SEEKING TRUTH AND ACTIONABLE KNOWLEDGE: HOW THE SCIENTIFIC METHOD INHIBITS BOTH

Chris Argyris

The stewardship of social scientists

Regardless of the particular theoretical bent or the bias each of us has for conducting research, all social scientists share, I believe, a common commitment to produce valid knowledge for understanding and explaining whatever universe of discourse we have chosen to study. The implementation off this commitment requires that we exercise three responsibilities. The first responsibility is to make as certain as we can that we are not unrealizingly kidding ourselves or others about the validity of the knowledge that we produce. We continuously strive to approximate truth by producing knowledge that is as free of distortion and error as we know how to produce. The second responsibility is to push back the frontiers of knowledge. We strive to expand knowledge and wherever possible, to make it additive. A third responsibility is to monitor continuously the ideas in good currency about the nature of sound theory and the methods to conduct empirical research.

I believe that much progress has been made and continues to be made regarding these three responsibilities. The focus of this paper is to ask, are there any ways in which the *effective* implementation of the above results in consequences that are counterproductive to the responsibilities just described. Are there inner contradictions in our practice?

I believe the answer is yes when we consider research whose objective is to understand some of the most fundamental processes used to control human beings in order to maintain social systems ranging from the family to organizations. Our research indicates that the theory of control used necessarily inhibits producing valid information if the subject matter is threatening; the very condition when valid information is especially needed. The connection with the conduct of empirical research is that social scientists often use a theory of control over subjects that is similar to the one used in the larger society. It too neces-

sarily limits learning. For reasons to be discussed below, social scientists have not explored the possibility that this theory of control will also inhibit the production of valid information, a consequence that is opposite to what we intend.

I begin with a quick summary of some findings about how the internal systems of organizations maintain themselves from embarrassment and threat. Next, I will present, again in summary form, a theoretical perspective to explain the findings. Then I will use these examples to illustrate how the correct use of rules for rigorous research can have the unintended consequences just described.

Organizational defensive routines

Organizational defensive routines are policies and actions that prevent individuals, parts, or the whole organization from experiencing threat or embarrassment, and simultaneously prevent them from identifying and reducing the causes of the potential embarrassment or threat.

Organizational defensive routines differ from psychological defensive routines along four dimensions. First, we have collected data on nearly five thousand people that show that they almost all behave consistently with their organization's defensive routines. Yet on the basis of personality research, we would expect a diversity of personalities in such a large group. Second, although individuals move in and out of organizations, the defensive routines do not change over time. Third, our research suggests that the source for creating these defensive routines are the social virtues taught to human beings early in life. Therefore, the organizational defensive routines are more likely a product of socialization. Finally because the actions used to create or to trigger organizational defensive routines are used by most people, their use cannot be attributed primarily to individual psychological anxiety.

Organizational defensive routines proliferate around dilemmas that contain important conflict. One of the most important dilemmas in most organizations, be they private or public, large of small, voluntary or nonvoluntary, is the age-old conflict of autonomy versus control. Subordinates wish to be left alone but held accountable. Superiors agree but do not want surprised. The subordinates push for autonomy asserting that letting them alone is the best sign that they are trusted by the top. It is fascinating to see that they push for a solution that combines trust with distancing. The superiors, on the other hand, push

for no surprises by using various kinds of information systems as controls. The subordinates see the control feature as confirming mistrust.

The point is not how to get rid of the dilemma. That will never occur; it is built into the concept of decentralization. The point is how to deal with it in a way that decentralization works. The most frequently used strategies that we have observed executives using to deal with this type of dilemma were mixed messages.

The top keeps communicating "We mean it: you are managing your show." The division heads concur that the message is credible except when the division or cooperate gets into trouble or when a very important issue is at stake. In the eyes of the divisional heads, corporate begins to interfere precisely when they want to prove their metal. In the eyes of corporate, they interfere precisely when they can be of most help, that is when the issue requires a corporate perspective.

Divisional heads described the mixed messages they received as:

"You are running the show, however ..."
"You make the decisions, but clear with ..."
"That's an interesting idea, but be careful ..."
"Be innovative, but don't get into trouble..."

The logic embedded in the mixed messages

Mixed messages contain meanings that are simultaneously:

ambiguous and clearly so
imprecise and precisely so

Anyone who deals with mixed messages experiences the dilemmas that are embedded in them. The designers know that designing a message to be clearly ambiguous requires skill and knowledge about the receiver. They know that to be vague and to be clear is inconsistent. Furthermore, to be clearly vague is not only inconsistent, but it is designed inconsistency. To design inconsistency makes the designer vulnerable unless the receiver does not question the inconsistency.

The logic and rules of mixed messages

There are therefore four rules about designing and implementing mixed messages. They are:

1) Design a message that is inconsistent.

2) Act as if the message is not inconsistent.

3) Make the inconsistency in the message and the act that there is no inconsistency undiscussible.

4) Make the undiscussibility of the undiscussible also undiscussible.

The logic produces paradoxes. To be effective, be inconsistent. To manage, act as if you are not managing. To make some issue discussable, make others undiscussible, and act as if this is not the case.

Not surprisingly, mixed messages escalate error and organizational defenses into defensive loops that act like organizational knots. Also not surprisingly, we find that organizational defensive routines become unmanageable and unchangeable. Most strategies used to reduce them activate and strengthen the defensive mechanisms. The two most frequent reactions that we receive when we insure about changing defensive routines are (1) cynicism and disbelief, and (2) fear that things may blow up, that we will be opening up Pandora's box.

Causes of organizational defensive routines

Thus, the strategies embedded in this logic are: when dealing with organizational defensive routines, be inconsistent, yet act as if you are not being inconsistent. Make the issues undiscussible and uninfluenceable, and act as if this is not the case. Thus the undiscussibility and unfluencability become undiscussible.

Why would human beings create universes with such monumental barriers to changing them? Very briefly, our research suggests that there are at least four major patterns of causes of phenomena such as mixed messages (Argytris and Schön, 1974, 1978; Argyris, 1980, 1982; Argyris, Putnam, and Smith, 1985). We are:

1) Human beings hold to kinds of theories of effective action. The first is the theory that they espouse. The second is the theory that they actually use when they act.

To date, it appears that most human beings' espoused theories vary widely. When dealing with issues that contain threat, their actions are often inconsistent with their espoused theories and they are unaware of the discrepancies. The discrepant actions and the unawareness can be explained by postulating a theory of action that they actually use. This theory-in-use has been called Model I (Figure I).

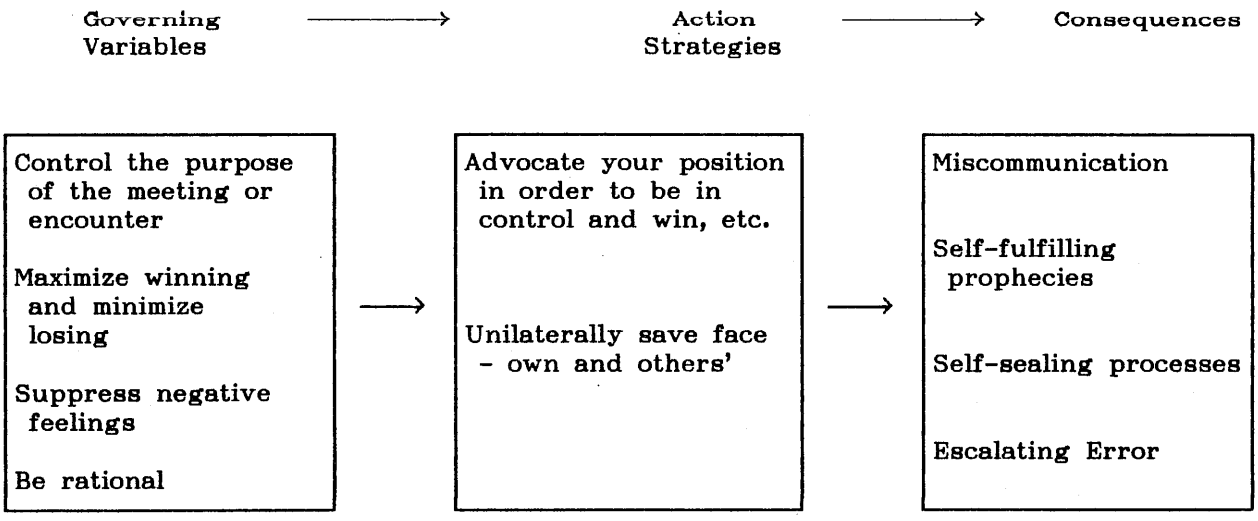

Figure 1. MODEL I THEORY IN USE

The opposite of Model I would be a theory-in-use with such governing variables as: (1) control should not be unilateral, (2) minimize losing for all, (3) suppress intellectual activities and express feelings, and (4) suppress rationality. The behavioral strategies would include nondirective activities and inquiry with little advocacy.

Most individuals observed to date use either Model I or the opposite of Model I. This tends to create defensive behavioral worlds where mistrust, self-fulfilling prophecies, self-sealing processes, and escalating error are predictable. Face-saving activities, such as mixed messages, become widely used, and their consequences become routinized.

2) Human beings programmed with Model I theory-in-use or the opposite create contexts within organizations that lead to limited search and limited learning. These O-I learning systems (Figure II) help to reinforce the anti-learning and overprotective consequences of Model I defensive routines.

3) Human beings create operating assumptions that they use as generalized rules as to how to act in a given situation where the potential for threat or embarrassment exist. These operating assumptions remain within the requirements of Models I and O-I. For example, there is an operating assumption that individuals should be caring and supportive. These are defined as saying to individuals those things they are able to or want to hear. The result is a lot of actions that "ease-in" such as mixed messages.

4) Being embedded in a universe of Model I theories-in-use, O-I limited learning systems, and operating assumptions that are consistent with the first two factors, leads human beings to use defensive reasoning when they design and implement their actions as well as respond to the reactions of others (which, in effect, require designs and implementation). By reasoning, I mean the thought processes individuals use to produce premises, make inferences, and reach conclusions. Defensive reasoning is characterized by premises that are tacit, inferences and conclusions that are tacit and framed in such ways that they are not publicly testable and disconfirmable.

To sum up the argument so far, individuals acting as agents for organizations or for themselves produce defensive routines that prevent or distort valid information, that are undiscussible, and whose undiscussibility is undiscussible. All of these features are protected by both the internal and external cultures of the organization. Individuals express hopelessness, fears, and anxieties about changing these features.

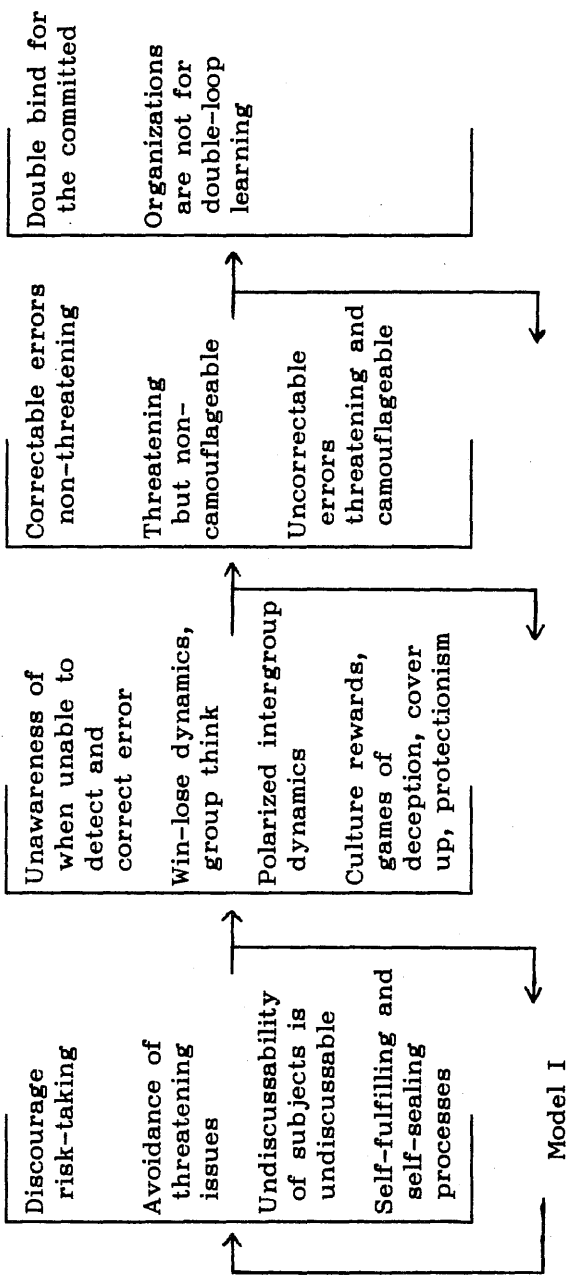

Figure 2. O-I LEARNING SYSTEM

Conducting research on organizational defensive routines

Let us now suppose that we wish to conduct empirical research on organizational defensive routines. No matter what our bias, there are at least three rules that we must follow if we are to act consistently with the ideas in good currency about conducting sound empirical research. They are (1) describe the universe as accurately as possible, (2) minimize threats to internal and external validity, and (3) remain as neutral as possible. How is that if we implement these rules correctly, we will limit our abilities to discover new knowledge, and we will increase the threats to validity?

Rule I: Describe the universa as is.

The first rule of normal science is to describe the universe as completely as possible and to produce explanations of what has been described. The modes of description and explanation may vary widely but not the requirement to produce descriptive knowledge.

My description of defensive routines above suggest that it is possible to describe them and to develop explanatory theories. The conclusion is only partially correct.

In order to provide a comprehensive description of organizational defensive routines, we must produce propositions about what happens when we try to change them. How do they respond to varying degrees of change? We must answer these questions in order to describe the universe as it is.

So far, we have found it almost impossible to study these questions by observing everyday life. Individuals do not strive to change them. This is not surprising if we recall (1) the societally taught logic that makes defensive routines undiscussible and their undiscussibility undiscussible, (2) the sense of hopelessness, and (3) the sense of fear and danger about changing them. The world is not likely to offer us examples to study, for in order to do so, someone must have decided to violate powerful cultural norms.

If changes in defensive routines are to be conducted, social scientists will have to help to create and implement them in actual settings. What would be required to conduct such research?

1) A model of a universe in which defensive routines would be rare. This would mean a normative model of a universe that may be espoused but does not exist.

2) A theory of intervention of how to get from the present universe to a new one.

3) A theory of instruction on how to teach the new skills required; how to create a new culture; how to create organizational contexts that would reinforce the new skills and cultural norms. Also a theory of instruction on how to unfreeze the old variables and how to deal with the bewilderment, frustration, and anger that individuals will experience as they realize that some of their most cherished skills are no longer valid.

But even if we met all these requirements, we would have to gain the cooperation of the subjects. Usually, we find that the first requirement is that they be assured that neither they nor their organization will be harmed. As they come to realize how much they will have to expose themselves, they usually up the ante of what they require of us. In addition to not being harmed, they also want to be helped. In effect, they change their role from subjects to clients. That request has nontrivial consequences for how we design the research.

If we examine most of the leading texts on conducting research, we will find, I believe, almost no attention paid to producing normative models of rare universes; to intervention theories and methods; and to theories of instruction of the kind described above. Moreover, little attention is paid to how to create a psychological contract with subjects as clients. Little attention is also given to the clinical skills required to help individuals deal with their feelings of embarrassment, frustration, and anger. At most, the texts speak of debriefing the subjects.

Rule II: Reduce threats to validity

Social scientists strive to reduce threats to validity by developing rules such as those formulated by Campbell and Stanley (1963). If we conceive of these rules as parts of a theory of action on how to conduct research, we may them ask what is the theory-in-use embedded in these rules as to how to manage subjects in ways to reduce threats to validity. Elsewhere I have suggested that there is embedded in these rules a theory of control that is highly similar to Model I (Argyris, 1980). For example, the researchers are advised how to reduce threats to validity by being in unilateral control of the research activities, by suppressing unwanted noise, by using face-saving devices whenever necessary in order that the subjects be kept appropriately in the dark about the experimental manipulation, the

hypothesis being tested, and the modes of analysis.

The result is that we are using research methods whose theory-in-use is consistent with the one used by most individuals to create defensive routines in the first place. If this is the case, then the research methods should also create, for the subjects, limited learning systems such as those described in Models I and 0-I. For example, the subjects are not supposed to inquire beyond that which the researchers design for them. The researchers design very little space for inquiry into the purpose, the experimental manipulation, who is in the experimental and control groups, etc.

What are the consequences of this for the study of how to reduce defensive routines? How would we create an experimental manipulation to reduce defensive routines without the subjects realizing that is our intention and hence knowing the nature of the experimental manipulation? If we tried to hide knowledge of the experimental manipulation, and if we acted as if we were not hiding that information, and if we made all of this undiscussible with the subjects (and all three ifs represent correct practice), then we are using defensive routines to conduct research on how to reduce them.

There is another problem. In order to gain the participation of individuals, the researchers would have to know how to produce a world where defensive routines are rare. They would also have to know the processes by which they could get from the Model I/0-I world to the new one *and* these processes would have to be consistent with the old world. If they cannot, then the clients have cause to be concerned about the credibility of the new world they try to create it for themselves or for others.

Rule III: Be neutral

The third problem created by combining the theory-in-use embedded in our research methods and the one used in everyday life is related to the criterion of neutrality. Social scientists are not supposed to take positions favoring one point of view over another. Social science research is supposed to encourage the systematic inquiry into any position.

If a Model I theory-in-use is embedded in research methods, then the generalization produced by such research will have Model I as part of the conditions under which they hold. If Model I conditions are embedded in the generalizations, and if other models are therefore precluded, then social scientists are not neutral. I conducted a review of research from social psy-

chology, sociology, and political science and found this to be the case (Argyris, 1980).

I will cite a few examples briefly:

* The studies on mass communication suggest that the presenter should present several alternative points of view if the group to be influenced is largely composed of people who are smart and one point of view if they are not (Aronson, 1965, pp. 47-88).

Assume this is true. Can you imagine the presenter giving the rationale to a "dumb" group or even to some smart groups? They would have to hide the rationale, act as if they are not hiding it, make the hiding undiscussible and its undiscussibility undiscussible.

* Let us take another example of the research on interpersonal touch and the foot-in-the door effect. Research has shown that the foot-in-the door procedure increases compliance for a desired request and that touch can also encourage compliance (Goldman, Kiyohare, and Pfannesteil, 1985). The authors cite Bem (1972) to explain the results. "An individual (into whose door someone has placed a foot) observing his or her own behavior while granting an initial, easy request decides that he or she is a cooperative individual who is helpful to others. This altered self-perception induces the individual to continue to be helpful and to comply with a second larger request" (p. 143).

Assume that the explanation is valid. An individual using the foot-in-the-door technique would have this explanation in mind when acting in this way. But, he would *also* have in mind keeping the explanation secret. If the individual ever told the recipient, "I am going to put my foot in the door, because (and then recited the explanation)", it might result in the door being slammed. The reason is that the actor has a theory about how to induce someone to do something whose empirical validity is reduced if it is said openly. The point is that the reason social psychologists found that these actions succeed is that they are consistent with a Model I world of unilateral control and unilateral censorship.

* Learning theorists speak of partial reinforcement schedules to motivate individuals. Consider the following scenario. Supervisor (A) has a partial reinforcement schedule designed to motivate subordinate (B). A gives the first reinforcement at the appropriate time. The reward surprises and pleases B. He tells some of his peers who assure his that he is "in" with the boss. This too is reinforcement but not part of the schedule. What if B remembered he did not thank A and returned to his office to do so? If he said thank you, what is A to say in return? If A said,

"you are welcome. It was a pleasure", that would be another reinforcement but not on the schedule.

Or, consider the scenario where A tells B that he has a partial reinforcement schedule to reward him. Does he tell the schedule ahead of time? What is the impact of having this knowledge on B? Would he not feel that he is part of a designed, mechanical process? How is this psychologically motivating?

In the first scenario, the partial reinforcement process follows the logic of defensive routines (making them undiscussible and their undiscussibility undiscussible). In the second, it makes the process public, but how discussable or influenceable is it? Can B alter it? If he cannot, will he not feel that his relationship with his superior is one-way?

None of this raises questions about the usefulness of learning theories. The point is that if they work, it is because the social scientists have embedded into their generalizations a Model I world.

* The lack of neutrality is also relevant for the inferences made by researchers from the data that they have collected. For example, Milgram explains the importance of unilateral obedience by saying it is required to maintain order in the universe (Argyris, Putnam, and Smith, 1985). This inference may well be valid for a Model I world. But, it places the researcher on the side of not exploring worlds where unilateral obedience is not necessary in order to maintain worlds.

* In the fields of organizational assessment, a central concept has been the fit between individual needs and organizational requirements. The assumption is that the better the fit, the more likely that individuals will perform better or have more positive attitudes, or minimize such behavior as absenteeism (Van de Ven and Joyce, 1981). There is a fundamental methodological assumption in such research. If individuals tell the truth about the fit as they experience it, the aggregate of such reports is a valid indicator of the fit.

The difficulty with this assumption is that we find many individuals hold beliefs about the fit that are incorrect, yet they are unaware of the gap. For example, most individuals do not see themselves as contributing to organizational defensive routines. Yet, when we observe them, they do. Hence their responses about the fit may be incorrect. If researchers are not aware of this systematic blindness on the part of the subjects, they may not only produce invalid descriptions of reality; they may create or reinforce injustices.

As an illustration, consider the study of several hundred young professionals. All had graduated at the upper levels of

their respective business schools. They worked for a large management consulting firm. They lamented the fact that their superiors were unable to create conditions for genuine participation, openness, risk taking, and trust. They had been taught to expect these capabilities from their superiors. Since the superiors were not acting in these ways, the subordinates concluded that their superiors were ineffective and should undergo training in leadership. An experiment was carried out where the subordinates conducted a study within their firm to illustrate their conclusions. They were then asked to feed back the results to the superiors. An analysis of the tape recording shows that the subordinates used the very behavior that they condemned while feeding back the results. For example, they described opinionated and unilateral actions on the part of superiors in ways that outside observers evaluated as opinionated and unilateral. The subordinates resisted this conclusion until they listened to the tape recording. They agreed and also reflected on their blindness (Argyris, 1982).

There is a second type of blindness that we have observed. Individuals are able to produce the actions they espouse but they are unaware that the actions are counterproductive to their own intentions. Again, the blindness appears to exist for most individuals. It is related to certain commonly accepted Model I norms.

For example, most of our respondents report that when dealing with organizational or individual defensive routines, they should act in ways that are supportive and show respect for others. Support is defined as communicating that which the others are able to hear without becoming defensive. Respect means not questioning the reasoning behind others' defensive actions.

The difficulty with these actions is that they can get people in trouble. For example, we have found that support gets translated into hiding negative judgements. People strive to communicate negative messages by being indirect and acting as if they are not. They may ease-in by asking questions. "How do you feel the session went with your subordinate?" "How much do you think he heard when you talked with him?" The hope is that the receiver will answer the question in a way that he will realize that he did not do well.

In our research, we found that if the individual does not cooperate by learning from the indirect mode, the "helpers" switch to showing integrity (which means telling it as they see it) and strength (which means not giving in). In short, when support and respect do not work, people switch to integrity and

strength. Usually those who have more positional power win out.

It is unlikely that we could discover these results unless we focused on the differences between espoused theory and theories-in-use. But helping people recognize the discrepancies is not enough in the sense that such awareness leaves them feeling helpless, bewildered, and frustrated. It is necessary for the researchers to help the clients overcome these problems. This requires a normative theory with different meanings of support, respect, integrity, and strength (Argyris, 1985).

Implications for researchers

More time and effort should be spent on learning how to produce normative models of rare universes that are empirically disconfirmable. In producing such models, researchers will find it necessary to make explicit the values that are embedded in their models and to provide a rationale for those values. All of us will find ourselves dealing with generating theories of morality.

In this connection, it is important to differentiate between normative theories and normative theories that become prescriptive for society. Researchers should focus on making their normative theories as comprehensive and as empirically valid as possible. Researchers should also study the processes by which individuals can use the theories in everyday life. The choice, however, as to whether or not the individuals will choose to use them is a citizen's choice.

Another implication is that the rules to reduce threats to validity will be refocused. For example, researchers would now make predictions about features of the theory-in-use and not about behavior. Researchers must study the actual behavior because the theory-in-use is inferrable from the behavior.

The refocus has certain interesting advantages. First, so far, most human beings use Model I and most limited organizational learning systems are consistent with Model 0-I. Although the specific actions used may vary widely, if we make our predictions about action strategies and their consequences, then our sample task is greatly simplified. So far we have not found any difference in theory-in-use related to age, gender, education, economic position, or race; nor in organizations that are small or large, young or old, profit or nonprofit.

Secondly, if we focus on theory-in-use, then we are making predictions about behavior that cannot easily be altered by individuals. This, in turn, makes it possible to become more open about experimental treatments without running the risk of con-

founding the responses. For example, people who wish to alter their Model I theory-in-use, who understand Model II, who accept Model II, are still unable to produce it. This is predictable. It is not possible for human beings to change their theory-in-use because they wish to do so. Behaving according to a new theory-in-use requires new skills and new values. It is doable, but it requires at least as much education and practice as is required to play a decent game of tennis.

Thirdly, if theories-in-use are not alterable by knowing a different one or by wanting to change, or by being rewarded to change, then issues such as learning or maturation cannot be threats to validity. It is not possible to produce that which they are incapable of producing (i.e., behavior inconsistent with their theory-in-use).

A fourth implication for reducing threats to validity is to provide the reader with samples of the observable data: that is the conversations people used. It is then possible for the readers to judge how the researchers made their inferences and drew their conclusions. It is also possible for them to challenge the validity of the research conclusion. The import of this rule can be illustrated by recalling that almost all paper and pencil instrument that use abstract categories (and almost all do) bypass the directly observable data. For example, it appears that leaders who are judged as "considerate" are ones who have been scored as being close to their subordinates and who leave them alone (Argyris, 1976).

The fifth implication is for social scientists to realize that it is unlikely that we can be neutral and use the technology of rigorous research. The technology for rigorous research is biased toward model I. Recall, for example, that the learning theorists and those producing the research on mass communication had embedded in their propositions conditions that were consistent with a Model I world. Scientists have a right to produce such generalizations. They also have an obligation to make the tacit choices explicit.

The sixth implication is that researchers should pay more attention to producing scientific generalizations that are reproducible under real time conditions by individuals who wish to do so. For example, let us assume that there is a curvilinear relationship between variable A and B. Let us also assume that Mr. X wants to have his action influenced by such a proposition. The problem arises as to how he would ascertain, in an on-line manner, where variables A and B are along a continuum. In order for human beings to use propositions, they must be producible under everyday life conditions. Otherwise, life will

pass them by while they are trying to diagnose the situation.

Perhaps more important is the likelihood that practitioners could find themselves using our generalizations correctly yet producing conditions that are counter to the ones the generalization states will occur. For example, assume that Stodgdill's handbook on leadership contains valid and usable information. The first thing we can say about it is that leaders are not likely to store all the information in their heads and retrieve it when they need it. If leaders use the handbook as an external memory, then the leaders face a new dilemma. Let us assume that there is a chapter on intergroup relationships. What is a leader supposed to do when he suddenly realizes that he is faced with an intergroup situation? Does he turn to the chapter to refresh his memory? What does he tell his group to do while he is exploring the chapter?

Harvard University

BIBLIOGRAPHY

Argyris, Chris, "Problems and New Directions for Industrial Psychology", in Marvin D. Dunnette (Ed.), *Handbook of Industrial and Organizational Psychology*, Rand-McNally, 1976.

Argyris, Chris, *Inner Contradictions of Rigorous Research*, New York: Academic Press, 1980.

Argyris, Chris, *Reasoning, Learning and Action*, San Francisco: Jossey-Nass, 1982.

Argyris, Chris, *Strategy, Change, and Defensive Routines*, Boston: Pitman Publishing, 1985.

Argyris, Chris, and Donald Schön, *Theory in Practice*, San Francisco: Jossey-Bass, 1974.

Argyris, Chris, and Donald Schön, *Organizational Learning*, Reading, Mass.: Addison-Wesley, 1978.

Argyris, Chris; Robert Putnam, and Diana M. Smith, *Action science*, San Francisco: Jossey-Bass, 1985.

Aronson, Elliot, *The Social Animal,* San Francisco: W.H. Freeman and Co., 1965.

Bem, D., "Self-perception Theory", in L. Berkowitz (Ed.), *Advances in Experimental Psychology*, Volume 6, New York: academic Press, 1972.

Campbell, D.T. and J.C. Stanley, *Experimental and Quasi-experimental Design for Research,* Skokie, Ill.: Rand-McNally, 1963.

Goldman, Morton; Odette Kiyohare; and Dorothy H. Pfannensteil, "Interpersonal touch, social labeling, and the foot-in-the door effect, *Journal of Social Psychology*, 125 (2), 1985, 147.

Van de Ven, A. and Joyce W., (Eds.), *Perspectives on organizational Design and Behavior*, New York: Wiley, 1981.