EVOLUTION OF METHODOLOGIC STRATEGIES AND TACTICS FOR FUNCTIONAL ASSESSMENT OF RETARDED BEHAVIOR

Beatrice H. Barrett

Working draft, prepared December 1980.

The decision to base our work in a laboratory instead of a clinical setting arose from observation of clinical practice with retarded clients and from the demonstrated clinical relevance of free-operant laboratory methods.

PREVAILING CLINICAL CLIMATE

Clinical practice, under the aegis of psychiatry, placed retardation low on its lists of interests. Devoid of the rich theoretical applications found in other clinical groups, people with low I.Q. scores offered meager training material in the psychodynamically oriented mental health world. After all, if there is little ego development, defense mechanisms are unnecessary. Projective tests yield barren protocols from those with little or no verbal repertoire. All that was applicable were intelligence tests. Designed to bifurcate the school population into retarded and nonretarded, these instruments could do little more than classify. Comparisons were limited because different scales were used depending on age and severity of handicap. Often the necessity of an infant test or a scored structured interview with an informant, e.g., Vineland Social Maturity Scale, automatically classified a non-infant as "untestable" by conventional procedures. Under this circumstance, a habilitative program was rarely considered. The common recommendation was placement in an institution where custodial care would be provided.

Regardless of classification "level," the single-score yield of standard intelligence tests masked individual differences within protocols. Homogenization, compounded with the inferential nature of the test score, produced little or no information on which to base an individually tailored habilitative plan. More importantly, the one-shot test, presumed to lose sensitivity if repeated too often, gave no opportunity for potential performance improvement to be shown.

EMERGING BEHAVIOR-ANALHTIC TECHNOLOGY

Free-operant conditioning methods offered a striking--and challenging--contrast. The earliest applications to human behavior analysis had demonstrated their unique sensitivity to individ .lly different behavior patterns among people with chronic psychosis (Lindsley 1956. 1960). Moreover, these methods had revealed previously unseen learning tential among mute autistic children (Ferster & DeMyer, 1961).

With topographically simp e movements, complex behaviors could be taught and analyzed. Originating in la ratory studies of animal behavior, the methods were culture-free and required n verbal behavior. Automated programming and continuous, automatic recording p ovided direct, permanent tracings of a subject's behavior without any of the usual hum n sources of unreliability. Moreover, single-subject experiments could be conduc ed over long periods of time without the sensitivity loss associated with other forms f repeated measurement. Drawn directly by a subject's interaction with an automatic lly programmed apparatus, the resulting records were unaffected by tester bias, te cher fatigue, or scaling transformations. The subject's behavior spoke for itself.

Beyond choice of methodo ogy, major strategic decisions governed what we attempted to do with it. In thi regard, an overview of other applications at the time is relevant. In formulating he strategies and tactics that guided our applications of laboratory operant condit: ning, we were greatly influenced by trends apparent in the then-small world of huma behavior analysis. Relatively little had been published at the time we began. Howe r, we obtained a good view of the field through reports and contacts at B.F. Skinne s biweekly Pigeon Lab meetings at Harvard, through correspondence with various nvertigators, through conversations with visitors to Ogden Lindsley's Harvard M ical School laboratory where we conducted our pilot studies, and through frequen and extensive visits to colleagues' laboratories. The picture that emerged reveale a sizable gap in which our work could begin to take shape. Briefly, it appeared s follows.

Careful programming of e perimental environments had already successfully trained primates and pigeons o perform tasks usually thought to be too complicated for lower organisms. Operating equipment in space vehicles (Rohles, 1966), performing binary arithmetic (Ferster & Hammer, 1966), guiding missiles to their targets (Skinner, 1960), detecting faulty diodes (Cummings, 1966) and imperfect pills, (Verhave, 1966), demonstrating conceptual skills (Herrnstein & Loveland, 1964), and engaging in the dyadic interactions of games such as table tennis (Skinner, 1962) had shown the power of operant methodology in synthesizing skills never before seen in lower organisms.

With the drama of these successes in animal laboratories, it was not surprising that the new skill-synthesizing technology would be applied to ameliorate deficits and deficiencies in human behavior, especially in those nonverbal individuals for whom psychiatry had little to offer. Acquisition of token-exchanging and matching-to-sample had just been demonstrated in long-term, intensive experiments with autistic children (Ferster & DeMyer, 1961). Subject-paced sequences of changes in the experimental environment had demonstrated that "retarded" children could rapidly acquire discriminated free-operant behavior (Bijou & Orlando, 1961). An analysis of some component behaviors in free-operant successive discrimination had also begun (Orlando, 1961). These studies, along with early explorations of schedule effects in institutionalized retarded children (Ellis, Barnett, & Pryer, 1960), demonstrated the sensitivity of free-operant methods to individually different behavior patterns.

But the time was right for demonstrations of behavior normalization. And such demonstrations were highly reinforced. Operant investigators were departing the laboratory for the more visible settings of classroom (Birnbrauer, Wolf, Kidder & Tague, 1965) and residential institutions (Spradlin & Girardeau, 1966). With few exceptions, what remained in the laboratory was increasingly fine tuning of the experimental environment to obtain the stimulus control necessary for "errorless" learning (Sidman & Stoddard, 1967). Based on principles derived from Terrace's (1961) experiments with pigeons, the methodology of stimulus control was being modified to the subject-paced format of programmed instruction. With trial-by-trial presentation similar to Ebbinghaus's memory drum, frame-by-frame error analysis was replacing analysis of freely emitted operant response frequencies. Frames (trials) which occasioned subject errors were viewed as a signal of faulty design. Sequences and/or stimulus arrays were revised to remove opportunities for mistakes so that subjects could progress without error toward criterion behavior.

. Regardless of the setting--whether laboratory, classroom, or ward--the more rapid the attainment of terminal behavior, the more successful the stimulus control techniques.

Reappearance of the ubiquitous trial (Shulman, 1970) and growing emphasis on refinements of errorless learning methodology had a number of spin-off effects on applications of operant methods in human behavior research:

<u>An individual's behavioral variability, shown as fluctuations through time, was discarded as a subject for analysis</u>. Well-honed stimulus control techniques showed their success within minutes. Pre- and post-tests showed the magnitude of effect. The temporal dimensions of human behavior became irrelevant and long-term analysis unnecessary. The methodology of stimulus control did not permit temporal fluctuations to be revealed.

<u>Transitional or prerequisite states shown by temporal changes in patterns of concurrent behaviors were being superceded by linearity of change in a single "correct" behavior which altered the stimulus array (presented in a different frame) on each occurrence</u>. With rapid linear attainment of terminal behavior as the criterion of programming adequacy, differential patterns of "correct" and other behaviors that might otherwise occur during the course of acquisition were obliterated. Iterative revision of stimulus arrays to eliminate subject error also removed opportunity to observe emergent/prerequisite stages in the acquisition process. Potentially identifiable stages in a subject's progress toward criterion behavior became embedded in the sequence of stimulus array changes--the behavior of an experimenter--rather

than revealed by recorded changes in subject behavior.

Detection of individual differences became unnecessary. The perfected errorless program would have the desired generality across subjects. Substantial intersubject differences were an indication that the experimenter had not achieved rigorous control. Reminiscent of the statistician's error variance, this view of intersubject differences could be relabeled "experimenter error." Good stimulus control techniques were, by definition, powerful enough to obliterate intersubject differences.

Behavior deficits disappeared; diagnosis became unnecessary. As a refinement of rapid stimulus control techniques became the hallmark of good experimental design, defective behavior lost its status as a subject matter for analysis. The more rapid and complete the experimental control, the more convincing the demonstrations that "defective" behavior could be normalized. And because acquisition could be demonstrated in persons traditionally considered "hopeless," the concept of behavior deficits came to be questioned, even attacked. With its finely faded, graduated stimulus changes, bahavior analysis convinced experimenters that deficits were not properties of organisms but evidence of defective environmental contingencies. Behavior deficits were redefined as attributable entirely to the experimenter, not the subject. It became unpopular to analyze the putative deficits of one's subjects because, in so doing, one would expose one's own experimenter deficiencies. Furthermore, "deficits" and "diagnosis" were tainted by association with the "disease" substrate of the medical model.

Rate of responsing was being abandoned as the basic datum of the science of human behavior. Too sensitive to "extraneous" variables, it was considered unsuitable in the quest for rapid stimulus control. The controlling relation between a response and a stimulus array or element thereof would only be muddied by the natural variability of freely emitted behavior.

Ironically, by the time new engineering technology produced a "silent high speed" cumulative recorder, subjects' fluctuations in behavior frequency had almost disappeared from laboratory study. Frames that did not occasion correct subject responses were omitted or revised. Sequences were either completed or not. A new version of the check list was emerging as the datum of human behavior analysis! The delicate sensitivity of Skinner's most important contribution was no longer required or desired. Its continuous recording functions were either applied to apparatus monitoring or replaced by noncumulative event recorders. Behavior change over time was being quietly phased out of the human behavior analysis domain.

Freely emitted behaviors were disappearing from laboratory analysis. Behaving choices were being restricted. With trial-controlled programming, which presented a new stimulus array after each response and which determined session duration by a preset number of frames (trials), subjects had fewer options. To complete the sequence was the objective. Each frame required a response. To ensure that competing behaviors would not produce long latencies, subjects were being pretrained to return the responding hand to the lap and sit quietly between trials. And these intertrial intervals were being shortened by faster-than-relay programming circuitry.

Subject exclusion was becoming popular. Reminiscent of classroom exclusion and the "easy subject" selection of instrumental conditioners, investigators set limits on the behaviors tolerated in their laboratories. The new stimulus-controlling apparatus was not built to withstand or measure disordered/assaultive behavior. So the one-response per-trial paradigm was limited not applied to subjects who could not be readily taught to use apparatus with care.

For our exploratory assessment purposes, however, the properties of Skinner's original measurement system were uniquely suitable and desirable. We incorporated as many of its capabilities as was feasible without computer-assisted programming and analysis. We had no way of knowing in advance what significant behavioral aspects of retardation might emerge. We wanted as many options for analysis as could be practically handled, for we saw no point in trying to outguess our subjects.

GUIDING METHODOLOGIC STRATEGIES

Laboratory environments designed for free-operant behavior analysis share common properties: 1) an enclosed space that determines where measured behaviors occur, 2) one or more devices, operable only by the behaviors under study, that provide ways for each participant to manipulate the immediate environment, 3) automated apparatus programs that present environmental events and arrange their relationship to measured behavior, and 4) a continuous, automatic recording system that separately quantifies and graphs the rates of each behavior the environment is designed to analyze. These are the properties of behavioral conditioning environments that have yielded sensitive and reliable data from organisms of many countless species.

Our application of behavior-analytic environments was guided by four interrelated methodologic strategies formulated early in the course of our work: individual differentiation, functional calibration, universal applicability, and prescriptive utility (Barrett, 1965).

Individual Differentiation

For optimal effectiveness, habilitative procedures must be designed for and selectively applied to specifically defined behavioral characteristics of individuals. When our work began, only the qualifying adjectives of our imprecise verbal system were available to describe behavioral differences among retarded persons--differences that had been masked by the global categories of norm-referenced inferential assessment. To delineate reliable individual patterns of abilities and deficits that can and should guide development of more precise habilitative endeavors, behavior-analytic environments must be sensitive enough to reveal individual differences and finely controlled enough to permit analyses of an individual's behavior.

Until we can quantitatively describe treatment-pertinent characteristics of an individual's behavior, assessment will continue to be limited by the vagaries of our verbal descriptions. And diagnosis of retardation will continue to be a process of homogenizing via the labeling that results from conventional group-comparative assessment procedures.

Functional Calibration

Extent of deviation from average may distinguish a clinical group. But when such deviation is defined only on the basis of a summary test score, it yields a comparison too gross to be helpful in fashioning effective individual treatment. A more useful approach accepts the predefined clinical group, then sensitizes an assessment system to reveal intragroup behavioral deficits or deficiencies that might become functional habilitative targets. This approach, in effect, tailors procedures to the clinical group itself rather than standardizing them on the normal population from which deviation has already been acknowledged. This assessment strategy does not preclude comparison with an average or normal group.

Accordingly, much of our work was devoted to determining an individual's degree of efficiency along a series of habilitatively relevant behavior dimensions, defined empirically by the measured behavioral extremes of the retarded population. Habilitative success could then be judged with respect to each person's progress along multiple dimensions relative to that person's status at the inception of treatment. This approach, called multiple baseline design, is one of the most powerful designs for evaluation of treatment effects on a number of individual behaviors as well as environments that were as useful with assaultive, self-injurious, incontinent, and nonverbal people as with the best behaved and most articulate of the retarded.

Prescriptive Utility

We also sought to develop methods that would yield reliable data on which to base individually tailored habilitative prescriptions. A single sample of a person's performance on a series of tasks yields only a static quantification of competence under whatever conditions were in effect during the test. A more informative assessment would include 1) the degree of competence under a standard set of conditions, 2) a functional description of performance changes in response to specific manipulations of the "test" conditions, and 3) the durability of changed performance following effective remediation. Records of the environmental conditions under which a behavior occurs or does not occur permit a functional, rather than topographical, description of the individual's behavioral assets and deficits. That is, such records permit a description of behavior in terms of what it does rather than what it is.

Some basic premises determined our selection of tactics. 1) Assessment should serve a prescriptive function; that is, it should determine or facilitate selection of effective treatments. 2) To facilitate selection of efficacious treatment, assessment should reveal a variety of potentially treatable deficits and deficiencies. 3) Treatment procedures may have both positive and negative side effects. 4) Assessment should reveal treatment-produced changes. 5) Adequacy of treatment may be evaluated by the extent to which it produced more "normal" behavior patterns both during its course and following its termination.

METHODOLOGIC TACTICS TO FULFILL STRATEGIC AIMS

Sensitivity to Differences in Individuals' Behavior Characteristics

Automated programming ensured uniformity of environmental events for all subjects by eliminating human errors resulting from boredom, fatigue, or unwitting prejudice.

Direct, continuous, automatic recording of ongoing behaviors not only eliminated problems of measurement validity, reliability, and objectivity, but produced permanent tracings of moment-to-moment changes in each subject's responding. Simultaneous recording of multiple behaviors also revealed temporal interactions, or behavior covariations, obscured by summary measurement (Barrett, 1977).

Standard session duration and uniform screening contingencies remained in effect until each subject's behavior reached its own level of session-to-session "stability." These tactics revealed individual differences in temporal dimensions of acquisition.

Contingencies permitting free emission of behaviors allowed the widest possible range of individual patterns to emerge. "Errors" or nonreinforced behavior could occur without penalty, concurrent with "correct," or reinforced behavior. The relationships between the two constituted the patterns of interest for further analysis.

Repeated measurements permitted each subject's behavior pattern to stabilize, thus permitting subjects to reach their highest level of accuracy.

For individual differences to emerge, intensive analyses of each subject's data, rather than extensive statistical analyses of group data, were undertaken.

Functional Calibration

Sensitizing our methods to a broad range of behavior deficits and deficiencies was approached through subject selection and choice of behaviors to be recorded and analyzed.

Selection of subjects. To examine a broad spectrum of behavioral deviation, we chose subjects already classified as retarded. As residents of a state institution, they represented a degree of deviation found unacceptable by community standards. Comparable behavioral data were already available from "normal" persons, but no data had been obtained from a broad range of the institutionalized retarded population. Ironically, the "higher level" residents occupied domitories "on the hill," nearer the schoolhouse! "Low level" residents were assigned to dormitoried "down the hill." Down-the-hillers were not offered education or training programs. Higher level people--if there were no behavior problems--were more likely to be involved in schooling, such as it was. Ward employees' statements about a resident's behavior were critical in determining eligibility for schooling.

While our sample was biased toward the younger residents, subjects came from a wide range of dormitories and represented the full range of psychometric scores, from "untestable" to what would not even be considered retarded today because of redefinition by the American Association on Mental Deficiency. Building matrons' willingness to cooperate by having the residents ready for their appointments and building physicians' clearance on residents' medical status also played a role in subject selection, as it did in decisions about resident inclusion in other institutional activities.

Our aim was to develop a pool of subjects representing the selection standards of the institutional subculture. For the purpose of delineating behavioral extremes as contrasted with norm-deviation extremes, functional criteria outweighed statistical criteria.

Guided by the notion of dehomogenization, we started with a broad selection of subjects but increasingly loaded out subject pool with those considered to be severely and profoundly retarded. Our aim was to look for a) comparison groups defined by functional institutional criteria and b) behaviorally defined groups that may or may not have shared similar psychometric or other institutionally applied criteria.

Choice of behaviors. We further attempted to sensitize our methods to behavioral extremes by measuring a range of behaviors and by selecting a variety of behavior measures for analysis. In exploratory work such as ours, the greater the range of measurement capability, the greater the probability of finding constellations that define individual "behavioral signatures."

Simple free operant, differentiated operant, discriminated operant, simultaneous differentiation and discrimination, and nonreinforced concurrents such as locomotor pacing, vocal stresses, and stereotyped rocking were among the behaviors we studied. Measures included hourly rates; intrasession rate changes; the number, duration, and distribution of pauses and spurts in responding; and relative rates of reinforced and nonreinforced responding (an index of accuracy or efficiency).

Tactics for Universal Applicability

To make the individually differentiating, functionally calibrated system universally applicable as well, we used the following tactics.

Selection of naturally occurring behavior topographies. To screen a wide range of retarded population required selection of movements already in the repertoires of our subjects. We sought an assessment system that could be applied to all persons within the target clinical population. Accordingly, we chose the manual movements of pulling, pushing, and touching; the foot movements of tapping or pushing; body weight shifts (standing); body rocking (sitting); head rocking; and vocal stresses. In keeping with the strategies of individual differentiation and functional calibration, these movements are all repeatable, observable, and readily transducible--and therefore can be reliably recorded. While topographically simple enough to be executable even by most multiply handicapped, profoundly retarded people, these movements are clearly within normal behavior repertoires as well.

The elemental physical properties of these movements make them more prevalent in the retarded population than the more complicated topographies often used in traditional assessment. Their functional properties, however, could be made increasingly complex by arranging appropriate contingencies by which they manipulated the behavior-analytic environment. Recording all movements in the same measurement units provided a single system that was directly applicable to all subjects--a "yardstick" that permitted direct comparison regardless of age or degree of retardation.

Selection of subjects representing the range of retardation was another tactic that helped us develop a system in which there were no "untestables." "Hyperactive," "self-destructive," "assaultive," "incontinent," "withdrawn," "vegetative," "psychotic," "autistic," "behavior problem," "severely disturbed," are some of the labels often applied to a large portion of our subjects. We asked building matrons to select some of their "best" and some of their "worst" residents. The usual mix contained more of the "difficult" (and usually "untestable") residents than the "easy" residents. However, loading our subject pool with challenges was consistent with our aims. Ward personnel were delighted to be free of the "trouble-makers" as often and for as long as possible and, being considered "hopeless," these residents were not included in other programs and activities.

As we moved into the most challenging end of the retardation spectrum, we were, quite by accident (i.e., due to administrative problems beyond our control), fortunate to find that we had a sizable number of severely and profoundly retarded residents from two buildings that were considered by the institution to house a homogeneous group--exactly what we needed for dehomogenizing the so-called low-level, subtrainable, custodial segment of the population.

Availability of a variety of antecedent and subsequent events, as well as a range of contingencies, furthered the cause of universal applicability. Lights of different sizes, different colors, different positions; tones of different pitches; and loud buzzers, loud bells, and taped verbal instructions were some of the programmed events that could function as potential discriminative stimuli should any one or combination prove to control responding.

An array of contingent subsequent events served as potential reinforcing consequences. Included were candy, potato and corn chips, bits of cookies, colored slides, turning a light on and off, taped music, sounds of screams, the separately controllable audio and video portions of commercial television, the separately controllable televised sight and sound of a wardmate or the subject's teacher, the televised image of the subject, pennies exchangeable for dimes to operate the vending machine, tokens exchangeable for pennies for the same purpose, and access to other laboratory enclosures to operate apparatus that delivered these various consequences.

Episodic scheduling of potentially reinforcing events ranged from a fixed ratio 1 (a consequence contingent upon each appropriate response) through various values of fixed ratios and variable intervals. Taped music and television were programmed conjugately so that the content was continuously available but its intensity varied directly and continuously with the rate at which a subject responded.

<u>Basic behavioral processes</u> were selected for their universal applicability. Learning to obtain food by executing an already acquired behavior (for example, pulling a plunger) is basic enough to be demonstrated throughout the phylogenetic range. Performing the behavior that brings reinforcement and ceasing to perform behaviors that do not bring reinforcement (i.e., response differentiation) is another example. Still another is learning to emit the food-getting response only when a "food-available" signal is on (i.e., successive discrimination). All are prerequisites for higher order behaviors.

Systematically assessing the acquisition and maintenance of these components of higher-order functioning precludes the possibility of mistakenly attributing deficits to more complex processes when, in actuality, the simplest elements for their formation have not yet been acquired. Moreover, by assessing critical components we increase the likelihood that a single assessment procedure will be applicable to individuals representing the range from profound retardation to "normal."

Tactics for Prescriptive Utility

In addition to its sensitivity to individual differences throughout the range of the clinical group on which it is calibrated, a prescriptive assessment system should not only pinpoint specific problems of individuals but also provide comparative information on the relative effectiveness of various treatments for those problems, both in a given individual and in others with the same problems. With these criteria in mind, we incorporated the following tactics.

A standard multiple-solution experimental design as a screening program and as a criterion "testing" environment. Nonrestrictive contingencies revealed a range of individual skill/deficit patterns and sifted these behavior patterns into quantitatively defined subgroups. These subgroups were functionally described by the number and nature of the behavioral processes each individual acquired during the screening phase (Barrett, 1978). Earlier research (Lindsley, 1958; Barrett, 1965) showed that individual behavior patterns characterizing the subgroups were not only reliable but dispersed the previously homogenized clinical group into a range from "normal" patterns to severely deficient and defective patterns. The experimental design of the screening environment thus pinpointed a variety of treatment targets. In addition, it provided a "normal" pattern that was used as a criterion performance against which to evaluate the adequacy of the treatment effects.

Acquisition screening status as criterion for treatment selection. Development of accurate prognostic-prescriptive information depends on the effectiveness and reliability of the treatment-deficit "match." Our screening paradigm sorted individuals into deficit-defined subgroups, thereby providing treatment groups homogeneous with respect to remediation targets. We planned the same treatment variations with individuals showing similar deficit patterns. With this approach we hope to find reliable effects from a given treatment for all individuals with a given deficit pattern.

Repeated opportunities in the standard multiple-solution screening environment were offered until each individual's behavior pattern "stabilized" at that individual's best (asymptotic) performance. Thus, the behavior patterns which grouped individuals for remediation were maximum acquisition patterns for each subject. In addition to forming the basis for treatment selection, individual maximum-acquisition screening patterns constituted the "baselines" for evaluating treatment effects. This tactic permitted us to distinguish treatment effects from "practice" effects or adaptation effects within the screening environment. Continuous measurement during the repeated-performance opportunities of the screening phase permitted direct comparison of individuals' initial patterns with their asymptotic—a first step in the development of prognostic assessment that should form the basis for design of maximally efficient remedial procedures. This tactic revealed the nature of whatever changes might occur between initial response to screening contingencies and the more enduring behavior patterns eventuating from more time to interact with these contingencies. Measures of individual variability are essential in selecting treatment and evaluating its effects. Repeated measurements provide a picture of each individual's established patterns of session-to-session variability, even at asymptotic performance. To be seen as effective, treatment must produce desirable deviations from these established patterns of fluctuation. In addition, individuals with wide session-to-session variability may require more time to show treatment-associated change than those with ideal "steady state" behavior. Intersession variability revealed in repeated measurements may thus give prescriptive information on the temporal dimensions of treatment effects.

Continuous multichannel measurement of ongoing within-session behavior provided on-line recording of the latency of onset of treatment effects as well as side effects on nontargeted behaviors, both of which are critical in evaluating the efficacy of many treatment procedures.

Continuous measurement also provided on-line evidence of intrasession variability--"local" fluctuations germane to the selection of time-determined treatment contingencies such as time-out from reinforcement, conjugate rate requirement, sensitive timing of antecedents to obtain paced control of response frequency, etc. In addition, continuous measurement enabled us to test treatment procedures designed especially for those whose rapid intrasession acquisition would have been masked by summary measures.

Repeated measurement sessions, augmented by continuous measurement during each session, thus yielded both the microscopic and the macroscopic view of the temporal characteristics of treatment effects.

Functional measurement--when multichannel, continuous, and used throughout repeated opportunities in a multiple-solution environment--permits ongoing evaluation of the behavior-related properties of that environment. Behavior-environment relationships are the focus of habilitative endeavor, yet rarely are the relationships directly recorded. Our screening-criterion paradigm separately measured the relationships between operations of each plunger under each of two light-panel conditions (see schema of recording system in Figure 3). These separately measured functional relationships constituted the patterns defining deficit as well as the dependent variables for evaluation of treatment effects.

Functional measurement focuses remedial-prescriptive environmental manipulations directly on the behavior functions of interest. Thus, confusion of behavior topography with behavior functions is less likely to confound the evaluation of remedial effects and, in turn, less likely to limit the empirical validation of prescriptive utility.

The operant equation as a heuristic device in selecting and evaluation behavior prescriptions. By definition, functional analysis distinguishes the manipulated physical properties of environmental events from their functions in altering or maintaining individual behavior. A behavior prescription specifies one or more changes in the environment to effect desired change(s) in client behavior. The ingredients of the prescription describe these changes in phsycial terms. The process of validation then measures the functions of these physical changes in producing the more "normal" behavior.

To aid in determining and evaluating the ingredients of a behavior prescription, we have found it helpful to assign a separate set of terms (Lindsley, 1964) to the physical and the functional components of operant behavior. The apparatus--the console with its plungers and lights and its surrounding enclosure, as well as the program which schedules both the alternation of lights and the delivery of goodies--can be described as consisting of four components:

--antecedent events (panel lights and their timing),

--movements (plunger pulls) which operate the apparatus,

--events which follow operations of the apparatus, (candies, tokens dropped into tray),

--the arrangement for delivery of the subsequent events (e.g., every tenth pull of left plunger when left light is on).

Acquisition is the behavioral transition that demonstrates conversion of physical components into functional components. Continuous measurement throughout the transition verifies that conversion.

Any one or more of the physical components of the immediate environment may be nonfunctional for a given individual. The repeated opportunities of screening let each of our subjects' measured behavior transitions show us which components of the physical environment are functional and which must be modified to become useful for that person.

When physical components become functional either through the process of acquisition or through remedial conversion, their designations become more familiar. Antecedent events become signals or cues or stimuli; movements become responses; subsequent events become accelerating or decelerating consequences and the arrangement for their delivery (occurrence) becomes a contingency.

Use of commonly available agents. To increase the likelihood of prescriptive findings that might apply in habilitative work outside the laboratory, we programmed potential reinforcing and discriminative events that closely approximate those commonly available in habilitative environments. For potential reinforcers we used food, tokens, projected slides, taped music, the video and auditory channels of commercial television, and televised self-images (the "mirror" one works to see). In addition to panel lights which differed in position, size, or color, we included as potential discriminative stimuli tones that shifted in frequency and verbal requests that were both taped and personally delivered. This choice of antecedent and subsequent events afforded comparison of both the discriminative and reinforcing functions of visual and auditory modes as well as social-nonsocial content for a given individual. While "fading" various dimensions of antecedent events to develop their discriminative function was technically feasible, we chose to use simple presence-absence and discrete, very obvious (to us) changes more like those found in the natural environment--and more like those that could be easily applied without sophisticated apparatus.

Sequencing the restrictiveness of potential remedial variables. To reveal individual differences in remediation requirements, the order of environmental variations introduced for remedial testing began with what we considered to be the least powerful or least restricting. Least restricting describes those variations which maintained the open or nonrestrictive contingencies necessary to reveal individual differences in the patterns of behavior that could produce programmed events. Increasingly restrictive contingencies and penalties were tested only after other procedures had failed to show desired behavior change(s). We were less interested in obtaining rapid effects than in augmenting our functional descriptions of behavior deficit severity to include the extent of restriction or amount of "control" necessary to normalize them. Distinguishing the necessary from the sufficient was essential to this endeavor.

Development of prescription/prognosis testing environments. The predictive validity of assessment findings and prescription effects in extra-laboratory settings is germane to the utility of laboratory prescriptive diagnosis. From the beginning we had hoped that our work would be useful in improving the care and treatment of our subjects. Toward this goal we considered a laboratory-practice feedback loop as a necessary vehicle for testing and furthering the clinical relevance of our methods and findings as well as for training of habilitative personnel in their application.

As a first step, we sought ways of obtaining validation information on our subjects' behavior in extra-laboratory environments. Broadly conceived as representing decreasing degrees of environmental control, we planned to develop data-based feedback systems from two other environments: a classroom specifically for our subjects and the wards where they lived. In these settings we hoped to determine the predictive or prognostic value of obtained individual differences in acquisition time, acquisition subgroup status, amenability to change during treatment, durability of post-treatment gain, relative reinforcer power, relative effectiveness of auditory and visual presentation modes and of social and nonsocial components of environmental events.

As an application-testing and training arena, we initiated in late 1965 what was, to our knowledge, the first daily operating classroom for severely retarded "custodial" children. Its philosophical basis, structure, curriculum, and data base were described in an earlier report (Barrett, 1971).