
Part I

**An Introduction
to Organizational Learning**

1

What Is An Organization That It May Learn?

Some Preliminary Distinctions

“Learning” may signify either a *product* (something learned) or the *process* that yields such a product. In the first sense, we might ask, “What have we learned?” referring to an accumulation of information in the form of knowledge or skill; in the second sense, “How do we learn?” referring to a learning activity that may be well or badly performed. When we speak of “drawing lessons from experience,” for example, we implicitly treat “lessons” as learning product and “lesson drawing” as learning process. The product/process ambiguity, which cuts across the many different meanings scholars and practitioners give to learning in general and “organizational learning” in particular, is important to every other question with which we shall be concerned—for example, what forms of organizational learning are desirable, and what constitutes evidence for their existence.

Generically an organization may be said to learn when it acquires information (knowledge, understanding, know-how, techniques, or practices) of any kind and by whatever means. In this overarching sense, all organizations learn, for good or ill, whenever they add to their store of information, and there is no stricture on how the addition may occur. The generic schema of organizational learning includes some informational content, a *learning product*; a *learning process* which consists in acquiring, processing, and storing information; and a *learner* to whom the learning process is attributed. Learning may be attributed to an agent inside or outside the organization, or even to the information itself, as when one says that “new ideas invade an organization.” We may also speak of the particular kind of learning that consists of “unlearning:” acquiring information

that leads to subtracting something (an obsolete strategy, for example) from an organization's existing store of knowledge.

Within this generic schema, an important species of organizational learning consists in an organization's improvement of its task performance over time. Such *instrumental learning* also rests on a schema, which refers to an action's effectiveness in achieving its intended objectives and to criteria and measures for assessing that effectiveness. The action must be of an identifiable type (for example, filling an order, sending a bill, or producing an airplane), for otherwise we could not speak sensibly of its repetition, implicit in the idea of improved performance. There must be an agent that deliberately seeks to improve performance and an intermediate process of deliberate thought and action ("trial and error," for example) through which improvement is achieved. Evidence of improvement consists in data that permit a comparison of performance at different points in time. Instrumental learning is the species of organizational learning favored by economists, implicit in their "learning curves," which originated in the field of industrial engineering.

From a normative perspective, however, instrumental organizational learning should be taken only as a point of departure. Instrumental learning may be good or bad depending on the values used to define "improvement." The distinction between single- and double-loop learning, which we introduce later in this chapter, differentiates instrumental learning within a constant frame of values from learning to change the values that define "improvement." The distinction also differentiates the values of learning outcomes from values inherent in the learning process.

Individual and Organizational Learning

To the distinguished social scientists who were repelled by the idea when we first broached it in the early 1970s, "organizational learning" seemed to smell of some quasi-mystical, Hegelian personification of the collectivity. Surely, they felt, it is *individuals* who may be said to learn, just as to think, reason, or hold opinions. To them, it seemed paradoxical, if not perverse, to attribute learning to *organizations*.

Yet in everyday conversation, as well as in scholarly discourse, it is increasingly common to find people attributing to teams, departments, or whole organizations, such activities as thinking, reasoning, remembering, or learning. They say, for example, "The marketing department realized that sales were about to decline," or "The

administration learned to consult the faculty before announcing a re-organization." In everyday conversation, such statements may be taken as shorthand for more complex, perhaps tacit, processes. But when scholars treat organizational entities as knowers or learners, they seem to be consciously sidestepping the problem of relating individual to organizational phenomena.

How can it seem natural and unproblematic to attribute thinking, knowing, or learning to an organization? Two strategies seem to be involved:

1. adopt the stance of a distant spectator so that the organizational unit or subunit can be seen as a monolithic entity and
2. treat that entity as an impersonal agent.

To treat an organizational entity as an impersonal agent is to adopt a kind of machine language, as when one says, for example, that "an organizational routine survived and established dominance over other routines," or that "general management selected one of the proposals generated by R&D." The increasing use of such language seems to reflect the rising influence of the computer. With the pervasiveness of computers in organizations comes a tendency to employ computer language to refer to phenomena that used to be attributed to thought, will, deliberation, feelings, or habits. People in organizations now say familiarly, "I am in sales mode," "I'm not programmed for this task," "This is our default option." If computers may be said to act, think, remember, or know, and if computational systems may be casually treated as intelligent or stupid, then why not organizations or parts of organizations? The power of the computer metaphor may underlie the growing tendency to treat organizations and their parts as impersonal agents.

A spectator's distant stance toward an organization is consistent with, and perhaps necessary for, its treatment as a monolithic, impersonal agent. For example, economists concerned with the theory of the firm and theorists of business strategy tend to see the business organization as though from a great distance, enabling them to see it whole, but also as a black box. They speak of firms as agents competing with other firms, adopting or changing strategies in order to gain competitive advantage in a market environment. All such distant theories of organizational behavior, operating at high levels of social aggregation, may make useful contributions to economic theory or even to policy analysis. What they do not do or seek to do is to describe and explain the processes within an organization that give

rise to patterns of activity seen, in the aggregate, as the organization's knowing, thinking, remembering, or learning. Nor do such theories say anything about how to achieve "productive learning," however that term may be defined.

If theorists of organizational learning seek to be of use to practitioners, they must somehow link organizational learning to the practitioners' thought and action. And even if they do not wish to be of use to practitioners, as we shall argue in Chapter 9, they should explore these linkages if only to provide a coherent and robust theoretical account of the aggregate organizational phenomena they *do* seek to explain.

If we shift from a distant spectator's viewpoint, from which organizations and their parts can be seen as impersonal agents, and come close enough to become aware of the individual and interpersonal processes that underlie an organization's behavior, we reawaken the paradox of organizational learning. Perhaps this term is a metaphor, on a par with "organizational" force or entropy. Does it make sense to say that organizations *literally* know, remember, think, or learn? How could we test whether they do so? A literal understanding of organizational learning requires that we spell out these processes and conditions.

However troublesome it may be, the paradox that hovers around the notion of organizational learning constitutes its potential interest to organization theory, for it pushes against the boundaries of our usual understandings. In order to explore organizational learning, we must rethink what we mean by organization. We must ask what an organization is that it may be said to learn.

Organizations As Collectivities

One might begin such a quest by arguing that, since organizations are collectivities made up of individuals, they learn something when their individual members, or a substantial fraction of them, learn it. But a moment's thought suggests that such an equivalence cannot hold.

In many cases when the knowledge held by individuals fails to enter into the stream of distinctively organizational thought and action, organizations know *less* than their members do. For example, a social service bureaucracy may continue to operate on the basis of categories of clients, such as "single parents," "abused children," or "dysfunctional families," even though social workers in the agency know very well that these categories fail to capture critically important characteristics of the people they serve. In some cases an organi-

zation cannot seem to learn what *all* of its members know. When a mistake becomes "too big to admit"—in the case of large-scale investments in technology or massive projects of economic development in the Third World—an organization may persist in a course of action that all of its members recognize as foolhardy. Conversely, there are situations in which an organization seems to know far *more* than its individual members. Structures, procedures, and memories built into the fabric of organizations such as the army or the telephone company may permit an organization to perform brilliantly, at least for a time, even when its individual members seem far from brilliant.

Such considerations may suggest that we should think of organizational learning in terms of the "organizational environments" within which individuals think and act. Organizations have been conceived as behavioral settings for human interaction, fields for the exercise of power, systems of institutionalized incentives that govern individual behavior, or socio-cultural contexts in which individuals engage in symbolic interaction. From one or more of these perspectives, we may be able to describe the conditions under which, within an organizational environment, the thought and action of individuals yield organizational learning. But such an approach still leaves us with the problem of linking individual to organizational processes.

We might consider solving this problem by treating organizational learning as the prerogative of a person at the top who learns "for" the organization as a whole. But the bosses of large and complex organizations often complain of being unable to communicate to their subordinates the lessons they themselves have acquired. Bosses may follow one another in rapid succession, while the organization beneath them remains very much the same. And when something that looks like organizational learning occurs, it seems, not infrequently, to have little to do with the person at the top.

Alternatively, we might think of clusters of individual members as the agents who learn "for" the larger organization to which they belong. For example, groups of middle-level employees may discover, in interaction with one another, how to solve a production problem or improve product quality. Yet the learning outcome generated by a group of individuals may not be diffused throughout the larger organization. And even when the results of a group's investigation are broadly diffused, they may not enter into the stream of debates and deliberations that affect an organization's policies, programs, or practices.

We are still left with the problem of determining under what conditions the thought and action of individuals become distinctively organizational.

Organizational Action

The idea of organizational action is logically prior to that of organizational learning, because learning itself—thinking, knowing, or remembering—is a kind of action, and because the performance of an observable action new to an organization is the most decisive test of whether a particular instance of organizational learning has occurred. How can we know what it means for an organization to learn, then, unless we know what it means for it to take action?

Does an organization “act” whenever one of its members acts? If so, there would appear to be very little difference between an organization and the collection of individuals who comprise its members. Yet it is clear that some collections of people constitute organizations and others do not. Furthermore, even when a collection of individuals clearly belongs to an organization, these individuals may do many things (such as eat, sleep, go for walks, gossip with their friends) which do not qualify as instances of organizational action.

Organizations are not merely collections of individuals, yet there is no organization without such collections. Organizational action cannot be reduced to the actions of individuals, even of all the individuals that make up the organization, yet there is no organizational action without individual action. When, then, does it make sense to say that a collection of individuals constitutes an organization that acts?

Consider a mob of students who are holding a spontaneous protest against their university’s financial aid policy. At what point do they cease to be a mob and begin to be an organization? The mob is a collectivity, a collection of people who may run, shout, and mill about together. But it is a collectivity that cannot make a decision or take an action in its own name, and its boundaries are vague and diffuse. The mob begins to resemble an organization as it begins to meet three conditions. The individual members of the mob must

1. devise agreed-upon procedures for making decisions in the name of the collectivity,
2. delegate to individuals the authority to act for the collectivity, and
3. set boundaries between the collectivity and the rest of the world.

As these conditions are met, members of the collectivity begin to become a recognizable “we” that can make decisions and translate their decisions into action.

When the members of the mob become an identifiable vehicle for collective decision and action, they become, in the ancient Greek sense of the term, a *polis*. Before an organization can be anything else, it must be “political,” because it is as a political entity that the collectivity can take organizational action. Then it is the individuals who decide and act, but they do these things on behalf of the collectivity, as its agents. And in order for individuals to be able to decide and act in the name of the collectivity, there must be rules that determine the boundaries of the collectivity, when a decision has been made and when authority for action has been delegated to individuals. Insofar as members of a collectivity create such rules, which we call “constitutional,” and become a *polis*, they have organized.

The rule-making that brings organizations into being need not be conscious, and constitutional rules need not be explicit. What is essential is that the members’ behavior be *rule-governed* in the crucial respects. If the students milling about in front of the dean’s office know how to organize, something they probably learned to do as children playing in a neighborhood, they will spontaneously work out enacted rules for decision, delegation, and membership. Their rules may well remain tacit; there will be no reason to make them explicit unless something happens (a surprise, a crisis, an influx of new people) that calls them into question. So long as there is continuity in the rules that govern the behavior of individual members, the organization will persist, even though some of its members may come and go. And the organization’s existence need not be compromised (might even be enhanced) by the presence of vagueness, ambiguity, or inconsistency in those rules.

By establishing rule-governed ways of deciding, delegating, and setting the boundaries of membership, a collectivity becomes an organization capable of acting. But if we wish to apply this theory of organizational action to the familiar organizations of our society, we must make some further distinctions.

The group of protesting students may form an organization that lasts no longer than their protest; their organization may come to an end, for example, when the dean agrees to hold a university-wide meeting on financial aid. Gian-Francesco Lanzara (1985) used the term, “ephemeral,” to describe such temporary, informal organizations. Ephemeral organizations may arise spontaneously in response

to a crisis, such as the Abruzzi earthquake in Lanzara's example, springing up overnight and disappearing just as quickly. Yet they function for a time as cooperative systems.

As Chester Barnard pointed out (1938), organizations are a species that belong to the genus of systems in which individuals cooperate to perform tasks that arise repetitively (such as making coffee and distributing it to the victims of an earthquake). Every cooperative system embodies a strategy for dividing up, according to one principle or another, the tasks it regularly performs and delegating the components to individual members, thereby establishing organizational roles. The organization's "task system," its pattern of interconnected roles, is at once a division of labor and a design for the performance of work. This design shares the properties of other designed artifacts. It is more or less complex; it involves a multiplicity of variables, values, and constraints; it is subject to variation and change; and it may be represented prior to its enactment—"planned in advance"—or designed and redesigned while in operation.

In contrast to ephemeral organizations, an *agency* is a collection of people that makes decisions, delegates authority for action, and monitors membership, all on a continuing basis. It is a collective vehicle for the regular performance of recurrent tasks. Households, the exemplar of organizations, are agencies, as are the durable, cooperative systems formed by individuals in folk societies. Amish communities, for example, form cooperative building teams that are regularly entrusted with building houses, silos, or corn sheds. Usually they operate without a formal plan or identified leaders. They work out their situation-specific task systems through talk and gestures on the spot, in the presence of site and materials. Such informal agencies are especially interesting because they suggest the existence of culturally specific schemas of organizing that are familiar to all members of the culture and capable of being reproduced again and again, with infinite variation.

Agencies also include the entities we more familiarly treat as organizations: business firms, churches, schools, armies, manufacturing plants, labor unions, social service institutions, or government bureaus. These are formal organizations, whose rules are to some degree explicit and grounded in the legal system of the society. These organizations are formally identified as "legal persons." Their task systems are complex, and they possess in varying degrees the properties that Max Weber attributed to bureaucracies: a clear distinction between persons and the roles they occupy, a complex and detailed

articulation of roles and rules, proceduralized task systems, hierarchical and pyramidally organized layers of authority (Weber, in A.M. Henderson and Talcott Parsons, 1957). Such complex task systems may be tightly or loosely coupled, rigid or variable. All of them, however, fall within our basic definition of the conditions for organizational action: they are cooperative systems governed by the constitutional principles of a polis.

Organizational Inquiry

If a collectivity meets these conditions, so that its members can *act* for it, then it may be said to learn when its members *learn* for it, carrying out on its behalf a process of inquiry that results in a learning product.

We use "inquiry" here not in the colloquial sense of scientific or juridical investigation but in a more fundamental sense that originates in the work of John Dewey (1938): the intertwining of thought and action that proceeds from doubt to the resolution of doubt. In Deweyan inquiry (which we discuss more fully in the following chapter), doubt is construed as the experience of a "problematic situation," triggered by a mismatch between the expected results of action and the results actually achieved. Such a mismatch—a surprise, as we experience it—blocks the flow of spontaneous activity and gives rise to thought and further action aimed at re-establishing that flow.

Inquiry does not become organizational unless undertaken by individuals who function as agents of an organization according to its prevailing roles and rules. Individuals may also inquire in ways that remain separate from the organization to which they belong. Consider, for example, an agency that produces television programs for use in schools. Some of the agency's staff members may wonder what classroom teachers actually do with their programs. Nevertheless, their explorations may remain unconnected to the stream of distinctively organizational activity because their superiors display no interest in learning what happens to the agency's programs so long as schools continue to pay their yearly fees. Conversely, what an organization learns may remain inaccessible to any particular individual, as when staff members are kept ignorant of the reasons for a change in programming policies.

When individual and organizational inquiry do intersect, individual inquiry feeds into and helps to shape organizational inquiry, which then feeds back to shape the further inquiry carried out by

individuals. If, for example, members of the educational television agency get their managers interested in probing how teachers use their programs, the agency may adopt new practices for deciding on program content, which may lead agency staff, in turn, to involve teachers in collaborative program design.

Organizational Knowledge

The output of organizational inquiry may take the form of a change in thinking and acting that yields a change in the design of organizational practices. For example, a telecommunications company may experience a large-scale network failure. Its investigation of the causes of the failure may result in redesigned procedures for developing and maintaining network software or for new ways of detecting threats to network reliability.

Under what conditions does such knowledge become “organizational?” We recognize two distinct but complementary answers to this question.

First, organizations function in several ways as *holding environments for knowledge*, including the knowledge gained through organizational inquiry. Such knowledge may be held in the minds of individual members. If it is held in only this way, it may well be lost to the organization when the relevant individuals leave; an organization may be devastated, for example, by the departure of the one person who really knows how the budget works. But knowledge may also be held in an organization’s files, which record its actions, decisions, regulations, and policies as well as in the maps, formal and informal, through which organizations make themselves understandable to themselves and others. Finally, organizational knowledge may be held in the physical objects that members use as references and guideposts as they go about their business. Roger Barker’s study of behavioral settings (1960) showed how everyday knowledge is embedded in once-familiar places, such as the corner drugstore—in the presence of the soda fountain, the prescription counter, the candy display. Everyone who belongs to the culture in which that setting has its place knows how to deliver the appropriate behavior. More recent studies by Sylvia Scribner and her colleagues (1982) show how workers may use objects in the workplace as things to think with. Scribner describes how veteran milkmen perform the calculations required to fill complicated orders far more rapidly than their school-trained coworkers by making use of visual patterns they discern in

standard milk containers. The results of inquiry carried out by individuals may be embedded in an organization’s holding environment for knowledge in any or all of these ways.

Second, *organizations directly represent knowledge* in the sense that they embody strategies for performing complex tasks that might have been performed in other ways. This is true not only of an overall task system but of its detailed components. Organizational knowledge is embedded in routines and practices which may be inspected and decoded even when the individuals who carry them out are unable to put them into words. For example, a sugar-refining company embodies in its practices particular answers to questions such as how to grow, harvest, and refine cane, or how to distribute and market refined sugar. In this sense, any given organization represents answers to a set of questions or solutions to a set of problems.

Such organizational task knowledge may be variously represented as systems of beliefs that underlie action, as prototypes from which actions are derived, or as procedural prescriptions for action in the manner of a computer program. We have chosen to represent such knowledge through what we call “theories of action,” which have the advantage of including strategies of action, the values that govern the choice of strategies and the assumptions on which they are based. We define a theory of action in terms of a particular situation, *S*, a particular consequence, intended in that situation, *C*, and an action strategy, *A*, for obtaining consequence *C* in situation *S*. The general form of a theory of action is: If you intend to produce consequence *C* in situation *S*, then do *A*. Two further elements enter into the general schema of a theory of action: the values attributed to *C* that make it seem desirable as an end-in-view and the underlying assumptions, or model of the world, that make it plausible that action *A* will produce consequence *C* in situation *S*.

Theory of action, whether it applies to organizations or individuals, may take two different forms. By “espoused theory” we mean the theory of action which is advanced to explain or justify a given pattern of activity. By “theory-in-use” we mean the theory of action which is implicit in the performance of that pattern of activity. A theory-in-use is not a “given.” It must be constructed from observation of the pattern of action in question. From the evidence gained by observing any pattern of action, one might construct alternative theories-in-use which are, in effect, hypotheses to be tested against the data of observation. In the case of organizations, a theory-in-use must be constructed from observation of the patterns of interactive

behavior produced by individual members of the organization, insofar as their behavior is governed by formal or informal rules for collective decision, delegation, and membership.

We can say that the values, action strategies, and assumptions embedded in the sugar refining company's routine activities constitute its theory-in-use for cane growing, for distributing and marketing, and for its other functions. Taken together, these components make up the organization's *instrumental theory-in-use*. This instrumental theory includes norms for corporate performance (for example, margin of profit or return on investment), strategies for achieving values (for example, strategies of plant location or selection of manufacturing technology), and assumptions that bind strategies and values together (for example, the assumption that maintenance of a high rate of return on investment depends on the continual introduction of new technologies). An organization's instrumental theory of action includes in its scope communication and control, allocating resources to particular functions, rewarding or punishing individual performance, constructing career ladders and regulating the rates at which individuals climb them, and recruiting new members and instructing them in the ways of the organization.

Like the rules for collective decision and action, organizational theories-in-use may be tacit rather than explicit and tacit theories-in-use may not match the organization's espoused theory. An organization's formal documents, such as organization charts, policy statements, or job descriptions, not infrequently contain espoused theories of action incongruent with the organization's actual patterns of activity.

Organizational theory-in-use may remain tacit because it is indescribable or undiscussable. It may be indescribable because the individual members who enact it know more than they can say and are unable, rather than unwilling, to describe the know-how embedded in their day-to-day performance of organizational tasks. It may be undiscussable because any attempt to reveal its incongruity with the organization's espoused theory would be perceived as threatening or embarrassing.

Whatever the reasons for its tacitness, an organization's theory-in-use largely accounts for its identity over time. Consider a large and enduring organization, such as the United States Army. Over a period of fifty years or so, the Army's personnel turn over completely. Still, we speak of "the Army." If it no longer consists of the same people, in what sense does it remain the same? In an attempt to

answer this question, we might examine artifacts such as uniforms and weapons. Over a period of fifty years these are likely to have undergone radical change. We might examine the societal functions carried out by the Army. But even if we discover that these functions have remained substantially in force for fifty years, we leave open the possibility that the organization now fulfilling them is radically different from its earlier counterpart. If, on the other hand, we study the fifty-year evolution of the theory-in-use implicit in the Army's military practices and discover that certain of its strategies, values, and assumptions—say, for command and control, promotion, and training—have remained relatively constant throughout that period, then we would have to treat these as most central to the Army's continuity of identity. The Army's theory-in-use, seemingly an abstract idea, would then be the "realest" thing about it.

One may describe an organization's theory-in-use from the point of view of an outsider or an insider to the organization. An outsider would have to observe how the organization's task system is enacted through the rule-governed behavior of its members. Insiders (as we will discuss at greater length in Chapter 8) have some access to the know-how through which they generate and control the practices appropriate to the organization's task system. This know-how may take the form of procedural knowledge, such as rules of thumb or the members' grasp of various categories of situations and behavior appropriate to them or their spontaneous perceptions of "the right thing to do now." Such representations constitute the organization's theory-in-use as seen from the inside.

Organizational Learning Again

Drawing on these ideas of organizational action, inquiry, and knowledge, we can now describe more precisely what we mean by organizational learning.

Each member of an organization constructs his own representation of the theory-in-use of the whole, but his picture is always incomplete. He strives continually to complete his picture by redescribing himself in relation to others in the organization. As conditions change, he remakes his descriptions; other individuals do likewise. There is a continual, more or less concerted meshing of individuals' images of their activity in the context of their collective interaction.

An organization is like an organism, each of whose cells contains a particular, partial, changing image of itself in relation to

the whole. And like such an organism, the organization's practice stems from these very images: its theory-in-use is dependent on the ways in which its members represent it. Hence, our exploration of organizational learning must deal not with static entities called organizations but, as Karl Weick pointed out (1969), with active processes of organizing. The members' evolving images of the organization shape the very object of their investigation.

Organizational continuity would not be understandable if it depended exclusively on multiple, parallel, private imaging. When organizations are large and complex, their members cannot rely entirely on face-to-face contact to help them compare and adjust their private images of organizational theory-in-use. Even in face-to-face contact, private images of organization often diverge. Individuals need external references to guide their private adjustments.

Such reference functions are fulfilled by organizational maps, memories, and programs. Examples of maps include diagrams of work flow, organization charts, and drawings or photographs of the workplace. An actual building may serve as a map of groupings of individuals and patterns of communication among them. Organizational memories include files, records, data bases, and financial accounts, as well as the physical objects (tools, products, or working materials) that hold organizational knowledge. Programs are procedural descriptions of organizational routines; they include work plans, policies, protocols, guidelines, scripts, and templates. Artifacts such as these describe existing patterns of activity and serve as guides to future action.

Organizational learning occurs when individuals within an organization experience a problematic situation and inquire into it on the organization's behalf. They experience a surprising mismatch between expected and actual results of action and respond to that mismatch through a process of thought and further action that leads them to modify their images of organization or their understandings of organizational phenomena and to restructure their activities so as to bring outcomes and expectations into line, thereby changing organizational theory-in-use. In order to become organizational, the learning that results from organizational inquiry must become embedded in the images of organization held in its members' minds and/or in the epistemological artifacts (the maps, memories, and programs) embedded in the organizational environment.

The learning products of organizational inquiry may take many forms, all of which, to qualify as learning must include evi-

dence of a change in organizational theory-in-use. Often such changes are mediated by lessons drawn from inquiry. These include

- a. interpretations of past experiences of success or failure;
- b. inferences of causal connections between actions and outcomes and their implications for future action;
- c. descriptions of the shifting organizational environment and its likely demands on future performance;
- d. analysis of the potentials and limits of alternative organizational strategies, structures, techniques, information systems, or incentive systems;
- e. descriptions of conflicting views and interests that arise within the organization under conditions of complexity and uncertainty;
- f. images of desirable futures and invention of the means by which they may be achieved;
- g. critical reflections on organizational theories-in-use and proposals for their restructuring; and
- h. description and analysis of the experiences of other organizations.

All such intermediate outcomes of inquiry qualify as products of organizational learning when they are accompanied by changes in behavior that signify changes in organizational theory-in-use and when they are embodied in the individual images (the memories, maps, or programs) that store organizational knowledge.

Near Misses

It may help, at this point, to clarify our intended meaning by considering some examples that are "near misses"—almost but not quite organizational learning. We have already mentioned the case in which members of an organization gain new insights that are not converted to action. There are also cases in which individuals' inquiry leads to both new understandings *and* action but remains outside the stream of distinctively organizational activity and produces no change in organizational theory-in-use; for example, an individual or small group becomes an underground champion of an innovation in organizational policy, technology, or practice. A case in point was McLain's unauthorized, secret development of the first prototype of the Sidewinder missile in a shed at the Naval Research Station at Indio, California, in the 1950s. Such a development may become

organizational and its learning product an organizational one, if (eventually as happened with McLain) the covert project is discovered, publicly legitimized, and formally adopted by the organization.

There are instances in which organizational inquiry produces a temporary change in organizational theory-in-use, but the new understandings associated with that change, held only in the minds of certain individual "carriers," are lost to the organization when they leave. This often happens in small professional organizations, such as research and development firms, design offices, or software companies, where staff members habitually move in and out of organizational homes, taking their ideas and capabilities with them. Sometimes members of the host organization recognize the risk of loss and seek deliberately to uncover and document the special insights and skills of these organizational birds of passage.

Finally, it is worth noting that not all changes in organizational theory-in-use qualify as learning. For example, changes in an organization's environment (such as a slackening of product demand) may trigger patterns of response that undermine organizational norms. Members may lose enthusiasm, become sloppy in task performance, or lose touch with one another. These kinds of changes are forms of deterioration. On the other hand, as in an example we will describe in Chapter 9, an episode of organizational inquiry may provoke deterioration in some regions of organizational theory-in-use even as it generates learning in another.

All such near misses suggest that our category of organizational learning has fuzzy edges. There are boundary instances in which it is difficult to determine with precision whether interactive inquiry is truly organizational or has truly changed organizational theory-in-use or whether its results have been truly embedded in the organization's memories, maps, and programs. Such vagueness may be inherent in organizational phenomena that are ill-formed or emergent or may reflect, on the other hand, a lack of information sufficient to permit a clear determination. They do not invalidate our definition as long as many examples do clearly fall inside it and as long as what information we would need in order to make a definite attribution of organizational learning is clear.

Productive Organizational Learning

There are several ways in which instrumental learning may be for ill rather than for good. Some of these are particular to organizational learning; others, applicable to learning by agents of any kind.

First, the *ends* of action may be reprehensible. The value we attribute to an increase in effectiveness or efficiency depends on how we answer the question, Effectiveness or efficiency for what? and how we evaluate the "what." This issue is critically important when the action in question emanates from an organization whose members are eager or unthinkingly compliant participants. During World War II, Eichman's bureaucracy learned over time to become more efficient at sending its victims to the gas chambers.

The value attributed to a particular instance of learning also depends on how we judge its validity. Learning seems to suggest the acquisition of valid, workable knowledge or know-how. But when we treat organizational learning as inquiry that leads to a change in theory-in-use, we open up the possibility that any given change may be based on a lesson that turns out to be false or unworkable. James March (1988) uses the term "superstitious learning" to refer to one such class of lessons: those based on the belief that because events have followed one another in time they are also related to one another as cause to effect. For example, corporate managers' may believe that a rise in profits following the institution of a new policy must have been caused by that policy, though it may have been due to nothing more than an improvement in market conditions. March suggests that managers are drawn to superstitious learning because it reinforces the myth of managerial control—a belief congenial to the norms of managerial stewardship but often contrary to fact.

Organizational learning that is valid or workable at the time of its first occurrence may lead to effects that are negative overall. To take a notable example, "competence traps" (also March's term) are situations in which an experience of perceived success leads an organization to persist in a familiar pattern of thought and action beyond the time and conditions within which it yields successful outcomes. The behavior that yields success at time, t , may not yield it at $t + 1$. Yet an organization lulled by its success and misguided by the lessons drawn from it, may persist in a familiar pattern of behavior long after it has ceased to work. In business strategy, General Motors, IBM, and Digital Equipment Corporation come to mind as recent examples of firms that persevered in following a once-winning strategy that had become a losing strategy, apparently blind to the fact that the competitive environment had shifted out from under it. Such examples should be understood in terms of the webs of interest organizations build up around familiar strategies, technologies, or structures, and the "dynamically conservative" processes (Schön, 1967) that

reinforce an organization's adherence to the lessons it has drawn from past experience.

Later we will have opportunity to see how people can learn collectively to maintain patterns of thought and action that *inhibit* productive organizational learning. For example, they may learn to respond to error by the use of scapegoating, games of unilateral control and avoidance of control, systematic patterns of deception, camouflage of intentions, and maintenance of taboos that keep critical issues undiscussable. Such patterns of thought and action, learned from experience, often have the effect of inhibiting the kinds of productive learning that yield improved performance or restructured values for performance. Yet members of the organization may develop an attachment to these patterns, even to the point of exclaiming, "It has taken us years to learn to live in this screwed-up world; don't make waves!"

If we were to use learning only in a positive sense, then we would have to qualify the learning involved in all such negative examples with adjectives like dysfunctional, pseudo, or limited. These semantic devices are misleading, since they tend to be applied to learning products after the fact; whereas we are often uncertain in any given situation of action, whether an alleged instance of productive organizational learning is valid and workable. The crucial point is that, as we try to understand or enhance organizational learning, we should keep in mind the variety of ways in which any particular example of it may prove to be invalid, unproductive, or even downright evil.

For these reasons, it is useful to distinguish three types of productive organizational learning:

1. *organizational inquiry*, instrumental learning that leads to improvement in the performance of organizational tasks;
2. inquiry through which an organization explores and restructures the values and criteria through which it defines what it means by improved performance; and
3. inquiry through which an organization enhances its capability for learning of types (1) or (2).

Single- and Double-Loop Learning

By **single-loop learning** we mean instrumental learning that changes strategies of action or assumptions underlying strategies in ways that leave the values of a theory of action unchanged. For example, qual-

ity control inspectors who identify a defective product may convey that information to production engineers, who, in turn, may change product specifications and production methods to correct the defect. Marketing managers, who observe that monthly sales have fallen below expectations, may inquire into the shortfall, seeking an interpretation they can use to devise new marketing strategies to bring the sales curve back on target. Line managers may respond to an increase in turnover of personnel by investigating sources of worker dissatisfaction, looking for factors they can influence, such as salary levels, fringe benefits, or job design, to improve the stability of their work force.

In such learning episodes, a single feed-back loop, mediated by organizational inquiry, connects detected error—that is, an outcome of action mismatched to expectations and, therefore, surprising—to organizational strategies of action and their underlying assumptions. These strategies or assumptions are modified, in turn, to keep organizational performance within the range set by existing organizational values and norms. The values and norms themselves (related in the previous examples to product quality, sales level, or work force stability) remain unchanged.

By **double-loop learning**, we mean learning that results in a change in the values of theory-in-use, as well as in its strategies and assumptions. The double loop refers to the two feedback loops that connect the observed effects of action with strategies and values served by strategies. Strategies and assumptions may change concurrently with, or as a consequence of, change in values.¹ Double-loop learning may be carried out by individuals, when their inquiry leads to change in the values of their theories-in-use or by organizations, when individuals inquire on behalf of an organization in such a way as to lead to change in the values of organizational theory-in-use.

¹ We borrow the distinction between single- and double-loop learning from W. Ross Ashby's *Design for a Brain* (New York: John Wiley and Sons, Inc., 1960). Ashby formulates his distinction in terms of (a) the adaptive behavior of a stable system, "the region of stability being the region of the phase space in which all the essential variables lie within their normal limits," and (b) a change in the value of an effective parameter, which changes the field within which the system seeks to maintain its stability. One of Ashby's examples is the behavior of a heating or cooling system governed by a thermostat. In an analogy to single-loop learning, the system changes the values of certain variables (for example, the opening or closing of an air valve) in order to keep temperature within the limits of a setting. Double-loop learning is analogous to the process by which a change in the setting induces the system to maintain temperature within the range specified by a new setting. See especially pp. 71-75.

Organizations continually engaged in transactions with their environments regularly carry out inquiry that takes the form of detection and correction of error. Single-loop learning is sufficient where error correction can proceed by changing organizational strategies and assumptions within a constant framework of values and norms for performance. It is instrumental and, therefore, concerned primarily with effectiveness: how best to achieve existing goals and objectives, keeping organizational performance within the range specified by existing values and norms. In some cases, however, the correction of error requires inquiry through which organizational values and norms themselves are modified, which is what we mean by organizational double-loop learning.

In any particular instance of double-loop learning, the resulting changes in values and norms may not be judged to be desirable: their desirability can be determined only through a situation-specific critique of the changes themselves and of the inquiry through which they are achieved. Nevertheless, it is through double-loop learning alone that individuals or organizations can address the desirability of the values and norms that govern their theories-in-use.

Consider a chemical firm which has set up a research and development division charged with the discovery and development of new technologies (an example we consider at greater length in Chapter 3). The firm has created its new R&D division in response to the perceived imperative for growth in sales and earnings and the belief that these are to be generated through internally managed technological innovation. However, the new division generates technologies that do not fit the corporation's familiar pattern of operations. In order to exploit some of these technologies, the corporation may have to turn from the production of intermediate materials, with which it is familiar, to the manufacture and distribution of consumer products with which it is unfamiliar. This, in turn, requires that members of the corporation adopt new approaches to marketing, managing, and advertising; that they become accustomed to a much shorter product life cycle and to a more rapid cycle of changes in their pattern of activities; that they, in fact, change the very image of their business. And these requirements for change come into conflict with another sort of corporate norm, one that requires predictability in the management of corporate affairs.

Hence, the corporate managers find themselves confronted with conflicting requirements. If they conform to the imperative for growth, they must give up on the imperative for predictability. If they

decide to keep their patterns of operation constant, they must give up on the imperative for growth, insofar as that imperative is to be realized through internally generated technology. A process of change initiated with an eye to effectiveness under existing norms turns out to yield a conflict in the norms themselves.

If corporate managers are to engage this conflict, they must undertake a process of inquiry which is significantly different from the inquiry characteristic of single-loop learning. To begin, they must become aware of the conflict. They have set up a new division that has yielded unexpected outcomes; this is an error, a surprise. They must reflect upon this surprise to the point where they become aware that they cannot deal with it adequately by doing better what they already know how to do. They must become aware that they cannot correct the error by getting the new division to perform more efficiently under existing norms; the more efficient the new division is, the more its results will plunge the managers into uncertainty and conflict. The managers must discover that it is the norm for predictable management which they hold, perhaps tacitly, that conflicts with their wish to achieve corporate growth through technological innovation.

Then the managers must undertake an inquiry that resolves the conflicting requirements. The results of their inquiry will take the form of a restructuring of organizational norms and very likely a restructuring of strategies and assumptions associated with those norms; these must then be embedded in the images and maps that encode organizational theory-in-use. There is in this sort of episode a double feedback loop which connects the detection of error not only to strategies and assumptions of effective performance but to the values and norms that define effective performance.

In such an example of organizational double-loop learning, incompatible requirements in organizational theory-in-use are characteristically expressed through a conflict among members and groups of members. One might say that the organization becomes a medium for translating incompatible requirements into interpersonal and intergroup conflict.

For example, some managers of the chemical firm may become partisans of growth through research; while others, committed to familiar and predictable patterns of corporate operation, become opponents of the new, research-based conception of the business. Double-loop learning, if it occurs, will follow from the process of inquiry by which these groups of managers confront and resolve their

dispute. They may respond in several ways, not all of which meet the criteria for organizational double-loop learning.

First, the members may treat the conflict as a fight in which choices among competing requirements are to be made, and weightings and priorities are to be set on the basis of dominance. The R&D faction, for example, may include the chief executive who is able to win out over the old guard because of his greater power, or the two factions may fight it out to a draw, settling their differences in the end by a compromise that reflects nothing more than the inability of either faction to prevail over the other. In both of these cases, the conflict is settled for the time being but not by a process that could be appropriately described as learning. If the conflict ends with a power play or a stalemate, neither side is likely to emerge with a new sense of the nature of the conflict, its causes and consequences, or its meaning for organizational theory-in-use.

On the other hand, the adversaries may engage their conflict through inquiry in any of the following ways:

- a. They may invent new strategies of performance that circumvent the perceived incompatibility of requirements; they may succeed in defining a kind of research and development addressed solely to the existing patterns of business that offer the likelihood of achieving existing norms for growth. They will then have succeeded in finding a single-loop solution to what at first appeared a double-loop problem.
- b. They may carry out a trade-off analysis that enables them to conclude jointly that so many units of achievement of one norm are balanced by so many units of achievement of another norm. On this basis, they may decide that the prospects for R&D payoff are so slim that the R&D option should be abandoned, and with that abandonment there should be a lowering of corporate expectations for growth. Or they may decide to limit R&D targets so that the disruptions of patterns of business operation generated by R&D are confined to particular segments of the corporation. Here there is a compromise among competing requirements, but it is achieved through inquiry into the probabilities and values associated with options for action.
- c. The incompatible requirements may be perceived as incommensurable. In such a case, the conflict may still be resolved through inquiry that gets underneath the members' initial commitments. Participants must then ask why they

hold the positions they do and what the positions mean. They may ask what factors have led them to adopt particular standards for growth in sales and earnings, with what rationales, and what are likely to be the consequences of attempting to meet the standards by any means whatever. Similarly they may ask what kinds of predictability in operations are of greatest importance, to whom they are important, and what conditions make them important.

Inquiry of type B or C may lead to a restructuring of corporate values and norms. Or it may lead to the invention of new patterns of incentives, budgeting, and control that take greater account of requirements for both growth and predictability.

In this type of organizational double-loop learning, individual members resolve interpersonal and intergroup conflicts that express incompatible requirements for organizational performance. They do so through organizational inquiry that creates new understandings of the conflicting requirements—their sources, conditions, and consequences—and sets new priorities and weightings of norms, or re-frames the norms themselves, together with their associated strategies and assumptions. In such a process the restructured requirements for organizational performance become more nearly compatible and more susceptible to effective realization. And the resulting understandings, priorities, and reframed norms become inscribed in the images, maps, and programs of the organization and are thereby embedded in organizational memory.

Additional Considerations

The distinction between single- and double-loop learning is complicated by several factors. As we consider these, we identify gradients of significance in organizational learning, become aware of zones of ambiguity at the boundaries of these two types of learning, and identify a variety of patterns of inquiry through which organizations may engage in double-loop learning.

First, the distinction between single- and double-loop learning is complicated by organizational size and complexity.

Organizational theories-in-use are structures composed of many interconnected parts. Some of these are local and peripheral, while others are core elements fundamental to the structure as a whole. In a chemical firm, for example, norms governing requirements for growth and predictability are fundamental to the theory-in-use of the whole organization. If these norms were to change, a great

deal of the rest of the theory-in-use would also have to change; this secondary change is why their conflict is so important for the whole firm. On the other hand, a particular norm for product quality could change without affecting much of the rest of the organization's theory-in-use. From such observations, we infer that double-loop organizational learning may be of greater or lesser significance for the organization as a whole, depending on the degree to which core values and norms are involved.

Furthermore, large organizations are composed of many layers. Such organizations can be described in terms of a *ladder of aggregation* that proceeds from individuals to small groups, to departments made up of many small groups, to divisions that are clusters of departments, to the organization as a whole, to the larger field in which the organization interacts with other organizations. These organizational layers exist not only as abstractions but as living entities each of which may be described as having interests, intentions, values, and theories-in-use of its own. From the point of view of each such entity, the rest of the organization is environment. An organization may be said to act, interact, inquire, and learn; so may the groups, departments, and divisions at different levels of aggregation within it.

Often the actions of intraorganizational units are crucial to organizational inquiry and consequential for single- and double-loop learning. Learning may be more or less contained within an organizational unit, depending on how tightly or loosely that unit is coupled with others. For example, a change in the technology of a production line may change values that guide behavior in that workplace without repercussions on the larger organization's theory-in-use. In contrast, a bank's decision to introduce just-in-time paper processing in one of its divisions may provoke a shift in the norms by which the bank's control system perceives, evaluates, and rewards production in all of its divisions. In this instance, what begins as single-loop learning at one level of aggregation stimulates double-loop learning at all levels. More generally, the type of organizational learning that occurs may vary with the level of aggregation at which it occurs and the tight or loose coupling of units within or across levels.

A second factor that complicates the distinction between single- and double-loop learning is the relationship between learning products and processes. We have so far defined single- and double-loop learning in terms of the *products* of organizational inquiry, distinguishing a change in organizational theory-in-use that affects only

strategies of action and assumptions from one that affects values. But we are also concerned with values and norms that govern *processes* of organizational inquiry, for these are critical to an organization's capability for improving its performance and restructuring the values that define improvement.

As we try to determine whether an instance of organizational learning is single- or double-loop, it is important to notice not only where inquiry begins but where it goes. For example, the managers of a firm may decide that in order to gain market share, their organization needs to become far more nimble and proactive in its responses to threats and opportunities. Starting with this goal, they may propose to create a flatter and more decentralized organization in which local units take on much greater freedom of action and display higher levels of initiative. As the managers begin to implement the new organizational plan, they may well discover that its success depends on creating an organizational climate tolerant of public dissent and debate, risk-embracing, and hospitable to decision making under uncertainty. These values may have been included in the organization's espoused theory of action but not in its theory-in-use. The original initiative, framed as one of increasing organizational effectiveness, turns out to have critical implications for double-loop changes in core values that govern theory-in-use for the conduct of organizational inquiry.

Consider a related example, which we will take up at greater length in Chapter 9: an organization introduces a program of Total Quality Management (TQM) (Argyris, 1994). As individuals working within this program search out the "root causes" of defects in a product or process, they may identify two different kinds of problems. Inefficiencies in production represent one kind of problem; the other is illustrated by a group of employees who stand passively by and watch inefficiencies develop and persevere. TQM may produce the simple learning necessary to effect a solution to the first problem, but it is not likely to prevent a recurrence of the second or cause the supervisors to wonder why they never acted.

Double-loop learning in organizational inquiry calls for an additional step or even several additional steps. It turns the question back on the questioner, exploring not only the objective facts surrounding an instance of inefficiency, but also the reasons and motives behind those facts. For example, a CEO who discovers in his organization the practice of requiring 275 sign-offs for the approval of an innovation, might ask, "How long have you known about these

requirements?" "What prevented you from questioning these practices?" Such double-loop learning depends on questioning one's own assumptions and behavior.

The distinction between double-loop learning outcomes for organizational theory-in-use and double-loop learning in processes of organizational inquiry is correlated with the distinction between *first- and second-order errors*. First-order errors in organizational theory-in-use are illustrated by excessive costs or too many sign-offs. The second-order errors that arise in processes of organizational inquiry, such as a failure to question existing practices, allow such first-order errors to arise and persist. Double-loop learning in organizational inquiry consists in the questioning, information-gathering, and reflection that get at second-order errors. When it is successful, it results in change toward values for inquiry that yields valid and actionable learning about second-order error. As we shall show in Part II, such changes are closely linked to an organization's "learning system" and the individual theories-in-use that both reinforce it and are reinforced by it.

Organizational Deuterolearning

An organization's *learning system* is made up of the structures that channel organizational inquiry and the behavioral world of the organization, draped over these structures, that facilitates or inhibits organizational inquiry. Together, structural and behavioral features of an organizational learning system create the conditions under which individuals interact in organizational inquiry, making it more or less likely that crucial issues will be addressed or avoided, that dilemmas will be publicly surfaced or held private, and that sensitive assumptions will be publicly tested or protected.

By "organizational structures," we mean

- channels of communication (forums for discussion and debate, formal and informal patterns of interaction);
- information systems, including their media and technologies (the computer, for example);
- the spatial environment of the organization insofar as it influences patterns of communication;
- procedures and routines that guide individual and interactive inquiry; and
- systems of incentives that influence the will to inquire.

Insofar as such structures facilitate organizational inquiry, we speak of them as *enablers*.

By the "behavioral world" of the organization, we mean the qualities, meanings, and feelings that habitually condition patterns of interaction among individuals within the organization in such a way as to affect organizational inquiry—for example, the degree to which patterns of interaction are friendly or hostile, intimate or distant, open or closed, flexible or rigid, competitive or cooperative, risk-seeking or risk-averse, error-embracing or error-avoiding, productive or defensive. A key feature of the behavioral world of an organization is the degree to which organizational inquiry tends to be bound up with the win/lose behavior characteristic of organizational games of interests and powers. These games are usually intertwined with organizational inquiry. At the extreme, as Michel Crozier demonstrated (1963), surprises that might trigger productive organizational learning may be overwhelmingly interpreted in the light of their meaning for the status of players within games of interests and powers.

An organization's learning system is interdependent with the theories-in-use that individuals bring to its behavioral world. Individual theories-in-use help to create and maintain the organization's learning system; this system, in turn, contributes to the reinforcement or restructuring of individual theories-in-use. For example, when individuals operate in terms of "mystery and mastery," keeping their intentions and strategies private while they seek to master their interactions with others, they tend to engender distrust, which may then be widely perceived as a consistent feature of the organization's behavioral world. And a behavioral world characterized by distrust tends to reinforce the disposition of individuals to act according to theories-in-use that feature win/lose behavior and unilateral self-protection.

A critically important kind of organizational double-loop learning, therefore, is the second-order learning through which the members of an organization may discover and modify the learning system that conditions prevailing patterns of organizational inquiry.

This is the organizational equivalent of what Gregory Bateson (1972) calls **deuterolearning** by which he means second-order learning, or "learning how to learn," organizational deuterolearning. We shall describe this in terms of a shift from O-I to O-II learning systems. Organizational deuterolearning is critically dependent on individual deuterolearning, which we shall describe in terms of a shift from Model I to Model II theories-in-use. These correlated shifts are at the heart of the concerns that have led us to write this book.

2

Turning the Researcher/ Practitioner Relationship On Its Head

What does an inquiry-based view of organizational learning imply for the relationship between practitioners and researchers? If we see practitioners as inquirers who are called upon in their day-to-day work to detect and correct errors in organizational performance, how should we think about their relationship to those who do academic research on organizational learning? Depending on our answer to this question, what are the implications for the appropriate roles, attitudes, and methods of academic researchers?

In this chapter we argue that when we see both organizational practice and academic research as forms of inquiry, we can reframe the conventional view of their relationship in a way that promotes both usable knowledge and robust research. We will no longer see this relationship in terms of practitioners' application of knowledge generated by researchers but as a collaboration between types of inquirers who occupy different roles and rely on different but complementary skills and methods. Central to this reframing is the Deweyan idea of inquiry and, within it, a recognition of the different ways in which researchers and practitioners treat causality and causal inference.

Deweyan Inquiry

Dewey's idea of inquiry derives in part from the writings of Charles Peirce, the founder of American Pragmatism; but whereas Peirce (1877) treated doubt as a property of individual consciousness, Dewey (1938) believed we doubt because we are in a situation that is inherently doubtful. He thought that inquiry begins with an indeterminate, problematic situation, a situation whose inherent conflict, ob-

scurity, or confusion blocks action. And the inquirer seeks to make that situation determinate, thereby restoring the flow of activity.

Inquiry for Dewey combines mental reasoning and action. The Deweyan inquirer is not a spectator but an actor who stands within a situation of action, seeking actively to understand and change it. When inquiry results in a learning outcome, it yields both thought and action, at least in some degree new to the inquirer.

Both doubt and its resolution are transactional properties of the relationship between the inquirer and the situation; the inquirer participates in constructing the situation to which he also responds. For example, an artist makes the painting in which, at any given moment, she finds requirements and possibilities that call for further making. The problems and potentials in an interpersonal relationship that generate feelings and call for new thought and action, are always, at least in part, of the participants' making.

The transaction between inquirer and situation is continuing and inherently open-ended. As inquirers seek to resolve what is problematic about a situation of action, they bring new problematic features into being. Inquiry "does not merely remove doubt by recurrence to a prior adaptive integration," as Dewey put it, but "institutes new environing conditions that occasion new problems." Within such a dialectic (a term whose organizational counterpart we shall explore in the following chapter), there is, in Dewey's words, "no such thing as a final settlement." Inquiry is to be tested by its success in resolving a problematic situation and by the value inquirers come to attribute to the new problems their resolution creates.

Detecting and Correcting Error

It is the detection of error, which we define as the mismatch of outcomes to expectations, that triggers awareness of a problematic situation and sets in motion the inquiry aimed at correcting the error. When the outcomes of our action are mismatched to expectations, the inquirer gets an experience of surprise—an experience essential, as Israel Scheffler wrote (1987), to the process of coming to think and act in a new way. The attempt to resolve a problematic situation frequently generates new sources of surprise.

Consider what happens when a telephone network suddenly breaks down. In the company responsible for the network, people look for ways to restore it to full operation. If they do not immediately figure out how to do so, they try to devise methods of experimentation that will not make things worse. They investigate the

causes of the breakdown and explore ways of preventing the future occurrence of such disruptions. How they understand the causes and how they frame the problem of reliability guide their invention of short-term fixes and longer-term strategies of prevention.

Not infrequently, their attempts to carry out their problem-solving strategies reveal flaws in their initial framing of the problem. If they locate the immediate cause of the breakdown in an error buried deep in millions of lines of network software, they may first direct their attention to the process of software-testing. Further inquiry may lead them to frame a deeper problem. They may discover that as they have tried to improve network performance, they have introduced systems that increase the complexity of the network software, making it more vulnerable to error. The discovery of this more fundamental problem, an inconsistency in the objectives for network design, may call for a more fundamental solution. Inquiry, triggered in the first instance by surprise, generates new surprises that call for new rounds of error detection and correction.

It is important to note, however, that the term, "error," which we define as "a mismatch of outcomes to expectations," tends in ordinary usage to suggest a *mistake*, an invalid strategy of action or assumption, for example, the mistaken belief that better software-testing tools will assure network reliability. Getting from an error to the mistake that underlies it requires a further process of inquiry, as in the telecom staff's analysis of the causes of the network breakdown. Errors are not in themselves mistakes; rather, they signal the presence of mistakes.

It is also important to notice that although error tends in ordinary usage to suggest a negative outcome, it may actually have a positive meaning. A surprising outcome of action is perceived, on occasion, as desirable. In the realm of science, a researcher may stumble onto a surprising phenomenon, such as the unexpected bloom of mold on a petrie dish that led to Alexander Fleming's discovery of penicillin, and thereby detect an error that leads to the development of a new technology. In the interpersonal domain, a manager may be pleasantly surprised to find a colleague actually receptive to direct confrontation on a threatening issue.

Whether inquirers perceive a surprising outcome of action as negative or positive, they try to correct the error, to realign outcomes and expectations so as to convert a mismatch to a match. In one case, they try to reshape what they perceive as an undesirable outcome in order to make it conform to their original expectations; in the other,

they realign their expectations and intentions in order to conform to the happy outcome. Hence, when we describe inquiry in terms of the detection and correction of error, we are not claiming that one learns more readily from failure than from success; we are emphasizing the role of surprise as a stimulus to new ways of thinking and acting.

Organizational Inquiry

Dewey saw inquiry as a social process. He meant not only that people usually think and act together in a social setting, but that the very process of inquiry, individual or collective, is conditioned by membership in a social system that establishes inquiry's taken-for-granted assumptions. Like Peirce, Dewey saw individual inquirers as members of a "community of inquiry," bound by contractual responsibilities. He wrote that an inquirer enters into a "contract such that...he is committed to stand by the results of similar inquiries." In just this way, we see organizational learning as a process carried out by members of an organization, working alone or in interaction with one another, within an *organizational* community of inquiry.

Inquiry becomes organizational when individuals inquire on behalf of the organization, within a community of inquiry governed, formally or informally, by the roles and rules of the organization.

It follows that individuals may inquire and learn in ways that are connected to and, at times, disconnected from the organization to which they belong.

When organizational inquiry leads to learning, its results are manifested in thought and action that are in some degree new to the organization. In instrumental learning, organizational inquiry yields new ways of thinking and acting that enable the improved performance of an organizational task. In this sense, the attribution of organizational learning is contingent on the presence of an observable change in behavior. To be sure, behavior may change in ways other than through learning, for example, through deterioration, forgetting, or random variation. But such observations show only that change in behavior is not a sufficient condition for learning. We argue that it is a *necessary* one, however. The action that resolves a problematic situation is what Dewey would call the "end-in-view" of inquiry, the purpose that sets it in motion; and it is by reference to such an action that we can judge whether organizational inquiry has been effective. It is not easy to imagine how we could confirm an occurrence of

organizational learning without observing a change in behavior. It is true that individuals who puzzle over a phenomenon may gain important insights that remain dormant for long periods of time. For example, members of a consumer product company may detect changes in the marketplace that portend a shift in the ground rules governing competition for market share, yet their insights may lie fallow for months or may never find their way into the design of new product strategies. Such insights fall short of learning, since they do not result in new forms of organizational action. They are best seen as representing a *potential* for organizational learning.

The Researcher/Practitioner Relationship

When we define organizational learning in terms of the inquiry practitioners carry out within an organizational setting, we point toward what they and academic researchers hold in common: both are inquirers, concerned with detecting and correcting errors, making sense of confusing and conflictual problematic situations. This is in contrast to a more conventional emphasis on a critical difference between the two groups. According to the conventional view, which emanates mainly from the research universities, research gives rise to special expertise of the sort that Everett Hughes (1959) called "esoteric knowledge." The relationship between researchers and practitioners is governed by "the Veblenian bargain" (Schön, 1983): from the practitioners, their problems; from the researchers, the expert knowledge whose application to those problems enables practitioners to solve them in a distinctively professional way. This view tends to take one of two forms, depending on whether the researcher's claim to expertise rests on

1. research-based theory or
2. expert intuitions.

Each of these interpretations leads to difficulties. When researchers see themselves mainly as sources of research-based knowledge, the consequence of their interactions with practitioners is likely to be rejection or dependency. Dependency is the likely outcome if practitioners pick up the experts' esoteric knowledge and become little scientists—most often, "little social scientists"—who use fragments of theories as ritual clichés, floating, without palpable connection to the ways in which work is actually done. This condition, which in organizations holds for much of the current use of such

terms as "organizational culture," has been described (DeMonchaux, 1992) as the "loss of the innocent eye." Rejection is the likely outcome if practitioners question how well researchers' theories fit the practice situation, how they stand in relation to theories held by the practitioners themselves, or whether the researchers' actual behavior is consistent with the theories they profess.

In the case of researchers seeing themselves as operating on expert intuition, rejection or dependency is again the likely outcome, but for somewhat different reasons. The researcher's intuitive expertise tends to be opaque to the practitioner, who must then choose, more or less blindly, whether to "buy" it on a dependent basis or reject it. We cannot easily imagine how a practitioner can learn from expertise that presents itself as intuitive. It is true that athletes and artists do seem at times to learn from one another by observation and imitation alone; but even if we grant that organizational practitioners may sometimes learn from researchers in this way, neither the practitioner nor the researcher is likely, so long as the expertise remains tacit, to reason critically about it, hence to make a reasoned choice to accept it within limits, in certain respects and not in others.

Whether research-based expertise takes the form of esoteric theory or intuition, the conventional model of expert-practitioner interaction ignores the *practitioners'* inquiry, their own theories and ways of reasoning or testing ideas. What the practitioner already knows is ignored, just as conventional models of good teaching ignore the pupil's spontaneous understandings. How, then, is a practitioner's capability for inquiry thought to be enhanced as a result of interaction with a research-based expert?

Practitioners As Inquirers

We propose to turn the conventional relationship between researcher and practitioner on its head. We see practitioners not as passive recipients of expertise, but as Deweyan inquirers. Hence, we ask, "What do these practitioners already know?" "How do they inquire and learn?"

We perceive striking similarities between the issues of greatest interest to those who practice and those who conduct research on organizational learning. Both groups have an interest in making sense of organizational experience in instrumental terms: they want to know what makes for effective organizational action. Because they share an interest in understanding how organizations work and how they may be changed, they want to learn about the causal connections

between organizational actions and outcomes. They have an interest in discovering what patterns of thought and action account for past experiences of organizational success or failure, and how individuals contribute to either outcome. If they are impressed by the surprises and puzzles that arise in the course of organizational experience, they seek to make sense of them, as well as of patterned regularity.

Organizational practitioners may be curious about the processes through which they carry out the day-to-day business of organizational inquiry. Often they are capable of reflection *on* organizational practice, which researchers tend to see as their special prerogative. Practitioners are sometimes curious about how success and failure are defined in organizational terms, how goals and priorities are set, and, how ends of action are chosen as desirable. Practitioners, as well as researchers, may be interested in threats to the validity of organizational learning, that is, the kinds of reasoning and the forms of behavior that lead them to draw distorted lessons from past experience. Often they see that by focusing on immediate issues of local importance they may become blind to the larger significance of their actions. Increasingly they suspect that existing structures and incentive systems may undermine their ability to function well in a changing environment. Many of them want to learn how to create new structures and incentives and how to acquire new skills, enabling them to increase the learning capability of their organizations.

The ways in which practitioners and academic researchers inquire into such issues are in some ways alike and in other ways radically unlike.

Geoffrey Vickers describes two types of inquirers, each representing a distinctive stance toward inquiry (Vickers, 1975). He uses the term "spectator-manipulator" to refer to distant observers who keep their subjects at arms' length, exempting themselves from the worlds they study, only occasionally perturbing those environments under carefully controlled conditions in order to observe their subjects' responses. In contrast, "agents-experient" locate themselves within the problematic situation as concerned actors "whose actions and appreciations may be partly guided and changed by better understanding of the situations which prove to be relevant to [their] concerns."

Organizational practitioners are, of necessity, agents-experient. Only in fantasy or by way of retreat can they afford the luxury of becoming spectators. They are in the situations they try to understand, and they help to form them by coming to see and act in them in new ways. Through their perceptions, words, and thoughts as well

as their actions, they help to construct the objects of their inquiry. They are designers, not in the special sense of the design professions but in a more inclusive sense: they make things under conditions of complexity and uncertainty. The objects they design include products and services, policies, marketing strategies, information systems, organizational roles and structures, jobs, compensation schemes, and career ladders. They may even become designers of whole organizations. Not least, they design their day-to-day strategies of action.

Practitioners share with academic researchers an interest in building explanatory models of organizational worlds. Like researchers, practitioners try to account for the data they consider relevant, and they often show a decent respect for disconfirming evidence. But practitioners' models must also serve the purposes of designing. However appealing models may be as tools of exploration or explanation, they are judged by how well they "work," in the sense of enabling practitioners to do something they wish to do. This decisively affects what criteria apply to the reasoning of practitioners, in what sense they experiment, and in what sense their experimentation may be appropriately called "rigorous." Like academic researchers, organizational practitioners have a lively interest in forming and testing hypotheses about their environments. But because practitioners are agent-experient designers whose actions serve the dual function of probing and influencing their situations of action, their inquiry is subject to a different "stopping rule." In at least one view of science (Popper, 1968), the scientific cycle of hypothesis forming and testing should continue for as long as members of the community of inquiry bring forward plausible competing hypotheses. For practitioners, on the other hand, that cycle comes appropriately to a close when their inquiry enables them to achieve their intended results and when they like, or can live with, the unintended side effects inherent in their designing.

In these respects, the norms of practice inquiry differ from those of academic research even though in other respects the two forms of inquiry hold fundamental processes and criteria of adequacy in common. Nowhere is this more evident or more significant than with respect to the meaning of causality and the nature of causal inference.

Causality and Causal Inference

Researchers and practitioners alike are unavoidably concerned with issues of causality and causal inference. When organizational researchers

try to understand variations in patterns of organizational growth and deterioration or when they study how technological innovation works, how incentive systems function, how risks are managed, or how ideas circulate and evolve, their understandings hinge on causal connections. Practitioners, for their part, try to understand things so that they can change them or adapt to them. They seek to discover the features of context and action that caused past successes or failures in order to design more effective systems and strategies. They are continually engaged in detecting and correcting error, as we shall see in the following chapter and throughout the rest of this book. They are vitally interested in both the causes of error and the causal efficacy of the actions they design to correct error. But researchers and practitioners characteristically operate on different models of causality and reason about causes in different ways—a fact of great significance for the nature of their possible collaborative inquiry into organizational learning.

The model of causality conventionally adopted by normal scientists centers on the idea of a “variable,” a named attribute extracted from the complexity of observed phenomena which is treated as essentially the same in whatever local context it occurs. It is this presumed constancy of meaning that allows scientists (social scientists, in this instance) to speak of variations in the local values of a variable. If variables were not seen as having constant meanings, we could not speak sensibly about variation in the values or arguments of the same variable; we would have to speak of a different variable whenever a variable assumed a different value. Herbert Simon expresses a similar idea when he writes that each value of variables, X and Y , standing for cause and effect, defines a “class of events,” and that each variable, therefore, comprises a set of classes of events (Simon, 1977). Building on this conception, Simon treats causality as a function of an effect, Y , on one or more causes, X , as expressed by the formula, $Y = F(X)$. Simon calls such a function “self-contained” when “one and only one value of Y is associated with each value of X .”

According to this model of causality, researchers who investigate a causal relationship aim at formulating general causal propositions in the form of “covering laws.” Examples of covering laws are: “The occurrence of aggressive behavior always presupposes the existence of frustration and, contrariwise, the existence of frustration always leads to some form of aggression,” or “State anxiety, defined as ‘subjective, consciously perceived feelings of tension, apprehension, and nervousness,’ is caused by perceptions of role overload” (quoted

in James, Mulaik, and Brett, 1982). Researchers look for evidence to show that the values of Y (state anxiety) can be determined completely by the values of X (role overload), given the values of X and the knowledge that X has occurred, independent of any other features of the contexts in which X and Y occur.

According to the normal-social-science model of causality, probabilistic covering laws may be inferred from data provided by either of two empirical methods: “contrived experiment” or “natural experiment.” In both cases researchers try to determine whether values of the effect variables are uniquely determined by values of the cause variables, relying on one or more of Mill’s methods: Agreement (X is regularly followed by Y), Difference (without X , no Y), or Concomitant Variations (variations in X are regularly followed by analogous variations in Y) (Mill, 1843). In a contrived experiment the researchers construct a setting in which they can control variations in the value of X in order to observe changes in the value of Y . In the method of natural experiment, or “quasi-experimental method” (Campbell and Stanley, 1963), the researchers observe a number of settings in order to measure the naturally varying values of X and Y , relying on observation of many local settings in order to avoid being misled by the peculiarities of any particular one. In both cases researchers try to maintain a respectful “research distance,” lest they contaminate their data or become affected by their subjects’ biases. They avoid referring to their subjects’ intentions, which they regard as subjective, idiosyncratic, and qualitative—unsuited to the generality, quantitateness, and context-independence that are essential to the normal-social-science model of rigorous causal inference.

In everyday practice, on the contrary, organizational practitioners think in terms of “design causality:” the causal relation that connects an actor’s intention to the action he or she designs in order to realize that intention. To explain the cause of an action in terms of design causality, we describe the intention we believe the actor is trying to achieve by means of the action. To put the same idea in different terms, we describe the reasoning that led up to the action, not the reasoning by which that action might be justified after the fact. Olafson (1967) calls this type of causality “cause by reason,” and Von Hayek (1948) calls it “sufficient reason.” Practitioners make reference to a second type of simple causality, “efficient cause”: the causal connection between an act and its consequences, intended or unintended. Furthermore, when individuals habitually interact with one another in an organizational setting, their designed behaviors,

together with the first- and second-order consequences of which those behaviors are the efficient causes, tend to create complex organizational systems that display "pattern causality."

Consider the behavior of grant-making officers in a philanthropic foundation. They know that funds are made available to them on a "use-it-or-lose-it" basis. Hence, they make sure they spend their allocated funds (design causality), no matter how shoddy the grants they may support in order to do so. In aggregate, such grant-making behavior produces a high level of questionable grants (efficient causality). Taking notice of this effect, the director of the foundation institutes increasingly rigorous systems of external evaluation (design causality), by which grant makers feel unfairly constrained (efficient causality) and which they seek to elude (design causality) with the further result that evaluations come to be considered unreliable (efficient causality). Once such patterns have been created, they tend to maintain themselves (pattern causality) through feedback loops that influence how individuals in the system think and act.

Consider a factory in which top management has instituted a system of "pay-for-performance." Supervisors in that factory learn to soften their criticisms of their workers, in order to avoid the complaints and grievance actions they have come to expect in response to their honest evaluations (design causality). As a first-order consequence (efficient causality), workers' performance ratings become inflated; as a second-order consequence (efficient causality), the wage bill grows. The rising wage bill, the unintended aggregate effect of each supervisor's deliberate actions, may alarm upper-level managers and become, in turn, a trigger for their new attempts to drive the wage bill down (design causality), meeting a fate similar to the one described previously. The system of pay-for-performance, originally intended to improve productivity, turns out to have the cumulative effect of rewarding mediocre performance and increasing the cynicism of the supervisors (pattern causality).

A practitioner-inquirer who operates on such a model of organizational causality tries to infer component causes of organizational events and to construct and test models of their interaction in causal patterns. The practitioner uses the method of *causal tracing*, observing how one phenomenon leads to another. Causal tracing depends for its feasibility on the inquirer's having a background model of the system's pattern causality. The process is like the reasoning of a skilled plumber: given a broad understanding of a system of pipes and how liquid flows through them, the plumber tries to identify the cause of a

leak by tracing it back (in space and time) to its sources within the system. Like plumbers, organizational inquirers can test their causal hunches by carrying out on-the-spot experiments. For example, the plumber may close a valve and observe whether the leak stops. In the instance of the foundation, an organizational inquirer might institute a change of policy (eliminating the "use-it-or-lose it" rule, for example) in order to observe consequent changes in the grant makers' behavior. Inquirers into organizational or plumbing systems can also test their pictures of a whole causal pattern by considering alternative models of it and trying to discriminate among them by finding data that only one of the competing models can explain. In the last analysis, practitioners test their causal inferences by determining whether they can use them to get design results they intend and like.

The normal social-science model of causal inference aims at generalizability: it employs general cause- and effect-variables and calls for observation of multiple instances in which those variables take on different values. The practitioner's model of causality is situation-specific. It deals with named phenomena, for example, "softening evaluations of subordinates," that need not (and usually do not) take the form of general, quantifiable variables. The normal social scientist employs contrived or natural experiments that cut across many different contexts. The practitioner traces causes in the light of a background model of a particular system and tests causal inferences in that system through on-the-spot experiment.

The general covering laws established through normal social science tend to have relatively little utility in practice because of the characteristic representation of a covering law. The functional relationships of variables that result from normal-social-science research tend to be precise, quantitative, probabilistic, abstract, and complex, making it difficult for practitioners to form images of research results that can be used to guide action in a particular local context. There is also a "gap of valid application" between the contexts of research and practice. In order to establish that the general causal relationships among variables, established in a research context, will hold at a particular time in a particular practice setting, the practitioner will have to recreate in that setting the conditions under which the research results were obtained in the first place. This means that he or she will have to construct operational definitions of the key terms and show that observers who do not know the propositions to be tested can use these definitions to make reliable observations of the relevant phenomena across a suitable range of variance. Moreover, the practi-

tioner will need to minimize threats to internal and external validity through the use of suitable controls, including keeping the experimental strategy secret from the subjects lest their awareness of it confound its results. Under practice conditions of real time, confusion, and pressure, such actions are difficult to carry out. Moreover, when implemented they are not neutral. As Argyris has shown (1980), actions of this kind imply an approach to management that is reminiscent of the managerial climate of hierarchical organizations. Such an approach places subordinates in a submissive, dependent role that is very likely to create strong feelings of ambivalence or outright hostility.

Finally, we have been impressed with the step-function difference between what happens when people discover problems and invent strategies of action, and when they actually try to produce their inventions under everyday conditions of real time and pressure—especially when it comes to situations of embarrassment or threat. Under these conditions, individuals frequently produce actions contrary to their inventions and are unaware of the discrepancy. A manager may want to persuade the middle-managers under her to stop withholding negative evaluations of their workers. But in order to avoid upsetting her subordinates, she may actually smooth over the negative attributions she makes about them. Her unawareness in such a case is likely to be due not so much to ignorance as to skillful adherence to theories-in-use learned early in life.

In ordinary organizational practice, phenomena like these are the rule rather than the exception. It is a rare normal social scientist who takes them into account or focuses on the skills practitioners would actually need in order to produce inventions based on normal science research. Hence, normal scientists are unlikely to generate research results that practitioners can actually use to produce inventions derived from those results.

The practitioners' causal inquiry does not yield general covering laws. Their situation-specific inferences of design, efficient, or pattern causality can be generalized only by a process we call "reflective transfer"—"transfer," because the model is carried over from one organizational situation to another through a kind of seeing-as; "reflective," because the inquirer should attend critically to analogies and disanalogies between the familiar situation and the new one. In reflective transfer, causal stories play roles similar to the roles of legal precedent in judicial decision making or precedents in architectural design. The utility of the prototype lies in its ability to generate explanation and experimentation in a new situation. When it is carried

over to a new situation, its validity must be established there by a new round of inquiry through which it is very likely to be modified. And the modified prototype that results from the new round of inquiry may serve, in turn, as a basis for reflective transfer to a new situation.

Practitioner/Researcher Collaboration in Action Research

What shall we say about the roles, methods, and attitudes appropriate to research on organizational learning once we recognize that the process of organizational learning is carried out by practitioners who are inquirers in their own right, that the topics of interest to organizational inquirers and academic researchers have large areas of overlap, and that the patterns of causal reasoning characteristic of the two groups of inquirers are in some respects alike yet in other respects, radically unlike?

For researchers who hope to produce knowledge useful to practitioners, the implications of our analysis are straightforward. These researchers should join with practitioners who seek to promote productive organizational learning and to understand the nature of their own learning processes and systems. The researchers should try to discover what practitioners already know how to do and to learn to appreciate the inquiry in which practitioners are already engaged, including the questions they know how to ask and the knowing-in-action they may take for granted and be unable to describe. These researchers should join with practitioners to help discover the hidden rationalities that are often built into everyday organizational practice, the productive forms of pattern causality of which practitioners themselves are often unaware. But this research function should be coupled with helping practitioners extend and enhance the inquiry they already know how to carry out. This means helping them to discover how they get stuck and what dilemmas underlie their getting stuck; how the same patterns of action that lead to success may also, on occasion, lead to failure; how practitioners can learn from failure; how they can enlarge a focus of attention that may be limited to the local and the immediate, thereby opening the field of design possibilities; how they can become aware of counterintuitive effects masked by their, perhaps tacit, background models; how they can reflect on and explore the impediments to productive organizational learning embedded in their limited organizational learning systems; and how they can become aware of their own contributions to the maintenance of those systems.

There are several reasons for this research focus. First, a researcher who is interested in the study of organizational learning ought to have an interest in studying how practitioners' inquiry contributes to that process. Secondly, an organizational researcher who wants to produce results useful to practitioner-inquirers should want to meet their understandings with his own. He needs to listen to them and get inside their ways of thinking and acting, with respect to both strengths and limits, in order to increase his chances of being listened to and of making his research relevant in their eyes. If practitioners are already inquirers, then an outside researcher's effort to involve them in research must take account of the inquiry in which they are already engaged. Finally, as Kurt Lewin pointed out many years ago (Lewin and Grabbe, 1945), people are more likely to accept and act on research findings if they helped to design the research and participate in the gathering and analysis of data.

Indeed, Lewin's conception of action research is prototypical of the kind of research we have in mind. Lewin's research career had two main sources: his training as a physical scientist and his experience as a Jew driven from his homeland by the Nazi menace. His research revealed his commitment to democracy (as in his studies of democratic versus authoritarian group climates), as well as to creativity, productivity, and inquiry orientation. He was, as Alfred Marrow called him, a practical theorist who often remarked that there was nothing so practical as a good theory. He sought to achieve desirable social results, for example, persuading children to drink orange juice and eat their spinach, promoting the sale of war bonds, and reducing discrimination based on race or religion. The method he evolved was that of involving his subjects as active, inquiring participants in the conduct of social experiments about themselves. He adopted the working hypothesis that people would tend to adopt beliefs in whose development and testing they had been active participants. In the course of such practical experiments, limited to particular social problems and situations, Lewin had the skill and imagination to discover ideas of wide-ranging importance. For example, his invention of the concept of "gatekeeper," which has long since entered into the language of both social scientists and lay persons, grew directly from his experimental studies of influence and persuasion in such contexts as the drinking of orange juice and the buying of war bonds.

A constellation of values and methods, similar to Lewin's, informed the early Tavistock studies and experiments, as at the long-

wall coal mining operations in Great Britain (Trist and Bamforth, 1951), and at Glacier Metals (Jacques, 1952). The theory and practice of participatory, semiautonomous work groups grew out of these early examples of organizational action research. Nearly a generation before Lewin's and the Tavistock Institute's development of action research, John Dewey advocated similar values of inquiry, participation, and collaboration. However, he based his arguments on philosophical rather than empirical grounds, and he sought to apply his philosophy through social experiments mainly in the fields of citizenship and education.

A researcher who embraces in this spirit a program of collaborative action research on organizational learning becomes, like the practitioners he joins, an agent-experient. He, too, places himself within the situations that he studies and must, in consequence, study himself. His commitment to the organization in question is likely to be less intense than the practitioner's. His interests in inquiry are likely to be overlapping, not identical, with the practitioner's. He is likely to miss some of what the practitioner's local knowledge enables him to see and to retain a partly skeptical stance toward the practitioner's claims to organizational knowledge and learning. Yet he may also attach greater appreciation than the practitioner does to the taken-for-granted practice knowledge that informs his everyday competence.

The action researcher will join the practitioners in their organizations and collaborate with them in conducting their design inquiry, entering into their underlying models of causality and causal inference. At the same time, the action researcher will seek to become aware and help the practitioners become aware of the limits of those models. The theories, models, exemplars, and heuristics that researchers bring to the practice situation they will use as lenses on the situation, to be tested for their use in making sense of it, but not as substitutes for what Kevin Lynch once called "the best kind of theories," those constructed in the situation itself.

Even from the perspective of a skeptical normal social-science researcher who wants only to study the phenomena of organizational learning and has no interest in producing knowledge useful to practitioners, it makes sense to reverse the usual relationship between practitioner and researcher. There are at least two reasons for doing so, which we take up at greater length in Chapter 9, but state briefly here:

1. Scholarly researchers into organizational learning will want to test the insights they gain from simulation studies or

distant empirical research. They have a scientific interest in filling the gaps that now exist between their relatively high-level formulations, for example, the theories of internal variation and selection within organizations which we discuss in Chapter 9, and the fine-grained processes by which such phenomena actually arise. This requires gaining access to the inner workings of organizations through the cooperation of practitioners who are, and often see themselves as, inquirers in their own right. Such cooperation is more likely to take place and to work when outside researchers join practitioner-inquirers in collaborative action research.

2. As a matter of scientific curiosity, scholarly students of organizational learning may wonder whether the patterns of limited learning they discover are law-like, somehow inherent in the structures of organizational life or whether they are, in a sense, artifacts of states of consciousness peculiar to the practitioners they are observing. Could these patterns be changed if practitioners became aware of them? How could such a possibility be tested unless the researcher were to help practitioners design and enact ways of circumventing patterns of limited organizational learning? Robust tests of scholarly models, for example, those that relate to competence traps, superstitious learning, or garbage-can phenomena, call for creating in organizations the conditions under which such models might be confirmed or disconfirmed in action. This, again, requires collaboration with practitioner-inquirers.

Appropriate Rigor in Collaborative Action Research

Practitioners can use their models of causality to make causal inferences, and to engage in reflective transfer of such inferences, because they are agents-experient who live in close proximity to the situations they seek to understand. It is this closeness that enables them to hold usable background models of phenomena, carry out causal tracings, and conduct on-the-spot experiments. On the other hand, the practitioners' closeness to the situation of action also presents a variety of threats to the development of valid, usable knowledge:

- Their busyness may deter them from engaging in inquiry that would otherwise be useful to them; they are often con-

strained by the need to leave off thinking and begin to "get things done."

- Their familiarity with their own patterns of action may prevent them from seeing what they actually do and leave them unable to describe the action knowledge on which they greatly depend.
- Their biases as interested actors in the situation may blind them to data that might have caused them to change their minds.
- Their focus on pressing business, local and short-term in nature, may prevent them from taking a potentially useful longer-term, wider-ranging view.
- Their very familiarity with their environment may make them blind to it. More specifically, they may be at least partly unaware of causal patterns that constrain productive organizational learning and unaware of how their transactions with the organizational environment contribute to the very impediments that they see as imposed on them from the outside.

What does it mean for practitioners to cope well with these sorts of vulnerabilities? How might they be helped to do so by entering into a collaboration with action researchers based in the academy?

In social psychology there is a long tradition of debate over the trustability of our everyday intuitions into the reasoning—the intentions, thoughts, and feelings—that motivate other people's actions. Zajonc (1989) raises this question and concludes that although our own experience

can be a rich source of ideas and a source of hypotheses... everything we know from the systematic study of social perception indicates that we should be very distrustful of our so-called intuitions.

He cites in support of this position the large number of studies that have revealed

a substantial self-serving bias in estimating probabilities of causal events.

Even if we take a much more sanguine view of the reliability of our everyday intuitions into human reasoning and intention, it is certainly clear that we sometimes find them to be mistaken. Reasonably competent organizational inquirers, ones who exhibited

appropriate rigor, would certainly try to test such intuitions, especially those that played important roles in their construction of causal stories. Consider, for example, the inference that supervisors in the pay-for-performance case smoothed over their negative evaluations "in order to avoid upsetting the workers," or that technical staff in the telecommunications company bypassed software reliability checks "in order to respond to management's pressure for fast installation." How might such inferences be tested? Certainly, one of the principal tests would consist of on-the-spot experiment, namely, asking for information. But this form of experimentation is also vulnerable to error for several different sorts of reasons.

Let us take a situation in which one person asks another, "What led you to do this?" and receives an answer to the question, inferring from it an interpretation of the actor's intentions. First of all, the informant may not be able to make an accurate reconstruction of the before-the-fact reasoning that led to her action. She may not remember it, or, in accordance with the very widespread tendency to engage in instant historical revisionism, she may reconstruct it more or less unconsciously to suit her present ideas, interests, or inclinations. Such distortions might be corrected, perhaps, by inducing the informant to make a careful reconstruction of the incident in which she was involved, or when possible, by comparing the responses of several different informants.

But remedial measures such as these are vulnerable to a second source of error, one that affects any attempt to inquire directly into reasons for action. An inquiry into an actor's reasons for acting in a certain way is itself an intervention, and when it takes place in an organizational setting, it is also an intervention into the life of the organization. Both of these factors can and usually do have powerful effects on the ways in which both inquirer and informant construe the meaning of their interaction, interpret each other's messages, act toward each other, and perceive each other's actions. These effects can complicate and often subvert the inquirer's quest for valid information.

One reason is that the individual who finds herself in the role of informant also seeks to discover the meanings of the situation in which she is involved and acts on the basis of the meanings she constructs. She may answer questions in the light of what she believes the inquirer expects of her. She may construe the situation as one that calls for putting the best possible face on prior actions. Her interaction with the interviewer may be designed, more or less consciously as a form of image management. To the extent that she feels threat-

ened or distrustful in the interview, she may deliberately withhold information she feels might be taken in a negative way.

All such interactions are also affected by what George Devereaux (1967), borrowing from quantum physics, once described as "complementarity." The interviewer's questions and his responses, verbal and nonverbal, are also intrusions into the interpersonal situation, affecting the informant's constructions of meanings and her willingness to give valid information. The act of inquiry influences the situation inquired into. So, for example, an investigator into the network failure may convey a threatening, judgmental attitude that makes his informants even more defensive than they might otherwise be.

What is true of the informant is also true of the inquirer. His interpretations of the messages he receives are also affected by the more or less conscious meanings he constructs for his interaction with the informant, his attributions of meanings to her, and his intimations of the way in which she perceives him. If he perceives her as likely to withhold negative information, he may interpret her answers in light of his suspicions about the information he believes her to be withholding.

These interaction effects are by no means peculiar to causal inquiry in organizations. They are also characteristic of the experimental environments of normal social science where they are equally capable of foiling the researcher's quest for valid information. As Harré and Secord (1972) noted, "Social psychological experiments are also social episodes." Indeed, the very controls by which social psychologists strive to make their contrived experiments rigorous are likely, as observed earlier, to introduce systematic distortions into their research results.

In organizational inquiry, however, all such interaction effects may be exacerbated by certain peculiarities of the organizational context. Organizational inquiry is almost inevitably a political process in which individuals consider, whether they choose to be decisively influenced by such considerations, how the inquiry may affect their standing or their reference group's standing, within an organizational world of competition and contention. The attempt to uncover the causes of a systems failure is inevitably a perceived test of loyalty to one's subgroup and an opportunity to allocate blame or credit. Such an inquiry is likely to trigger familiar games, for example, allocation of blame and avoidance of blame, exercise of control and avoidance of control, winning credit and preventing others from winning credit. Within such games, strategies of deception, pre-emptive blame,

stone-walling, fogging, and camouflage, including camouflage of these very strategies, frequently inhibit inquiry into the causes of organizational events and the reasoning of the actors involved in them. We describe such phenomena later in this book from the point of view of what we call "limited learning systems."

Given the personal, interactive, and organizational phenomena that can inhibit the quest for valid information about design causality, how ought one carry out organizational inquiry so as to increase the likelihood of producing valid information? This global question sets a critically important direction for a possible social science that might take seriously the problem of enhancing organizational inquiry. It has at least two main parts. First, how can we build more accurate and usable accounts of the personal, interpersonal, and organizational patterns that inhibit causal inquiry in organizations? Second, from the point of view of the process of inquiry itself, what theories of action, strategies, values, and underlying assumptions are most likely to enable an inquirer to elicit information, interpret it, and test interpretations so as to form valid inferences about design causality?

These questions form the basis of the approach we call "the theory of action perspective" or "action science," a development of Lewinian action research. It focuses on the problem of creating conditions for collaborative inquiry in which people in organizations function as co-researchers rather than merely as subjects. And it does this, as already noted, on the assumption that people are more likely to provide valid information about their own intentions and reasons for action when they share control of the process of generating, interpreting, testing, and using information. The theory-of-action approach posits the existence of a behavioral world created by the parties to an interaction and identifies the characteristics of behavioral worlds that may inhibit or encourage valid inquiry. It explores the features of theories-in-use that are conducive to exchange of valid information in behavioral worlds of interpersonal inquiry, emphasizing the importance of making private attributions public, treating these attributions as disconfirmable, and subjecting them to public test. It operates from the assumption, for which we think there is considerable evidence, that theories-in-use tend to exert a contagion or mirroring effect. We believe that individuals become more effective inquirers when they employ theories-in-use which, if mirrored by their informants and co-researchers, would be likely to produce valid information.

These points, crucial to any future collaboration between organizational practitioners and academic researchers, will be pursued in the remainder of our book, beginning with the following chapter in which we explore the long-term evolution of an industrial firm and a consultant's attempts to foster productive organizational learning at a crucial juncture in that evolution. In the current chapter, we have called attention to questions of method, attitude, and relationship that become central to research into organizational learning once we see it in terms of organizational inquiry, recognize practitioners as inquirers, and turn the researcher-practitioner relationship on its head.

We believe that organizational learning occurs within the context of such dialectical processes which stem from two conditions of organizational life. First, organizations are necessarily involved in continual transaction with their internal and external environments which are continually changing in response to both external forces and organizational actions. Second, organizational objectives, purposes, and norms are always multiple and potentially conflicting.

As a consequence, it is no accident that organizational solutions give rise to further problems; they may be expected to do so, given the dialectical context of organizational inquiry.

This, then, sharply raises the problem of criteria for the evaluation of organizational change and learning, a problem central to our inquiry which will occupy us especially in Part III. At this point, however, we may note the following:

- The achievement of stable solutions is not an appropriate criterion for organizational learning; it is in the very nature of organizational problem solving to change situations in ways that create new problems.
- Organizational effectiveness, as measured by the achievement of espoused purposes and norms, is an incomplete criterion for organizational learning. It is appropriate in situations where error correction can occur through single-loop learning alone. It is insufficient where inconsistencies in organizational theory-in-use set requirements for double-loop learning.

“Good dialectic” is the term we use to describe processes of organizational inquiry which take the form of single- and double-loop learning and where both single- and double-loop learning meet standards of high-quality inquiry.

The achievement of good dialectic requires organizational deuterolearning. It requires that the organization’s members reflect on and inquire into their organizational learning system and its effect on organizational inquiry.

Part II

Defensive Reasoning and the Theoretical Framework that Explains It—Model I and O-I

4

Defensive Reasoning In Individuals

In this chapter, we illustrate what we mean by “defensive,” in contrast to “productive,” reasoning, and we outline in some detail the methods we use to study how individuals reason.

Our methodology operates on an interdependent, double-track strategy through which we attempt to engage with practitioners in collaborative action research.

First, we try to describe the reality of a particular context as accurately as possible, organizing our descriptions in the form of propositions that are generally applicable to many such contexts. In any particular context, we try to test such propositions and, therefore, prove them false. We do not lower the standards of falsification because our tests are conducted in field-organizational settings. We also frame our research propositions so as to make them usable by practitioners, not only to provide an additional test for them but to contribute to the practical effectiveness of knowledge in a world we seek to understand.

Secondly, we try to help practitioners become more reflective inquirers into their practice so that they can monitor it and, by detecting and correcting errors wherever they occur, increase the likelihood of producing what they intend. We seek to help practitioners understand their world in such a way that they can produce conditions for organizational learning, especially double-loop learning.

We see our double-track strategy as a model not only for effective research but for reflective practice especially regarding double-loop learning.

A Generic Dilemma In Double-track Research

Our initial premise is that human beings design their actions and implement their designs. We call these designs theories of action,

differentiating, as we pointed out in Chapter 1, between the theories of action individuals espouse and the ones they actually use their theories-in-use.

Both of these types of theories are learned early in life and supported by features of societal and organizational cultures. Although human beings' sense of competence, independence, and self-esteem are based on both types of theories, we consider theories-in-use to be more powerful in explaining and changing behavior, especially in relation to double-loop learning.

Almost all of the individuals we have studied hold theories-in-use that are systematically counterproductive for double-loop learning, especially when the issues are embarrassing or threatening. Moreover, these same theories-in-use, when skillfully used, make the actors unaware of the counterproductive features built into them. Since the theories-in-use are sanctioned and supported by organizational and societal cultures, individuals have little reason to be aware of or to explore this predicament. Indeed, as we shall show, practitioners may even interpret attempts to explore the predicament as bewildering, if not inappropriate.

As researchers, we are therefore likely to be faced with a dilemma: Individuals may unknowingly provide us with distorted information, and these same individuals may hesitate to engage in the dialogue that is required to explore the possibility of such distortions. If we persist in exploring these issues, practitioners may become defensive—their defensiveness leading, in turn, to new distortions, both recognized and unrecognized.

This research dilemma is systematic. It is unlikely to be overcome by the use of better sampling procedures. Nor can it be overcome by the use of the accepted methods of normal social science because embedded in the execution of these methods, with their reliance on research strategies of secrecy and unilateral control, is the same theory-in-use as the one that causes the dilemma in the first place (Argyris 1980, 1993).

Features of Our Research Method

Our approach to the research dilemma begins with the task of constructing the theories-in-use that underlie it, those theories-in-use that inform the reasoning and action of practitioners. This requires the collection of relatively observable data. Observations of actual behavior, especially the tape recording of conversations, is the dominant mode. We may also use questionnaires, projective tests, or

structured interviews; but if we use such instruments, we recognize that they are likely to give us insights into espoused theories and not theories-in-use.

The observations made and the conversations recorded should be connected to objectives and actions to which individuals are highly committed, for example, observations of meetings about non-routine issues that tend to stimulate feelings of embarrassment or threat. Such events are intimately tied to an individual's sense of competence, confidence, and self-esteem. A slightly less powerful set of data may be obtained in classrooms and workshops, as long as those activities raise problems that the participants consider important and persistent features of their everyday working lives.

This is one reason why our research methods focus on intervention and change. Practitioners examine such research proposals with care, especially when they deal with double-loop issues. The practitioners realize that research of this kind may generate a variety of costs for them, triggering defenses at all levels of their organizations.

Such defenses should not be avoided or suppressed. When they occur, they become additional data that can be used to test diagnoses of individual, interpersonal, and organizational phenomena. In order to deal with defenses effectively, however, the researcher must possess the necessary skills, the same skills the practitioners will have to learn if they are to deal with defenses in their organization. The theory of action the researcher uses to obtain valid information becomes a model available for use by practitioners.

If a researcher or practitioner is to act effectively in the service of double-loop learning, skills are necessary but not sufficient. The inquirer also needs an *actionable* theory of organizational learning, one that may be used to generate and test specific hypotheses in a wide variety of settings, as well as in the individual case.

"Individual case" typically denotes an *n* of one. In our research, however, the individual case functions as a setting for multiple observations. Most of our research is conducted in one organization, but we treat that organization as a setting in which we conduct many observations of individual and interpersonal behavior.

Our observations are guided by our theories of individual and organizational learning. For example, we are interested in observing behaviors such as evaluating or making attributions. We make a priori predictions about the impact of such behaviors. These predictions vary depending on whether evaluations and attributions are crafted in ways that discourage or encourage inquiry, especially that which

leads to double-loop learning. We are able to score transcripts to assess the frequency with which such behaviors occur, making use of interobserver reliability studies and other appropriate measures. We can then make a priori predictions of the behaviors' consequences for such features of inquiry as self-fulfilling or self-sealing processes. These, as we will illustrate in the following chapter, enable us to make a priori predictions about the impact of individuals' behavior on organizational phenomena such as interdepartmental relationships or organizational defensive routines.

A Relatively Simple Paper-and-Pencil Method for Obtaining Valid Information About Individual Theories of Action

Throughout this book, we will illustrate research methods that we claim produce valid knowledge about individual and organizational learning, in such a way as to engage, not bypass, the research dilemma we have described earlier. The first such illustration is a unique type of case that we ask individuals to prepare for classroom or workshop settings.

Typically, we send a letter to each prospective participant in which we describe the format of the case, asking them to send their cases to us at least a week before the seminar. The format instructions are:

1. Identify an important problem that you have tried to solve or will try to solve in the near future. There are no limits on the subject selected except that you should evaluate it as crucial to your own, your department's, and your organization's performance.
2. Describe the steps you took (or plan to take) in order to resolve the problem. With whom did you meet? What was the purpose of the meeting as you recollect it (or as you expect it to be)?
3. Divide the next several pages in half. In the right-hand column, write the conversation, as you can best recollect it. Begin with what you said, what the other(s) said, then what you said, and so on for about three pages. If it is a session that has not yet been held, describe what you plan to say, what you expect others to say, and so on. In the left-hand column, write any thoughts and feelings you had as the conversation proceeded (or that you believe you will have when you meet sometime in the future).

Typically, we receive cases of three to five pages (a few have been as long as ten pages). A brief illustration of a case is presented below. This case was written by a CEO who had been telling his immediate reports that the time has come for "calling a spade a spade." He urged them to be forthright and straightforward with their employees because the firm was in serious financial trouble.

Case Illustration: The Easing-in CEO

Thoughts and Feelings Unsaid

I hope we'll work cooperatively. I fear we won't.

What is wrong with him? He's missing the most important point.

He doesn't want to accept ownership; he wants to make me responsible.

The Conversation

CEO: *I'm sure that you and I share the same goals. We need to rethink our latest cost estimates.*

Other: *The latest estimates are not the most reliable...*

CEO: *I can see that some costs can be reduced (by your recommendation), but that still leaves us with a very large problem.*

Other: *The original estimates were produced by others. We never really agreed to them.*

CEO: *We will have to use these estimates. That's the reality.*

The participants' cases usually contain several features. First, the left-hand column usually contains important negative evaluations and attributions, indicating that the writers are experiencing the "other" as being ineffective and performing poorly. Second, the case writers typically do not state these evaluations and attributions in a candid and forthright manner as in this example, the writer had advised his subordinates to do for their subordinates. Third, the conversations that are crafted bypass any strong feelings, and the bypass is covered up. Fourth, the writers appear unaware of any discrepancy between what they are espousing and how they actually behave.

Reflecting on the Case Approach in the Light of the Research Dilemma

Our instructions to case writers ask them to identify crucial problems they are experiencing or expect to experience. We ask them to write a short description of the strategy they used (or would use) in order

to deal with these problems. The written case tells us what the respondents' objectives are and gives us insight into the processes they state they will use in order to implement their objectives. This information gets at their espoused theory of action.

The conversation the participants write provides directly observable data from which we infer their theory-in-use. The description of thoughts and feelings they did not (or would not) state gives us insight into their self-censoring processes, another key feature of their theory-in-use.

As we shall see, it is possible to use such cases to identify discrepancies between the writers' espoused theories and the theories-in-use built into their actions. Awareness of such discrepancies makes it possible for the writers to assess the degree to which they have been skillfully unaware that their behavior is counterproductive to their intentions. This, in turn, can become the basis for inquiry into the nature of a theory-in-use that leads them to act skillfully yet produces ineffective actions.

We will address the question of the validity of such data as we go along, because we test validity in different ways under different conditions. All we are claiming at this point is the following:

1. The written conversations and thoughts and feelings present what the practitioners wrote. (We are not claiming that these are the words they actually used or would use in the events to which their cases refer. Nor is it necessary for us to make such a claim.)
2. These cases (and the dialogues tape-recorded during their discussion) are adequate to infer the writers' respective theories-in-use as manifested by the data of their cases. (We will eventually answer the question of the generalizing of these inferences to situations beyond the case.)

Case 1: Strategic Case Management Seminar

We draw our first example of the research method described above from a seminar that took place at the Harvard Graduate School of Business Administration in 1993. Thirty-seven upper- and some middle-level financial executives attended. Five were female. Fifteen came from foreign continents (e.g., Africa, Australia, Asia, and South America). The focus of this seminar was on new concepts and procedures used in strategic cost management and

on the human problems likely to arise when such concepts and procedures are implemented in organizational settings.

We were not able to discuss all the cases the financial executives wrote because there was not enough time during the week-long seminar. At the outset, the faculty member (Argyris) used the full set of cases as a vehicle to provide the entire group with an overall picture of the underlying action strategies they used. Three lists were developed. List #1 contained examples of comments quoted verbatim from the "left-hand columns," illustrating the writers' views of the individuals with whom they were dealing:

List #1

1. *Don't let these guys upset you.*
2. *Say something positive.*
3. *This is not going well. Wrap it up and wait for another chance.*
4. *Remain calm. Stick to the facts.*
5. *He is clearly defensive.*
6. *He's playing hardball because he is afraid of losing power.*
7. *She is overblowing the systems issue to avoid having to change.*
8. *He is baiting me now.*
9. *Will he ever be able to change?*
10. *This guy is unbelievable. He will never change.*
11. *You are nowhere as good as you think you are.*
12. *The trouble with you is that you do not really understand accounting as a managerial function.*

These comments illustrate the following patterns, which we find as a general rule in a wide variety of settings whenever we use this case method:

The left-hand columns contain thoughts and feelings that are critical for learning to occur. Yet they are systematically covered up.

Advocating, evaluating, and attributing are the three action strategies subjects most often use in dyadic relationships, in groups or intergroup settings, or when they deal with organizational policies, practices, rules, and norms.

The classroom session began with the executives (all of whom had written cases) reading List #1. After a few minutes, the faculty member asked them to describe their reactions to the list. He asked, "What does this list tell you about the individuals who wrote the comments? What inferences do you make as to what is going on?"

The executives responded easily and quickly. Eight examples of their comments, taken from the transcript, were as follows on the next page.

List #2

1. They were opinionated.
2. They talk as if they are right.
3. They are frustrated and angry.
4. They are entrenched.
5. They are avoiding conflict.
6. They are not listening.
7. They are fearful.
8. They exhibit lack of empathy.

The faculty member wrote these responses on the board. He then asked the participants to reflect on the nature of their comments. The executives responded that their responses indicated an overall negative reaction. The comments were primarily negative evaluations and attributions of defenses in others. Moreover, the class comments indicated that the executives thought the writers of List #1 (whom they knew to be themselves) appeared closed to learning.

We see in these reactions the following general patterns:

Evaluations and attributions are made in ways that do not encourage testing. The writers appear to act as if their diagnosis is valid and does not require testing.

The writers appear closed to learning or, at least, they see learning as unnecessary. Yet all of them attended the seminar and wrote the case with the expressed purpose of learning how to be more effective in dealing with the human side of enterprise.

The class comments on List #1 led to reflection on a different issue. One executive said that what surprised her was the negativeness of the first list. She recognized her comment in List #1 and it, too, was negative. Yet, she added, she was certain that her intent was to be positive. She guessed that this was probably the intent of others in the class. Several class members responded affirmatively.

The faculty member then asked the executives to analyze List #2, their comments about List #1, as he had written them on the board. The executives responded that these comments, too, were negative. They were evaluations and attributions crafted in ways that did not encourage inquiry. This also surprised them.

Here, we find another general pattern:

There appears to be a systematic discrepancy between the writers' expressed aspirations to learn and help others to learn and their actual behavior, which is largely counterproductive for learning. The

individuals are systematically unaware of the ways in which they produce their unawareness.

The faculty member then said that the dialogue so far illustrated some of the main findings that had been obtained worldwide from nearly 6000 individuals of both sexes, ranging widely in majority or minority status, education, wealth, and organizational rank. What the class participants were experiencing was not unique. It seems that individuals throughout the world deal with difficult, embarrassing, and threatening issues in a similar manner. For example, they make evaluations and attributions that are crafted in ways that do not encourage learning. They are predisposed to be unaware of the discrepancies they produce, such as aspiring to be positive yet being negative.

At this point, one executive said that he agreed that a negative pattern did exist in the self-censored left-hand columns and in the responses the faculty member had written on the board. He was certain that part of the "free-flowing negativeness" (as he called it) was due to the fact that the participants thought that they were talking privately. They would have crafted their conversations differently if those they were evaluating were present.

This led to several attempts by the executives to show in roleplaying how they would craft their conversations differently. Indeed, they were different. They were more diplomatic and easing-in. They bypassed the threatening issues and acted as if they were not doing so. The faculty member said that he obtained the same results from all other groups. Unawareness, he suggested, was highly skilled, in that it was in the service of crafting conversations that were intended to be positive; unawareness of this type was not caused by ignorance. The faculty member hoped to show it was caused by a master design in the heads of individuals, through which they attempt to act positively, yet the results of the use of that design are consistently negative.

The faculty member then handed out List #3. This list contained examples from the participants' cases of left-hand column thoughts and right-hand column conversations. The conversations were crafted in ways that were diplomatic and smoothed-over, and they bypassed the meanings in the left-hand columns. Thus their cases written several weeks before were similar to their roleplaying in class. For example,

List #3**Thoughts and Feelings Unsaid**

You guys come up with more excuses that make no sense. You do this all the time.

Conversations

You still have the ability to offer different combinations of products.

If we gave you everything you ask for, we would lose our shirts.

The research we have done shows that there is a consumer movement toward my views. Your sales will not be harmed.

How can I convince the group of the necessity to change while we're on top?

Although we are the leaders, it is becoming more difficult to remain responsive and react quickly enough. Our product development process has to become more effective and efficient.

Winning the Nobel Prize will not help the company. Perhaps it's time to expand development staff and downsize research staff.

I am sure that you all realize that we work in a for-profit industry and must be realistic-oriented. Yet, it should be possible to find commercial value even in basic research.

A further general pattern may be inferred at this point:

Individuals, having realized that features of their actions are counterproductive to learning, are unable to produce actions that encourage learning.

Several members of the class then pointed out that when they are on the receiving end of such positive statements as those in the right-hand column of List #3, they know that the diplomacy is used to cover up negative thoughts (like those shown in the left-hand columns). They also admitted that they themselves keep such thoughts private. Often they try to inquire discreetly, but this rarely seems to work because the other party senses that their discreet inquiry is a cover-up. More often, they bypass inquiry during their meetings but later spend hours holding private conversations to try to find out what was really going on.

Findings such as these have been produced for decades in executive program classrooms and workshops held not only in the United States but in Europe, South America, the Near East, and the Far East. How do we explain them? What are their consequences for organizational learning? We explore these questions in the following chapter.

5

The Case of the CIO: Primary and Secondary Inhibitory Loops

Case Description

The chief information officer (CIO) of a large electronics firm was told by the CEO, his superior, that an important organizational problem existed and had to be corrected.¹ The problem was that the Information Technology (IT) group was too large and too expensive. Moreover, its service to the line organization was inadequate.

The CEO reminded the CIO that this was not the first time he had spoken of this problem. He was becoming impatient. He warned that if costs did not go down and if the quality and efficiency of service did not become better, he would be forced to take drastic action that could include finding a new CIO. The CIO called a meeting of his immediate reports to take corrective action.

CIO and His Immediate Reports

How the Meeting Began

The CIO opened by telling his subordinates that he had received a "read-our-lips" order from line management: cooperation was nonexistent, and the information professionals were providing minimal value added, despite higher budgets.

He then said, "I want to discuss with you our ability to react to users' needs and the fact that we are always having difficulties with line

¹ This case was presented and discussed at an executive seminar conducted by Argyris at the Harvard Graduate School of Business Administration in 1993.

management. They are, after all, our customers. We must be concerned about meeting their needs." The information professionals responded as follows:

We are concerned about their needs. The big trouble is that they do not know what they want.

When they do [know what they want], they have no idea how long it will take to provide them with high-quality services. They want everything yesterday.

We have "had it up to here" with line management's complaints. The problem would be easily solved if the line gave us the people and resources we truly need.

The CIO expressed empathy with their frustration and anger and suggested that they might begin to turn things around by developing "a credible plan to respond to [customer] needs." The professionals responded in the following way:

There is no sense in planning; our users don't plan. Anyway, we are convinced that just about the time we think we are on top of things, they will make more demands and complain about what we are failing to do.

The CIO replied:

But since we do not have a solid plan, we cannot review the way we are managing our resources...As I see it, we have two choices. The first is to do what we are doing—and I believe that will be disastrous. The second is to break out of this mold and change the way we do business.

Members of the group countered by arguing that there was no way to change line management. As one said, "If you want to try, good luck." The CIO replied, "If planning isn't the way to go, how do you propose to solve the problem?"

The information professionals responded with increasing emotion. They said, in effect,

- 1. the problem is not solvable because line management makes impossible requests, and*
- 2. the information professionals are already killing themselves.*

"That's why the good people are leaving," said one individual. "I agree," said another, adding, "It is not fixable."

Virtually at the end of his patience, the CIO exclaimed: "We have to fix it because we have no choice! Otherwise we are not being responsible."

What Is Going On Here?

Clearly the information technology professionals are expressing frustration with and mistrust of the line executives, as well as their own superior. Their conversation is crafted in a way that makes a dialogue difficult. For example, they advocate their positions and make evaluations and attributions about line management in ways that do not encourage inquiry or testing. These are examples of their unillustrated, untestable evaluations and attributions about line management:

- Line does not know what they want.*
- Line makes demands with unrealistic deadlines.*
- If we meet their demands, they will follow up with more unrealistic demands.*
- The problems are unfixable because of line management's recalcitrance.*

The CIO's Reaction.

The CIO wanted to get the subordinates to be cooperative, and he also wanted to minimize the likelihood that they would see him as unfair and judgmental. Unlike his subordinates, he censored his evaluations and attributions and acted as if this were not the case. Asked to write out his private thoughts and feelings, he offered the following:

- These guys act like a bunch of babies.*
- They do not realize how insensitive and opinionated they are.*
- Sometimes I feel that I should read the riot act to them. They've got to wise up or all of us will lose.*

When asked what led him not to make these thoughts and feelings public, he looked astonished, "If I said these feelings and thoughts, all I would have done was add fuel to the fire." He was correct. His private thoughts and feelings were crafted in the same counterproductive manner as were his and his subordinates' public conversation.

The use of self-induced censorship in order to create conditions for dialogue is rarely successful. For example, when some of the professionals were asked if they had any idea of their bosses' private thoughts, they responded with words that were almost identical to the ones the CIO used. When they were asked what led them not to say so, they responded with the same look of astonishment. "Are you kidding," said one of them, "that would make things worse." Quite likely, the subordinates were also carrying on internal monologues that were not vocalized. Thus we have people holding

private conversations about each other and thinking that the others do not hear these conversations. In actuality, the others do hear them but act as if they do not.

Reflecting On the Action of the CIO and His Reports

The CIO decided to begin the meeting with a “take charge” attitude. He told his group about negative evaluations by the top, and warned them that the time had come for corrective action. He then requested a constructive dialogue about what could be done to correct the situation. The subordinates also had their own “take charge” attitude. They bypassed the CIO’s requests, arguing that the problems were caused by top management.

The CIO responded in two ways. First, he avoided publicly expressing his negative feelings, fearing that doing so could make the situation worse. We would agree. If he were to make public his negative evaluations and attributions, he would be likely to activate the same kind of defensiveness that his subordinates’ negative evaluations and attributions had activated in him. The very way he framed his private thoughts and feelings was counterproductive to learning. The irony was that his private thoughts were consistent with the views of the CEO. For example, both saw the IT professionals as uncooperative and acting childishly.

Thus the first action strategy the CIO used in the name of producing a positive dialogue increased the amount of information that was withheld, suppressed his personal feelings, bypassed the feelings of his subordinates, and acted as if he was not doing so.

The second way the CIO responded was to take a rational approach to the problem. He asked the group to develop a credible plan to respond to the needs of line management. The subordinates rejected this suggestion on the grounds that it was irrational: the line managers did not know how to plan, were not likely to be satisfied with a sound plan, and would only escalate their demands and criticisms.

So far, all three levels of participants seem to have the same strategy. All believe that they should take charge and warn the others that their actions are not acceptable. This activates a barrage of evaluations and attributions on all sides, crafted in ways that do not encourage learning. For example, the subordinates evaluate the line as unable to plan and attribute to them the intention of making life difficult for the IT group. The CIO privately felt the same about the IT

group but decided that in order to have a constructive discussion, he should remain rational and focus on developing new plans. Again, he appears to be struggling to remain at the rational level of planning, suppressing his private thoughts and feelings, in order to avoid provoking emotional responses that he sees as unconstructive.

This strategy fails to achieve its intended objectives. First of all, the subordinates respond with even more emotion and make more negative evaluations and attributions which they communicate with a sense of certainty. (If the CIO were to ask them openly how they knew their diagnosis was correct, they would probably say, “Trust us, we work with them, we know!”) Secondly, when some of the professionals were asked if they had any idea of their boss’s private thoughts, they used words that were almost identical to the ones the CIO used about himself. When they were asked, then, what led them not to discuss their views, one of them said, “That would make things worse. All you would get is a blowup.” Like the CIO, the subordinates were carrying on private conversations, making private attributions to their boss, which they also covered up.

Reflecting on these consequences, it seems fair to infer the following: The participants experienced an interest in solving the business problem, but their ways of crafting their conversation, combined with their self-censorship, led to a dialogue that was defensive and self-reinforcing. When this happens, the participants have to focus on two major problems:

1. the business problem they set out to solve in the first place and
2. the problem triggered by defensiveness—mistrust, its cover-up, and the cover-up of the cover-up.

The latter problem, we suggest, takes up a lot of the players’ span of attention. Not only do they have to listen to the other party and think about their response; they must strive to do all this and at the same time keep track of their cover-ups.

The Primary Inhibitory Loop

We began with a business problem identified as high IT costs and poor IT performance. All three levels of players agreed with this description of the problem, but they disagreed about its causes.

The CIO went into his meeting with three assumptions about the action strategies he should adopt in order to make the meeting effective:

1. I (CIO) should take charge of the meeting, if it is to meet my objectives. The definition of victory is that the subordinates develop a plan that will convince the senior line executives that they intend to cut costs.
2. It is important for me to involve my subordinates in developing a corrective plan. They will then be more committed to its implementation.
3. If my subordinates become emotional and do not act like grown-ups, I must suppress the latter and defuse the former by focusing on rational actions, such as planning. The way to suppress my negative feelings is to experience them privately, censor them from the group, and act as if I am not doing so.

The subordinates appear to have used action strategies based on similar assumptions. They, too, felt that they should take charge of the meeting. Their definition of victory is that the CIO will realize that the major obstacle to solving the problem is the line managers. They, too, seek to involve the CIO so that he will become a more effective spokesman for their views with top management. Finally, they assume that their responses are rational. Whenever they add emotional decibels, it is to make sure the CIO hears what they say.

These shared strategies and assumptions lead each side to craft conversations that upset the other side, leading, in turn, to expression and bypassing of the emotional dimensions of important issues, which then become undiscussable. A conversation that is intended to be positive actually produces defensive reactions in all players who deal with their defensive reactions in ways that reinforce and escalate defensiveness, again, with positive intentions (consequences similar to the ones created by the financial executives in the class described in the previous chapter).

We call these self-reinforcing patterns of action strategies and antilearning consequences **primary inhibitory loops**. The loops are primary in the sense that they are informed by the participants' theories-in-use during face-to-face discussions (especially when these are laced with embarrassment or threat). Within such loops, defensive, dysfunctional responses (based on theories-in-use like those shared by the CIO and his subordinates) are triggered by and, in turn, reinforce **conditions for error**, properties of information that tend to ob-

scure error and make it uncorrectable. Not being able to discuss important issues is one example; other such examples are vagueness and ambiguity. In the CIO's meeting with his subordinates, the actual causes of line/staff troubles remain vague (they are not clearly specified or illustrated) and ambiguous (the different interpretations of line/staff conflict are not clarified or resolved). Attributions that the CIO and his subordinates make to each other or to line management remain untested, and, to the extent that they cannot be tested so long as the issues and feelings associated with them are undiscussable, they become untestable.

A fuller, though not exhaustive, list of conditions for error, together with corrective responses to them, is given below:

Conditions for Error	Corrective Responses
Vagueness	Specify
Ambiguity	Clarify
Untestability	Make testable
Scattered information	Concert
Information withheld	Reveal
Undiscussability	Make discussable
Uncertainty	Inquire
Inconsistency/incompatibility	Resolve

In organizational settings, conditions for error trigger defensive reactions like those of the CIO and his subordinates; these reactions, in turn, reduce the likelihood that individuals will engage in the kind of organizational inquiry that leads to productive learning outcomes. Vagueness and ambiguity in organizational theory-in-use yield organizational situations that individual members find threatening. Uncertainty over the nature of troublesome situations, over what is to be done and by whom, or over criteria for performance, increase individual feelings of defensiveness and anxiety. When important information is withheld, when important issues are treated as undiscussable, individuals tend to feel mistrust and uneasiness. Incompatibilities in organizational theory-in-use tend to be expressed in interpersonal conflicts, which individuals then live out in terms of win/lose games.

Many readers will not find the existence of such patterns to be news. Indeed, many CIOs who participated in executive programs

have confirmed that they have experienced similar situations. Executives representing other managerial functions also recognized that the case material illustrates familiar patterns of interdepartmental competitiveness and mutual mistrust between line and staff.

Why do such counterproductive dialogues, with their primary inhibitory loops, occur so systematically? To our knowledge, no one argues for them; and texts and courses on organization design, leadership, conflict resolution, and employee involvement or empowerment, see these loops as violating both the letter and the spirit of management and organization theory. Why, then, do such counterproductive dialogues occur and persist when, most often, the participants go into them with constructive intent to solve a business problem? One answer to this question is related to the assumptions and skills individuals learn to use early in life to deal with issues that are embarrassing and threatening. But individuals are only part of the answer. Organizational phenomena also reinforce such counterproductive dialogues. If we study everyday activity in organizations, we would be hard put to separate the individual from the organizational. The two factors interpenetrate to such an extent that one must conclude there exists a circular relationship of pattern causality where each factor "causes" and reinforces the other.

We believe, then, that there are three levels of explanation for the patterns represented by primary inhibitory loops: the first is individual; the second, organizational; the third, an interaction of the two. Let us begin with the individual level.

Model I Theories-in-Use

As we pointed out in Chapter 4, human beings hold two types of theories of action about effective behavior. One is the theory of action that is espoused; the other is the theory that is actually used, the theory-in-use. We have found that when human beings deal with issues that are embarrassing or threatening, their reasoning and action conform to a particular model of theory-in-use which we call Model I. Model I informs the actions that enter into primary loops, as described above, with the effect of inhibiting double-loop learning. Neither the CIO nor his subordinates espoused such a theory of action, yet all of them used it.

In the table that follows we present a schema of Model I. The first column of the table lists "governing variables," or values, that actors strive to satisfy through their actions:

1. *Define goals and try to achieve them.* Participants rarely tried to develop with others a mutual definition of purposes,

Model I Theory-in-Use

Governing Variables	Action Strategies	Consequences for Behavioral World	Consequences for Learning, Effectiveness
Define goals and try to achieve them.	Design and manage the environment unilaterally (be persuasive, appeal to larger goals, etc.)	Actor seen as defensive, inconsistent, incongruent, controlling, fearful of being vulnerable, withholding of feelings, overly concerned about self and others, or underconcerned about others.	Self-sealing. Decreased long-term effectiveness.
Maximize winning and minimize losing.	Own and control the task (claim ownership of the task, be guardian of the definition and execution of the task).	Defensive interpersonal and group relationship (depending on actor, little help to others).	Single-loop learning.
Minimize generating or expressing negative feelings.	Unilaterally protect yourself (speak in inferred categories accompanied by little or no directly observable data, be blind to impact on others and to incongruity; use defensive actions such as blaming, stereotyping, suppressing feelings, intellectualizing).	Defensive norms (mistrust, lack of risk taking, conformity, external commitment, emphasis on diplomacy, power-centered competition and rivalry).	Little testing of theories publicly. Much testing of theories privately.
Be rational.	Unilaterally protect others from being hurt (withhold information, create rules to censor information and behavior, hold private meetings).		

nor did they seem open to being influenced to alter their perception of the task.

2. *Maximize winning and minimize losing.* Participants felt that once they had decided on their goals, changing them would be a sign of weakness.
3. *Minimize generating or expressing negative feelings.* Participants almost unanimously declared that generating negative feelings showed ineptness, incompetence, or lack of diplomacy. Permitting or helping others to express their feelings tended to be seen as poor strategy.
4. *Be rational.* This is the counterpart to value 3. It is an injunction to be objective, and intellectual, and to suppress feelings. Interactions should be construed as objective discussions of the issues, whatever feelings may underlie them.

The second column identifies action strategies that participants adopted in order to satisfy these governing variables:

1. *Design and manage the environment unilaterally.* Plan actions secretly and persuade or cajole others to agree with your definition of the situation.
2. *Own and control the task.* The CIO privately decides to resolve the task assignment through a full staff meeting and tries to get others to see things his way.
3. *Unilaterally protect yourself.* Keep yourself from being vulnerable by speaking in abstractions, avoiding reference to directly observed events, and withholding the thoughts and feelings that lead you to do what you do. In order to achieve mastery of the situation, keep your own thoughts and feelings a mystery.
4. *Unilaterally protect others from being hurt.* Withholding valuable and important information, telling white lies, suppressing feelings, and offering false sympathy are examples of this strategy. The speaker assumes that the other person needs to be protected and that the strategy of protection should be kept secret; neither assumption is tested. Thus the CIO protects his subordinates from his negative feelings about them, and they do the same for him. In doing so, the CIO also protects himself from their negative reactions to his feelings, and they protect themselves from his negative reactions and feelings.

To the extent that one behaves according to any of the four action strategies, one will tend to behave unilaterally toward others and protectively toward oneself. If successful, such behavior controls others and prevents one from being influenced by them. But as a consequence, the actor tends to be seen as defensive (since he or she is defending), and interpersonal and intergroup relations tend to become more defensive than facilitative, more a matter of win/lose than of collaboration. These effects tend to generate mistrust and rigidity.

Given these governing variables and strategies, there is likely to be little public testing of the assumptions embedded in theories-in-use, because such testing would require confronting one's own defensiveness and the defensiveness of others. Neither the CIO nor his subordinates could test their assumptions about each other's mistrust, for example, without confronting their own defensiveness and without risking the negative reactions that would be likely to follow.

If there is little genuine public testing of one's theory-in-use and if one must nevertheless act, then one will act on an untested theory-in-use. Since behaving according to one's theory-in-use will influence the behavioral world, self-sealing will probably occur. Thus the CIO eventually feels a sense of mistrust, cynicism, and a lack of confidence in his subordinates. The subordinates feel the same about the CIO. None of these evaluations and attributions are tested publicly. Hence, they become sealed in self-reinforcing, defensive loops. In our earlier book, *Theory in Practice* (Argyris and Schön, 1974), we describe the resulting situation as follows:

...Lack of such public testing risks creating self-sealing processes...the individual not only helps to create behavioral worlds that are artifacts of his theory-in-use but also cuts himself off from the possibility of disconfirming assumptions in his theory-in-use and thereby cuts himself off from the possibility of helping to create behavioral worlds that disconfirm his starting assumptions about them. However, public testing of theories-in-use must be accompanied by an openness to change behavior as a function of learning. The actor needs minimally distorted feedback from others. If others provide such feedback—especially if they do so with some risk—and if they experience that the actor is not open to change, they may believe that they have placed themselves in a difficult situation. Their mistrust of the actor will probably increase, but this fact will be suppressed. The result will be the creation of another series of self-sealing processes that again make the actor less likely to receive valid information the next time he tries to test an assumption publicly... (p. 78)

Because double-loop learning depends on the exchange of valid information and public testing of attributions and assumptions, Model I tends to discourage it. Because long-term effectiveness depends on the possibility of double-loop learning, Model I tends toward long-term ineffectiveness.

All of these consequences of governing values and action strategies of Model I feed back to reinforce those values and strategies. In a world of defensiveness, escalating errors, and self-fulfilling processes, it is understandable that individuals should protect themselves by striving even harder to be in unilateral control, to win and not lose, to deal with the defensiveness of others by attempting to be, and encouraging others to be, "rational," and to suppress, as best they can, their own and others' negative feelings, as in the left-hand columns of the participants' cases.

Another result of Model I is that social virtues such as concern, caring, honesty, strength, and courage become defined in ways that support Model I theory-in-use. For example, concern and caring come to mean: "Act diplomatically; say things that people want to hear"—meanings that lead to action strategies such as easing-in, covering-up, and telling white lies. Strength becomes defined in terms of winning, maintaining unilateral control of the situation, and keeping private one's feelings of vulnerability.

There is another factor that powerfully reinforces Model I, increasing the likelihood of antilearning processes. Individuals are highly skilled in their execution of Model I. Skillful actions usually "work," in the sense of achieving their intended objectives; they appear spontaneous, automatic, and effortless; they are taken for granted; and they require little conscious deliberation. These features combine to make it less likely that the actors will reflect on and learn about their Model I behavior.

We have found the distinguishing features of Model I theory-in-use not only in executive sessions but in countless meetings in all sorts of organizations. Moreover, in any setting where the actors use Model I theories-in-use, we have never observed consequences opposite to those listed above. And as already noted, the overwhelming majority of the people we have studied (close to 99 percent) use Model I theory-in-use in threatening or embarrassing situations.

This brief sketch of Model I will be illustrated and extended in the cases that follow. The elements of Model I interact in ways that are far more complex than we have so far described. Similarly, there are many ways in which the behavioral world created by Model I behavior feeds back to reinforce that behavior.

The Secondary Inhibitory Loop

We use the term **secondary inhibitory loop** to refer to the behavioral loops—causal connections between action strategies and antilearning consequences—that are supra-individual, pertaining to interactions of groups within organizations. These loops are secondary in the sense that primary loops lead to them, although they become self-reinforcing once they have been set in motion.

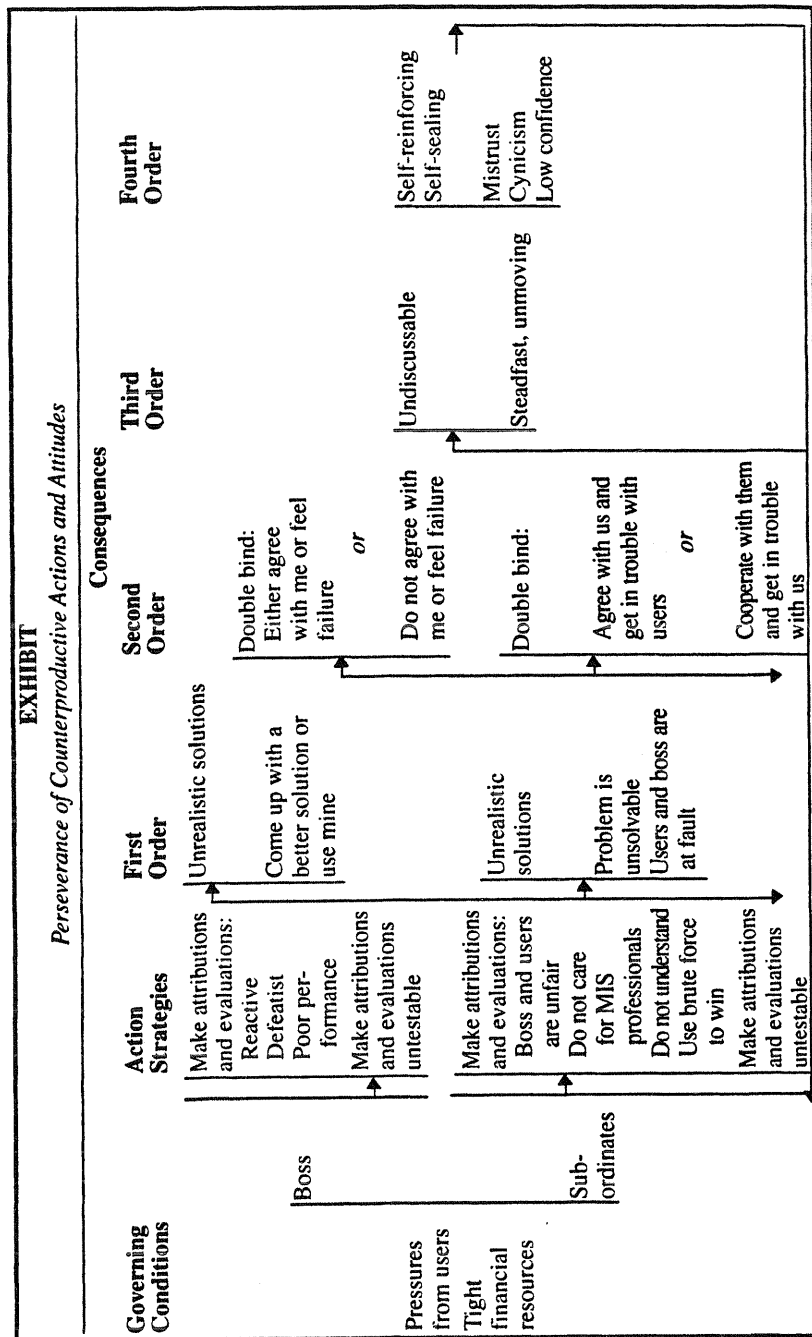
Let us return to the case of the CIO to illustrate how face-to-face interactions lead to secondary inhibitory loops that make antilearning actions and attitudes persevere, even though most participants wish they did not.

In the exhibit that follows (on the next page), we present a map of the pattern of relationships among primary and secondary inhibitory loops. We begin with the governing conditions that set the pattern in motion: the dissatisfactions of the users and the tight financial conditions that no longer permit high IT costs. Next we have the action strategies the CIO and his subordinates used during their meeting. These action strategies led to certain first-order consequences. One was that the boss evaluated the subordinates' conclusions as unrealistic and challenged them to come up with a better solution or to use his solution. The subordinates, in turn, evaluated the boss's solution as unrealistic and said that the problem would remain unsolvable unless line management changed.

These interactions led to second-order consequences in the form of two sets of double binds. One affected the group of subordinates; the other affected the CIO and would have affected line managers had they been present.

We now have the beginnings of the transition from primary to secondary inhibitory loops. For example, line managers judge IT staff as unrealistic and lacking in cost-consciousness; the staff attributes the underlying problems to line managers themselves. Both views become embedded in the organizational norms that govern relationships between line and staff. The double binds that follow from such judgments and norms feed into a pattern of intergroup conflict typical of line-staff relationships.

The ramification of secondary inhibitory loops continues. Sensitive issues of intergroup conflict become undiscussable, and their not being discussed becomes undiscussable. Each group sees the other as unmovable, and both see the problem as uncorrectable. Again the literature of line-staff relations is replete with illustrations of such effects.



All of these consequences feed back to reinforce the primary inhibitory loops illustrated in the CIO's meeting with his subordinates, creating an organizational pattern that seals in the counterproductive processes. This leads each participant to feel mistrust of the others, low confidence in interactions with others, and cynicism about the likelihood of resolving intergroup conflicts. These feelings are legitimized because they are seen as typical of the organization.

Our map shows that any dialogue around the business problem of reducing IT's costs and improving its performance will be negatively influenced by the primary and secondary inhibitory loops. If the top line management is unaware of the detailed processes shown in our map because they have been shielded from them, they see only the persistence of high costs and poor performance. This leads them to mistrust the CIO and feel doubt, even cynicism, about his ability to solve the business problem. (As it turned out, the CIO in our case was eventually fired and a "tough line officer" was placed in charge.) On the other hand, if the top line management were not shielded from the detailed processes shown in our map, they would probably still have similar reactions. In either case, little corrective action would be taken to improve the organization's capability for double-loop learning. Indeed, several months after his conversation with the CIO, the CEO faced a similar pattern of defensive interactions, this time between Research and Development (R&D) and Finance. Finance claimed that R&D cost too much; R&D counterclaimed that Finance had no real understanding of research and development. Again, the CEO solved the problem by taking a tough stance on controlling research costs.

We have now shown how primary and secondary inhibitory loops can produce consequences counterproductive to learning at all levels of organization. The dynamic interaction within an organizational setting of primary and secondary inhibitory loops, together with their antilearning consequences, is what we mean by a **limited learning system**.

Secondary Inhibitory Loops and Organizational Defensive Routines

Among the most important components of secondary loops are organizational defensive routines. These are actions and policies, enacted within an organizational setting, that are intended to protect individuals from experiencing embarrassment or threat, while at the same time preventing individuals, or the organization as a whole, from

identifying the causes of the embarrassment or threat in order to correct the relevant problems.

All organizational defensive routines are based on a logic that is powerful and profound in its impact on individuals and organizations. The logic can be expressed in terms of four rules:

1. Craft messages that contain inconsistencies.
2. Act as if the messages are not inconsistent.
3. Make the ambiguity and inconsistency in the message undiscussable.
4. Make the undiscussability of the undiscussable also undiscussable.

An example of a message conforming to these rules is a chief executive who says to his immediate subordinates,

We encourage everyone to be innovative and risk oriented. This is what we mean by empowerment. Of course, we also expect you to keep out of trouble.

When individuals communicate mixed messages, they usually do so spontaneously and with no indication that their message is mixed. If they appeared to be hesitant because of the inconsistencies in their message, it could be seen as a weakness. It is rare, indeed, for an executive to design and state a mixed message and then ask, "Do you find my message inconsistent and ambiguous?" The message is made undiscussable by the very naturalness with which it is delivered and by the absence of any invitation or disposition to inquire about it. Moreover, the very undiscussability of a mixed message constitutes a source of threat or embarrassment. In a Model I organizational world (which we call O-I), discussion of the undiscussability of a mixed message would trigger as much defensiveness as the mixed message itself.

Individuals follow such rules all the time, and they do so without having to pay attention to them because they have become highly skilled at enacting them. The irony is that this skillfulness is inextricably intertwined with incompetence, because the skillful use of mixed messages leads to a range of unintended and counterproductive consequences. For example, the CIO and his subordinates created a dialogue in which crucial messages were covered up, and the cover-up was not discussable. This led to increasing emotionality, as well as to double-binds, which, in turn, led to or reinforced existing feelings of mistrust, cynicism, and lack of confidence in the other parties to the dialogue.

We suggest that it is not possible to deal effectively with any subject if it is not discussable and if its undiscussability is also undiscussable. Under the rules that govern defensive routines, individuals with a high sense of integrity and willingness to accept personal responsibility will feel that they are in the following double-bind:

If we do not discuss the defensive routines, then these routines will continue to proliferate. But if we do discuss them, we are likely to get into trouble.

One colorful senior executive said that in his organization these double binds go under the name of "s--- sandwiches."

The result of such double-binds is that defensive routines are protected and reinforced by the very people who would like to get rid of them. But because their protection is covert and undiscussable, defensive routines appear to other people as self-protective and self-reinforcing.

Whenever actions are self-protective and self-reinforcing, they can easily become self-proliferating. The irony here is that the self-proliferating features of defensive routines are activated especially when someone tries to engage them directly. Once individuals realize that danger, they shy away in the name of progress and constructive action.

Under these conditions, defensive routines flourish and spread into organizational loops that are known to all and manageable by none. Indeed, executives have told us that the thought that defensive loops could be managed is unrealistic, futile, or romantic. A few have wondered if such management might not be dangerous because, as one put it, "Wouldn't it mean that we would have to give up whatever we have to protect us?"

These reactions make sense in the world as it is. They are also self-fulfilling and self-sealing—self-fulfilling because they create the conditions under which it would be naive or dangerous to engage them; self-sealing because they also create conditions under which it is unlikely that the self-fulfilling prophecy will be interrupted.

Hence, we have one of the most important causes of organizational rigidity and stickiness: defensive routines that get stronger and stronger while the individuals responsible for them believe it is unrealistic or even dangerous to do much about them.

Reactions to Defensive Routines. Because defensive routines and the secondary inhibitory loops associated with them, are accepted as inevitable, natural, and immune to management or influence, it is

not surprising that the most common reaction to them is a sense of helplessness. Employees in industrialized societies appear as fatalistic about them as peasants do about poverty.

The inevitability of defensive routines, sanctioned by the prevailing culture, also has a personal side. Individuals do not take responsibility for creating or maintaining defensive routines. They are willing to say that they are personally influenced by defensive routines but are unable or unwilling to see how they may create or reinforce them.

One way to live with having little choice about defensive routines is to develop a cynical attitude about them. Cynicism leads to pessimism and doubt. For example,

Nothing will change around here.

They don't really mean it.

I doubt if anyone will listen.

Hang on. Don't get fooled. Next year there'll be a new fad.

Cynical attitudes make it more likely that individuals will ignore or sneer at evidence of positive intentions. The cynic automatically mistrusts other people and sees the world as full of evidence that nothing will change.

It is a short step from cynicism to blaming other people in the organization for any difficulties that may arise, and people will have plenty of evidence that someone else is to be blamed. They can easily infer the existence of defensive loops, they can see individuals acting consistently with them, they can see the cover-ups, and they can see that promotions often go to individuals who bypass the defensive routines.

Finally, people often give to others advice that reinforces defensive routines. For example, "Be careful. You'll get yourself in trouble if you try to change...That is a legacy from way back." So now we have the very individuals who feel helpless and cynical and disposed to blame others for taking initiative, becoming "positive" by advising others to respect defensive routines and inhibitory loops—the very phenomena that make it difficult for people to take constructive initiatives in organizational life.

To continue the propositions that we began in the previous chapter, we suggest the following:

Individual and supra-individual unities exist in circular, interdependent relationships with each other. When embarrassment or threat are involved, these relationships interact to create self-fueling, limited-learning processes.

For double-loop learning to occur and to persist at any level in the organization, the self-fueling processes must be interrupted. In order to interrupt these processes, individual theories-in-use must be altered.

How the Technical-Objective Dimension of Organizational Life Is Smothered By Individual-Organizational Defensive Routines

In our research and consulting, we are faced with issues that vary in their degree of objectivity, that is, the degree to which they are seen as matters of fact or as subjects of merely technical theories. Typically, when we deal with issues that are defined by the use of technical theories, we expect that their degree of objectivity will make it difficult for Model I and O-I defensive features to dominate the dialogue. This is not necessarily the case.

Technical theories are theories of action that specify measurable objectives and procedures for achieving those objectives. Some of the technical theories familiarly encountered in organizations include the following:

- accounting
- finance
- economics of the firm
- information technology
- marketing
- competitive strategy
- research and development
- certain theories related to human resource functions, such as incentive schemes, personnel selection, and training.

All technical theories are characterized by a set of common features. First, there is an aspiration to reduce to a minimum the gap between the technical espoused theory and the technical theory-in-use. The technical professional tries to perform in accordance with the prescriptions of established technical theory. Second, there is an explicit emphasis on the use of productive reasoning. Those who construct technical theories aspire to make the premises and inferences from them as explicit as possible. They aspire to derive conclusions that are testable, especially falsifiable. Technical theories are, therefore, written to be causally rigorous because that is required if the procedures are to specify in causal terms what actions will lead to what consequences. Technical theories are causally rigorous in another way. If two professionals follow a set of technical specifications correctly, they will collect the same data, make the same analyses, and produce the same technical conclusions. If this does not happen,

it is possible to trace backwards to find where one or both professionals deviated from the established procedures.

It seemed plausible to us that these features of technical theories would be so powerful that they would not succumb easily to the individual-organizational defenses we have described earlier. We thought that the robustness of their intellectual objectivity should make it easier to reduce conditions for error such as vagueness and ambiguity.

So far, however, our expectations have not been confirmed. What often happens is that the objective features of technical theories are bypassed and submerged by Model I defenses. For example, a superior (S) wrote a case about his relationship with a subordinate (O) regarding the latter's inadequate performance around certain technical issues in an information management system.

S and O Case

Thoughts and Feelings

I am concerned (angry) about what is a continual problem.

I sense that he is avoiding responsibility for the problem.

My feeling is that the group should accept responsibility.

I again see the avoidance. I want to lead him toward a solution where he can take responsibility.

I feel that he cannot concede the point and will move to other issues as an escape.

I begin to feel frustration. I'm canceling out the additional excuses.

Actual Conversation

S: We need to find ways to have your group deliver part of the product on time.

O: It's simple. We cannot test our debugger until the compiler has finished all of its testing.

S: Are there any tests that can be run before the compiler is ready?

O: There are limited tests that can be run. But the cause of the most difficulties is in the compiler/debugger interaction.

S: But it is possible to capture correct compiler output and run your tests against that.

O: Sure, we could do that, but it would not catch places where the compiler has changed. Besides it would also take more disk space. It is simple; we are dependent on the compiler.

S: First of all, disks are cheap. If you need more space, we can get it. Second, there are other components that interact with the compiler that do not have the same problems with delivering.

I am led in another direction.

O: The other components do not interact as closely. Look at the last release. The compiler added new features, and we did not find out until the end.

In reading the right-hand column, the dialogue appears to be about the delays around the compiler and the debugger. The performance of both machines and their interaction is specifiable in technical terms, namely, in terms related to the domain of computer and information technology. Indeed, the disagreement between S and O, as it appears in the right-hand column is crafted primarily in terms of technical issues.

Let us now expand our view. S, who wrote the case, wrote that he was frustrated with O's performance. S doubted O's explanation for delays, namely, that the debugger could not be tested until the compiler finished its testing. This claim could have been tested because the technical theories involved specify the performance features of each machine and their interrelationships. But this technical test was not requested or required by S.

One reason that S did not force such a test was that he believed that the important issue was that O and his group were avoiding their responsibility. S was faced with a leadership and group-performance issue. This claim is illustrated by the left-hand column comments. Yet the case suggests that S acted in ways that suppressed the primacy of the interpersonal-organizational defensive issues. He appeared to hope that by making the technical issues primary he could, through appropriate questioning, eventually surface the leadership and group performance issues. S's strategy was, therefore, to make secondary what he believed was primary and to coverup that he was using such a strategy.

O, on the other hand, crafted his conversation to deal with the technical issues. He was able to distance himself from the interpersonal-organizational dimension that upset S. This resulted in a counterproductive dialogue. S began by noting that disks were cheap (technical). If S would provide more space (technical), and since other components interact with the compiler (technical), then the technical problems could be resolved. O found reasons why S's technically based solution was inadequate. S saw O's emphasis on technical issues as further evidence that O was acting irresponsibly. O could argue that he was doing so because he had not been told of S's view of O's irresponsibility.

S covered up by focusing only on the technical issues, and he acted as if this were not the case (Model I). S never engaged O about his unhappiness over O's avoiding responsibility. O, in his responses, remained at the

technical level and appeared to be designedly sidestepping and acting as if this were not the case (again, Model I).

It is our hypothesis that each individual in crafting his arguments retrieves knowledge from his mind that is related to technical theories of action (compiler/debugger interaction) and states it forthrightly. Each has little choice but to be forthright in this respect, because it is hard to distort technical features based on publicly stated theories without giving the other party the opportunity to falsify the claim.

S also crafts his conversation in ways that cover up his feelings and acts as if he is not doing so. The difficulty with this strategy is that it makes it easy for O to remain at the technical level and, if he were sidestepping, to act as if he were not.

We have a conversation, therefore, that is unlikely to resolve the problem that S believes is crucial (O's avoiding responsibility). If O is sidestepping because he believes that S's requests are unfair, then that problem will also not be solved. S and O can end the conversation by privately attributing negative evaluations to the other party, each feeling that he is dealing with a difficult individual. What results from such a conversation are the self-fulfilling prophecies and self-sealing processes predicted by Model I, the patterns of an O-I limited learning system.

Conclusion

Individuals are programmed with Model I theories-in-use. When faced with embarrassing or threatening issues, they act in ways that enhance conditions for error. For example, important features of issues become undiscussable, and their undiscussability is also undiscussable. This exacerbates the degree of inconsistency and incongruity, the vagueness, and ambiguity that surround the issues. These consequences lead to organizational behavioral worlds that are dominated by organizational defensive routines. Such defensive routines reinforce the counterproductive learning consequence of Model I theories-in-use and O-I learning systems. At the same time they also create such a degree of interpenetration between individual and organizational defensiveness that it becomes difficult to disentangle the causal roles of these two levels of phenomena. The result is for individuals to experience mistrust, distancing, and cynicism about the potentiality for productive organizational learning around issues that are embarrassing or threatening.

This, in turn, results in a low likelihood that high-quality inquiry will occur. The conditions for what we have called "good dialectic" will be suppressed. There will be a tendency to minimize conflicts and disagreements by bypassing them and covering up the bypassing. It will be difficult to reflect on and produce an accurate history of the problem in question. It is likely that the participants will have difficulty even in agreeing on what actually happened. It is even more likely that they will not try to test their interpretations of events. If they do try, their tests are likely to be poor ones because they will be crafted to avoid threatening the reasoning used by the individuals who made the interpretations in the first place.

At the heart of explaining human behavior are the concepts of reasoning and causality. Human beings use reasoning to diagnose what is going on, to design actions, and to produce their designs. The concept of causality plays a key role in all of these processes because human action is intended to be effective. Effectiveness, in turn, requires having some concept of "If A..., then B..." when diagnosing, inventing, and producing.

Model I theories-in-use and organizational defensive routines combine to sanction the use of defensive reasoning. Defensive reasoning consists in making one's premises and inferences implicit and invulnerable to public testing. This leads to conclusions that are testable only within the constraints of the logic used by the actors who crafted the conclusion. We call this the use of self-referential logic because the testing is not designed to utilize logic independent of the logic used to create the conclusion in the first place. Under these conditions, defensive reasoning becomes sanctioned as the correct reasoning to use. But, as we have seen, this is a recipe for exacerbating conditions for error and for diminishing the condition for good dialectic. We have a paradox: the behavioral strategies that are defined as effective also reduce the likelihood of productive learning at all levels of the organization.