We present the results of our review of some forty community-level interventions undertaken in the developing world over the past twenty years in order to reduce malnourishment in children. We argue that such interventions, if they are considered as social experiments, cannot be assimilated to models of quasi-experimental method. We propose an alternative model of experimentation, which we call "reflection-in-action", which seems to us better suited to account for the kinds of validity and rigor attainable in situations such as these.

Social Experimentation as Reflection-in-Action

Community-Level Nutrition Intervention Revisited

DONALD A. SCHÖN
Massachusetts Institute of Technology

WILLIAM D. DRAKE
University of Michigan

ROY I. MILLER
Community Systems Foundation

Social experimentation is central to the idea of liberal social reform. Reform suggests progress, which suggests societal learning, and our main image of societal learning is that of experimentation. Patrick Moynihan saluted the American Bicentennial, for example, by asking, "What have we learned? It is two centuries now since the American people commenced what they very well understood to be an experiment in liberty" (Moynihan, 1976). To ask what we have learned is to assume a "we" capable of learning, a "something" capable of being learned, and an image of "learning" attributable to a whole society. To see the American people as having "commenced an experiment in liberty" is to see them as having deliberately embarked on a course of *exploratory action*, a *social* experiment involving human beings whose thoughts and feelings might affect the outcomes of action, and a collective *intervention* aimed at changing things for the better. In this threefold sense, an American experiment in liberty is an intervention experiment.

Knowledge: Creation, Diffusion, Utilization, Vol. 6 No. 1, September 1984 5-36

For many years now, the idea of intervention experiment has been an idea in good currency. Republican as well as Democratic administrations have couched their proposals for reform in the language of experimental social problem-solving. Yet whenever governmental actions have been taken seriously as experiments, controversies have arisen over the interpretation of their results. What changes actually occurred? Were they really attributable to governmental intervention? What lessons ought to have been drawn? Among social scientists, such controversies have led to the view that governmental actions conceived as experiments should be rigorously designed as such. Yet on the rare occasions when this has been attempted, the interpretation of experimental results has been no less controversial. Indeed, the feasibility of intervention experiments has become a subject of controversy in its own right.

Those who advocate rigorously designed social experiments rely on models of inquiry derived from laboratory experiments in the social sciences. Their view of experimental rigor emphasizes quantitative methods, randomized samples, and experimental controls. They assume that social reality can be decomposed into measurable variables whose causal connections are lawful and predictable. They hold that social researchers must free themselves from their biases in order to make their experimental procedures and results replicable. Through iterative experimentation, they believe, it is possible to discriminate among competing theories of social reality so as to arrive at cumulative, consensual, and convergent knowledge. Of course, they recognize that the experimenter never completely controls his environment and they sometimes respond to administrative and political constraints by seeking out limited domains of intervention suitable for rigorous experimentation.

The critics of this view of social experiment note that policy conflicts often reemerge in debates over the interpretation of experimental findings. They observe that policy issues have a way of slipping out from under research; by the time the results of an experiment are available, policymakers are no longer interested in them. And some of these critics eschew rigorous controls and quantitative precision in favor of qualitative, narrative methods of inquiry appropriate to description of the intervening social processes that link inputs to outputs (see Rein and Weiss, 1970). They advocate action research in which practitioners double as experimenters, short-circuiting the tenuous process by which social experimentation is supposed to influence policy.[2] But action researchers tend to leave hard questions dangling. What model of knowledge underlies their method of inquiry? Do they aim at general, policy-relevent conclusions? If so, how do they deal with the confounding "uncontrollables" that plague social experiments conceived on the model of the natural sciences?

The idea of intervention experiment poses a dilemma of rigor or relevance. We must choose, apparently, between rigorous social experimentation that cannot be applied to matters of real-world importance and methods of experimentation that are applicable but hopelessly nonrigorous.

More than twenty years ago, Donald Campbell not only described this dilemma but proposed a solution to it (Campbell and Stanley, 1963: 20). He imagined a spectrum at one pole of which he placed the "true experiment" of the natural sciences and at the other, the "dirty" methods of action research. Between these extremes, he proposed a middle ground of "quasi-experimental method"—a patched-up approximation of true experiment which, even if it fell short of the ideal, might still produce generalizable causal inferences useful in the formation of public policy. For this heresy, he was roundly taken to task by social scientists committed to the natural science model.

Over the last twenty years, a great deal has happened to justify a reconsideration of Campbell's quasi-experimental model. Several large-scale interventions designed as rigorous social experiments have led to ambiguous results, refueling the controversy over social experimentation. And in the same period, positivist assumptions have fallen into disrepute.[3] There has been a growing interest in the continental exploration of hermeneutics, phenomenology, and critical method.

In this article we shall revisit the vexed questions of intervention experiment, quasi-experimental method, and action research, focusing on the domain of interventions aimed at reducing malnourishment in the developing world, a domain torn by controversies of its own. In the years following World War II, the United States undertook massive programs of food distribution and supplemental feeding—apparently unexceptionable interventions that were subsequently criticized as causes of dependency and black-marketeering. As the problem of malnourishment became increasingly visible and urgent, it triggered a "Rashomon" of conflicting diagnoses. Depending on the profession or ideology of the observer, malnourishment has been seen as a problem of diet, agricultural productivity, water quality, health care, population control, land ownership, economic policy, or social justice. Attempts to synthesize these several diagnoses have been inconclusive. In the design of nutrition interventions, some physicians and engineers have tried to conform to canons of experimental rigor, while other workers have seen themselves as compassionate fighters in a war against hunger. Even for the former, the rural villages and urban squatter settlements of developing countries, unstable and relatively uncontrollable, have presented massive impediments to rigorous experimentation.

Over the past four years, we have carried out a study of community-level nutrition interventions throughout the world (reported in Drake et al., 1980). Intending initially to learn what kinds of intervention worked, and under what conditions, we found it necessary to reflect on our underlying models of efficacy and rigor in social experimentation.

True Experiment and Quasi-Experimental Method

In Campbell's well-known analysis, intervention experiment consists of manipulating certain variables in order to observe the effects of manipulation on other variables. In an experiment on the effects of alternative teaching methods on children's reading scores, for example, the experimenter tests an intervention hypothesis of the form "X produces D," where "X" is an experimental treatment (a particular use of phonics, for example) and "D" is an intended difference between pre- and posttest observations of a dependent variable (reading scores, for example). An inference from experimental findings has "internal validity" if and only if, in the specific experimental instance, results to the experiment are shown to be incompatible with all other plausible accounts of the observed change in dependent variables—that is, when the intervention hypothesis has been shown to be more resistant to refutation than its competitors. Threats to internal validity include confounding events that occur between pre- and posttests, maturation of persons or systems observed, and selection biases (Campbell, 1969). The method of "true experiment" is intended to counter such threats by the use of control groups and randomization. True experiments must conform to the following principles:

(1) Predesign: Hypotheses and conditions of experiments must be specified prior to intervention so as to make it possible to identify and randomize the selection of control and experimental groups.
(2) Isolation: Both experimental and control groups must be isolated from all changes in the environment except those already included in experimental design.

(3) Constancy: Experimental and control groups must be kept constant throughout the experiment, except insofar as they are intended to change.
(4) Quantitative precision: In order to permit discrimination among rival hypotheses, both hypotheses and results of the experiment must be expressed in quantitative terms.
(5) Distance: Both the subjects of experimental treatment and the practitioners who deliver it should be kept unaware of experimental design and distant from analysis of its results, lest they depart from design or distort analysis.

Quasi-experimental method is intended for use in the imperfectly controlled world of social interventions where true experiment is impossible. Its strategy is "to generate . . . as many *plausible* rival hypotheses as possible and then to do the supplementary research that would discriminate among those hypotheses (Campbell and Stanley, 1963: 5). The experimenter makes time-series observations where the effects of experimental treatment may show up as a "discontinuity in the measurements," employs nonrandomized control groups, and "patches" experimental designs by adding features to "control specific factors, more or less one at a time" (1963: 53).

Both "true" and quasi-experiments are intended to provide a basis for generalized policy advice. The experimenter seeks not only internal validity but "external validity"—that is, valid generalizability to settings other than the experimental one. The experimenter operates under a *schema of generalization* (our term), seeking to establish the truth of an intervention hypothesis for all contexts similar in the relevant aspects to the experimental one.

Campbell's treatment of experimental validity rests on certain assumptions about the kind of knowledge we can get about social reality and about the features of social reality that enable us to know it. He assumes, first of all, that social phenomena are lawful, that "the closer two events are in time, space, and measured value on any or all dimensions, the more they tend to follow the same laws" (Campbell and Stanley, 1963: 18). Although in any particular instance we may guess wrong, the laws governing social phenomena are assumed to be *in* the phenomena, "there" to be discovered.

On this basis, Campbell also believes in the possibility of *objective* knowledge of social phenomena—knowlege where truth is independent of the values or biases of the researcher. The researcher does not construct the laws he or she discovers, nor are they linked in any special way to the experimenter. Indeed, any such linkage appears in his or her view as a distortion of experimental method. If the experimenter is to remain an external observer of the consequences of his or her actions, he or she must be, in Geoffrey Vicker's phrase, a spectator-manipulator.

Just as laws are taken to be in the situation—waiting as it were, to be discovered—so are the things and relations to which laws refer, for example, "teachers," "students," "teaching methods," and "reading scores." Campbell's view of social experimentation rests on the assumption that the things and relations of an experimental situation are given. The problem is not to discover the variables but to find out how they are causally connected to one another. The task of experimental design begins when the description of the situation and the problem of intervention are already given. True and quasi-experimental methods aim at objective knowledge of the lawful patterns of variables taken to be inherent in social phenomena.

The Study of Community-Level Nutrition Interventions

Over the past four years, we have studied a large number of community-level projects aimed at reducing malnourishment, mainly among children between the ages of zero and six, in a variety of developing countries. In order to explore what could be learned from these projects, we considered them as intervention experiments.

Using conventional measures of malnourishment, we first tried to determine whether in the course of each project a change in the nutritional status of the target population had occurred, and if so, whether it could be validly attributed to project intervention. We posed plausible alternate explanations of the change and devised patching-up experiments to discriminate between the intervention-hypothesis and its rivals. In short, we tried to apply to our sample a method of experimental inquiry very much along the lines of Campbell's quasi-experiment. What we learned has a direct bearing on the applicability of Campbell's particular formulation and on the more general questions of efficacy, validity, and rigor.

We selected our sample by scanning the published literature, project reports, and lists of projects funded by agencies concerned with the problem of malnourishment in the developing world. Initially, we searched for projects at the village- or regional-level, which included as an objective the reduction in malnourishment among children below primary school age. We found about 25 projects of this type for which there existed documentation sufficient for analysis of some kind; and of these, only 8 met minimal criteria for evaluation as intervention experiments. The availability of time-series measurements of nutritional status was the most decisive of our selection criteria.

This is the sample of projects we finally selected for analysis:

(1) Candelaria—A program featuring home visits by volunteer "barefoot doctors" in the town of Candelaria, Columbia. Services include education on nutrition, hygiene, and the utilization of health services; weighing of children; and referrals to the medical establishment.
(2) Candelaria Revisited—A resurvey of the village of Candelaria two years after termination of the program.
(3) Primops—An extension of the Candelaria concept to an urban setting in Cali, Columbia.
(4) Esperanca—A program based on the establishment of health posts in rural villages in the Central Amazon region of Brazil. Services included the provision of basic health care by a visiting doctor and/or a locally trained "barefoot doctor", nutrition and health education, and child weighing.
(5) SCR-Honduras—A program following the Save the Children community development scheme, CBIRD, carried out in the Pespire region of Honduras. Food supplements were administered in parallel with more general development efforts.
(6) SCF-Indonesia—Another Save the Children program, this time in the special province of Aceh in northern Sumatra.
(7) Thailand—An experiment to test the effect of rice fortification on nutrition and health in the Chaing Mai region of Thailand.
(8) Kottar—A community development oriented program run by the Kottar Social Services Society in Tamil Nadu, India. Services include food supplementation, nutrition, and health education, immunizations, and more general community development assistance.

Later on in the project, we considered a few other similar projects reported in the published literature. For our main sample of eight projects, however, we relied on primary data in the form of protocols or IBM punch cards and we carried out independent field evaluations.

We should note that, although *we* tried to analyze all of these projects as experiments, the project leaders gave different weightings to the importance of experimental design and control. In only one case (Thailand) did the project leader conceive of his activity primarily as a hypothesis-testing experiment. In all other cases, although there was a strong interest in learning from project experience, the primary interest of the project staff was to improve children's nutritional status.

The main results of our analysis:

(1) In only two of the eight projects could we establish with reliability that the target population had experienced a positive change in nutritional status. In all other cases, our efforts to infer such a change were blocked by confusing, unreliable, or meaningless data, by inadequate measures or measurement methods, or by confounding changes in the composition of the target group.

(2) In no case could we make an unambiguous, internally valid attribution of change in nutritional status to the project's intervention. On the basis of available data, our supplementary research was insufficient to rule out some plausible alternate explanation of the change.

(3) The extent to which project staff had tried to achieve the conditions of rigorous experimental design seemed to make no difference to the interpretability of their results. The more rigorously designed projects produced data as indeterminate as the data produced by the most "informal" ones.

Thus, as Campbell might have predicted, we found it impossible to treat community-level interventions as "true experiments." We found, to a greater extent than his writings suggest (but perhaps not surprisingly, in view of the settings of our projects), that "dirty data" and inadequate measures were a major obstacle to the interpretation of experimental results. Also, we found in the more rigorous as in the less rigorous designs that the patching-up methods of quasi-experiments were in some cases able to rule out some plausible rival hypotheses but in no case sufficient to permit an unambiguous interpretation of the data.

These findings raise a critically important question about the extent to which the experiments produce indeterminate results because the experimenters failed to conform to the canons of quasi-experimental rigor. Would it have been possible to learn more from the experiments had the researchers been less sloppy, or was the indeterminacy of experimental results inherent in the intervention situations themselves?

Clearly, some obstacles to the interpretation of experimental results could have been eliminated by tighter research practices. Some inadequate measures could have been improved, some dirty data could have been cleaned (although even here, as we will see, improving the quality of data requires a very different relationship between researcher and practitioner than the one advocated by Campbell). But these improvements in research practices would not have eliminated certain fundamental obstacles to interpretation. Indeed, more "rigorous" experimentation would have introduced new sources of indeterminacy.

Thus, the thrust of our argument is not to urge that community-level nutrition interventions conform more rigorously to the model of quasi-experimental models. Rather, we shall question the model's appropriateness and propose an alternative to it.

Let us consider, to begin with, the phenomena of "dirty data" and "inadequate measures."

Dirty data. All of our projects were plagued by faulty practices in the collection, storage, and processing of data. As a result, project findings were inconsistent, confusing, or uninterpretable. In the most egregious case, a tropical bug ate machine-readable holes in the computer cards used for data storage! While this was a rare event, there were other less exotic problems in the design of formats for data collection, the reliability and consistency of reports, the use of numerical scales and standards, and the interpretation of subject's reports.

Often, we found that it was possible to detect and correct errors in a data set long after the data had been collected. By testing the range assumed by relevant variables and examining their internal consistency within and between observation periods in a longitudinal sequence, we were able to eliminate many mechanical errors without loss of valuable information.

In the Primops projects, for example, the time series observations on each individual were linked by a single identification number. The age of each child at each observation was recorded, as was the date of the observation. By comparing the change in age between successive observations with the calendar time between those observations, we were able to identify cases where two observations for a single identification number could not possibly describe a single individual. Some of these errors could be corrected (long after the fact) in Michigan by careful scrutiny of the data. For example, when more than two observations appeared for a single child, the age for the one erroneous entry would be set to conform to the sequence created by the other, apparently correct, entries. However, without returning to Cali, Columbia, and visiting the children, many of these errors could not be corrected.

In general, we found, the greater the distance between the collectors and the users of data, and the longer the time interval between collection and use, the dirtier the data and the more difficult it was to detect and correct errors. Indeed, when practitioners and participants felt disconnected from the researchers and their purpose (as required by the principle of "experimental distance"), they tend to give short shrift to the task of data collection. When the *promotoras* could see no immediate, practical utility for the data they were asked to collect, for example, they tended to regard data collection as a meaningless diversion from the task of health service or food distribution, whose utility they *could* understand.

Inadequate measures. Normal methods for assessing the nutritional status of chidren use anthropometric measurements of height and weight. Typically, the height or weight of a preschool child is compared to a standard height or weight derived from observations of healthy children of the same age as that child. Alternatively, the weight of the child may be compared to the weight of healthy children of the same height as that child. A deficiency in either of these scores is taken as an index of malnourishment.

In the field application of these apparently simple methods, we found six sources of error that interfered significantly with our efforts to estimate change in the nutritional status of a population of children over time.

(1) Where measures of health do not coincide with measures of size, a child may be misclassified as malnourished. It has been shown, for example, that no child who suffered from chronic malnourishment, especially during the age periods of human growth, is incapable of the catch-up growth necessary to regain normalcy as defined by a growth standard derived from healthy children. Some researchers have claimed, in addition, that improved nutritional intake may contribute to increased metabolism or level of activity rather than to growth.

(2) Whether a child is classified as malnourished may vary with the selection of anthropometric standard. International standards derived from healthy children in developed countries often prescribe more rapid growth than the locally generated standards of a developing country. In our Kottar analysis, for example, we showed, using a local standard, that the percentage of children suffering from second or third degree malnourishment dropped over a two and one-half year period from 50% to 42.5%. When we switched to an international, sex-differentiated standard, we showed a smaller decline, from 50.4% to 45.5%.

(3) Similarly, the picture of change in nutritional status can vary with the choice of a cut-off-point for defining malnourishment. Because the nutritional scores of a population tend to cluster around traditionally accepted cutoff-points, small shifts in those points cause rather large numbers of children to cross the boundaries between categories of malnourishment.

(4) There is a debate over the definition of "improved nutritional status in a community." Some analysts use mean percentage of stan-

dard, while others emphasize the need to show the greatest improvement among those initially worst off. Often, the selection of a statistical method is tantamount to a definition of "improvement" and, in these cases, the choice of statistical method governs the description of change in nutritional status.

(5) Depending on the selection of the variables used to monitor change in nutritional status, estimates of change may vary. Weight for age, height for age, and weight for height—the most common anthropometric ratios—measure different aspects of malnourishment. The last measures acute malnourishment; the second, chronic undernourishment; and the first, a composite of the two. Because height is the numerator of one score and the denominator of the other, it is possible and not surprising that instances exist whereas height-for-age scores increase, weight-for-age scores decrease, and vice-versa.

(6) Finally, estimates of change in nutritional status are distorted by the usual method of accounting for the relationship between malnourishment and life cycle. The typical child in the developing world experiences a gradual deterioration of nutritional status from birth to some age of "maximum" risk (usually between 18 and 24 months) and then improves. Because participating children grow older in the course of an experiment, analysis must take account of the typical pattern of growth. We have recommended mapping changes in nutritional status by the use of a "characteristic curve"—a graph of malnourishment against age in a population at a single point in time (Drake et al., 1980: 97). However, even with a characteristic curve, the picture of change can be altered by redefining age categories or selection of different cutoff-points or standards.

The use of anthropometric measures of malnourishment is likely to remain troublesome. The development of other methods (blood serum tests, for example) may eventually supplant anthropometrics, although, at present, most nutrition intervention projects cannot bear their additional costs. If it were possible, through field research, to learn more about the response of each type of anthropometric score to "prove" interventions, it might also be possible to set precedents for the use of these scores to assess changes in nutritional status in the field. For reasons that we shall shortly explain, we are unlikely to be able to establish "proven interventions" of this kind. Nevertheless, it is clear that *some* obstacles to the valid detection of change in nutritional status may be eliminated through the use of better measures and measurement methods, just as some (but not all) dirty data may be eliminated through the postexperimental application of better methods of data cleaning.

Even when a positive change in nutritional status has been detected, however, there remains the task of explaining it. Usually, such changes are ambiguous, in the sense they they are plausibly attributable to factors other then the experimental intervention. Conversely, an actual effect of intervention may be masked by other changes in the environment. In order to tell, therefore, whether indeterminate experiments might be clarified by a more rigorous application of quasi-experimental methods, we must explore the sources of ambiguity.

Ambiguous data. We found several sources of ambiguous data: the changing make-up of the population under study, concurrent changes in the project's environment, and the participants' or practitioners' awareness of the experimental process.

In every one of our data sets, there was evidence of change in the composition of experimental or control groups. In Primops, for example, there was a marked tendency for children to disappear from the data set from one round of data collection to the next. Very likely, their "disappearance" was due to some combination of in- or outmigration, intrabarrio movement, death, defection from the sample, or recording error. What was striking, however, was that in each round more of the malnourished disappeared than did the wellnourished. The appearance of improvement in nutritional status from round to round may well have been due to this selective disappearance. Such changes in composition of the sample are unavoidable so long as it is impossible to isolate the experimental group from ordinary patterns of population movement.

We encountered several kinds of experiment-confounding changes in the project environment. Changes in climate were important. In Esperanca, for example, time-series observations of matched pairs of experimental and control villages showed a greater change in nutritional status associated with experimental treatment. But because parts of the baseline survey were administered in different seasons at six-month intervals while the resurvey was done in a single season, the observed difference may have reflected ordinary seasonal shifts in nutritional status. In Kottar, we could not distinguish the effects of intervention from the impact of the easing of a highly localized drought.

Changes in economic climate were also important. Inflation or depression (or recovery from them) may affect nutritional status, perhaps in combination with changes in local economy. In Thailand, the rice fortification experiment was seriously hindered for a time when the rising black market for rice, across the Laotian border, led to illicit export of the product intended for local consumption.

Changes in the infrastructure, or in the availability of other services, may overlap the effects of intervention. It was not possible, for example, to distinguish the effects of teaching the *promotoras* in Candelaria

Revisited from earlier improvements in water and sewage disposal systems. Changes in social norms and structures may also affect nutritional status. In Kottar, project leaders saw their nutritional interventions as incentives for community organization. There, we could not validly attribute change in nutritional status to supplementary feeding or nutritional education alone, because participating families were also learning at the same time to make better use of other resources, including their own labor.

Changes in nutritional status may result from the practitioners' awareness of the experimental process. In Honduras, for example, nine months into the feeding component of the intervention, a medical consultant who had been monitoring the data urged that food be targeted to the *moderately* malnourished. In his opinion, the most seriously malnourished, who suffered from diseases beyond food deficiency, would derive less benefit from the food. But the net result of his policy was a spurious improvement in aggregate scores of nutritional status.

We were able to reduce some of the ambiguity of our experimental results by quasi-experimental patching up. In our analysis of the Kottar data, for example, we were able to rule out the hypothesis that change in nutritional status reflected a bias in the selection of new participants over time. Our supplementary research had only a limited impact on ambiguity, however, because it occurred at the end of the experiment when it was no longer possible to alter the data-gathering process. Neither in Kottar nor in any other project could we rule out all plausible rival hypotheses.

A deeper consideration of our sample suggests, moreover, that *even if patching experiments had been initiated in the midst of the projects, some indeterminacy of experimental results would still remain.* The situations of community-level nutrition intervention are in certain crucial respects inconsistent with Campbell's fundamental assumptions about the lawfulness and the objective knowability of social reality. These situations are inherently vulnerable to unpredictable instability and, therefore, to real uncertainty. They are often unique; general principles derived from them cannot be validly applied to other cases. And because some of them are understandable only in terms of the experimenters' transactions with them, they are better regarded as holistic transformation than as experimental manipulations of discrete variables.

We will consider each of these features in turn.

Uncertainty. Campbell calls attention to the fact that the contexts of experiments may be unstable, changing in ways that confound experimental designs.[4] In our sample of intervention experiments, we found many examples of instability. Over the long time periods of intervention, contexts tend to change out from under the experimental design. Often, moreover, the changing variables that confound experiments are ones the experimenters have not thought to name or include in their design. The context of experiment is *unpredictably* unstable. We have already mentioned the effects of the black market for rice on the carefully designed rice fortification experiment in Thailand. In Indonesia, to take another example, the staff of the Save the Children Foundation reported a major flood, an outbreak of cholera, and military uprising, all of which disrupted the flow of services to the target population. Such an unpredicted, unnamed instability of context not only confounds the experimenter's design but causes him, at least for a time, to be unable to make *sense* of the situation. It is this phenomenon that we will call "uncertainty."

When a situation become uncertain in this sense, its initial *description* becomes problematic. It is not only that the experiment is confounded by an unexpected change in the environment (although this is true) but that the experimenter no longer feels confident that he or she knows the names of the variables to be manipulated or observed. The experimenter is confronted not with a "problem" (for which the suggested intervention hypothesis may be a solution) but with a problematic situation. He or she cannot revise the experiment until the situation has been redescribed and the problem under investigation has been reframed. From the point of view of the validity of experimental inference, what this means is that an intervention hypothesis that seemed internally valid at one time may become internally invalid at a subsequent time, not because an important variable has been neglected, but because the whole situation has changed.

The situations of community-level intervention are inherently vulnerable to an unpredictable instability (and the experimenter is vulnerable to real uncertainty) because they have to do with human beings in interaction with one another. The patterns of the situation, and of the larger environment in which it is embedded, are dependent on what Philip Herbst has called "behavioral worlds." If social contexts manifest patterns of relatively long-term stability, Herbst points out, it is because indiviudals create and maintain their stability.

> A necessary condition for stable behavioral relationships to manifest themselves is that parametric steady-state conditions are firmly established and maintained. Ideal conditions of this type can be found under everyday conditions where persons have built up a stable cognitive and

> behavioral structure. . . . We know that if a person starts a new job or becomes a member of a new organization it generally takes weeks and often months before these conditions are achieved [Herbst, 1970: 52].[5]

In the social context of a community in a developing country, for example, a new practice, such as the consumption of soya or the prolonged breast-feeding of infants, may take on the characteristics of a stable pattern of behavior, if the inhabitants of the behavioral world of the community choose to make it do so. In spite of the potentially destabilizing effects of events beyond their control, the members of a community may choose to maintain a stable pattern of behavior—depending always on the meanings they construct for these events.

> No event has intrinsic characteristics of stress, strain, output, etc. but events acquire these characteristics in so far as a behavior structure of mutually dependent elements is evolved which operates so as to maintain its survival [Herbst, 1970: 52].

On the other hand, when meanings change, along with patterns of thinking, feeling and deciding, the stable "laws" of social phenomena may suddenly destabilize.

> A behavior system will cease to function if its organizational structure is dissolved or destroyed, the systems boundaries no longer operate, and no output is produced which is needed to maintain the essential transactional process with the environment [Herbst, 1970: 53].

The uncertainty characteristic of the contexts of nutrition intervention experiments in developing countries is an indication that the "lawfulness" of social phenomena is very different from the lawfulness of phenomena studied in the natural sciences. The former is an artifact of behavioral worlds, self-created by their inhabitants, and maintained—to the extent that it is maintained—by their patterns of thinking, feeling, and deciding.

Uniqueness. In our study of community-level nutrition interventions, we found, as Herbst had found in his longitudianl studies of work groups, that behavioral worlds may vary significantly from case to case, or even in the same case from an earlier world to a later time.

Candelaria has been shaped by twenty years of intense interaction with teams of professionals from a nearby medical school. In the village of Villa Rica, Colombia, where the mostly black inhabitants earn a meager living by working as hired hands on local plantations, there is a geographic split between residents of higher and lower status that affects every attempt at intervention. In Buenos Aires, a mountain village not far away, the social patterns of family and community life changed radically in mid-experiment in response to the start of construction work on a dam.

We agree with Herbst that the potential for uniqueness is inherent in the character of behavioral worlds:

> The basic difference between non-living matter and living beings is that the former is subject to laws whereas the latter create the laws that determine their behavior. Every person as a result of his aim-directed behavior builds a behavioral universe and the laws in terms of which it operates [Herbst, 1970: 53].

Because individuals build the worlds in which they live, and groups of individuals build community worlds, it is understandable that their patterns of stability may vary from case to case. For any given intervention hypothesis that purports to describe a general pattern of reactions to intervention there is good reason to expect the next situation to be, in some important way, an exception.

It does not follow from this that social contexts of intervention that are unique in some respects must be unique in all respects. The people who participate in nutrition experiments have a physiological as well as a social dimension. The categories of anthropometric measurement of human beings, or nutritional contents of foods, may be validly generalizable across social contexts, even when the social meanings of measurement or food are found to vary from one context to another. As we will explore more fully later on, nutrition intervention experiments may lend themselves to a combination of experimental methods, some of which are adapted to the uniqueness of behavioral worlds and others, to the replicable categories of nutrition science.

What does not follow from the uniqueness of behavioral worlds, however, is that we cannot, with any confidence, claim to generalize to "like situations" an intervention hypothesis inferred from a particular set of experimental results.

Transaction. The methods of true and quasi-experiment depend on a view of the experimenter as one who manipulates a few variables in order to observe the effects of his manipulation on other variables. According to this view, the experimenter stands outside the experimental situation. Any change the experimenter induces in the situation, apart from his or her intended manipulations, counts as a distortion—an effect of "testing" or "reactivity."

We have observed, on the contrary, that the interventions in our sample always produced changes in the situation beyond the intended experimental treatment. In Honduras, for example, a supplementary feeding program attracted an influx of new families whose demands overwhelmed the project's limited resources. In the more apparently successful of our cases, moreover, the specifically nutritional component of the intervention was a small part of a broad-guaged transformation of the behaivoral world of the community. In Candelaria, health professionals flooded the residents with services, with wide-ranging effects on the patterns of family life. In Kottar, the project leaders undertook a regional program of community organization that built up a network of local cooperatives and effected widespread changes in attitudes toward regional government. It might be argued, indeed, that these interventions were successful *because* they changed the nature of whole communities. Certainly, in both cases, the experimenters' ability to collect reasonably accurate time-series measurements of the nutritional status of children depended on their success in instituting a new pattern of regular child visits to health clinits—a radical change in an important aspect of family life.

As a result of these holistic changes, neither Candelaria nor Kottar was, by the end of the experiment, the "same community" that it had been at the beginning. Such changes are not properly described by an intervention hypothesis of the form, "in situation S, intervention X produces a difference, D, between pre- and postintervention observations." Postintervention, "S" is no longer "S" but "S^1". We would do better to say that the experiment had transformed the situation into one that manifested the pattern subsequently observed. In such cases, the intervenor does not "confirm an intervention hypothesis"; rather, his or her ability to effect a presumably desirable transformation is revealed.

Similarly, the intervenors were significantly changed by their interventions. They became deeply involved in community life, entered into close relationships with local practitioners and participants, and—in at least one case—underwent a radicalizing change of view that led them to recast the problem of malnourishment as a problem of political and economic organization. If we wished to be faithful to the experience of both the experimenters and the participants, we would have to place the intervenors *in* the situation they were trying to understand and change. As Geoffrey Vickers has put it, they were not "speculator/manipulators" but "agents experient." They engaged in a *transaction* with the situation in the course of which both they and the situation as a whole were transformed in ways both intended and unintended.

From the point of view of the validity of true or quasi-experimental inference, we would have to say that these sorts of interventions—our most likely candidates for "success"—were not experiments at all. There was no manipulation of a few variables, holding the rest of the situation constant. Rather, we should say that the intervenor framed the problem of the situation in question found in the community, constructed an image of a desirable future for that community, and then embarked on a broad-guaged strategy of intervention through which he or she tried to make that vision come true. The resulting transformation is, inevitably, bound up with the person or persons who undertook the intervention. The uniqueness of the changing situation is, at least in part, of the intervenor's own making.

In summary, then, we have attempted to apply quasi-experimental methods to a sample of community-level nutrition interventions, having chosen from a verly large sample those that seemed most amenable to being treated as experiments. For the most part, we found the results of the experiment indeterminate. In only two cases were we able to ascertain that a positive change in nutritional status had occurred and, even in these, we could not attribute the change unambiguously to the intervention.

The sources of indeterminacy were mixed. In some cases, the difficulties of interpreting data were attributable to sloppy practices. We found that some dirty data could be made interpretable postexperiment, and that some methods of measuring nutritional status could be improved. We also found it possible, by supplementary research undertaken after the experiment, to eliminate some plausible accounts of observed changes. But we found residual dirty data, measurement error, and ambiguity of results, which we could not eliminate by any of these after-the-fact methods.

More fundamentally, we found that certain inherent features of our intervention situations are inconsistent with the philosophical assumptions on which quasi-experimental method depends:

(1) Our situations are vulnerable to unpredictable instabilities which not only confound the experiment but bring the experimenter into a state of real uncertainty. Under these circumstances, the ontology of the situation becomes problematic. The inquirer cannot continue to apply quasi-experimental methods until he has redescribed the situation and reframed the problem he is trying to solve. *These* tasks cannot be undertaken *via* quasi-experimental method; they are necessary conditions for the further application of the method.

The potential for instability is inherent in the behavioral worlds of the situation and its larger social environment. The stable patterns of behavior that sometime characterize behavioral worlds are dependent on the thinking, feeling, and choosing of the individuals who create and maintain those patterns; and they may destabilize when individuals come to think, feel, and choose differently.

(2) Intervention situations are often unique. Behavioral worlds vary significantly from time to time and from case to case. If the "next case" is likely to be significantly different from any given intervention experiment, then we cannot assume the lawfulness of social reality on which external validity depends.

(3) Researchers, practitioners, and participants create for themselves a behavioral world whose properties affect the results of experiment. The principle of experimental distance, which demands that practitioners and participants remain apart from and ignorant of the goals of the experiment, contributes to the production of fragmentary and erroneous data.

(4) In many intervention situations, and especially in the ones where a positive change in nutritional status is detectable, the intervenor produces a holistic transformation. It is not appropriate here to say that a change in an experimental variable has made a difference in a dependent variable. The whole situation has changed through the interventor's interaction with it.

Possible Responses to Our Findings

If findings such as these are accepted (and some of them, at least, are familiar to other workers in the domain of malnourishment), several responses are possible.

Some researchers, wedded to the ideal of true experiment and plagued with difficulties in their attempts to use quasi-experimental methods, have called for still more stringent efforts at randomization of experimental and control groups, with a correlated increase in the complexity of experimental design:

> Therefore, any experimental design which does not randomly distribute the intervention and its controls within a village or region must have sufficient villages or regions covered by each treatment (replicates) so that one can estimate the probable contribution of non-specific influences at the village or regional level. Adjacent villages and regions must have different treatments, and the villages and regions should be so stratified that any other random non-specific influences are controlled for. Designs which show differences between required replicates must remain suspect [Habicht and Butz, 1979: 150].

But, given our findings, such efforts are doomed to failure. If anything, they would increase our dependency on principles of predesign, constancy, and distance that are impossible to maintain under the real-world conditions of community-level intervention. It is no accident that the most rigorous designs in our sample produced results as ambiguous as the most informal.

Indeed, a more stringent attempt at rigorous experimental control is likely to reduce the reliability of interpretation. To the extent that researchers try to kep subjects and practitioners from knowing about the experiment, in order to avoid the effects of reactivity or instrumentation, they create a pattern of deception that erodes the trust essential both to effective delivery of the intervention and to reliable data collection. If they try to exercise strict control over the distribution of treatments perceived as scarce goods, they tend to foster competition, envy, and hostility and, what is more, they engage in practices of dubious morality. If they try to enforce experimental controls by the use of coercive methods, their findings are likely to reflect the unintended consequences of their coercion.

Such conclusions might lead to a second response: Practitioners of nutrition interventions in the developing world should give up their hopes for learning based on rigorous experiment.

There is a third response. If we have been unable to implement the models of true or quasi-experiments, perhaps we should reconsider their appropriateness to the conditions of nutrition intervention in the developing world. Perhaps we must adopt a view of experimental efficacy, validity, and rigor better suited to the unpredictable instability, the uniqueness, the transactional properties and the dependence on local practitioners and participants, which we have found to be characteristic of our sample of community-level interventions.

In the following section, we shall outline such an alternative model of experimental inquiry. It differs from Campbell's quasi-experimental method in that it is not a proposal to "soften" or dirty the clean, hard method of natural science experimentation. Rather, it is a proposal to rethink the meaning of a good intervention experiment and the kind of knowledge we can get from it.

Reflection-in-Action

When we consider a community-level nutrition intervention as reflection-in-action, we consider it as a process of iterative experimentation, undertaken on the spot. The inquirer begins with an initial description of the community situation and an initial framing of the problem he is trying to solve. On the basis of these, he or she designs an initial strategy of intervention and begins to carry it out. The design will include a data gathering system, the terms of which will monitor, on-line, the consequences of his or her actions. Data are rapidly captured and fed back to the inquirer for interpretation. As they are interpreted, he or she becomes aware of consequences that may be expected or unexpected, desirable or undesirable. More narrowly, this information may be used to detect and correct faulty data gathering procedures, to identify ambiguities in the data, and make mid-course corrections in experimental design or redesign strategies of intervention. More broadly, as new features of the situation become apparent—some of which may be induced by his intervention—the intervenor may reframe the problem of the situation. The iterative process of describing, reframing, and redesigning may continue throughout the life of the intervention. Let us begin with the narrower view of reflection-in-action and then go on to the broader one.

As we have already seen, sloppy or inconsistent data gathering procedures may produce spurious changes in indices of nutritional status. When fieldworkers fail to "zero" a scale properly, for example, they may produce a consistent overstatement of children's weights. Or when weight-for-height is used as an index of malnourishment, a .3 kilogram overestimate may cause 15% of a group of malnourished children to appear normal. In Primops, the percentage of observations falling outside the predefined range increased steadily with each new batch of data from a low of 6.9% to a high of 24.1%—a change attributable, we believe, to the *promotoras'* decreasing interest in data gathering. In Kottar, we found obvious key-punching errors in 40% of the anthropometric data. The method of reflection-in-action has a twofold effect on these sources of error. Because it brings data analysis into the field at frequent intervals in the course of experiment, it permits the detection and correction of procedural mistakes before they produce an irreversible impact on the quality of data. Further, because it involves local practitioners in data analysis and redesign of experiment, it exposes them to the uses of the data and the disruptions produced by errors. Practitioners may then become highly motivated to improve their procedures. In Honduras, for example, a quick and dirty analysis of data on nutritional status, undertaken in the field in the early stages of the project, led to a wholesale revamping of the data gathering system.

Reflection-in-action cannot prevent changes in the composition of experimental or control groups that give rise to spurious analytic results, but it can give researchers a chance to become aware of these changes while it is still possible to ascertain their magnitude and explore the reasons for them. As local practitioners learn to shift their emphasis from service delivery to experimentation in effective reduction of malnourishment *through* service delivery, they are likely to become more interested in tracking the original sample of participants, charting their progress, and understanding the causes of their defection. They may learn to see changes in the sample, which might otherwise confuse estimates of progress, as occasions for redesign and retesting of interventions. When researchers work closely in this way with practitioners and participants, they choose to accept and deal with the risks of Hawthorne effects in order to avoid the costs of experimental distance. Similarly, reflection-in-action offers advantages over quasi-experimental method in its treatment of ambiguous data. For one thing, practitioners and participants are closer to the project environment than are researchers and, as a result, they are more likely to identify changes in environment and behavior that affect nutritional status. Hence, they are often better equipped to generate plausible rivals to the intervention hypothesis. In the Honduras food supplement experiment, for example, local practitioners made the researchers aware that an apparent improvement in nutritional status might be due to a new screening policy, instituted by a medical adviser, that allowed only malnourished children to enter the sample. Second, because reflection-in-action encourages the early discovery of ambiguous data, it permits experimenters to carry out on-the-spot experiments that generate new data in the field while the intervention is still underway. In Kottar, for example, had inquirers been aware of the problem of distinguishing the effects of intervention from the effects of recovery from drought, they might have used a strategy of *differentiated environment*. In mid-experiment, they could have divided the group receiving experimental treatment into subregions of high and low drought. In Primops, where we learned after the fact that it was impossible to distinguish the effects of improved sewage disposal from the effects of the *promotoras'* teaching, it might have been possible in the course of experiment to employ a strategy of *differentiated treatment*. Of two groups of children subject to the benefits of improved sewage disposal, only one would be selected to receive the intervention.

Of course, these strategies of on-the-spot experiments would have their limits. In the effort to differentiate environment, experimenters might introduce new sources of variation; children in a low drought region might turn out, for example, to be surrounded by a richer supply of foods. In the effort to differentiate treatment, experimenters might select a treatment group that differs in *several* ways from the control group. In the first case, one might further differentiate the sample by distinguishing in the low drought region between poor and less poor families. In the second, one might randomize the selection of the treatment and control groups. In the field, however, there are limits to the number of groups that can be set up and observed. Repeated differentiation of the sample may produce cell sizes too small for significant analysis, and the establishment of control groups may be politically or morally unacceptable.

In situations like these, *qualitative description* may help to discriminate between an intervention-hypothesis and its rivals. A richer description of the context and process of intervention may provide evidence in favor of one of the contending hypotheses. Recovery from drought might be reflected, for example, in an observable change in the variety and quantity of foods available to the families. Or, where the problem is one of distinguishing the effects of education in nutrition and hygiene from the effects of improvement in the economy, experimenters might observe changing patterns of diet, hygiene, and infant diarrheas at the family level. Here, too, however, supplementary research may uncover new sources of ambiguity. A better description of the intermediate effects of intervention may suggest new rival hypotheses to account for an observed change in nutritional status.

Because the several strategies of on-the-spot experiment may fail to discriminate among plausible hypotheses, and may even multiply sources of ambiguity, it is reasonable to ask when it is legitimate to bring hypothesis testing to a close. How much discrimination among hypotheses is enough? This question calls for a reexamination of the functions of experimentation.

We recognize three such functions. In the first, *exploratory experimentation,* we probe something to see how it will respond. In the second, *intervention-testing experiment,* we experiment to test the efficacy of an action: When we act in a certain way, do we get what we intend, or like what we get? In the third, *hypothesis-testing experiment,* we try to confirm or disconfirm a hypothesis by testing whether the consequences of action deduced from the hypothesis occur. An exploratory experiment succeeds when it leads to the discovery of something new. An intervention-testing experiment succeeds (or is "affirmed," as we shall say) when the inquirer likes what he or she gets from it, taking its intended and unintended consequences as a whole. A hypothesis-testing experiment succeeds when it discriminates among rival hypotheses.

Understandably, given his natural science view of social experimentation, Campbell focuses only on hypothesis testing. When intervention experiments are considered from the perspective of reflection-in-action, however, they are seen to combine the three functions. The hypothesis-testing experiment is also an intervention by which the inquirer tries to change the situation for the better and as a probe with which he or she hopes to discover new phenomena. In reflection-in-action, therefore, we evaluate an intervention experiment on the basis of three criteria: Has the intervention been confirmed? Has it led to the discovery of something new? Has it discriminated among competing hypotheses? The three criteria are logically distinct. An intervention may reveal new phenomena or may be affirmed, even when it fails to produce its intended consequences. Conversely, an intervention may succeed in discrimination among rival hypotheses while producing an undesirable change in the situation. Under these circumstances, the inquirer's evaluation should reflect the relative importance he or she attaches to the various interests in the experiment. Is the inquirer more interested in changing the situation, in understanding the causes of its change, or in discovering something new about it?

Suppose, for example, that a supplementary feeding intervention yields no observable change in a community's rate of malnourishment. The intervenors may then discard their initial intervention hypothesis in favor of a new one: Perhaps, even with supplementary feeding, patterns of food sharing and substitution leave the level of consumption of nutrients unchanged. If new observations are inconsistent with this hypothesis, the intervenors may look to parasites in the community's drinking water. If they find parasites in the water, clean water may become the focus of a new intervention. If clean water coupled with supplementary feeding yields an improvement in nutritional status, and no further decline in the overall situation occurs, the intervenors may bring the learning sequence to a close. Here, it is the logic of affirmation that answers the question, How much discrimination among alternate hypotheses is enough? The answer is, Enough to produce an intervention that can be affirmed. Of course, if the intervenor thinks of the present project as a preparation for future ones, he or she may continue hypothesis-testing experimentation beyond this point, seeking to rule out plausible rival accounts of his or her success or to pinpoint features of the intervention that worked.

Thus, a reflection-in-action hypothesis-testing experiment is bound by the inquirer's appreciations. Initiated by the detection of something troubling or promising, it is terminated by changes the inquirer finds on the whole satisfactory—even when the store of rival hypotheses has not been exhausted—or by the discovery of new features that give the situation a new meaning and change the questions to be asked.

Turning now to the broader meaning of reflection-in-action, we recall that intervention experiments are always embedded in a particular way of framing the problematic situation. Even when inquirers begin with a common objective function (for example, "over a five-year period reduce by 20% the rate of malnourishment in children between the ages of zero to six"), they must still select things and relations in the situation to which they will pay attention; and they must frame the situation in a way that gives direction to inquiry. Often the problem framing is tacit. The inquirer may become aware that the problem has been framed in a particular way only when he or she is surprised by an unexplained and unwanted turn of events. Then he or she may become aware of his problem-setting, and may attempt to reframe the problem. This inquiry takes the form of a frame-experiment.

Consider a case in which field researchers in a developing country begin by framing the problem of infant malnourishment in terms of nutrient deficiency. They invented a solution to this problem—a new infant formula containing supplementary nutrients essential to infant growth, which they planned to distribute through normal marketing channels. Several years later, mothers in many developing countries were found to believe that it was more fashionable or "modern" to feed their infants the new formula than to breast-feed them. At the same time, researchers found that the nutritional status of infants actually declined. Not only did infants lose the immunizing benefits of mother's milk, but because of the poor quality of water used to prepare the formula, they suffered from a higher rate of diarrheas. As a result, researchers began to rethink the problem thay had been trying to solve. Some of them reframed the problem as one of encouraging mothers to shift back to earlier habits of breast-feeding; their solutions took the form of education and public information campaigns. Others focused attention of the multinational companies that manufactured and sold infant formula. They framed the problem as one of profiteering and misinformation, and their solution took the form of efforts at regulatory control and political pressure.

In this case, an initial intervention produced new information that the researchers interpreted as calling not for a new data-gathering system and not only for a new intervention hypothesis, but for a new problem setting. They focused on new things and relations in the situation (mother's attitudes toward breast-feeding, the amorality of multinational corporations) and they reorganized the directions of their inquiry. They might then test their reframing of the problem against the following criteria:

- Do their new descriptions correspond to the facts? For example, have mothers actually shifted in large numbers from breast-feeding to infant formula? Is the water used to prepare the formula really of low quality? Has the rate of infant diarrheas really increased? Negative answers to these questions would be inconsistent with the new problem settings; affirmative ones would be consistent with them (but might also be consistent with others).
- Can researchers solve the new problems they have set? Can they, for example, persuade large numbers of mothers to return to breast-feeding? If not, their reflection-in-action may lead, appropriately, to yet a new problem setting.
- Does the new problem setting make a confusing and uncertain situation coherent to the inquirers? Does it open up inquiry to new directions of interventions?

The interpretation of a frame-experiment is always relative to an inquirer's appreciative system—that is, to the system of values and norms that guided his or her evaluations.[6] Hence, frame experiments yield personal knowledge that may be objective in some senses but not in others. "Objectivity" may mean "neutrality" (freedom from interpretive bias), "interpersonal objectivity" (holding true for one person as well as for another), or "independence of think-so" (grounded in evidence beyond mere opinion). Practitioners of true and quasi-experimental methods aim at neutral, interpersonally objective judgments; hence their emphasis on experimental distance and their selective inattention to their own problem settings. The practitioner of reflection-in-action recognizes that differences in problem settings are only partly resolvable through experiment and that differences of interpretation are only partly resolvable by reference to facts. Hence, he or she tries to reflect on the tacit appreciations that undelie his or her own interpretations and problem settings; and in relationships with local practitioners and participants, he or she does not try to eliminate Hawthorne effects but to become aware of them by helping to create a behavioral world conducive to the exchange of valid information.

We have already noticed, in Candelaria and Kottar, how the mothers' regular visits to local health clinics and the systematic measurements of children were inherently connected to the nutrition intervention. The creation of a regularized, predictable behavioral world was as indispensable to the effects of intervention as to the researchers' ability to detect those effects. In Kottar, the establishment of such a world was also very likely inseparable from the network of relationships built up by the two principal figures. These Hawthorne effects could not be eliminated without eliminating the essence of the intervention. In reflection-in-action, inquirers would not try to reduce such efforts but to gain valid knowledge of them. They would try to be explicit about the appreciative systems underlying their judgments and the person-dependent features of their interventions. Within such an explicit appreciative framework, it is possible to make interpretations of think-so, but such judgments are neither neutral nor interpersonally objective. In Kottar, the inquirers might discover, independent of think-so, whether their interventions produced a positive change in nutritional status, but they would neither dissociate their interventions from their own persons nor their judgments from their particular ways of framing and transforming community situations.

The absence of interpersonal objectivity is not troublesome to reflection-in-action because this method of experiment does not involve a search for general propositions applicable to similar cases. Here, the inquirer treats each episode of intervention as a unique case. Nevertheless, he or she aims at the *reflective transfer* of learning from one unique episode of reflection-in-action to others. Rather like a good medical clinician, he or she seeks to build up a usable repertoire of unique cases.

Once a case has entered into the inquirer's *repertoire*, he or she may be able to *see* a second unique case *as* the first, *doing* in the second *as* he or she has done in the first; the first functioning as an exemplar for inquiry into the second.[7] The two cases are not subsumed under a general proposition. When confronted with a new situation, the repertoire of cases derived from his or her own experience or from the recorded experience of others can be scanned to see if the new situation is similar to one or more of these. He or she need not, at this point, be able to say "similar with respect to what." But from understanding the first case, a variation appropriate to the second can be constructed—an initial understanding that functions as a starting point for a new round of reflection-in-action.

The norms of inquiry for reflective transfer of learning are only roughly similar to those of Campbell's external validity. Three main questions arise:

- What are the criteria of "fit" between an element of repertoire and a new situation?
- What does it mean to carry out a process of reflective transfer rigorously and well?
- What are the features of the understanding of a unique case that make it suitable for reflection transfer?

"Fit" cannot be reduced to effectiveness in improving nutritional status. The question is whether a case lends itself, in a particular situation, to a variation that would make a good starting point for inquiry. It would not do so if there were a gross dissimilarity between old and new situations. It would be inappropriate to see a new community situation as Candelaria, for example, if in the new situation the quality of water and sewage treatment were already relatively good. Further criteria of fit have to do with the inquirer's ability to detect intimations of useful similarities between old and new cases before he or she can articulate them or put them to the test. There is a craft of nutrition intervention, analogous to the craft of clinical medicine, that has to do with the inquirer's ability to form and test such judgments. It is important, for example, that the inquirer be aware of problem framing in the earlier case and attentive to differences of context in the two cases.

It is also important to notice that reflective transfer takes account of the uniqueness of the behavioral worlds of intervention situations. But there are also some dimensions of those situations—for example, those related to the anthropometric measurements of nutritional status—which lend themselves to the methods of true or quasi-experiment and to generalized application from case to case. An inquirer who treats his or her repertoire of cases only as a source of exemplars for the initial understanding of new situations may, at the same time, apply the method of characteristic curves to every situation encountered.

Finally, when we substitute reflective transfer for the schema of generalization, we also change the way in which intervention experiments lead to policy advice. In Campbell's formulation, intervention experiments aim at general policy recommendations. General policies are thought to originate at a center from which they flow to a distributed population. In reflection-in-action, practitioners are not implementers of a centrally defined policy but researchers, policymakers, and implementers of their own policies. The contexts of intervention are not instances of a type of situation but unique situations that bear, at best, a family resemblance to another. Reflection-in-action lends itself to "inductive planning" where policy is made and implemented in a highly distributed way, and practitioners learn from themselves and one another. In such a distributed system, a policy center would not make policy but would facilitate distributed learning by documentating, collecting, and disseminating cases of intervention and by helping to build and maintain the network of practitioners. (for further discussion of inductive planning, see Schön, 1971).

Conclusion

From our study we have concluded that the method of quasi-experiment is only partly applicable to community-level nutrition interventions. We found it possible, after the fact of experiment, to clear up some, but by no means all of the sources of ambiguity due to dirty data and confounding changes of context. We found that attempts to make a closer approximation to the methods of "true" experiment produced no further reduction in indeterminacy; indeed, in some cases, they added to it. We found, finally, that the epistemological assumptions underlying the methods of true and quasi-experiment are not true to the experience of nutrition interventions. Social phenomena are lawful only within the self-created and self-maintained conditions of unique behavioral worlds. Problems and variables are not given with the situation but are constructed by the inquirer who frames and reframes them in the course of a transaction with the situation in which he or she contributes to the creation of the phenomena, including their uncertainty and uniqueness, which he or she also observes.

In reflections-in-action, we have proposed a method of experiment appropriate to these findings. It is at once a proposal for the practical use of local information systems, a methodological proposal for on-the-spot experiments and mid-course corrections, and an epistemological proposal for objectives and norms suited to the actual experience of intervention in situations of instability, uncertainty, and uniqueness.

In our view, the inquirer should bring to the intervention situation an initial understanding of the situation, a framing of the problem, and an intervention-hypothesis, all of which will be subjected to iterative revision through the learning sequence of on-the-spot experiment. He or she monitors experimental data on-line, and conducts mid-course analyses in order to revise data collecting procedures, discriminates among rival hypotheses, redesigns the interventions, and restructures the description and framing of the problematic situation. He or she tries to use hypothesis-testing experiments to arrive at internally valid inferences, but only within the bounds of efforts to affirm the intervention. His or her conclusions are always relative to the uniqueness of the case and to the framing of the problem. He or she seeks to work closely with local practitioners and participants in order to draw on their insights and gain their involvement in the experiment. Objectivity in the sense of neutrality and independence of person is not the goal. Once the appreciations that shape interpretations and behavior have been determined, he or she is able to affirm an intervention in a unique case, which may become part of a repertoire of cases, a source of exemplars for the construction of initial understandings of other unique cases. The inquiry does not yield general policy advice but adds to a repertoire of cases from which other inquirers may learn as they frame the unique problems of other family-resembling situations.

Notes

1. Controversies over the "Negative Income Tax" and "Housing Allowance" experiments are cases in point. A very interesting recent discussion of the former is Leland Neuberg's Notes on the Conceptual Basis of Social Controlled Experimentation (manuscript in preparation).

2. The tradition has been most prominently represented in the United States by Kurt Levin and some of his students and intellectual descendants.

3. There is not a single major thesis advanced by either nineteenth-century Positivists or the Vienna Circle that has not been devastatingly criticized when measured by the Positivists' own standards for philosophical argument. The original formulations of the analytic-synthetic dichotomy and the verifiability criterion of meaning has been abandoned. It has effectively shown that the Positivists' understanding of the natural sciences and the formal disciplines is grossly oversimplified. Whatever one's final judgment about the current disputes in the post-empiricist philosophy and history of science . . . there is rational agreement about the inadequacy of the original Positivist understanding of science, knowledge and meaning [Bernstein, 1976: 207].

4. He limits his use of the term "instability" to instability of measures, "fluctuations in sampling persons or components, autonomous instability of repeated or 'equivalent' measures" (Campbell, 1969: 411). However, he includes instability of environment in his category, "history: events other than the experimental treatment, occurring between pre-test and post-test and thus providing alternate explanations of effects" (1969: 411).

5. Herbst's usage of the term, "behavior world," is paralleled in Argyris and Schön, *Theory in Practice,* where the authors describe the behavioral worlds created and maintained by individuals who regularly bring certain theories of interpersonal action to their encounters with one another. In these authors' examples, such worlds may be characterized by features such as high or low defensiveness, openness, warmth, win/lose behavior, and the like. Argyris and Schön share, with Herbst, the emphasis on the self-created character of behavioral worlds, and on the *work* that goes into their creation and maintenance.

Other researchers have also described behavioral worlds (with or without explicit use of that term), and have emphasized one or more of their features. Bateson (1972) describes the patterns of "zero-order learning, that is, of continual stability-maintaining error-detection and correction, which characterize human organization.

6. The term is borrowed from Geoffrey Vickers.

7. In *Essential Tension,* Thomas Kuhn has spoken of the function of exemplars in the process of scientific evolution, as in the process by which a student learns a science. Kuhn believes that a physics student, for example, learns certain canonical problems (e.g., the

problem of calculating the accelleration, etc. of a ball rolling down an inclined plane). He is then able to perceive similarities between these problems and others, before he can say "similar in respect to what." What Kuhn earlier called a "paradigm" he now believes to be better described as a set of exemplars, shared by the members of a community of inquiry.

References

BATESON, G. (1972) Steps Toward an Ecology of Mind. New York: Ballantine.

BERNSTEIN, R. J. (1976) The Restructuring of Social and Political Theory. New York: Harcourt Brace Jovanovich.

CAMPBELL, D. T. (1969) "Reforms as experiments." Amer. Psychologist 24, 4.

———and J. C. STANLEY (1963) "Experimental and quasiexperimental designs for research," in Handbook of Research on Teaching. Skokie, IL: Rand McNally.

DRAKE, W., R. MILLER, and M. HUMPHREY (1980) Final Report: Analysis of Community Level Nutrition Intervention Programs. Ann Arbor, MI.

HABICHT, J.-P. and W. P. BUTZ (1979) "Measurements of health and nutrition effects of large-scale nutrition intervention projects," in R. E. Klein et al. (eds.) Evaluating the Impact of Nutrition and Health Programs. New York: Plenum Press.

HERBST, P (1970) Behavioral Worlds: The Study of Single Cases. London: Tavistock.

MOYNIHAN, D. P. (1976) "The American experience," in The American Commonwealth (tenth anniversary issue of The Public Interest). New York.

REIN, M. and R. WEISS (1970) "The evaluation of broad-guaged social programs." Admin. Sci. Q. 15, 1.

SCHÖN, D. A. (1971) Beyond the Stable State. New York: Random House.

DONALD A. SCHÖN has worked as a researcher and practitioner in the fields of technological innovation, organizational learning, and professional effectiveness. He has served as an industrial consultant, a government administrator, and as president of a nonprofit social research consulting firm. He was invited in 1970 to deliver the Reith lectures on the BBC. His books include Invention and Evolution of Ideas *(1963),* Technology and Change *(1967), and* Beyond the Stable State *(1971); he is coauthor with Chris Argyris of* Theory in Practice: Increasing Professional Effectiveness *(1974) and* Organizational Learning *(1978). He is currently Ford Professor in the Department of Urban Studies and Planning at MIT. Dr. Schön's most recent book,* The Reflective Practitioner, *was published in January 1983 by Basic Books.*

WILLIAM D. DRAKE has worked as a practitioner and researcher on issues surrounding planning and implementation. He chaired the Ann Arbor Metropolitan Transportation Authority during its formulative years, and was instrumental in initiating an interdisciplinary Ph.D. program in Urban and Regional Planning at the University of Michigan. Currently he is Professor of Resource Planning and Conservation at Michigan and President of Community Systems

Foundation. He is coauthor of a seven-volume work, Analysis of Community-Level Nutrition Programs.

ROY I. MILLER is a Research Associate of Community Systems Foundation in Ann Arbor, Michigan. For the past five years, Dr. Miller has worked on developing and applying techniques for monitoring and evaluating nutrition interventions. As an evaluation specialist, he has worked with individuals administering food aid programs in Honduras, India, and Sri Lanka. He is a coauthor of the seven-volume work Analysis of Community-Level Nutrition Programs.