3

Theories of Action

The idea of an action science that we have been describing is grounded in our practice as researchers, educators, and interventionists working within the theory of action approach (Argyris and Schön, 1974, 1978; Argyris, 1976, 1980, 1982, 1985). It may well be that other research programs and theoretical approaches can provide alternative ways of conducting action science. Our particular approach, however, provides the perspective from which we have been able to envision an action science. It is also the perspective from which we criticize and provide alternatives to examples of mainstream research in Part Two, and from which we describe the process of learning the skills necessary to conduct action science in Part Three. Hence, this chapter presents the theoretical orientation which informs our work.

The theory of action approach begins with a conception of human beings as designers of action. To see human behavior under the aspect of action is to see it as constituted by the meanings and intentions of agents. Agents design action to achieve intended consequences, and monitor themselves to learn

if their actions are effective. They make sense of their environment by constructing meanings to which they attend, and these constructions in turn guide action. In monitoring the effectiveness of action, they also monitor the suitability of their construction of the environment.

The complexity of the design task far exceeds the information-processing capabilities of the human mind (Simon, 1969). Designing action requires that agents construct a simplified representation of the environment and a manageable set of causal theories that prescribe how to achieve the intended consequences. It would be very inefficient to construct such representations and theories from scratch in each situation. Rather, agents learn a repertoire of concepts, schemas, and strategies, and they learn programs for drawing from their repertoire to design representations and action for unique situations. We speak of such design programs as *theories of action.*

Theories of Action

These theories may be thought of as a very large set of complexly related propositions. The form of a proposition in a theory of action is, "In situation *s,* to achieve consequence *c,* do action *a*" (Argyris and Schön, 1974). From the perspective of the agent who holds the theory, it is a theory of control. It states what the agent should do to achieve certain results. From an observer's perspective, to attribute a theory of action to an agent is to propose a theory of explanation or prediction. In the language of the previous chapter, it is to make a dispositional attribution. The example we used is, "John follows the rule, 'If I am about to deprecate someone, first deprecate myself.' " But from John's perspective, this is a theory of control. We can see this by making explicit the intended consequence of enacting the rule, which, let us suppose, is to avoid making the other person defensive. Hence a proposition of a theory of action can be understood both as a disposition of an agent and as a theory of causal responsibility held by an agent.

Espoused Theory and Theory-in-Use. There are two kinds of theories of action. Espoused theories are those that an indi-

vidual claims to follow. Theories-in-use are those that can be inferred from action. For example, when asked how he would deal with a disagreement with a client, a management consultant said that he would first state what he understood to be the substance of the disagreement, and then discuss with the client what kind of data would resolve it. This was his espoused theory. But when we examined a tape recording of what the consultant actually did in that situation, we found that he advocated his own view and dismissed that of the client.

The discrepancy between what people say and what they do is an old story. It is sometimes expressed in the saying, "Do as I say, not as I do." But the distinction between espoused theory and theory-in-use goes beyond this common conception. It is true that what people do often differs from the theories they espouse. We are saying, however, that there is a theory that is consistent with what they do; and this we call their theory-in-use. Our distinction is not between theory and action but between two different theories of action: those that people espouse, and those that they use. One reason for insisting that what people do is consistent with the theory (in-use) that they hold, even though it may be inconsistent with their espoused theories, is to emphasize that what people do is not accidental. They do not "just happen" to act in a particular way. Rather, their action is designed; and, as agents, they are responsible for the design.

Espoused theory and theory-in-use may be consistent or inconsistent, and the agent may or may not be aware of any inconsistency. The agent *is* aware of the espoused theory, by definition, since it is the theory that the agent claims to follow. Recall in this connection our previous discussion of tacit knowledge and rule-governed behavior. As many approaches to social inquiry emphasize, human beings can be understood to act according to rules that they cannot state.

Theories-in-use are the often tacit cognitive maps by which human beings design action. Theories-in-use can be made explicit by reflecting on action. But we should note that the act of reflection is itself governed by theories-in-use. Becoming an action scientist involves learning to reflect on reflection-in-

action, making explicit the theories-in-use that inform it, and learning to design and produce new theories-in-use for reflection and action.

Nested Theories. Theories of action can be articulated at different levels of detail. Consider a theory-in-use description of the common interpersonal strategy that we call "easing-in": "If you are about to criticize someone who might become defensive and you want him to see the point without undue resistance, do not state the criticism openly; instead, ask questions such that if he answers them correctly, he will figure out what you are not saying." This is a relatively high-level design program. The action injunction to ask questions of a certain kind could be performed only with the help of a great many detailed propositions about the production of grammatical sentences, motor skills for speech, and so forth. Similarly, the recognition of a particular situation as one in which another person may become defensive requires the coordinated performance of complicated perceptual routines and higher-order judgments about human defensiveness.

A full specification of the theories of action held by any individual would be enormously lengthy and complex. Yet in order to understand theories of action it is necessary to make them explicit. What is required are models—simplified representations—chosen to illuminate those features of theories of action most relevant to particular fields of inquiry. Linguistics, for example, selects those features relevant to the production of grammatical sentences.

Action science concentrates on the level of abstraction at which agents in everyday life reflect on their actions. The spotlight of attention shifts depending on the concerns of agents in particular situations, but the example of the questioning strategy of easing-in illustrates a characteristic level of abstraction. Action science would probably not focus on the particular intonation patterns by which speakers of English communicate doubt, although the action scientist might indeed inquire into whether hearers attributed doubt to a speaker or into the reasoning that led an agent to communicate doubt.

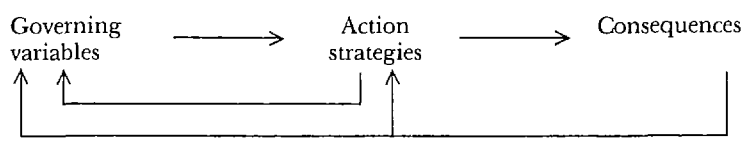
The models we create in action science are shaped by our

interest in helping human beings to make more informed choices in creating the worlds in which they are embedded. Because we are interested in helping human beings design and implement action, our models should be "connectable" to concrete situations. We seek both generalizability and attention to the individual case. This is difficult to achieve, but probably no more difficult than for all human beings who must do this as agents in everyday life.

As action scientists we are concerned with the effective functioning of interventionists in behavioral systems, which range from individuals to groups, intergroups, organizations, and communities. We are therefore centrally concerned with the features of theories of action that promote or inhibit learning in behavioral systems. The models we will describe were developed in light of earlier theories about what it is in the way human beings interact that leads to escalating error and the inability to learn (Argyris, 1970). The models provide ways to test and extend the theory and, at the same time, help client systems to reflect on their theories-in-use and to learn new theories-in-use.

Modeling Theories-in-Use. Models of theories-in-use can be constructed according to the schematic frame shown in Figure 1. Governing variables are values that actors seek to satisfy.

Figure 1. Theory-in-Use Model.



Each governing variable can be thought of as a continuum with a preferred range. For example, individuals would not want anxiety to get too high; but they might not want anxiety to drop to zero, either, lest they feel bored. We commonly speak of actions as directed toward an end state, as if there were only one relevant governing variable; but in fact, human beings live in a field of many governing variables. These variables can be ig-

nored as long as their values are within a satisfactory range; when one falls out of that range, however, the actor takes steps to bring it back to a satisfactory level. Any action can have an impact on many governing variables. Agents typically must trade off among governing variables, because actions that raise the value of one may lower the value of another.

Action strategies are sequences of moves used by actors in particular situations to satisfy governing variables. Action strategies have intended consequences, which are those that the actor believes will result from the action and will satisfy governing variables. Consequences feed back to action strategies and governing variables.

Actions have consequences for the behavioral world, for learning, and for effectiveness. Consequences may be intended or unintended, productive or counterproductive. Consequences that are unintended may nevertheless be designed, in the following sense: Action intended to achieve particular consequences may, by virtue of its design, necessarily lead to consequences that are unintended. For example, the questioning strategy of easing-in typically creates the very defensiveness that it is intended to avoid, because the recipient typically understands that the actor is easing-in. Indeed, easing-in can be successful only if the recipient understands that he is supposed to answer the questions in a particular way, and this entails the understanding that the actor is negatively evaluating the recipient and acting as if this were not the case.

The consequences of action depend on the theories-in-use of recipients as well as those of actors. One's theory-in-use includes a vast store of information about what people are like and how they will respond in various situations. We can argue that consequences are designed, whether they are intended or unintended, when they necessarily follow from the action and the actor's presuppositions about the theories-in-use of recipients.

Single-Loop and Double-Loop Learning. When the consequences of an action strategy are as the agent intends, then there is a match between intention and outcome, and the theory-in-use of the agent is confirmed. If the consequences

are unintended, and especially if they are counterproductive, there is a mismatch or an error. The first response to error is typically to search for another action strategy that will satisfy the same governing variables. For example, if the agent wants to suppress conflict (governing variable) and to this end avoids saying anything that might be controversial (action strategy), but others raise threatening issues anyway (mismatch), the agent may try the strategy of talking volubly about issues on which everyone is likely to agree. In such a case, when new action strategies are used in the service of the same governing variables, we speak of *single-loop learning.* We do so because there is a change in action but not in the governing variables.

Another possibility is to change the governing variables themselves. For example, rather than suppress conflict, the agent might choose to emphasize open inquiry. The associated action strategy might be to initiate discussion of conflictual issues. In such cases we speak of *double-loop learning.* We suggested in the previous chapter that the deliberative process appropriate to double-loop learning is similar to what Rorty (1979) calls abnormal discourse. It is concerned not with choosing among competing chains of means-ends reasoning within a given set of standards, but with choosing among competing sets of standards ("frames" or "paradigms").

Having made a clear distinction between single-loop learning and double-loop learning, we must note that in fact the two kinds of learning exist on a continuum. Values and strategies may be nested, and learning that is double loop with respect to particular actions may appear single loop with respect to more encompassing governing variables. For example, both the effort to suppress conflict and the effort to discuss conflict openly might be in the interests of "getting others to do what I think best." Double-loop learning, on this account, might involve designing ways to jointly decide whether to discuss conflict. Another way of thinking of this nested quality is in terms of second-order standards by which alternative frames or paradigms may be evaluated.

There are several cues by which double-loop problems may be identified in practice. Dealing with double-loop problems requires dealing with the defenses of human beings. Thus

situations in which participants give cues that they or others might feel embarrassed or threatened are likely to require double-loop learning. Problems that are undiscussable are likely to be double loop; and the undiscussability itself is most certainly a double-loop problem, as is the cover-up of the undiscussability. Problems that persist despite efforts to solve them are likely to have double-loop issues embedded in them. For example, a business might continue to lose money because it has higher costs than its competitors or because its market is shrinking. But if the business is trying the wrong solutions or if there are no solutions within its capability, why are these errors not discovered? Persistent errors in learning point to double-loop problems. These are problems that require inquiry into governing variables if they are to be solved in such a way that they remain solved.

There are instructive parallels between the theory of action approach and family systems theory. Thus, in discussing families, Watzlawick, Weakland, and Fisch (1974) distinguish between first-order change and second-order change. First-order change occurs when people decrease deviation from a set norm. This is an application of the cybernetic principle of negative feedback mechanisms (Watzlawick, Beavin, and Jackson, 1967, p. 31). For example, a thermostat turns on the furnace when the temperature falls below a certain point. In our language, the thermostat is a device for single-loop learning.

Second-order change is necessary, argue Watzlawick, Weakland, and Fisch, when the structure of the family system must undergo change. Indeed, they argue that families develop problems (and wind up in the therapist's office) when attempts at first-order change are made in situations in which the system's structure itself has to undergo change. They refer to this problem cycle as the "more of the same" syndrome. For example, Father may nag Johnny about his schoolwork, which leads the rebellious Johnny to play hookey more often, which leads Father to escalate his nagging and threats, and so forth. The therapist's role is to bring about second-order change—change that alters the structure of the system, often by interventions that reframe the meaning of what is going on.

In our approach to action science we focus on double-

loop learning, the theories-in-use that inhibit it, and the ways in which theories-in-use that enhance a system's capacity for double-loop learning may be learned. We find that individuals and organizations are generally competent at single-loop learning, but are generally incompetent at double-loop learning. Family systems theorists seem to agree that human beings have little competence for double-loop learning and that the interventionist provides help by enabling double-loop learning to occur. But there is an important difference between family systems theory and our approach. We are concerned with increasing the client system's capacity for double-loop learning. The family therapist is concerned with bringing about a double-loop change, but not with increasing the family's competence to engage in double-loop learning more generally. For example, a therapist might help a family to solve a problem such as chronic bickering; but members of the family do not learn the skills to correct such situations without intervention by an outsider. Watzlawick, Weakland, and Fisch recognize this, but it does not concern them. A family might have to return to the therapist when a new problem arises; but the therapist, who is not concerned with helping the family gain insight into its problem but rather with correcting the problem, is supposed to be able to work quickly. It may in fact be more efficient to go to a therapist periodically than to learn skills for double-loop learning on one's own; and so this may be an appropriate strategy in family therapy, especially with families that are not articulate and reflective. But our practice is with organizations and with helping professionals, both counselors and consultants. These are clients for whom it is important not only to solve particular problems but to increase their capacity for double-loop learning more generally.

Model I Theory-in-Use

Argyris and Schön (1974) developed a model, or an ideal type, that describes features of theories-in-use that inhibit double-loop learning. While espoused theories vary widely, research indicates that there is almost no variance in theory-in-

use (Argyris, 1976, 1982). More precisely, the theories-in-use of virtually everyone we have studied are consistent with the master program called Model I. There is considerable difference within Model I in the weightings individuals give to particular governing variables, as well as in the particular strategies individuals favor, but these lower-order variations appear to be governed by the Model I master program (see Table 1).

The four governing variables of Model I are (1) achieve the purpose as the actor defines it; (2) win, do not lose; (3) suppress negative feelings; and (4) emphasize rationality.

The primary behavioral strategies in Model I are to control the relevant environment and tasks unilaterally and to protect oneself and others unilaterally. Thus, the underlying behavioral strategy is unilateral control over others. Characteristic ways of implementing this strategy include making unillustrated attributions and evaluations, advocating courses of action in ways that discourage inquiry, treating one's own views as obviously correct, making covert attributions, evaluations, and face-saving moves such as leaving potentially embarrassing facts unstated.

The consequences of Model I strategies include defensive interpersonal and group relationships, low freedom of choice, and reduced production of valid information. There are negative consequences for learning, because there is little public testing of ideas. The hypotheses that people generate tend to become self-sealing. What learning does occur remains within the bounds of what is acceptable. Double-loop learning does not tend to occur. As a result, error escalates and effectiveness in problem solving and in execution of action tends to decrease.

In claiming that human beings are programmed with Model I theory-in-use, we are making predictions about the kinds of strategies they will and will not use, and the kinds of consequences that will and will not occur. These predictions have been tested in dozens of client groups that included thousands of individuals, and to date they have not been disconfirmed (see Argyris, 1982, chap. 3). Most people hold espoused theories inconsistent with Model I; and, when confronted with our predictions about the strategies they will use, seek to dem-

Table 1. Model I Theory-in-Use.

Governing Variables	*Action Strategies*	*Consequences for the 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).	Actor seen as defensive, inconsistent, incongruent, competitive, controlling, fearful of being vulnerable, manipulative, withholding of feelings, overly concerned about self and others or underconcerned about others.	Self-sealing.	Decreased effectiveness.
Maximize winning and minimize losing.	*Own and control the task* (claim ownership of the task, be guardian of definition and execution of task).	Defensive interpersonal and group relationship (dependence upon actor, little additivity, little helping of others).	Single-loop learning.	

Minimize generating or expressing negative feelings.	*Unilterally protect yourself* (speak with inferred categories accompanied by little or no directly observable behavior, be blind to impact on others and to the incongruity between rhetoric and behavior, reduce incongruity by 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).	Little freedom of choice, internal commitment, or risk taking.	

Source: Argyris and Schön, 1974.

onstrate that our predictions are not valid. But even when Model I has been explained and people are trying to produce action that does not fit the model, they are unable to do so. This result holds whenever people are dealing with double-loop issues, which is to say whenever they are dealing with threatening issues. At best, they are able to produce strategies consistent with opposite Model I, the mirror image of Model I.

The governing variables of opposite Model I are (1) participation of everyone in defining purposes; (2) everyone wins, no one loses; (3) express feelings; and (4) suppress the cognitive intellective aspects of action. The associated behavioral strategies include emphasizing inquiry and minimizing unilateral control (Argyris, 1979).

Opposite Model I is more common as an espoused theory than as a theory-in-use. Elements from it are practiced in T-groups and nondirective therapies. Usually elements of opposite Model I are embedded in an underlying Model I theory-in-use in which unilateral protection of self and others is prominent, and competitiveness and unilateral control are present but camouflaged. It is often used in oscillation with Model I; the agent tries some elements from opposite Model I, and, if they don't seem to be working, switches to Model I. The consequences of opposite Model I for the behavioral world, for learning, and for effectiveness are similar to those of the more overt variety of Model I. In our discussion of the learning process in Part Three we will illustrate these ideas with concrete examples.

Action takes on the features represented by Model I especially in situations that agents perceive as potentially threatening or embarrassing. It is in such situations that agents are most oriented to controlling others and to protecting themselves. Self-protection frequently takes the form of attributing responsibility for error to others or to the situation rather than to oneself. The very situations that most require double-loop learning are the ones that most evoke Model I action—action that inhibits double-loop learning.

Model O-I: The Behavioral World. Individuals are embedded in a behavioral world or culture. This behavioral world has a dual nature. On the one hand, it is created by the actions

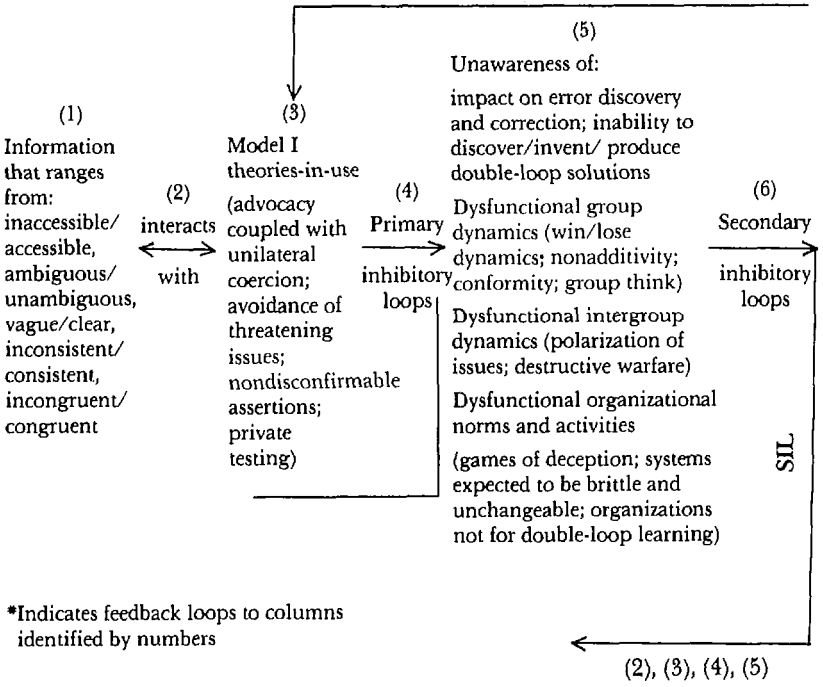
of the individuals who live it. On the other hand, it has an objective existence independent of the actions of any individual. Theories-in-use, in guiding all deliberate behavior, also guide the construction of the behavioral world. At the same time, the behavioral world guides the socialization of individuals with particular theories-in-use, and creates conditions in which theories-in-use are effective or ineffective.

The consequences of Model I theory-in-use, as we have described, include defensiveness, low freedom of choice, and self-sealing processes. If it is the case that Model I identifies features of theories-in-use that are common to virtually everyone, it follows that the behavioral worlds of groups, families, and organizations will have features that correspond to Model I. The interaction of people programmed with Model I theories-in-use generates pattern-building forces (Hayek, 1967, p. 33) that create a characteristic behavioral world.

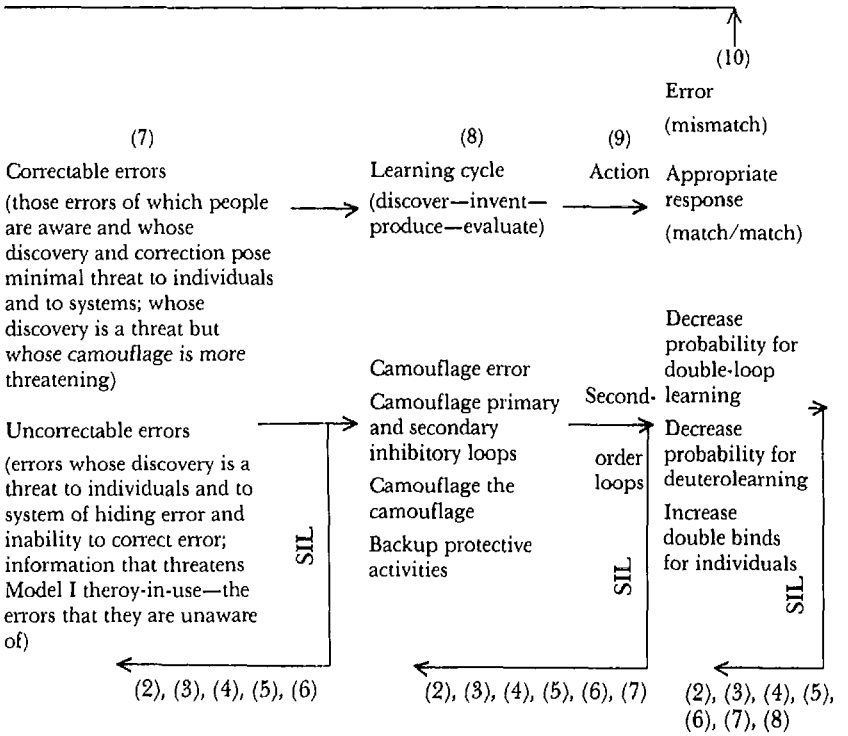
Argyris and Schön (1978) created a model of the behavioral world that is congruent with Model I theory-in-use. Model O-I ("O" signifies "organization") is a model of a *limited learning system* (Figure 2). The model states that when individuals programmed with Model I theory-in-use deal with difficult and threatening problems, they create primary inhibiting loops. That is, they create conditions of undiscussability, self-fulfilling prophecies, self-sealing processes, and escalating error, and they remain unaware of their responsibility for these conditions. Primary inhibiting loops lead to secondary inhibiting loops such as win-lose group dynamics, conformity, polarization between groups, and organizational games of deception. These secondary inhibiting loops reinforce primary inhibiting loops and together they lead people to despair of double-loop learning in organizations.

Under these conditions, organizations can correct errors that do not threaten underlying their norms, and they may also seek to correct errors that cannot be camouflaged. But they become unable to correct errors when this would require questioning and changing underlying norms. They also spin out elaborate webs of camouflage, as well as camouflage of the camouflage, and they engage in backup protective activities such as

Figure 2. Model O-I Limited Learning Systems.



Source: Argyris, 1982.


(10)
Error
(mismatch)
(7)
(8)
(9)
Correctable errors
(those errors of which people are aware and whose discovery and correction pose minimal threat to individuals and to systems; whose discovery is a threat but whose camouflage is more threatening)
Learning cycle
(discover—invent—produce—evaluate)
Action
Appropriate response
(match/match)
Decrease probability for double-loop learning
Camouflage error
Camouflage primary and secondary inhibitory loops
Camouflage the camouflage
Backup protective activities
Second-order loops
Decrease probability for deuterolearning
Increase double binds for individuals
Uncorrectable errors
(errors whose discovery is a threat to individuals and to system of hiding error and inability to correct error; information that threatens Model I theroy-in-use—the errors that they are unaware of)
SIL
SIL
SIL
(2), (3), (4), (5), (6)
(2), (3), (4), (5), (6), (7)
(2), (3), (4), (5), (6), (7), (8)

compiling special files "just in case the boss asks." All this creates double binds for committed individuals. On the one hand they see errors and unproductive activities that, as responsible members of the organization, they feel an obligation to correct; but on the other hand, to bring these threatening issues to the surface may be perceived as disloyal and as a threat to the organization.

Relation to Other Descriptive Theories. There is substantial agreement among scholars from several theoretical perspectives on the empirical nature of the phenomena that we explain with Models I and O-I. The descriptions produced by leading researchers in social psychology, sociology, and organizational behavior are consistent with our descriptions of the Model I world. Where we differ is in the generative or causal mechanisms that we identify and in the attention we give to possibilities for changing the world as it exists.

One of the most popular topics in contemporary social psychology is the reasoning processes of everyday life (Nisbett and Ross, 1980; Einhorn and Hogarth, 1981). Much of this research has focused on the seeming errors and biases of social cognition. Not only are people prone to inferential errors, but they also reason in ways that cause errors to persist. New data are interpreted in terms of previously formed opinions, to the point that disconfirming data are ignored or misinterpreted. People act in ways that elicit data that confirm their opinions, and express great confidence in reasoning that has little validity. They are predisposed to attribute the behavior of others (but not their own) to dispositional traits.

Such findings are consistent with what we would expect from individuals programmed with Model I theory-in-use. These individuals see their own views as obvious and do not test them publicly. They make attributions about others from scanty data, act in ways that elicit data that they then interpret as confirming their attributions, and remain unaware of their responsibility for generating these data.

One of the most influential theorists of everyday interaction has proposed that people manage the presentation of self so as to establish a definition of the situation that controls how

others will respond (Goffman, 1959). When events occur that disrupt the definition of the situation presented by a participant, the interaction breaks down. Participants often experience such breakdown as confusing, embarrassing, shameful, or infuriating. Hence, participants cooperate (collude?) to avoid disruption: "The person shows respect and politeness, making sure to extend to others any ceremonial treatment that might be their due. He employs discretion; he leaves unstated facts that might implicitly or explicitly contradict and embarrass the positive claims made by others. He employs circumlocutions and deceptions, phrasing his replies with careful ambiguity so that the others' face is preserved even if their welfare is not" (1967, pp. 16-17).

In our language, these are Model I strategies for the unilateral protection of self and others. The ways in which such strategies reinforce the self-sealing, error-reinforcing features of everyday reasoning should be obvious. Leaving unstated facts that might be threatening or embarrassing is a sure way to reduce the likelihood of double-loop learning.

In the field of organizational behavior, one prominent theme was first sounded by Cyert and March (1963), who pictured organizations as fields of conflict and bargaining among coalitions. Different groups have different preference orderings, and organizational goals are established through political maneuvering. The goals of an organization are frequently inconsistent, and the inconsistency is maintained by attending to the goals favored by different groups sequentially rather than simultaneously. Organizational goals are frequently nonoperational in the sense that they neither require nor preclude any particular behavior, because "nonoperational objectives are consistent with virtually any set of objectives" (p. 32). Hence by allowing organizational objectives to be inconsistent and nonoperational, it is possible to appear to satisfy more of the demands of different groups.

In our view, this is an accurate description of part of what takes place in a Model O-I world. There are inconsistencies in organizational theory-in-use, which are perceived in win-lose terms. Groups learn to protect themselves, to form coalitions

with other groups to enhance their positions, and to withhold or distort information that may increase their vulnerability. These are the strategies characteristic of a limited learning system. They are rational within the constraints of an O-I world, and at the same time they inhibit the double-loop learning that might permit a better overall solution.

Model II Theory-in-Use

The action scientist is an interventionist, seeking not only to describe the world but to change it. More precisely, he or she seeks to help members of client systems reflect on the world they create and learn to change it in ways more congruent with the values and theories they espouse. The normative perspective that guides the action scientist is found in Model II (Argyris and Schön, 1974). Model II as an espoused theory is not new; indeed most people readily espouse it. But as a theory-in-use it is rare. The action scientist intends to produce action consistent with Model II, because it is by so doing that the counterproductive features of Models I and O-I can be interrupted. Just as action consistent with Models I and O-I creates threats to validity and inhibits learning, action consistent with Models II and O-II is hypothesized to enhance validity and learning. Model II provides an image of the theory-in-use that the action scientist as interventionist seeks to help clients learn.

The governing variables of Model II (Table 2) include (1) valid information, (2) free and informed choice, and (3) internal commitment. These are the features of the alternative worlds that action science seeks to create (Argyris, 1980). Creating conditions in which these values are realized is the primary task of the interventionist (Argyris, 1970).

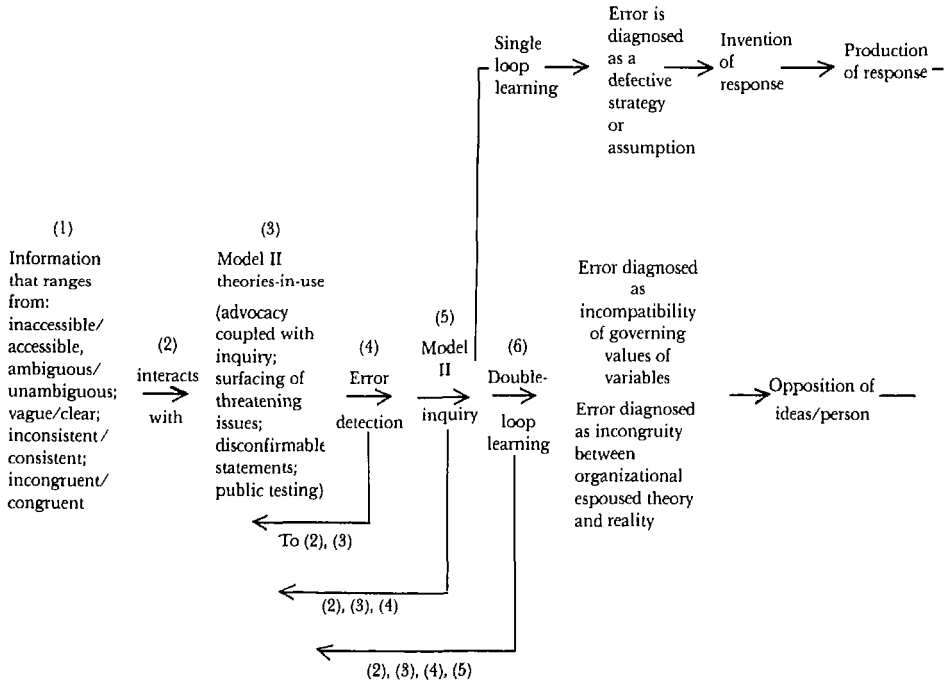
The behavioral strategies of Model II involve sharing control with those who have competence and who participate in designing or implementing the action. Rather than unilateral advocacy (Model I) or inquiry that conceals the agent's own views (opposite Model I), in Model II the agent combines advocacy and inquiry. Attributions and evaluations are illustrated with relatively directly observable data, and the surfacing of conflict-

Table 2. Model II Theory-in-Use.

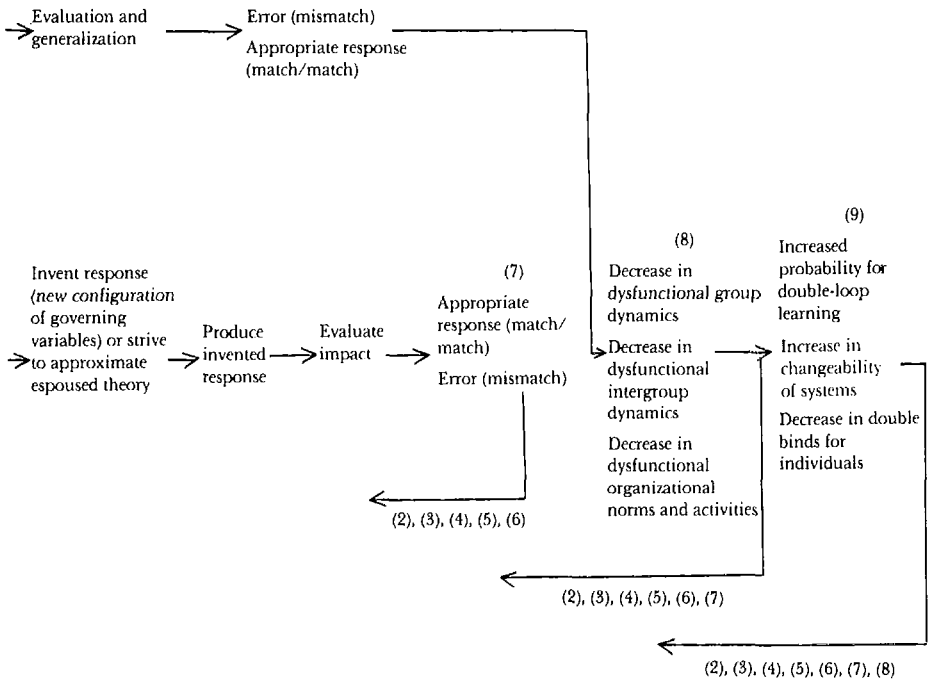
Governing Variables	*Action Strategies*	*Consequences for the Behavioral World*	*Consequences for Learning*	*Consequences for Quality of Life*	*Effectiveness*
Valid information.	*Design situations or environments where participants can be origins and can experience high personal causation* (psychological success, confirmation, essentiality).	*Actor experienced as minimally defensive* (facilitator, collaborator, choice creator).	Disconfirmable processes.	*Quality of life will be more positive than negative* (high authenticity and high freedom of choice).	
Free and informed choice.	*Tasks are controlled jointly.*	*Minimally defensive interpersonal relations and group dynamics.*	Double-loop learning.	*Effectiveness of problem solving and decision making will be great, especially for difficult problems.*	Increased long-run effectiveness.
Internal commitment to the choice and constant monitoring of its implementation.	*Protection of self is a joint enterprise and oriented toward growth* (speak in directly observable categories, seek to reduce blindness about own inconsistency and incongruity). *Bilateral protection of others.*	*Learning-oriented norms* (trust, individuality, open confrontation on difficult issues).	Public testing of theories.		

Source: Argyris and Schön, 1974.

Figure 3. Model O-II Learning Systems: Facilitating Error Detection and Correction.



Source: Argyris, 1982.


Evaluation and generalization
Error (mismatch)
Appropriate response (match/match)
(9)
(8)
Increased probability for double-loop learning
Decrease in dysfunctional group dynamics
Invent response (new configuration of governing variables) or strive to approximate espoused theory
Produce invented response
Evaluate impact
(7)
Appropriate response (match/match)
Error (mismatch)
Decrease in dysfunctional intergroup dynamics
Increase in changeability of systems
Decrease in double binds for individuals
Decrease in dysfunctional organizational norms and activities
(2), (3), (4), (5), (6)
(2), (3), (4), (5), (6), (7)
(2), (3), (4), (5), (6), (7), (8)

ing views is encouraged in order to facilitate public testing of them.

The consequences of Model II action strategies should include minimally defensive interpersonal and group relationships, high freedom of choice, and high risk taking. The likelihood of double-loop learning is enhanced, and effectiveness should increase over time.

Model O-II (Figure 3) describes the behavioral world created by individuals interacting on the basis of Model II theory-in-use. When members of organizations deal with difficult and threatening problems using Model II theory-in-use, they are engaging in Model II inquiry rather than creating primary inhibiting loops. Previously undiscussable issues will be brought to the surface, assumptions will be tested and corrected, and self-sealing processes will be interrupted. Both single-loop and double-loop learning can occur. Dysfunctional group and intergroup dynamics should decrease, and there should be less need for camouflage and games of deception.

At the espoused level, Models II and O-II sound like motherhood and apple pie. The trick is to produce them in the real world. This is quite difficult, both because people have been socialized to produce Model I and because the world continues to operate largely according to Model I, even when some people try to act according to Model II. Part Three of this book will be devoted to describing and analyzing the process of learning Model II at the level of theory-in-use.