

Linguistic Research Institute

Report No. 22A

REPRESENTATION, EVALUATION, AND RESEARCH

Peter G. Ossorio

1979

c Copyright 1979
All Rights Reserved

Linguistic Research Institute
Boulder, Colorado
Whittier, California

Linguistic Research Institute

Report No. 22A

REPRESENTATION, EVALUATION, AND RESEARCH

Peter G. Ossorio

1979

c Copyright 1979
All Rights Reserved

Linguistic Research Institute
Boulder, Colorado
Whittier, California

In What Actually Happens I made the point that a new beginning was needed if we were ever to achieve a behavioral science. I also made the point that such an enterprise called for the assured compatibility of explanation, methodology, and historical description, and that the only way to assure that compatibility was to formulate a single conceptual framework for all three.

Given that one of the principal dogmas of la dolce vita academica is the categorical separation of method and theory, it is not surprising that the primary reactions among academicians and experimental practitioners range from blank incomprehension to incredulity to simple rage and lashing out. Even the two prima facie examples in What Actually Happens, where a substantive problem was dealt with by recourse to a 'methodological' formulation and a methodological problem was dealt with decisively by recourse to a 'substantive' formulation, have drawn little more than efforts to explain them away or dismiss them as being very special cases rather than exemplars of a general possibility.

The present paper involves a third and fourth example of this general possibility. In both cases, methodological formulations (research paradigms) are derived from 'substantive' formulations. In point of fact, as I mentioned in WAH, it is in some ways misleading to speak only of crossing over from substantive to methodological and vice versa. In most cases it would be apt to say that the conceptual formulations in question can be applied equally to either substantive or methodological issues. We speak of crossing over primarily when

the requirement of compatibility is relevant.

Although this paper has that polemic aspect, it is not merely a polemic exercise. If anything were needed to convince us that it is not merely in principle, but in a simple practical sense as well, that we need an explicit, systematic, coherent development of the rationale for scientific practice, we have only to take a hardheaded look at the situation in dolce academica, where the bureaucratic and political structure of 'behavioral science' has almost completely obscured and subverted its original intellectual aspirations. What we find there is, except in detail, what we might expect in any large and established bureaucracy, namely, ad hoc conventions, folklore, customs, territorialities, authoritative pronouncements, cliques, self-perpetuation, log-rolling, and a hypertrophied development of administrative procedures, which in this case are codified in mathematical and quasi-mathematical models for experimentation and statistical analysis. What we do not find, though there are islands of rationality here and there, is an intellectually responsible, coherent, and substantively adequate body of thought concerning the rationale and technology of the scientific study of behavior.

One of the advantages of making a fresh start is that we need not be forever enmeshed in fruitless dialogues whose essential failure is guaranteed by the inadequacy of the general intellectual framework administered by dolce academica (the "Tar Baby problem" in WAH). However, two simple examples of what one might take issue with may be informative.

(a) The convention of 5 percent and 1 percent significance levels is a fairly obvious case of bureaucratic convenience taking

precedence over task-oriented rationality. The pretence that 5 percent significance is decisive evidence but 5.01 percent significance is no evidence at all has a special kind of absurdity. And the ad hoc conventions associated with it lead to further absurdity. For example, I may have three groups A, B, and C, with a significant difference between A and C but non-significant differences A-B and B-C; under these conditions I am required to conclude that A is different from C, but B is not different from either A or C. This absurdity results from what amounts to substituting my degree of certainty about the differences for my quantitative estimate of what the differences are. And although this procedure might well be formally repudiated when stated so baldly, in academica it is enforced administratively with a heavy hand by editors and editorial reviewers in actual practice. In the face of such conventions, one might well feel the need to come back through the Looking Glass to the real world.

(b) Much of what is now called "methodology" and "experimental design" is based on the idea that one samples from a population and generalizes one's results to other samples or individual members of that population. Yet it is almost never possible, even in a practical sense, much less in a methodologically rigorous way, to specify what population the subjects in a psychological experiment are to be taken as a sample from. The corresponding difficulty arises in justifying one's generalization of the results. (These issues can in general be paraphrased as issues concerning the representativeness of a sample with respect to a "population.") The convention in dolce academica has been to give an elliptic description of one's sample

as the description of the population one generalizes to or, more prudentially, to say nothing. For example, if my subjects are a particular cohort of sophomores in psychology classes in a particular institution, I will gravely announce that I generalize my results to "college sophomores" or "college sophomores at this institution." But nobody believes for a moment that my subjects are representative of such 'populations.' In short, as a member of dolce academia I must generalize, for experiments are not yet formally recognized as ends in themselves, but I cannot do so legitimately. After almost one hundred years of psychological research we have introduced the neologisms "internal validity" and "external validity" so as to limit the scope of the problem to the latter. Somehow, the introduction of these terms has not really solved the problem of generalization. It has, perhaps, made it easier to recognize the enormous difference between the existing administrative regulations concerning what qualifies as doing an experiment properly and the genuine methodological requirements for effective research concerning people and their behavior. (For an enormously different approach to generalization, see "Explanation, Falsifiability, and Rule-following.")

What we shall require, and what I propose to accomplish in a limited way here, is an explicit, systematic, and coherent development of the methodology of scientific practice in the study of persons and their behavior. Since that scientific practice encompasses an organized variety of behavioral forms, our task is to represent those forms normatively; hence there is the possibility of deriving our representations as special cases from more general formulations of

persons and their behavior. The specific task for an explicit methodology is to formulate as clearly as possible the structures of behavior and decision which are relevant for investigating persons and their behavior and formulating (and justifying) our understanding thereof. We will deal with two such structures here. They are in addition to the Pragmatic Paradigm presented in Persons and the Demonstration Research Paradigm presented in "Explanation, Falsifiability, and Rule-following."

The point of this is not to provide pre-packaged answers which are to be accepted without question nor to present procedural prescriptions which are to be implemented without question. On the contrary, it is to facilitate both (a) a descriptive appraisal of whether a given set of behaviors exemplifies a given methodological structure, and (b) a critical appraisal of whether the procedures which fit a given structure are relevant or appropriate to the task or question at hand on a given occasion. If behavioral science is to be a rational form of behavior, such critical appraisal cannot be abdicated in favor of political-administrative fiats and clerical-mathematical procedures.

In this respect I propose to (a) review two conceptual schemas which are used in Descriptive Psychology to represent real world phenomena, (b) delineate two general research paradigms which exemplify these conceptual schemas, and (c) illustrate these research paradigms with some special reference to evaluation research. I have chosen evaluation research because it appears that, more, even, than behavioral research generally, this field is afflicted by scientific

posturing, shamanism, and politics as usual.

Although I referred earlier to making a fresh start, when it comes to evaluation research it would be more accurate to speak of having made an early start. Much of the Descriptive rationale for evaluation research was developed in the course of a research program (Ossorio, 1964, 1966a, 1966b, 1967, 1968, 1971) which began in 1961, long before the emergence of Evaluation Research as a distinctive social enterprise. It would be surprising if the later developments in that field bore no resemblance to the developments in Descriptive Psychology. Such similarities are not to the point here. I am not concerned with the fragmentary ideas and techniques which have emerged here and there in the Evaluation Research literature, nor is the present report designed to contribute to that literature. I am concerned with delineating a comprehensive and coherent methodology which has evolved quite independently within Descriptive Psychology.

I. The Actor-Observer-Critic (AOC) Schema

This schema distinguishes three methodologically fundamental behavioral statuses, or ways of functioning. They are characterized individually as follows.

A. In functioning as Actor, the individual acts in response to his circumstances in accordance with his nature. This form of functioning is, therefore, characterized as (1) before-the-fact, (2) spontaneous, (3) creative, and (4) value-giving. It should be noted that although Actor, Observer, and Critic are distinguished as methodological roles rather than full-fledged human roles, there is

a natural tendency to think in the latter terms. In this case it is helpful to think of Actor functioning as involving both observation and appraisal but of a purely assimilative sort. That is, as Actor, one notices and makes use of things only insofar as they fit into, or affect, one's ongoing project. It is their place in one's life and activities which gives value to those things, hence the characterization "value-giving."

B. In functioning as Observer-describer, the individual takes note of (discriminates, describes, distinguishes) the course of events. Although particular attention may be paid to those things which are relevant to the behavior of the Actor, this form of functioning is objective, i.e. normative, rather than egocentric or merely subjective, as in Actor functioning. It is therefore characterized as (1) after-the-fact, (2) passive, (3) reflective, and (4) either value-neutral or value-finding.

C. In functioning as Critic-appraiser, the individual begins with the results of Observer-describer functioning and makes an appraisal of whether the course of events is satisfactory or unsatisfactory. If satisfactory, the Critic takes no positive action but appreciates/enjoys the success. If unsatisfactory, the Critic generates (a) a diagnosis, i.e. an account of what it is that has gone wrong and/or (b) a prescription, i.e. a practical guide in regard to what to do differently so as to improve matters. This form of functioning is, accordingly, characterized as (1) after-the-fact, (2) reflective, and (3) judgmental, verdict-giving, or value-finding.

There are two forms of the AOC schema. In the "methodological form," it has been used, among other things, to generate a criterion of adequacy for general theories of behavior (Ossorio, 1973). In this connection it may be noted that Critic-appraiser functioning is a special case of Observer-describer functioning, and the latter is a special case of Actor functioning.

In the "functional form" of the schema, the three statuses are related in the form of a feedback loop. As shown in Figure 1, what the critic judges to be the case is an essential part of the circumstances to which the actor then responds.

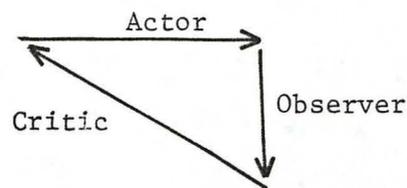


Figure 1. Functional AOC Schema

Formulas and schemas are, of course, not descriptions. The schematic differentiation of these ways of functioning is not a statement to the effect that there are different processes going on or that the person is engaged in three behaviors simultaneously, etc. In point of fact, a person is normally functioning in all three ways simultaneously. In any continuously functioning feedback loop, all the elements are operating simultaneously, yet the feedback function depends on a temporal succession and coordination among the three essential elements.

One of the primary values of the Actor-Observer-Critic schema is that it provides a systematic representation (in conjunction with

other conceptual-notational devices) of a person as being self-regulating while nevertheless being responsive to both personal and circumstantial states of affairs. Unlike common 'psychological' approaches such as Attribution Theory or Operant Conditioning Theory, which portray self-influencing mechanisms, the AOC schema permits us to portray the kind of self-regulation which is the primary expression of human rationality, though, of course, it does not force us to do so. It is this feature which finds a fairly direct exemplification in one of the research schemas, the Precaution Paradigm. In this paradigm, the collection of data is a case of rational precaution-taking or assurance-seeking. The methodology in this case has to do with the rationale of the precaution-taking; the technology, if there is one, has to do with the implementation (operationalization, realization) of the precaution-taking. In neither case is mathematical modeling (e.g. sampling theory, statistical analysis, measurement) or standard 'experimental design' necessarily involved.

II. The Precaution Paradigm

Some version of the classical experimental designs involving a control group and experimenter manipulation of independent variables is commonly considered to be the standard of rigor in establishing general conclusions on an empirical basis. Correspondingly, "applied research," including validation or evaluational research, is generally seen in either one of two lights. (a) In the first case, it is seen as merely the application of general "knowledge" which was itself acquired in this foolproof (rigorous) way. That is,

it is seen as a mere demonstration that some general principles arrived at via 'rigorous' methodology do apply to some phenomenon of interest in the real world. (But recall the problem of generalization and "external validity.") (b) In the second case the classic experimental procedures are held up as the ideal to be striven for in the empirical examination of the real world phenomenon itself. It follows in this case that to the extent that the classic design is not adhered to (and usually, it cannot be) the results are of dubious value and the conclusions drawn are unsound (though possibly true).

It should be noted however, that classical experimental methods are designed to support general conclusions. The primary interest is never on the actual subjects as such. (This is why the illegitimate generalization is required.) In contrast, much 'applied' research, and certainly most evaluation research, is designed to support particular conclusions. That is, what is of primary interest is not whether a certain kind of program, procedure, instrument, etc. is generally successful in this or that way or degree, but rather whether a historically particular program (or set of programs), procedures, etc. is in fact successful in this or that way or degree over a particular span of time.

Such a statement points up the contrast between the methodological structure of the precaution paradigm and the intuitive judgment of the traditionalist. For, in accordance with (b) above, the latter would be very likely to object. Thus:

Wil: That just shows how all this philosophizing can lead you astray. Of course all investigations aim at some degree

of generality. Just because you're not talking about a real experiment, that doesn't mean evaluation doesn't aim at some generality. Even a particular evaluation has to assume that if a program is effective it would also be effective with a similar population in similar circumstances. Otherwise, you'd just be engaging in an empty exercise of assigning praise or blame. Surely you don't want to do that.

Gil: You do have a penchant for playing "Now I've Got You, You Son of a Bitch," don't you. First off, let me say that what I was describing was evaluation research, not "an investigation." Secondly, let me tell you flatly that when I, as an administrator, commission a job of evaluation research or do it myself, what I want to know is how well this program is doing, and I don't give a tinker's dam about some other program, and I sure as hell don't care what someone might say about some purely hypothetical 'similar' program in some purely hypothetical 'similar' circumstances with some purely hypothetical 'similar' populations. I deal with the real world, not with mystic verbal or Pythagorean formulas. Thirdly, I can tell you just as flatly that, whether or not the evaluation involves praise or blame, the basic fact is that I'm going to act on that appraisal. I'm going to change the program, leave it alone, or axe it. You'd better believe that's not a meaningless exercise. By the way, tell me again about how you guarantee the external validity of an experiment and how you know what to generalize to. If you want a meaningless exercise, I give you that one for free.

The dispute and polemics are themselves not merely ad hoc. They, too, can be clarified by reference to the AOC schema. The traditionalist stance clearly reflects a combination of Observer-describer functioning and the "diagnostic," or verdict-rendering aspect of Critic functioning. Thus, Wil thinks primarily of judgmental descriptions as the end product and worries about their generality. Just as clearly, the precaution paradigm codifies a stance that reflects primarily Actor functioning and the "prescriptive" aspect of Critic functioning. Thus, Gil regards his later actions as the primary product of the evaluation and worries about their justification.

Even when more general conclusions are desirable in evaluation research, the emphasis is properly on a case by case approach because ordinarily the relative influence of historical-situational contexts is so marked that simple generalizations about procedures (programs, etc.) across contexts would be rash or foolhardy (even if it were a case of 'applying rigorously established findings'). One can always sum across replications of a fine-grained analysis; one cannot recover individual patterns from group data.

Clearly, then, the specifications for a viable evaluation methodology would include (a) being capable of rigorous implementation in the real world setting, (b) providing fine-grained or single case conclusions, and (c) lending itself to systematic variation and replication for supporting more general conclusions. The Precaution Paradigm is responsive to these requirements.

Let us first examine the fine-grained module. In this case we presuppose a real world setting where there is a procedure (instrument, program, etc.) which is engaged in and which is a candidate for evaluation. Our first move is to construct a representational formula (cf. the Process Description below) for the phenomena in question. This will have a general form of the following sort: A uses/does B with C in ways D (e.g. to accomplish G), which can go wrong in ways E_i , as indicated by observations F_{ij} .

For example, a therapist (A) may use a projective test (B) with a client (C) by interpreting M responses (D) to arrive at a psychodiagnosis and differential treatment (G). This might go wrong in that the client was misdiagnosed and the treatment would be ineffective (E_i), as indicated by the continued presence of the initial symptoms (F_{ij}). Or again, an agency worker (A) may use a "mastery of English" entry on an application form (B) with an unemployed refugee (C) to direct him to a laborer's job and a vocational training class (D). This might go wrong in that the refugee would perceive himself as having lost face and subsequently become depressed (E_i), as indicated by self-report of feelings and symptoms and failure to hold the job or progress in training (F_{ij}).

Given the representational formula, we next focus on the ways in which A could go wrong by using/doing B with C in those ways. At this point there arises the issue of real, or practical, doubt versus idle skepticism. In this regard, we introduce two test questions. The first is, which of these ways of going wrong am I most worried about? Given an answer to that, the second question is, is that possibility important enough so that it's worth taking precautions

against going wrong in that way? If the answer here is "yes," the next question is, which of the indicators of its going wrong that way are both convincing enough and accessible enough to make them worth getting (via the relevant observations)? If there are such indicators, the observations are made. These observations provide either a reassurance that the project is not going wrong that way or a warning that it probably is going wrong in that way. One of the implications of speaking of "indicators" rather than "measures" here is that neither measurement nor any quantitative procedure is essential here, although both are allowed for.

The foregoing constitutes the required methodological unit or module. In this regard the following may be noted.

- (a) The sequence, and the logic of the sequence, is that of Actor-Observer-Critic, and the sequence is one of an error-detecting feedback cycle.
- (b) What qualifies as "going wrong" (the "diagnosis") is context-bound insofar as it depends on the specific purpose(s) for which B is used/done and the specific setting, persons, et cetera which are involved.
- (c) Correspondingly, the indicated corrective measures (the "prescription") will be context bound. (In the context of program or treatment evaluation, an overall evaluation will correspond to a single feedback cycle whereas a pragmatic evaluation will involve repeated feedback cycles and modification of the program during its progress if the feedback information so indicates.)
- (d) The functional AOC schema is both an error-detecting and a success-detecting feedback loop. Although one might define success logically

in this context as simply the absence of failure, in fact there may be, and usually are, positive indicators of success. Thus, there is a corresponding module based on a modified formula in which we refer to ". . . can go right in ways E_i , as indicated by observations F_{ij} ." And, of course, we may combine the two, i.e., ". . . can go right or wrong in ways E_i etc."

- (e) What qualifies as success will in general be no less specific and context-bound than what qualifies as failure.
- (f) If there are no observational indicators of B having gone wrong in a given way in the real world settings, we may conclude that there is no real problem of its going wrong in that way.

The methodological module lends itself to systematic variation and replication because the representational formula amounts to a parametric analysis of the kind of use (instrument, program, etc.) which is to be evaluated. That is, A, B, C, D, E, and F will be parameters of that kind of use. Thus, one may replicate across A, i.e., persons or groups engaging in the procedure; across B, i.e., different forms, instances, or variations of the procedure; across C, i.e., different recipients or groups of recipients of the procedure; across D, i.e., the various ways of using/doing B, with special reference to the purposes, circumstances, settings, etc.; across E, i.e., ways of going right or wrong; and across F, i.e., different indicators of success or failure. And, of course, one may replicate across any or all combinations of these various possibilities.

By way of elaboration we may note the following.

- (a) Although the paradigm presented above involves the representation of a procedure in a historical setting, the procedure may be either actual or hypothetical, e.g., merely intended or planned. Correspond-

ingly, and particularly in the latter case, the ways in which the procedure might go right or wrong might be exemplified in some other setting, e.g., an experimental setting. For example, if the failure envisioned in the historical setting were the misdiagnosis of the client, that failure might be equally well exemplified in an experimental setting with clients selected specifically for the purpose of conducting the test. (But it also might not be. The problem of the "external validity" of experiments is a real one and there are no general solutions to it.)

- (b) The modular approach makes relatively clear something which is often glossed over in the classic experimental tradition, namely that it is impossible to take all logically possible precautions against being wrong or to obtain all logically possible assurances of being right. I say "glossed over" not in the sense that anyone would deny it overtly, but rather in that (a) taking more precautions is fairly automatically counted as being "more rigorous" than taking fewer precautions, and (b) pointing out a possible precaution which in fact was not taken is almost automatically counted as a legitimate criticism. These actions on the part of experimental and editorial practitioners speak louder than their words. These actions are compatible with the principle that one can and should take all possible precautions. They are not compatible with the principle that precautions reflect a critical appraisal and must be justified in each case.
- (c) Classic experimental design can be seen as a case of (1) taking certain standard precautions, whether relevant and important or not, and (2) taking them in advance of serious real world practice

('application'). Such a strategy might have some real value, but it might also be counter-productive, since the inability to meet the a priori requirement of certain precautions is likely to result in (a) the failure to do needed research or (b) the stigmatizing of such research as unsuccessful.

- (d) Experimental procedures, like any procedures, will exemplify B in Formula (1). We may represent the real world phenomenon by saying that A uses experimental procedures (B) with subjects and problems (C) in certain ways and with certain purposes in particular settings (D) such that A can go wrong in ways E_i as indicated by observations F_{ij} . This holds equally for the use of traditional experimental paradigms and for the newer paradigms such as the Precaution Paradigm. The use of experimental procedures is thus not something which per se carries any guarantee of any kind of success or avoidance of failure. And we cannot, with respect to a given kind of use of experimental procedures take all possible precautions etc. Experimental procedures provide a framework for exercising human judgment and competence, not a way of doing without it.

This conclusion may appear to be truistic. True. In classic 'methodology' it is merely an ad hoc, commonsense truism, not part of the methodology. In Descriptive Psychology, it is a methodological truism. In classic 'methodology' the necessity for exercising judgment in research appears extraneously in the form of exhortations to be knowledgeable or to be careful in this or that way. In Descriptive Psychology that necessity is an explicit part of the general behavioral concepts of which methodological concepts are special cases, and so is in no way extraneous. But then, classic

'methodology' is essentially only an elaborated technology, so such differences are to be expected.

III. Process Representation

The systematization of reality concepts (object, process, event, state of affairs, relationship) via a transformational calculational system and the systematic explicit representation of objects, processes, et cetera are presented in "What Actually Happen" (Ossorio, 1975, 1978). The following is a brief summary and review.

The two most relevant transformational rules ("Transition Rules") dealing with the general concept of a process are as follows.

Rule 4. A process is a sequential change from one state of affairs to another.

Rule 5. A process is a state of affairs having other, related, processes as immediate constituents.

Because these recursive rules deal with the general concept of a process they apply to all processes and will therefore not distinguish one process from another. In order to have a general method for giving explicit representation to a particular process or kind of process we require (a) a parametric analysis of "process" and (b) a systematic notation or format for giving process representations. Both of these requirements are met by the "Basic Process Unit" (BPU) shown in Table 1.

The Basic Process Unit has a gross structure of "Name" and "Description." The former identifies the process and the latter gives the explicit representation.

In the explicit "Description" we may distinguish a "gross structure"

(Stages, Options) and a "fine structure," or "State of Affairs Structure," (Elements, Eligibilities, Individuals, Contingencies, and Versions).

Rules 4 and 5 are most clearly expressed in the fact that the process is represented as a sequence of Stages each of which is itself a process. The fact that the latter processes (stages) can each have some number of distinguishable exemplars is expressed in the association of some number of Options (exemplars) to each Stage. It is the Stage-Option structure which codifies the recursiveness of Transition Rules 4 and 5, for each Option is itself a process and can therefore be represented by a Basic Process unit involving a new set of Stages and Options, and the latter can in turn be so represented, and so on ad infinitum. This enables us to represent processes of any magnitude in any degree of detail, down to continuous processes.

Processes generally involve object constituents as well as process constituents. These objects have certain relationships (their having these relationships is a state of affairs) which change over time and the changing of these relationships over time is (the same state of affairs as) the occurring of the process. "Element" refers to the formal ingredients (objects) of the process. "Individuals" refers to historical individuals in the abstract, and "Eligibilities" assigns Individuals to Elements. For example in the play "Hamlet," Hamlet, Polonius, the skull, and the castle are all Elements. Person X, Person Y, prop A, and prop B are abstract historical individuals. Either Person X or Person Y may be Eligible to play Hamlet or Polonius (but not the skull or the castle), whereas Prop A is eligible to play the part of the skull and Prop B is eligible to play the part of the castle. Both Individuals and Elements are needed in the formal specification because there need not be a one to one relation between them. For example, Person Y may

be eligible to play Hamlet in Scene I and to play a spearbearer in Scene II. Actual performances of "Hamlet" require actual individuals in place of the abstract Individual even though "Hamlet" as a kind of process involves no reference to particular persons.

The occurrence of a process involves the occurrence of one of the Options for Stage 1 followed by the occurrence of one of the Options for Stage 2, and so on. In the process representation restrictions on the occurrence of particular Options are given by Contingency statements. Contingencies may be Attributional or Co-occurrence contingencies.

In the case of a Co-occurrence contingency the occurrence of a given Option in a given Stage is incompatible with (or necessitated by) the occurrence (or non-occurrence) of some other Option in some other Stage (or a combination of such options). For example, in a chess game, White's possible fifth moves are highly restricted by which of the possible first, second, third, and fourth moves have actually occurred, and many moves which are in principle possible in chess as fifth moves (are Options in Stage 5 of that process) are not possible in this game (this Version (see below) of that process).

In the case of an Attributional contingency the availability of a given Option in a given stage is contingent on some attribute of the individual who is serving as a given Element. For example, throwing a 90-yard pass in football is one of the formal possibilities, but it could only occur if the passer had an exceptionally strong arm.

Statistical Contingencies represent empirical correspondences rather than conceptual requirements. They may be of either a Co-occurrence or Attributional sort. For example, a 90-yard pass would be unlikely if the line of scrimmage were more than 30 yards from the offensive team's goal

line (which reduces to a Co-occurrence Contingency). Similarly, a checkmate in four moves would be unlikely to be available to a player who did not see it ahead of time. And a refugee for whom attending class represented an extreme loss of face would be unlikely to attend class.

When a process takes place it takes place in one of the ways in which it can take place. The specification of Stages, Options, Elements, Individuals, Eligibilities, and Contingencies is a way of specifying the conceptual restrictions (optionally, augmented by empirical restrictions) on what sort and sequence of happenings would qualify as an occurrence of (an exemplar of) the process in question. Each such distinguishable exemplar is a Version of the process. Each of the ways in which a given process can take place is a Version of that process. The occurrence of the process on a given occasion is (the same thing as) the occurrence of one of its Versions on that occasion. In contrast, occurrences of the same process on different occasions will generally involve the occurrence of different Versions on the different occasions. And different Versions need not resemble each other in any other way than in being Versions of the same process.

The problem of "generalizing" or "applying" the results of laboratory or 'analogue' research is in part the problem of trying to draw conclusions about one process (Version A of Process X) on the basis of observing another process (Version B of Process X) which is not unlikely to be different in many crucial respects.

The requirements for describing a given process as such are in principle not different from the requirements for a parametric analysis of "process," i.e., for specifying the ways in which one process (or kind of process) can be the same as another process (or kind) as such or different from it. Thus, the reference to Stages, Options, Elements, Individuals,

Eligibilities, Contingencies, and Versions constitutes a parametric analysis of "process."

One important feature of the process representation provided by the BPU format is that it exemplifies a holistic, or "top down," approach. Description consists of specifying various facts about the process in question. Each fact serves to further distinguish that process from other processes of the same general sort, and there is no definitive 'complete' description. Thus, the BPU form of representation begins to be informative as soon as any information is available concerning the process in question.

Systematically incomplete process descriptions are codified by Means-Ends Descriptions or Task Analyses ("What Actually Happen," Chapter III). In a Means-Ends Description we merely specify the Elements of the process (or of each of the Stages) which contribute to the outcome of the process (including desired changes, if any). In a Task Analysis we merely specify what sets of states of affairs would qualify as successful Outcomes (in effect, we specify different 'Versions' of a successful outcome).

Those who are familiar with the range of conceptual-notational devices for real world representation will recall that a Process Description is appropriate for representing actual processes both in the abstract and in various historical exemplifications. A process representation is not *generally* the device of choice, for example, if one wishes to represent machinery or "systems" of any kind. For those, a Configuration Description, based on the Basic Object Unit would be the appropriate resource, for in such configurations the structure takes priority over the processes it undergoes in that, in general, without the structure there would be no such processes. And again, conceptual schemata are not descriptions or statements about how things are. Rather, they must be used in making

constructions which can be used to give descriptions or make statements, but conceptual constructions can be used in many other ways.

IV. The Simulation Paradigm

This research paradigm is designed to enable us to (a) formulate our understanding of a phenomena (and/or our guesses about it) by generating a process representation of it and (b) test that understanding by predicting certain facts about the phenomenon on the basis of other facts. What makes these predictions possible is the structure of the process description including, importantly, Co-occurrence and Attributional Contingencies. Because of the predictive implications of the contingencies which are stated in "if-then" form, the obvious technical implementation for this research paradigm is computer simulation, hence the designation "Simulation Paradigm."

In the context of the BPU process representation, the Simulation Paradigm involves the following features.

- (a) We begin with the general notion of a process and ask, "What is the phenomenon?" The answer here is given by specifying it as a process, e.g., "providing (certain) mental health services to children," and distinguishing various possible exemplifications (Versions) as categorized by the investigator. Classes of exemplifications correspond to the "outcome variables" of the classic evaluation design.
- (b) We ask, "What makes a difference in how the process goes?" The answer is given in terms of (1) the parameters of the process or any of its elements, and (2) the values of those parameters. The first of these two is a set that corresponds to the interventions or other predictive (causal or noncausal) variables in the traditional outcome study. There is no substantive limitation on the type of parameter which may be in-

volved. Thus, aspects of institutions, programs, persons, and situations or relationships may all enter the picture as "what makes a difference."

- (c) We ask, "Where does it make a difference?" In answer, we merely specify a locus in the process representation. Whatever makes a difference has to make a difference somewhere.
- (d) Finally, we ask, "What difference does it make there?" The answer will have the general form, "Depending on whether the value of parameter Q is x or y, the exemplification (of, e.g., "providing (these) mental health services to children") will belong to classes A, B, C," There is no restriction on the kind of functional relation which may appear here. (The statement of these relations is an integral aspect of the Process Description; formally, it consists of specifying attributional and co-occurrence contingencies.) For example, it may be a simple linear function, but it may be a non-numerical decision table or a logical "either . . . or . . . and . . . unless . . . if" kind of function. The latter is of particular interest in connection with the problem of strongly interacting variables (it is the potential for non-numerical values which leads us to speak generally of "parameters" rather than the traditional "variables"). In this case, the outcome is expressed as an explicit joint function of the values of the several "interacting variables," and that function will commonly have "or," "unless," and "if" components. In some of the very simple special cases, the joint function will be identical to a representation within the familiar linear additive model or other models used in

multivariate analysis. (Any type of multivariate analysis can be accomplished within the Simulation Model.)

The simulation model has the following features of interest.

- (a) It permits (indeed, requires) explicit statements of functional relationships between predictor and outcome parameters, but "default" relationships may be entered in the absence of relevant information or hypotheses.
- (b) There is no limit to the form which these functional relationships may take or to the nature of the parameters involved (e.g. personal, institutional, physiological, etc.).
- (c) It permits a clean separation between the conceptual, psychological model of the phenomenon and a mathematical model of the statistics.
- (d) It permits the detailed examination of the effects of any subset of predictor parameters, since the remaining parameters may be held constant, e.g. by using group means for individuals, hence allowing no variation in that parameter, or by using default relationships. (Such analysis would normally call for computer simulation implementation.)
- (e) Because it is, in effect, a predictive test of the entire set of relationships simultaneously, the sample size required for testing does not increase exponentially as it does in 'purely empirical' multivariate designs. (A complete factorial design, far from being purely empirical, represents the a priori application of a mathematical model. Yet it is one of the favorite paradigms of practitioners who insist they are "merely finding out" about the phenomenon, or "letting the data speak for itself.")

(f) There are two characteristic features of the paradigm which are worth noting. The first is that the actual Process Description will in general have to be relatively accurate in order to be of practical value in the sense of generating predictions which are both accurate enough and differential enough. This does not imply, however, that it has to be complete in the sense of incorporating all the concrete details of the phenomenon; since the Process Description is a "top down" approach one only goes down to the level of detail which is needed (or which is available), and, as we have noted, default relationships can be introduced where information or hypotheses are lacking. Thus, this approach makes possible simulation studies which are not possible under the traditional conventions of "bottom up" construction which do require all the detail. For example, in program or organizational evaluation we can generally obtain the required degree of information by interviewing a number of knowledgeable people; in traditional Evaluation Research, simulation requires so much information that, essentially, it is impossible to do.

The second is that in simulation studies there is no built-in analytical procedure for establishing the 'best fit' to a set of sample data. (Note that what we ordinarily call "best fit" in academica is almost wholly the product of convention. It is the best fit, given (a) a conventional criterion, e.g. the least squares criterion and (b) a conventional a priori, non-psychological model, e.g. a linear-additive model.) Inspection methods and systematic variation of hypothesized relationships are available, however, and

post hoc revision with cross-validation is a standard procedure. These latter methods provide for a much richer analysis of a set of data than is possible within the limitations of standard experimental designs and the associated statistical analyses. Traditional techniques are special cases within the range of possibilities generated by the simulation paradigm, since the linear additive model or any other mathematical model can be written in the form of functional relationships which are part of the simulation paradigm. Correspondingly, traditional statistical and curve-fitting analyses will be special cases of the "systematic variation of hypothesized relationships" noted above.

Here, as in general, the thrust of the Descriptive formulation is not to deny the possible utility of well known statistical and mathematical procedures. Rather, it is to provide a sufficiently general and fundamental methodology so that we can suit procedure to purpose rather than being constrained to a blind, uncritical, or fanatic adherence to the current folkways of academica to the detriment of our understanding of our subject matter.

- (g) Those who are familiar with the customs in Descriptive Psychology will almost automatically think of the preceding presentation as a Paradigm Case Formulation. There are clearly all manner of variations and extensions to be had on the basis of the Simulation Paradigm presented above, and there is no need to belabor the obvious.

There is one extension which may be worth noting. In What Actually Happens, I noted that "systems" representations will be given

by particular cases of Configuration Descriptions, i.e. Process Descriptions with Object components or Object Descriptions with Process components. In the case where the object of our study is a child's first year in school, a Process representation is clearly the most direct form of representation; the relevance of relationships among the child, the teacher(s), the parents, the other children, et cetera can appear in the form of Contingencies and functional relationships. In contrast, if our interest was in the classroom or the school or the school system as a system, an Object description would be the most direct formulation and the relationships among child, teachers, peers, supervisory staff, parents, etc., would appear explicitly within the Object description. Even here, however, it is the processes which take place involving the Elements of the system which would be the primary vehicle for simulation studies. Thus, the logic of the simulation paradigm is easily extended to the study and representation of systems of various sorts.

V. Evaluation Research

Evaluation research and routine program evaluation occur within the context of some basic facts which impose strong limits on the manner and ease with which evaluation can be effectively accomplished and the uses to which a given evaluation can legitimately be put. Among these brute facts are the following.

- A. Improvement does not occur in "pure form." Just as an actual success is never merely a success but is also a particular accomplishment such as

winning a race, so an actual improvement on the part of some particular individual is always also a more specific change in personal characteristics, behavior, relationships, achievements, etc. But it is the fact of improvement that we are interested in, not the particular changes per se.

- B. There is no specific change which is per se necessarily or intrinsically an improvement. The decision as to whether a given change will qualify as an improvement requires evaluative human judgment. This judgment will, in part, reflect the target individual's personal characteristics and his life setting, as well as the purposes or norms with respect to which the issue of improvement arises for the person making the judgment.
- C. When improvement is exhibited by different individuals, in general, it is exhibited in different specific ways, even when it is "the same" improvement. (The differences will correspond to the differences between different Versions of the same process.) Thus, in principle, there is no way to specify rigorously and in objective terms the specific changes which would qualify as improvement for an entire group of individuals. At its worst, such a stipulative attempt would approximate the irrelevance of flipping a coin to decide the question of improvement. At best, it might be good enough for some purposes for some people.
- D. In principle, there is no change of a specific sort which all observers would agree qualified as an improvement for a given individual. In general, a given change in an individual will affect different significant persons differently and they will, properly, make different judgments of "improvement." Judgments of this sort will, therefore, carry weight only among persons who are in agreement in this respect. The likelihood

of agreement may be increased in several ways:

1. By moving towards "lowest common denominator" phenomena: "If he's learned to dress himself (or "if he can now hold a job," "if she can now sit still in class") surely that's an improvement."
2. By specifying norms or frameworks: "From the point of view of reality contact, his being more aggressive at work is an improvement."
3. By restricting the implications of the judgment or uses to which it is to be put: "For purposes of deciding whether to terminate."
"From the standpoint of how to advise the parents," etc.
4. By stipulating that one party, e.g., the client, the parent, the school or agency representative makes the decision.

It is easy to assimilate the foregoing to traditional views and hence to misunderstand its import. Thus:

Wil: You're telling me that it's impossible to specify evaluation criteria in advance. But in that case all you could provide with your so-called evaluation research would be post hoc, accidental, arbitrary, and probably self-serving conclusions. That would kill evaluation research as a scientific enterprise and put it on a par with reading tea leaves.

Gil: Hold on! The sky isn't falling. I didn't say you couldn't specify evaluation criteria in advance. Indeed you can. What you can't in general do justifiably is specify criteria independently of context. That would be as foolish as trying to specify what things would actually be dangerous independently of context.

We manage the concept of "danger" quite well, thank you, even though our judgments of what is actually a danger to whom are context-bound, and the same for concepts like "good," "successful," "valuable," et cetera, which are involved in evaluation. Being context-bound is not a handicap: that's the kind of concept they are. If you think that specifying criteria in advance depends on specifying criteria universally, which in turn depends on specifying them in a context-free way, then you're in trouble. Your own rationale makes what you do illegitimate.

Wil: But if not on those grounds, then on what basis can one specify criteria in advance?

Gil: On the basis of being knowledgeable about the situation and about what is at issue, keeping in mind that different decisions in the matter are possible, because for different people or from different viewpoints there are different interests and different issues involved. Good bureaucrat that you are, you have a passion for uniformity and formality, and a distaste for simple reality. But passing regulations as to which things are to be called "dangerous" and which things are not will not keep you from getting killed off by one of your 'non-dangerous' things, nor will setting up arbitrary conventions about which criteria are the mark of success in organizations, programs, treatment, et cetera succeed in making silk purses out of sow's ears. As a bureaucrat, you can set up those conventions for your purposes, but as someone who has to cope with actual clients, actual program effects, et cetera I can't afford to fool myself

that I can define good results into existence just by calling them that.

If I don't discount my knowledgeability, I am quite willing to say in advance, "In this situation, for my purposes, X will count as a warning of possible failure and Z will count as a definite failure, et cetera." What would be plain stupid is (a) to say "In all situations, for every purpose, X will count as the mark of success, etc." or (b) to suppose that I need a universal of that sort in order to make or justify my judgment on a particular occasion. As a bureaucrat you're supposed to follow orders and not do anything for which you could be held personally responsible. As a citizen I have to make decisions and take actions for which I am personally responsible. You live a very sheltered life.

But also, if the situation calls for it, I am quite willing to say after the fact that I was mistaken in using X as the mark of success because, as it turned out, there were identifiable factors in the situation which made X unsuitable as a mark of success. Both that decision and the decision in advance to count X as a mark of success are subject to challenge and justification. I wouldn't try to evade the responsibility by quoting regulations and pleading that I was only following my orders.

E. In a real world setting the effects of a given influence, e.g., mental health treatment, are inevitably confounded with an unknown number of important events and other influences. Changes from before to after cannot simply be attributed to a given influence. The experimental devices of random assignment to groups or extensive control of individual cases are palliatives, not solutions, and they are almost never available in field settings, since therapeutic, administrative, political, or other considerations will properly take precedence.

It is in part because the effect of such extraneous influences is subject to re-evaluation that the choice of criteria is also subject to re-evaluation. The simulation paradigm provides an explicit framework for taking such influences into account insofar as we actually can, both before and after the primary data collection. By doing so it sensitizes us to this kind of possibility even when we do not represent the "influence" in question in our simulation. For example, I might designate "getting and holding a job" as a mark of improvement, and count a particular set of clients as not having improved because they do not meet this criterion. But if economic conditions have substantially worsened and unemployment is substantially higher now than previously, am I just being arbitrary in deciding to give that 'criterion' less weight now? Should I have foreseen this possibility in the original selection of criteria? (Recall the issue, above, of taking all possible precautions in advance.)

Perhaps it is easier now than previously to see the danger of talking in terms of "criteria" here. The term suggests both something

separate from the phenomenon and something foolproof, and the combination creates problems. (a) We do not normally think of a phenomenon as its own criterion, so that reference to a criterion suggests that we have to go by something other than the phenomenon of interest. But what is at issue is whether the clients have improved. If getting and holding a job is involved in having improved, it would make some sense to speak of that as a "criterion" but then we might better just speak of their having improved. If getting and holding a job were only empirically connected to having improved, it would be misleading to call it a criterion. In Descriptive Psychology we will generally refer directly to the phenomenon or to (mere) indicators rather than to "criteria," and the choice among them will codify the nature and extent of our willingness to change our minds. This is preferable to the illusion or pretence of certainty carried by "criterion."

- F. Evaluative phenomena, being historical rather than universal, will, in general, be different across time and place. A treatment procedure which is effective here now is likely not to be effective somewhere else (different clients, different problems, different milieu) or here ten years later (different staff, changing times leading to changing problems, etc.). Thus, the utility of a given evaluation is limited. In an organization, evaluation is likely to be required on a continuing or periodic basis.
- G. All of the foregoing were phrased in terms of the evaluation of "improvement." Corresponding statements could be made in respect to other evaluative concepts such as "adequate functioning," "psychopathology,"

"mental health," and "life problem." There is no set of behaviors, objective indicators, et cetera which is logically equivalent to any of these evaluative phenomena.

- H. There is no set of procedures which intrinsically or necessarily qualifies as performing an adequate program evaluation. The decision as to whether a given set of procedures does so qualify on a given occasion requires human evaluative judgment. The adequacy of a program evaluation procedure provides the same general problem of evaluation as does the degree of improvement of clients under treatment.

Consider the case of new programs for children and elderly persons initiated by an urban mental health center. The systematic evaluation described below is designed to be responsive to the general considerations noted above and to be of practical value for a variety of purposes, including decisions regarding changes in program procedures and decisions regarding allocation of resources to different aspects of the Child and Elderly programs.

A. Evaluation Structure

The overall evaluation has three major components, which are here designated as predictive, procedural, and retrospective.

1. The predictive component resembles the classic outcome of treatment design in which indicators of improvement are designated initially and are assessed before and after treatment.
2. The procedural component is related to traditional "process" studies of psychotherapy in that it involves a process description of the

"treatment" or "program" and an evaluation of its appropriateness independent of outcome.

3. The retrospective component is a closure-achieving procedure. It involves:
 - a. An evaluative review of the original assessment of the treatment planning;
 - b. An evaluation of improvement unrestricted by the predictively designated indicators of improvement; and
 - c. An evaluation of the extent to which improvement or lack of improvement could be attributed to factors other than the treatment program.

B. Evaluation Procedures

The following is a narrative outline of the procedures involved in the evaluation. The integration of treatment and evaluation is such that all but the last section (the retrospective components) is also essentially an outline of the treatment process.

1. Intake

In this phase, an intake worker obtains standard background information and a statement of the problem, and makes a routine assessment of personal characteristics, including behavioral tendencies and personal resources and deficiencies.

2. Treatment Planning

An ad hoc treatment team is formed for the client. The treatment team will include at least (a) the primary therapist or caregiver, (b) another clinically knowledgeable person, (c) the continuity of

care person, and (d) for designated cases, a member of the evaluation team. The treatment team reviews the assessment information and does the following.

- a. Requests further assessment, if needed
- b. Judges the degree of need for treatment ("severity")
- c. Decides on strategy and type of treatment
- d. Specifies the nature of the problem
- e. Specifies the particular ways the problem is manifest ("ad hoc indicators")
- f. Specifies prima facie indicators for improvement. It is primarily the "severity" rating and these indicators which appear in the predictive "before and after" analysis. Among the indices which are likely to be used are those below. Note that these will in general be different for each client.
 - (1) Specific achievements, e.g., expresses affection toward her son, establishes a friendship with someone, learns class lessons without disrupting class.
 - (2) Changes in personal characteristics, e.g., becomes more responsible, less passive-aggressive, more tolerant of other people's shortcomings, less anxious, etc. Along with the specification of such changes is a specification of the preferred way of establishing these changes, e.g., therapist judgment, self-report, standardized test, special interview.
 - (3) Judgments by two significant figures in the client's life in regard to either characteristics or changes. Among such possible significant figures are the client, the therapist, a

family member, an employer or teacher, a friend, a spouse or spouse-equivalent, an agency representative. Part of the specification here is what judgments on the part of those persons would confirm the evaluation of improvement. (In light of the introductory comments, it may be noted that sometimes the confirming indicator would be a negative judgment by a significant figure.)

(4) The length of time required for effective treatment.

- g. Specifies, if possible, a set of criteria or prima facie indicators for classifying treatment as successful and for terminating treatment accordingly.

3. Treatment Review

At least once during treatment (perhaps at an interval of one-third the time estimated for effective treatment) the treatment team reviews the treatment procedures. The review is based on the treatment plan and the therapist's progress notes together with any supplementary information which may be available from the evaluation team. At this time, the treatment team decides (a) the degree to which treatment is in accordance with the plan, (b) the degree to which the treatment now appears to be appropriate, and, if indicated, (c) how to proceed with a modified treatment. The evaluation, together with a parallel evaluation by the evaluation team after termination constitutes the procedure component of the evaluation.

4. Termination Evaluation

This evaluation is conducted by the treatment team and may be initiated by either the primary therapist or the continuity of care per-

son. The evaluation is based on the treatment planning and review data, summaries by the primary therapist and the continuity of care person, and assessment (of improvement indicators) data provided by the evaluation team member. The treatment team makes a new rating of need for treatment, reviews its previously formulated criteria for success, and recommends termination or continuation. In the latter case the procedure is repeated when termination is again in question.

5. Follow-up

At a suitable interval after termination, normally six, nine, or twelve months, the indicators of improvement, or a subset thereof, are again assessed. A comparison of this data with the termination assessment provides a descriptive characterization of the stability of improvement.

6. Retrospective Review

The retrospective review is conducted by an evaluation team which is formed ad hoc for each client. Only the continuity of care person will be common to the treatment team and evaluation team. The evaluation team will include at least (a) a member of the evaluation staff, (b) the continuity of care person, and (c) a mental health professional who would be competent to serve as the primary therapist for the client in question. The evaluation team will have access to all the information available for the given client. The evaluation team makes judgments in regard to the following.

- a. The appropriateness of the assessment procedures and of the diagnostic conclusion or problem formulation.
- b. The adequacy of the treatment plan.

- c. The validity of the improvement indicators.
- d. The degree of need for treatment at termination (independently of the treatment team decision and the predictive criteria of of success and indicators of improvement).
- e. The degree of improvement shown.
- f. The extent to which improvement or lack of improvement is attributable to treatment or other influences.
- g. A post hoc reformulation of problem formulation and treatment plan. ("If we'd known then what we know now, . . .") for reference and future practice.

C. Integration of data

Each of the before and after indicators of improvement will be obtained directly in quantitative form or will be transformed into quantitative form, so that the before and after differences may be evaluated as to their statistical significance by means of, e.g., t-tests. Since these are only indicators rather than genuine criteria, these results will be only suggestive rather than decisive (fortunately, since they are likely to be contradictory).

The pattern of results which would most clearly support an overall evaluation of "improved" or "successful" would be the following.

1. All the improvement indicators show a statistically significant positive change.
2. Severity ratings show a change from "needs treatment" (to some degree) to "doesn't need treatment."
3. Either all significant figures agree that improvement has taken place or else the pattern of disagreement among these judges is predicted and explained adequately.

4. The treatment plan was implemented and appropriate, as rated in the treatment review procedures.
5. Retrospective review specifies that
 - a. Assessment and treatment were appropriate.
 - b. The client improved and was not in need of treatment at termination.
 - c. The client's improvement is attributable primarily to treatment rather than to other influences.

The major likelihood, of course, is that the results will be more or less equivocal rather than conforming to the "ideal" case above. Thus, the final evaluation problem will be how to count different patterns of less than completely unequivocal results. No rigorous general principle or procedure for accomplishing this evaluation is possible. It is possible, however, for the evaluation committee to formulate ad hoc procedures for generating composite "degree of success" indices for particular purposes. (This is in accordance with the principle of limiting use of results as a way of increasing the likelihood of agreement.) It may also be possible to specify general categories of data which would have utility across different evaluations. Such categories would be useful for summarizing the results of several evaluations or for comparing them.

D. Modifications

The general structure and procedure described above are primarily designed for the evaluation of improvement of individual clients undergoing comprehensively planned treatment for a substantial period of time and under conditions ideal for evaluation. Certain modifications would be

called for under other conditions. Among the modifications of present interest are the following.

1. Conditions for evaluation are almost never ideal. Because of the amount of professional time involved in the evaluation procedure it almost certainly will not be feasible to carry out all of the procedures with every client. In that case it would be appropriate to select every nth client at random for the complete evaluation procedure. A ratio of one in five or one in ten would seem to be reasonable in this regard.
2. Some telescoping of the procedure would be required in cases where treatment was not continued for a substantial period of time (e.g., on site treatment or early termination).
3. The evaluation of the Child and Elderly programs has both qualitative and quantitative aspects. The evaluation design described above is primarily responsive to the qualitative aspect; if clients improve, then, qualitatively, the program may be judged to be a success.

To a large extent this evaluation is an exemplification of the Precaution Paradigm. The whole notion of evaluation research, as it has evolved, is that one needs to take precautions against (a) wrongly assuming that what one is doing is successful, (b) continuing to invest resources in activities which are ineffectual, (c) failing to take advantage of opportunities to improve performance, or (d) other ways of going wrong, or conversely, that one ought to proceed with some assurance that one is doing it (whatever "it" might be) correctly, completely, successfully, et cetera.

Thus, for example, the details of intake and treatment planning (1 and 2, above) reflect the following concerns, precautions, and reassurances.

- (a) The initial assessment provides some assurance that treatment planning can proceed in accordance with local norms. Planning treatment at all is a precaution against providing inappropriate treatment.
- (b) The composition of the review team provides some assurance that the relevant interests and points of view (including those of the evaluator) will be reflected in the decision making.
- (c) The procedures of the treatment team constitute several precautions against inappropriate treatment.
- (d) The procedures of the treatment team provide some assurance that data needed for evaluation will be available.
- (e) The ad hoc character of the team (constituted specifically for each client) is a precaution against the bureaucratic pressures and tendencies toward uniformity. Procedures (f) and (g) assure that data for the prospective study will be available and provides some assurance that the relevant observers (including the care-givers) will be sensitized to recognize improvement or lack of improvement and that, correspondingly, their eventual judgments in this respect are less likely to be capricious.

In the review procedure (3, above) Precaution and Simulation Paradigms are both reflected.

- (a) The introduction of a review and its timing in the treatment process constitute a precaution against (1) letting inappropriate or ineffective treatment continue longer than necessary and (2) failing to take advantage of opportunities to improve effectiveness.

- (b) Reviewing the process provides some assurance that the conceptualization of the way in which improvement could be expected to take place was sufficiently cogent and accurate to warrant the confidence placed on it.
- (c) The use of the Simulation Paradigm would be an explicit and systematic elaboration of the preceding point. The process in question would, say, be the process of therapeutic interaction. Various classes of Versions, including those corresponding to successful and unsuccessful outcomes could be distinguished, and therapeutic and non-therapeutic contingencies formulated, and so forth.

The termination evaluation procedure (4, above) provides some assurance that the various relevant interests and viewpoints will be reflected in the termination decision. The reference back to the previously formulated 'criteria' for termination is a precaution against capricious decision making (for example, it recalls for the team the *basis for* choosing those 'criteria'). Reference to the process review provides some assurance that the apparent improvement is consistent with the reasons for expecting it. The composition of the team is a precaution against self-serving evaluation or merely self-fulfilling treatment planning. ~~The~~ follow up (5, above) is a precaution against premature judgments of sufficient improvement.

The retrospective review (6, above) is a precaution against making or having made incorrect or less than optimal decisions with respect to the client or with respect to the evaluation. It also provides some assurance that experience with this client will contribute rationally to future practice. (Appeals to sampling theory and 'generalization')

for this purpose in this context would generally be merely bizarre.) It provides some assurance that all the relevant facts, and not merely those which were initially anticipated, are taken into account in arriving at conclusions which have a rational prescriptive value (in the Critic sense) for further action.

The evaluation research practitioner who is thoroughly steeped in dolce academica will v^ery likely protest that I have failed to distinguish 'carefully' between the program and the evaluation. That is not accidental. I have tried to illustrate the general principle that the difference is one of function rather than of procedure as such and that in the limiting case there may be complete overlap between the two. Our social practices did not evolve and are not exemplified in an ecological vacuum, and our participations in them have no a priori guarantee of success. Consequently, our participations in those practices do not, in general, consist of plunging blindly ahead (pure Actor functioning), so that a very special and categorically distinct set of activities is required to set us on the right path or keep us on it. Rather, precautions, assurances, and self-regulation are general prerequisites for normal human behavior. If prudential considerations often call for a greater degree of explicitness and systematization than is "built in," that could provide a *raison d'etre* for a distinctive discipline called "evaluation research." But the rationale for evaluation research lies in the logic of precaution-taking and assurance-seeking and self-regulation and not in the transcendental visions of sampling theory, 'experimental design,' et alia.

The research design described above is, at the given level of description, more rigorous than any actual evaluation research known to me and more rigorous, flexible, and resource-full than commonly accepted

examples and norms for "good practice" in evaluation research. For example, most evaluators would not consider the possibility of separate prima facie indicators for each client, either because that would violate a current interpretation of "objectivity" or because it is assumed not to be situationally feasible, et cetera. Too, most "good" evaluation designs do not include the retrospective component, probably because that would run counter to the predictive conventions of 'experimental design' or 'the scientific method.' And again, current evaluation technology allows for multiple outcome measures only insofar as they overtly agree; genuine and legitimate disagreement is so far a methodologically unmanageable concept. In contrast, the methodology of disagreement is a routine aspect of Descriptive Psychology.

But, to repeat, the point of the present report is not to contribute new techniques to a field of endeavor already burdened with a procedural orientation and a swollen grab-bag of procedure-schemas. Rather, it is to show how, at least within a Descriptive Psychology framework, it is possible to be methodologically grounded and procedurally justified. There is something beyond ad hoc and shamanistic appeals to authorities, "the literature," "the scientific method," or "customary good practice in evaluation research." To be sure, many current practitioners would say they see no value in such "philosophizing" unless it pays off in new procedures and statistically significant results. But it is characteristic of Descriptive Psychology to be concerned with such matters.

7.00 20

TABLE 2 *Basic Process Unit (BPU)*

P-NameA:	The process "Name" of process A.
P-DescriptionA:	The "Description" of A. It specifies:
I. P-Paradigms:	The major varieties of P-NameA. This is a technical option. If only one paradigm exists, it will be the same as P-NameA. For each paradigm, the following is specified:
(a) Stages I-K:	These are "Names" of subprocesses within A. They are systematically specified, e.g., as P-NameA11, P-NameA12, . . . , P-NameA1K for Paradigm 1. For each stage, specify:
(1) Options 1-N:	These are the various exemplars of the process (stage) in question. That is, these are the various ways in which that process could happen. Each Option is systematically indexed as P-NameA111, P-NameA112, . . . , P-NameA11N. Each of these can now be expanded (decomposed) on the model of P-NameA.
(b) Individuals	
(c) Elements	
(d) Eligibilities	
(e) Contingencies	
(f) Versions	

Then P-NameA is t
so on.

In contrast, Indi
gencies are designed
process P-NameA. B
related processes as
may also have other
only by virtue of Ru
ents, if only by virtu
involve object consti
and stand in certain
should be the same
already defined by re

Since the process
stituent processes, t
(which might be sys
in our ordinary way
of some subprocess.
the concept of "pr
structure.

For example, if I
a set of subprocesses
hand; (2) he point
trigger of the revol
wounding the bear.
stituents of the pro
hand, (c) the revol
bear. These would l

Certain relation
ample, "raised the
structure (a state of
and the revolver. It