we observed at the start of the previous chapter, understanding of individual skills informs understanding of organizational behavior in two ways. First, because individuals exercise skills in their roles as organization members, the characteristics of organizational capabilities are directly affected by the characteristics of individual skilled behavior. We have noted some of these connections. For example, an organization's capabilities require the exercise of individual skills that may involve a large component of tacit knowledge; this directly implies limits on the extent to which the organization's capabilities can themselves be articulated, and there are attendant implications for the character of the replication task. Then, too, the inflexibility of behavior displayed by large organizations is attributable in part to the fact that individual skills become rusty when not exercised; it is therefore hard for an organization to hold in memory a coordinated response to contingencies that arise only rarely.

Here we make explicit the other sort of contribution that understanding of individual skills makes to understanding of organizational functioning: the contribution at the level of metaphor. Routines are the skills of an organization. The performance of an or-

ganizational routine involves the effective integration of a number of component subroutines (themselves further reducible), and is ordinarily accomplished without "conscious awareness"—that is, without requiring the attention of top management. This sort of decentralization in organizational functioning parallels the skilled individual's ability to perform without attending to the details. A routine may involve extensive direct interactions with the organization's environment and the making of numerous "choices" that are contingent both upon the state of the environment and the state of the organization itself, but these choices involve no process of deliberation by top management. The intervention of top management in the detailed functioning of lower levels is ordinarily symptomatic of an attempt to modify routine or of difficulties with the functioning of existing routines—just as conscious awareness of detail and attempts at articulation are symptomatic of new learning or of trouble in the case of individual skills.

In a number of respects, organizational behavior seems to be subject to magnified versions of problems and pathologies that afflict individual skilled behavior. The scale and complexity of a large organization make impossible the degree of centralization of control represented by the brain of an individual human being. This relative weakness of centralized analysis and control in organizations, when compared to individuals, is the obvious explanation for the relative severity of the difficulties that organizations encounter in areas where centralization is for some reason important. Thus, for example, we noted that limits on articulation in the case of individual skills derive partly from the "whole versus parts" problem of reconciling an exhaustive account of details with a coherent view of the whole. Much more severe limits on the articulation of organizational knowledge arise from the same cause, because although attending to details is something that can be shared and decentralized, the task of achieving a coherent view of the whole is not. Similarly, improvisation of a coordinated response from a system requires centralized control of the system. Organizations are poor at improvising coordinated responses to novel situations; an individual lacking skills appropriate to the situation may respond awkwardly, but an organization lacking appropriate routines may not respond at all.

Organizations can get a great deal accomplished that they do not know how to do, by drawing on the capabilities of other individuals and organizations. In doing so, however, they exercise planning routines that involve the manipulation of symbols representing highly complex entities. Like individuals, organizations may make ineffective use of the array of capabilities available in their environments, or be victimized by hucksters, because of limitations on their plan-

ning vocabulary—particularly when they do not themselves possess even the rudiments of the capabilities they seek to acquire.

The basic metaphor can be elaborated and extended in a number of other directions, but we will leave these byways unexplored. The important contribution of the metaphor is the insight it provides into the role of bounded rationality in organizational behavior. We observed in our discussion of individual skills that bounded rationality imposes a tradeoff between capability and deliberate choice. That tradeoff exists for organizations as well, but the relative weakness of centralized control in an organization makes the terms of the tradeoff much less favorable to deliberate choice. One cannot infer from the fact that an organization functions smoothly and successfully in a particular range of observed environments that it is a rational and "intelligent" organism that will cope successfully with novel challenges. If anything, one should expect environmental change to make manifest the sacrifice of flexibility that is the price paid for highly effective capabilities of limited scope.

5. OPTIMAL ROUTINES AND OPTIMIZATION ROUTINES

Orthodox economists ordinarily profess a complete lack of interest in the processes by which firms actually make decisions. From their perspective, the fact that our discussion to this point has been concerned with *how* organizations function means that it offers no clue as to "whether firms really maximize profits," since that question relates to "what they do"—that is, to the transactions they engage in, not to how they decide to do it. Insofar as their point relates to the possible optimality of particular actions in particular circumstances, we agree with it. Indeed, the evolutionary model of the following chapter illustrates the possibility that firms modeled according to the spirit of our own view of decision process may wind up taking profit-maximizing actions in selection equilibrium. However, if their claim is that firms *consistently* optimize, even under completely unanticipated circumstances, then we obviously disagree. And we would argue that evidence relating to decision processes is highly relevant to that issue.

We will not go into the subtle questions of methodological principle involved in this area. However, one rather simple point illuminates the nature of the clash between the orthodox view that firms optimize and the evolutionary view that they function according to routine. Imagine a firm that functions with a completely inflexible routine, totally unresponsive to its changing environment. It purchases inputs at constant flow rates and converts them into outputs

which it sells at constant rates. The profitability of this operation varies as the environment changes, but imagine that it is always positive. Orthodoxy can accept this firm's behavior as profit maximizing, since the behavior is interpretable as reflecting optimization over a production set that contains only the single input-output list corresponding to the firm's routine—or perhaps that list and some others that are strictly inferior to it.

The key point here concerns the empirical basis of the claim that only that one pattern of behavior is available to the firm. If one accepts the methodological principle that "what the firm actually does" in market transactions is the only relevant evidence on the alternatives available, then the orthodox claim that this inflexible firm is an optimizer is safe from refutation. But if other sorts of evidence are admissible—for example, evidence that the firm's inflexibility reflects the existence of a delicate truce in an extremely severe case of latent intraorganizational conflict, or evidence on what other firms do—then the claim that this very rigid firm is an optimizer may well be refuted. More generally, the hypothesis that routinized behavior patterns really reflect optimization after all is likely to be more vulnerable to evidence that provides some sort of independent check on the alternatives that might be considered available than it is to evidence on the market transactions arising from the routine itself.[15]

Although a highly defensive and skeptical stance toward decision process evidence is typical, occasionally evidence of this sort is put forward in support of orthodox theory. Thus, for example, the fact that a particular firm has sophisticated accounting techniques, employs formal optimization procedures in some part of its decision making, or has a permanent in-house operations research unit may be adduced as evidence corroborative of the general proposition that firms optimize. Of course, the first question to be raised about this evidence is how representative it is, and whether orthodox analysis is to be understood as relevant only to the historical periods, economies, industries, and firm-size ranges in which these features of firm decision processes are typical. Beyond that, we emphasize that this sort of evidence fits into the evolutionary framework as useful information on the details of the routines that some firms follow.

We would conjecture, for example, that firms that have operations research (OR groups not only go about making decisions in different ways from firms that do not, but that the decisions themselves are likely to differ. Whether a firm has an OR group and systematically does OR as part of its higher-order decision making is a question

15. We return to these issues in Chapter 7. The questions of methodology are addressed more extensively in Winter (1975).

that we view very much in the same light as the question of whether a firm does or does not use the oxygen process for making steel. Both questions are about the routines employed by firms. The exercise of an OR capability indicates that a firm has that capability in very much the same way that exercise of the oxygen process for making steel means that the firm has that particular capability.

However, the fact that a firm has an OR group that builds models and that this group is influential in decision making does not imply that the firm's actual decisions are "truly" optimal. Indeed, we would view particular attention of the OR group on a certain area of decision as an indication that the firm presently is not satisfied with its current routines in that area. Presuming the OR group comes up with a proposal for reform, we would regard it meaningless to say that the new policy is truly optimal; only God knows what policy truly would be optimal. There is no guarantee that the policy that would be optimal within the operations research model is even superior in the actual economic environment to the policy that is being replaced.

Also, and relatedly, knowledge of the fact that the firm goes through explicit maximization calculations to guide its decision making does not mean that the orthodox economist can on the basis of his own model make good predictions of what the firm will do. His model and that used by the operations research group may differ in important respects. It does mean, however, that if the economist knew the model used by the firm's operations research group, that information might help him predict and explain the firm's actions. The economist would then have direct information on the ***routine*** employed in decision making by the firm. And that, of course, is the heart of our theoretical proposal: the behavior of firms can be explained by the routines that they employ. Knowledge of the routines is the heart of understanding behavior. Modeling the firm means modeling the routines and how they change over time.

6. ROUTINES, HEURISTICS, AND INNOVATION

Both in customary usage and in our technical use of the term, "innovation" involves change in routine. We have stressed the uncertainty that inevitably surrounds technical innovation — the implementation of a design for a new product, or of a new way to produce a product. A similar uncertainty surrounds other kinds of innovation — the establishment of a new marketing policy, or a new decision rule for restocking inventories. In general, two kinds of uncertainty surround these innovations. The precise nature of the innovation actu-

ally arrived at is usually not closely predictable at the start of the endeavor that culminates in the innovation. And the consequences of employing the innovation — changing the routine — in general will not be closely predictable until a reasonable amount of actual operating experience with it has been accumulated. There is, however, more to be said about the relations of routine behavior and innovation than to observe that these concepts are commonly (and appropriately) regarded as opposed ideas. Our final task in this chapter is to explore some of the subtler connections between routinization and innovation, and ultimately to indicate how the existence of innovative activity relates, in our evolutionary theory, to the general image of firm behavior as governed by routine.

Puzzles from Prevailing Routines

It is sometimes remarked of an important research achievement that the hard part was in locating the right question; finding the answer to that question then proved to be relatively easy. One way in which the routine functioning of an organization can contribute to the emergence of innovation is that useful questions arise in the form of puzzles or anomalies relating to prevailing routines. The concreteness of such questions and the obvious existence of an application for the answers is an important point in their favor as guides to problem-solving activity.

Consider the foreman of a work team responsible for a particular operation (set of routines) who observes that a machine is not working properly. He routinely calls in to the maintenance department, which in turn routinely sends out a machine repairman. The machine repairman has been trained to diagnose in a particular way the troubles that such a machine might have. He goes down a list of possible problems systematically, and finds one that fits the symptoms. He fixes the part so that the machine again may play its role in the overall work routine. He may also, however, report to the foreman that this particular kind of trouble has become very common since the supplier started using aluminum in making the part in question and that perhaps the machine should be operated in a different manner to avoid the difficulty.

Or consider a sales manager who observes a significant and sustained decrease in total sales of a particular item. He routinely calls in his young assistant — a recent graduate of a master's program in management — to do a study of the problem. The assistant, with a bit of clerical help, scans what has been happening to sales in particular regions and by particular salesmen. He ascertains that almost all of the decrease has occurred in the Southeast. He may go on to check up

on the activities of the salespeople concerned with the Southeast and may recommend some replacement of personnel. He may suspect that some important change in demand conditions has occurred and propose a new market survey to discover its nature. Or he may propose that a new advertising campaign, addressed to customers in the Southeast, may be needed.

These examples illustrate, on the one hand, the routine functioning of organizations. The responses described fall into the typical pattern in which a crisis or "exception" condition in one part of the organization is part of the routine content of jobs of other personnel. On the other hand, it is significant that the problem-solving responses routinely evoked by difficulties with existing routines may yield results that lead to major change. The effort triggered by the repairman's suggestion may lead to a radical improvement in the method of operation of the machine, or to a decision to switch to machines of quite a different sort, requiring numerous adaptations elsewhere in the routine. The market survey proposed by the young assistant may indicate that the trouble in the Southeast is only a symptom of a market change that is likely to become pervasive, and may thus trigger redesign of the product to meet the specific challenge that the survey identified. Problem-solving efforts that are initiated with the existing routine as a target may lead to innovation instead.

Existing Routines as Components

Schumpeter identified innovation with the "carrying out of new combinations" (Schumpeter, 1934, pp. 65–66). This phrase gives useful emphasis to the fact that innovation in the economic system—and indeed the creation of any sort of novelty in art, science, or practical life—consists to a substantial extent of a recombination of conceptual and physical materials that were previously in existence. The vast momentum of scientific, technological, and economic progress in the modern world derives largely from the fact that each new achievement is not merely the answer to a particular problem, but also a new item in the vast storehouse of components that are available for use, in "new combinations," in the solution of other problems in the future.

Innovations in organizational routine similarly consist, in large part, of new combinations of existing routines. An innovation may involve nothing more than the establishment of new patterns of information and material flows among existing subroutines. It may involve the replacement of an existing subroutine by a new and dif-

ferent one that performs, in relation to the rest, the same function that the old one did. Some parts of the innovative routine may rely on physical principles only recently discovered and now implemented through novel types of equipment and newly developed skills—but surrounding this novel core there may be many layers of complementary activity governed by the same routines that have prevailed for many years.

When an effort is made to incorporate an existing routine as a component of innovative routines, it is helpful if two conditions are satisfied. One is that the routine be reliable—that is, fully under control. The attempt to develop an effective new combination ordinarily involves a substantial amount of trial-and-error search, in which obstacles to effective performance are detected, diagnosed, and solved. It is helpful if the familiar elements of the new combination do not themselves contribute problems, particularly if the problems from that source would complicate the task of detecting and solving the problems arising from the novel elements. The second condition is that the new application of the existing routine be as free as possible from the sorts of operational and semantic ambiguities of scope that we discussed in connection with individual skills. Ideally, the existing routine may require only symbolic representation in the design effort for the new combination. For example, the existing routine for shipping the product to wholesalers may be as unambiguously applicable to the new product as it was to the old. In that case, the design effort for the new routine can handle the transportation problem simply by using the phrase "ship to warehouses," and the details of the shipment process need not be examined. But perhaps the new product is in some way more delicate than the old—more vulnerable to temperature extremes or to vibration. Then ambiguity may arise as to whether the existing shipping routine will suffice. If there is reason to doubt that it will, the problem of getting the product to the warehouses in good condition becomes interdependent with the rest of the design problem, and the simple symbolic reference to shipment will have to give way to consideration of details. The existing shipping routine may have to be tried out to see how it affects the new product; it may require modification, or perhaps the design of the product will have to be altered to make it less delicate.

These two conditions suggest an important qualification to the general notion of an opposition between routinization and innovation. Reliable routines of well-understood scope provide the best components for new combinations. In this sense, success at the innovative frontier may depend on the quality of the support from the "civilized" regions of established routine.

Heuristics and Strategies as Routines

Our final point concerning the relationship of routine behavior to innovation is centered on a simple distinction between organizational *activity* directed to innovation (or problem-solving more generally) and the *results* of such activity. The fundamental uncertainty surrounding innovative activity is uncertainty about its results. True, there may be considerable uncertainty when theactivity is initiated, about the details of the activity itself—particularly since those details may ultimately be recognized as an approach to some type of success that is not knowable in advance. But there may also be strong patterns of a highly predictable nature in the activity—and to the extent that this is so it seems reasonable to describe the activity as "routinized." A particularly clear illustration of the significance of the distinction is the case of systematic sequential search of a well-defined population for an element with attributes that make it the solution to a well-defined problem. When and whether a solution will be found may be quite uncertain, but the search itself follows a routine with a simple structure: select element, test for desired attributes, terminate with success if attributes are present, select next element if they are not.

Routinized arrangements for producing innovations and solutions to problems take a variety of forms, among which are some very familiar features of the organizational scene. Given a problem, direct a subordinate to look into it—or appoint a committee or a task force, or bring in a consultant with a good reputation. Given a decision to devote 4 percent of $100 million of sales to R&D, it is almost certainly possible to acquire some sort of facility, a research director, and some scientists, and go to work. In broad terms, at least, the art of deploying resources to *try* to bring about some result or other is not esoteric. Whether useful results are actually achieved is another matter. In fact, results that are more or less useful are often achieved—and it is an important feature of these problem-solving situations that the superior results that in some sense "could" have been achieved are usually not available as a standard of comparison.

The theory of heuristic search provides a helpful framework for thinking about these issues.[16] A *heuristic* is "any principle or device that contributes to the reduction in the average search to solution" (Newell, Shaw, and Simon, 1962, p. 85). Some heuristics are applicable across very wide ranges of problems—"work backward from the goal''—while others are relevant only in highly specific problem contexts. Devices like directing a subordinate to look into a problem,

16. See Newell, Shaw, and Simon (1962) and Newell and Simon (1972)

or appointing a committee, can be viewed as general types of managerial problem-solving heuristics. But every field of specialized competence contains a wide range of heuristics that are particularly appropriate to that field. The operations researcher will build an optimization model of the problem. The mechanical engineer will look at the mechanized aspects of the production process, and look for ways to mechanize it further. The chief executive officer whose background is in finance will bring a different set of heuristics to his job than one whose background is in production. The manager who transfers to a new organization will bring with him some of the heuristics that seemed to work in his previous employment.

The broad ideas that shape the most critical high-level decisions of a business enterprise may also be viewed as heuristics—they are principles that are believed to shorten the average search to solution of the problems of survival and profitability. Much discussion of heuristics of this sort has been carried on under the rubric "corporate strategy." Indeed, according to the concept of strategy that has been developed by a number of investigators associated with the Harvard Business School,[17] the fundamental heuristic imperative for top management is: "Develop a strategy." Other heuristics are involved in the implementation of that basic one—for example, "Assess the company's strengths and weaknesses in relation to the competition." A related idea is that the firm should adopt an organizational structure appropriate to its strategy.[18] More generally, principles that offer guidance for the selection of organizational structures may be viewed as another class of high-level managerial heuristics.

We propose to assimilate to our concept of routine all of the patterning of organizational activity that the observance of heuristics produces, including the patterning of particular ways of attempting to innovate. To the extent that such patterning persists through time and has implications for profitability and growth, it is part of the genetic mechanism underlying the evolutionary process. But we emphasize, once again, that viewing innovative activity as "routine" in this sense does not entail treating its results as predictable.

In many ways our position regarding these matters is consistent with that of Whitehead (1938), who proposed that sometime during the nineteenth century man invented the art of inventing, and is also consistent with the Schumpeter of *Capitalism, Socialism, and Democracy* (1950), who proposed that sometime during the twentieth century the modern corporation "routinized innovation." Neither Whitehead nor Schumpeter, we think, would deny the role of genius

17. See Caves (1980).

18. This idea is particularly associated with Alfred Chandler (1962).

or luck, or argue that systematic differences in innovative competence do not exist. But their views are quite compatible with the proposition that organizations have well-defined routines for the support and direction of their innovative efforts.

7. SUMMARY: ROUTINES AS GENES

Theorists should aim to tell the truth in their theorizing, but they cannot aim to tell the whole truth. For to theorize is precisely to focus on those entities and relationships in reality that are believed to be central to the phenomena observed—and largely to ignore the rest. To advance a new theory is to propose a shift of focus, to recognize as central considerations that were previously ignored.

In this chapter, we have focused upon the realities of organizational functioning that form the foundation of our evolutionary theory. Foremost among those realities are the factors that tend to limit the individual firm to the exercise of a distinctive package of economic capabilities that is of relatively narrow scope. Essential coordinating information is stored in the routine functioning of the organization and "remembered by doing." As in the case of individual skills, the specificity of the behavior involved is simply the obverse of its effectiveness; also, much of the knowledge that underlies the effective performance is tacit knowledge of the organization, not consciously known or articulable by anyone in particular. These cognitive factors are reinforced by motivational ones associated with the control of intraorganizational conflict. Prevailing routines define a truce, and attempts to change routines often provoke a renewal of the conflict which is destructive to the participants and to the organization as a whole.

As a first approximation, therefore, firms may be expected to behave in the future according to the routines they have employed in the past. This does not imply a literal identity of behavior over time, since routines may be keyed in complex ways to signals from the environment. It does imply that it is quite inappropriate to conceive of firm behavior in terms of deliberate choice from a broad menu of alternatives that some external observer considers to be "available" opportunities for the organization. The menu is not broad, but narrow and idiosyncratic; it is built into the firm's routines, and most of the "choosing" is also accomplished automatically by those routines. This does not mean that individual firms cannot be brilliant successes for a short or long period: success and failure depend on the state of the environment. As long as the world rewards great tennis playing, great tennis players will succeed in the world,

regardless of their talents as physicists or pianists. Efforts to understand the functioning of industries and larger systems should come to grips with the fact that highly flexible adaptation to change is not likely to characterize the behavior of individual firms. Evolutionary theory does this.

As a second approximation, firms may be expected to behave in the future in ways that resemble the behavior that would be produced if they simply followed their routines of the past. Just what "resemble" means here is an important and complex question. It is a question that is particularly illuminated by inquiry into the factors that hold behavior to the channels of routine, since whatever change takes place may be expected to follow the path of least resistance. But to assess where the resistance is likely to be least requires a discriminating analysis of the relative strengths of different sources of resistance. This is the great challenge of the subject of "organizational genetics"—to understand how the continuity of routinized behavior operates to channel organizational change. Our discussion of routines as targets and as components addresses this problem in a preliminary way, but the subject has barely been defined and the real work remains to be done. The particular models that follow are built on very simple assumptions regarding these matters, particularly the assumption that capacity expansion can be achieved with faultless replication of routine, and similarly that contraction of a firm is simply a scaling down of the same routinized pattern of operation. The discussion above provides support for these assumptions as a starting point for model building, but it contains some important caveats that should be kept in mind in future work. It also makes even more suspect the assumption that imitation of another firm's routines can be accomplished perfectly. However, for the limited purposes of these particular models, use of a weaker assumption would do more to complicate the analysis than to change its substantive content. The important consideration captured by the models is that imitation, though costly and imperfect in the individual instance, is a powerful mechanism by which new routines come to organize a larger fraction of the total activity of the system.

In the contemporary economy, some portion of business behavior is closely calculated by sophisticated optimization methods. Another portion is innovation activity shaped by the creative problem-solving insights of scientists, engineers, and managers. A full account of business behavior has to deal with these sophisticated portions, and the imagery of routinized behavior does not have the clear validity and power here that it has in discussing, say, a family firm whose product mix has remained unchanged for generations. We have argued, however, that the notion of routine behavior does have

application in this sophisticated realm, though in a qualfied sense. For example, the skills of the highly trained operations researcher, scientist, or manager are reflected in characteristic, highly patterned forms of problem-solving activity. The scope of the expertise involved in each case is defined by a certain class of problem-solving techniques and heuristics. For this and other reasons, even the sophisticated problem-solving efforts of an organization fall into quasi-routine patterns, whose general outlines can be anticipated on the basis of experience with previous problem-solving efforts of that organization. But the patterning of the problem-solving activity is reflected only vaguely in the immediate outcomes of that activity and even less clearly in the gross changes in firm behavior that these problem solutions may trigger. From the viewpoint of an external observer who has no access to the sophisticated workings within the organization, the results are hard to predict and on that ground are best regarded as stochastic. This is the approach we take in the evolutionary models that follow.