

SOME GENERALIZATION AND FOLLOW-UP MEASURES ON
AUTISTIC CHILDREN IN BEHAVIOR THERAPY¹

O. IVAR LOVAAS, ROBERT KOEGEL², JAMES Q. SIMMONS, AND
JUDITH STEVENS LONG³

UNIVERSITY OF CALIFORNIA, LOS ANGELES

We have treated 20 autistic children with behavior therapy. At intake, most of the children were severely disturbed, having symptoms indicating an extremely poor prognosis. The children were treated in separate groups, and some were treated more than once, allowing for within- and between-subject replications of treatment effects. We have employed reliable measures of generalization across situations and behaviors as well as across time (follow-up). The findings can be summarized as follows: (1) Inappropriate behaviors (self-stimulation and echolalia) decreased during treatment, and appropriate behaviors (appropriate speech, appropriate play, and social non-verbal behaviors) increased. (2) Spontaneous social interactions and the spontaneous use of language occurred about eight months into treatment for some of the children. (3) IQs and social quotients reflected improvement during treatment. (4) There were no exceptions to the improvement, however, some of the children improved more than others. (5) Follow-up measures recorded 1 to 4 yr after treatment showed that large differences between groups of children depended upon the post-treatment environment (those groups whose parents were trained to carry out behavior therapy continued to improve, while children who were institutionalized regressed). (6) A brief reinstatement of behavior therapy could temporarily re-establish some of the original therapeutic gains made by the children who were subsequently institutionalized.

The first succinct attempt to understand the behavior of autistic children within a behavioristic framework was carried out by Ferster (1961). Ferster presented a very convincing argument of how it was that, based on a general deficiency in acquired reinforcers, one might expect the very impoverished behavioral development one sees in autistic children. The

¹We express our thanks to the parents who entrusted their children to us, and for the help and encouragement they have given. The research has been supported by PHS Research Grant No. 11440 from the National Institute of Mental Health. Many persons have helped in this research; in particular, we are grateful for the help that Gail Abarbanell, M.S.W., Lorraine Freitas, M. S., Meredith Gibbs, Laura Schreibman, Ph.D., Joan Meisel, Ph.D., and Linda Silverman gave in directing and managing the Clinic, U.C.L.A. Monographs of this article are available for \$1.00 from the Business Office of the *Journal of Applied Behavior Analysis*, Department of Human Development, University of Kansas, Lawrence, Kansas 66044. Ask for Monograph #2.

²Now at the Institute for Applied Behavioral Science, University of California, Santa Barbara.

³Now at the Department of Psychology, California State College, Los Angeles.

primary contribution of Ferster's theoretical argument lies in the explicitness and concreteness in which he relates learning principles to behavioral development. Shortly after he presented his theoretical notions about autism, Ferster and DeMyer (1962) reported a set of studies in which they exposed autistic children to very simplified but controlled environments where they could engage in simple behaviors, such as pulling levers or matching to sample for reinforcers that were significant or functional to them. The Ferster and DeMyer studies were the first studies to show that the behavior of autistic children could be related in a lawful manner to certain explicit environmental changes. What the children learned in these studies was not of much practical significance, but the studies did show that by carefully arranging certain environmental consequences, these children could in fact be taught to comply with certain aspects of reality.

The first systematic attempt to use behavior modification procedures on more general, so-

cially practical behaviors of an autistic child was reported by Wolf, Risley, and Mees (1964). They worked with a 3.5-yr-old boy who did not eat normally, lacked normal social and verbal repertoires, and evidenced extreme tantrums and self-destructive behaviors, often leaving himself bruised and bleeding. By systematically controlling the child's environment, these investigators were eventually able to bring the child's responding toward a more normal level of functioning. Tantrum behavior was treated by a combination of mild punishment and extinction. They also reported on certain training procedures that helped the child to communicate more effectively verbally. At about this time, several other studies appeared where psychologists reported success in helping autistic children acquire certain basic and important repertoires, particularly in the area of imitation and language (Hewitt, 1965; Metz, 1965; Lovaas, *et al.*, 1966a).

These behavioristic attempts to treat autistic children carried with them a promise of help and a certain optimism for the autistic child. This contrasted with the general hopelessness that had grown out of the failure that the psychodynamic therapies had encountered in trying to help these children. Kanner, who was the first person to describe and label these children as "autistic", also reported on the failure of psychodynamic therapies to effect change (Kanner and Eisenberg, 1955). Brown's 1960 study supported Kanner's data that the children were unaffected by psychotherapy. Later, Rutter (1966) provided a comprehensive review of investigations dealing with sizable groups of autistic children. The results of the studies that Rutter reviewed are quite consistent with one another and are quite pessimistic regarding prognosis. They may be summarized as follows: (1) Of those children who originally had IQ scores below 50, almost none acquired speech nor received any schooling, and three-fourths were in long-term hospitals at follow-up. If the child was mute and had no appropriate play before the age of five, the prognosis was particularly bad. (2) When

marked improvement has taken place, it has generally become evident before the age of 6 or 7 yr. From middle childhood on, the course has been fairly regular, with a continuation of improvement or deterioration evident by then. (3) In almost all cases, there were declines in IQ. (4) Improvement was unrelated to whether or not a child had received therapy. When improvement has taken place, it has been described as "spontaneous", that is, independent of a professional prescribed treatment. Havelkova (1968) reviewed several other recent studies. The results have been consistent with those reviewed above.

In contrast to these very pessimistic observations, the early studies that used behavior therapy were quite optimistic. But since this form of intervention is quite new, it remains to be shown how effective it really is with autistic children. The design of the early studies left many questions unanswered. Most of the studies reported work on single subjects, which beg the question of generality across children. Little if any systematic data were presented on the extent to which the treatment effects generalized across environments, neither were data reported on response generalization. Except for the follow-up data on one child (Wolf, Risley, Johnston, Harris, and Allen, 1967), there are no data that allow one to assess how well the behavioral intervention held up over time.

The primary purpose of the present paper is to present some measures of generalization and follow-up data on 20 children that we have treated with behavior therapy during the last 7 yr. We hope to provide the reader with an approximation of changes one might expect to see in autistic children undergoing behavior therapy. However, it is also our belief that the results presented here probably underestimate the benefits of such therapy for autistic children because the results were influenced by our extensive efforts at measurement and replication as well as therapy.

We will try to evaluate the treatment effects along three dimensions: (1) stimulus generaliza-

tion, the extent to which behavior changes that occurred in the treatment environment transferred to situations outside that situation; (2) response generalization, the extent to which changes in a limited set of behaviors effected changes in a larger range of behaviors; and (3) durability or follow-up, how well the therapeutic effects maintained themselves over time (Baer, *et al.*, 1968).

METHOD

Subjects

We have treated a total of 20 children, all of whom have been diagnosed as autistic by at least one other agency not associated with this project. The majority of the children had been given more than one label, usually also being referred to as retarded and brain damaged. Our experience and that of others (*cf.*, Rutter, *ibid.*) suggests that there is considerable behavioral heterogeneity among autistic children. Therefore, it may be appropriate to describe the children we have treated in more detail. First, we have treated the very undeveloped children, that is, children who would fall within the lower half of the psychotic continuum, and whose chances of improvement were considered to be essentially zero. Most of the children had at least one prior treatment experience (up to 4 yr of intensive, psychodynamically-based treatment) which had not effected any noticeable improvement. Most of the children have been rejected from one or more schools for the emotionally ill or retarded because their teachers could not control them, in addition to which their behavior was often so bizarre that it was disruptive for the other children in the class. Clinically speaking, with three or four exceptions, they seemed void of anxiety, and none had any awareness that he was considered abnormal.

Generally, the children we have treated can be described along the following dimensions: (1) *Apparent sensory deficit*, indicating that when asked to complete the

Rimland Diagnostic Checklist for Behavior-Disturbed Children, Rimland, 1964), most of the parents report that their children (a) at one time appeared to be deaf; and (b) seemed to look through or walk through things as if they were not there. Furthermore, many of the parents indicated that at one time they sought professional opinion about their children's hearing and/or vision, only to be told that the child had "normal" hearing and vision. (2) *Severe affect isolation* was a predominant feature. This means that parents indicated on the Rimland Checklist that their children (a) fail to reach out to be picked up when approached by people; (b) look at or "walk through" people as if they were not there; (c) appear so distant that no one can reach them; (d) are indifferent to being liked; and (e) are not affectionate. (3) Our sample showed a high incidence of *self-stimulatory behavior*, that is, behavior that appears solely to provide the children with proprioceptive feedback (*e.g.*, rocking, spinning, twirling, flapping, gazing, *etc.*). A more detailed description of this type of behavior is given below in the method section (under instructions for observer identification). (4) *Mutism* occurred in about half of the children in our sample. These children produced no recognizable words (their sounds consisted primarily of vowels). (5) *Echolalic speech* was present in the remaining children. These children echoed the speech of others, either immediately or after a delay, often giving the impression of non-related inappropriate speech (a more complete description of these behaviors is also given below in the instructions for observer identification). (6) In all children *receptive speech* was minimal or missing entirely. Some of the children would obey simple commands (such as "sit down", or "close the door"), but *all* failed to respond appropriately to more complex demands involving abstract terms such as prepositions, pronouns, and time. Most often they responded to speech in a very generalized manner. For example, they would close the door when they heard the command, "Close the door.", as well as when they heard com-

mands like "Point to the door.", or statements such as "There is a window and a door." *etc.* (7) There was also an absence of, or only minimal presence of, *social and self-help behaviors*. For instance, most of the children could not dress themselves; most were unaware of common dangers (*e.g.*, crossing the street in front of oncoming cars); most could not wash themselves or comb their hair; some were not toilet trained, *etc.* (8) A small number of these children were *self-destructive or self-mutilatory*. All displayed severe aggressive, tantrumous outbursts, scratching and biting attending adults when forced to comply with even minimal rules for social conduct. Some smeared their feces.

Treatment

When one decides to treat a child within a reinforcement theory paradigm, then one can facilitate the behavioral development of autistic children in two ways. One way would be to concentrate efforts on facilitating the autistic child's acquisition of social reinforcers, rather than on building behaviors. If his developmental failure was based on a deficiency in social and other secondary reinforcers, as Ferster claimed it was, then an intervention at this level would seem to strike at the base of the problem. A treatment program centered on the establishment of a normal hierarchy of social reinforcers would give the child's everyday social environment (his parents, teachers, peers, *etc.*) the tools with which to build and modify the myriad behaviors necessary for the child to function effectively within that environment. In a sense, the person's behavioral changes would "take care of themselves", provided that he returned from treatment to a normal environment with a normal reinforcement hierarchy.

When we first began to treat autistic children, we explored this alternative of enriching and normalizing reinforcing stimuli for these children. We did succeed at establishing certain social stimuli as reinforcing, using either pain reduction (Lovaas, *et al.*, 1965*b*) or food presentations (Lovaas, *et al.*, 1966*b*). Although we

produced some durable reinforcers, they were too discriminated (situational) and the procedures too cumbersome to be of much practical significance.

We turned, therefore, to the second alternative; building behaviors directly relying on already effective, largely primary reinforcers such as food, essentially circumventing social stimuli. The use of primary reinforcement has several disadvantages, as compared to social, secondary ones. For example, in using primary reinforcers, special environments need to be established to develop and maintain the new behaviors. Since we have inadequate information about how to construct such environments, the gains that the child may make would probably fall short of the ideal. Despite these restrictions, however, it is worthwhile to assess how much one can accomplish using a limited range of reinforcers. Therefore, we describe the program we did develop.

Because the children were replete with interfering self-stimulatory, self-destructive and/or tantrum behavior when they entered treatment, we immediately attempted to reduce the frequency of such behavior. The procedures employed to extinguish and suppress pathological behavior (including biting and scratching of self and others, feces smearing, *etc.*) rely heavily on several operations: (1) contingent reinforcement withdrawal, that is, the adult simply looked away from the child when he was engaged in undesirable behavior, left the child in his room, or placed the child in an isolation room (separate from the treatment room); (2) contingent aversive stimulation, for example, a slap or painful electric shock; or (3) reinforcement of incompatible behavior, such as sitting quietly on a chair. The rationale for the suppression of self-stimulatory behavior lies in the observations we have made indicating an apparent attenuation of the child's responsivity while he is engaged in self-stimulation (Lovaas, Litrownik, and Mann, 1971). Simply stated, when the child is engaged in self-stimulation, it is difficult to teach him something else. The reasons for suppressing self-destruction, feces

smearing, *etc.*, are perhaps obvious, and our intervention model does not prescribe the therapeutic benefits of their expression. A detailed presentation of data and method for suppression of self-destruction may be found in Lovaas and Simmons (1969).

Simultaneously with the suppression of undesirable behavior, the therapist attempted to establish a kind of primitive stimulus control. Usually, the therapist demanded some simple behavior from the child, such as looking at the therapist, or sitting down when the therapist asked. These behaviors could be easily prompted if the child did not already know how to respond. Usually, the therapist's first attempts to establish stimulus control elicited tantrums and self-destructive behavior; therefore, we combined the suppression of undesirable behavior with the attempt to establish stimulus control.

Once these introductory steps had been taken, we introduced our central training program in which language training alone consumed about 80% of the child's total training. The heavy emphasis on language training was undertaken partly for academic reasons. We wanted to know how much could be accomplished using operant procedures. This was not necessarily the most beneficial therapeutic approach for all the children. Many of them have benefited more from a program emphasizing non-verbal communication.

If the child was mute, we began a verbal imitation program to facilitate his phonetic development (Lovaas, *et al.*, 1966a). Briefly, verbal imitation was established in five steps: (1) The child received reinforcement for vocalizing in order to increase the frequency of speech sounds. (2) We then established a temporal discrimination. The child received reinforcement only for those vocalizations that were emitted within a 5-sec period after the therapist made a vocalization. (3) The therapist now began to demand similarity of vocalizations between himself and the child. For example, the therapist gave reinforcement for a sound (for example, "ah") only after the therapist had first emitted that sound

himself. (4) After the child reliably emitted one sound, the therapist introduced a second sound (such as the consonant "mm") and reinforced reproductions of that sound. These first two sounds were then presented in a random order so that the child was required to discriminate between the two vocalizations. (5) A third sound was presented, requiring increasingly fine discriminations. In such a manner we attempted to build imitative behavior, which we conceived of as a discrimination where the child's response resembled its stimulus (the adult's response).

If the child was echolalic (or once a mute child had about 10 imitative words), we introduced a program designed to make speech meaningful and functional. For example, as soon as a child was taught the label for a particular food, he could eat only if he asked for the food by name. The child was gradually moved through a series of steps designed to establish increasingly proficient use of language, including training in semantics, such as use of abstract terms (pronouns, time, *etc.*), and syntax, such as the correct use of tense, *etc.* Some of the later levels were never reached by the mute children, but were usually obtained with the echolalics. A more detailed description of the language program exists on film (Lovaas, 1969) and in written outline (Lovaas, *in preparation*).

At the same time we were involved in building speech, we also initiated programs designed to facilitate the acquisition of other social and self-help skills. These programs focused on those behaviors that made the child easier to live with, such as friendly greetings and other indications of affection, as well as dressing, good table manners, brushing the teeth, *etc.* We have outlined a procedure based on non-verbal imitation (Lovaas, *et al.*, 1967) that has been particularly useful for these purposes.

Throughout, there was an emphasis on making the child look as normal as possible, rewarding him for normal behavior and punishing his psychotic behavior, teaching him to please his parents and us, to be grateful for what we would do for him, to be afraid of us when we were

angry, and pleased when we were happy. Adults were in control. In short, we attempted to teach these children what parents of the middle-class Western world attempt to teach theirs. There are, of course, many questions that one may have about these values, but faced with primitive psychotic children, these seem rather secure and comforting as initial goals.

We selected reinforcers on the basis of their value for a particular child. Many children would work only for food and required an occasional slap on the buttocks if the therapist was to control undesirable interfering behavior. For other children, symbolic approval and disapproval were effective in maintaining the child's behavior throughout the working sessions. As we became familiar with the idiosyncracies of the various children, the reinforcers seemed easily accessible and their selection was fairly simple, despite their limited range. However, scheduling these reinforcers was a much more difficult task. A relatively untrained person can build simple behaviors, like eye-to-face contact or raise the frequency of vocal behavior. But it is unlikely that a person will be able to build complex speech unless he is familiar with discrimination learning procedures. Most people who work with autistic children are not. Therefore, it seems likely that there will be few studies in the near future to replicate the present one.

Measurement

We have employed two measures of generalization of change during treatment. First, we have attempted to assess changes in the children's behavior using a *multiple-response recording*. Secondly, we have assessed changes in the children's Stanford Binet and Vineland Social Maturity scores. The multiple-response recordings constitute the main focus of our measures and were designed to provide information both on stimulus and response generalization. The Stanford Binet and Vineland do provide similar measures, but they give less specific information. We shall first present a description of the multiple-response recordings.

Multiple-response recordings. We have previously published (Lovaas, *et al.*, 1965a) information on apparatus that allows for simultaneous recordings of several commonly occurring and everyday behaviors in free-play/observation settings. Essentially, certain behaviors (both normal and pathological) are defined for an observer who records their frequency and duration on a button-panel, which in turn is coupled to a computer tape, allowing swift calculation of the frequency, duration, and interaction of the various behaviors.

The kind of child one is studying helps decide what kinds of behaviors to record. In the case of severely psychotic children, this is somewhat simplified because of their limited behavioral repertoires. We eventually selected five behavioral categories. The presence or absence of behaviors in these categories are used to describe autistic children, and we have found they can reliably be recorded. (1) *Self-stimulation*, which denotes the stereotyped repetitive behavior that appeared only to provide the child with proprioceptive feedback (*e.g.*, rocking, spinning, twirling, flapping, gazing, *etc.*). (2) *Echolalic speech*, which was defined as the child's echoing the speech of others, either immediately or after a delay, giving the impression of non-related inappropriate speech, with pronoun reversal, incorrect use of tense, *etc.* We also included bizarre words and word combinations in this category. (3) *Appropriate speech*, which was defined as speech related to an appropriate context, understandable, and grammatically correct. (4) *Social non-verbal behavior*, which denoted appropriate non-verbal behavior that is dependent upon cues given by another person for its initiation or completion (*e.g.*, responding to requests, imitating, *etc.*). (5) *Appropriate play*, which denoted the use of toys and objects in an appropriate age-related manner.

Two of these behaviors (self-stimulation and echolalia) are pathological. Their presence, and the relative absence of the remaining three "normal" behaviors, forms part of the behavioral complex diagnostic of autism. The instructions

for recording, and hence rather complete definitions of these various behaviors are given below.

Instructions for rater identification. You will be watching for five kinds of behaviors. These will be the only behaviors you will have to record, so part of the time you may not be pressing a button at all. If you are uncertain about what is going on, you may also not be recording. *The best rule is, if you can't make a decision, don't record anything.* Each of the behaviors will be carefully defined and you will be given examples of what they are and what they are not. Each key on the panel is labeled with the name of one of the behaviors. Each time you notice the child engage in one of these behaviors, press down the corresponding key, and hold it down until the child has terminated that behavior.

1. *Self-stimulation.* The best way to describe the various forms this kind of behavior may take is to begin with the head. The child may roll his eyes, cross them, look out of the extreme corners of them or squint them, contracting the muscles of the face all the way to the ears. He may stare intensely at lights, objects, or at parts of his own body (such as his hands). He may suck his tongue and lips or stick his tongue out repeatedly. He may put objects in his mouth. He may rock his whole head from side to side or allow it to fall forward, turning it slightly to the side with his eyes turned up or to the corners. He may cock his head and hold a particular position for long periods.

There are several typical forms of self-stimulation performed with the arms and shoulders. The child may move the arms up and down the sides of his body, flipping his hands from the wrist. He may flap his arms from the shoulder with his hands limp. He may hold his hands in very contorted positions, often staring at them intensely.

Using his whole torso, the child may assume rigid or contorted postures, or he may engage in body rocking. Rocking usually occurs in some sitting position and is a forward and back motion more often than side to side. He may twirl himself, rub, scratch, or tickle various parts of his body, or he may masturbate. He may jump repeatedly or run from wall to wall. He may walk on his toes.

In his interaction with physical objects, he may trace them, running his fingers along the edges, rub them, spin them, use them to make tapping noises or flip them back and forth. For example, he may spin a cup like a top, or he may hold a piece of string and rhythmically swing it to and fro.

All of these behaviors have the common appearance of producing sensory input for no other purpose than the stimulation itself. There is, however, a fine line between some of these behaviors and some of the more primitive forms of play. You will have to use your judgment to some extent to decide if the child is looking at something or staring at it, playing with it or tapping it. Often, you may be able to tell self-stimulation from appropriate play, by the fact that the child may begin to repeat a gesture or an act over and over again during self-stimulation.

Self-stimulatory behavior occurs to some extent in the repertoires of all children, and even to some extent in adults. You may find yourself recording something you feel looks like any child might do. Record it anyway. The difference is that you may see more of it than you might in a normal child.

These behaviors may or may not be accompanied by sounds or words the child may use. This makes no difference as the sounds or the speech will be treated separately.

2. *Echolalic speech.* Recording this behavior signifies that the child is using words in an unmeaningful or inappropriate manner. This type of behavior may present itself in one of several ways.

First, the child may appear to be repeating a word or several words to himself. The technical name for this type of speech is delayed echolalia. He may say things that sound like commands or statements he once heard, but which have nothing to do with his present activity, or the context in which he is operating. He may use phrases like some of the following: "Hello John.", "No, John.", or "How are you, John?". He may go to the door and say "You want to go out?". Although this last statement does have relevance to the situation, such phrases will also be included when they sound like the imitation of what another person has said to the child at some other time. He may also simply repeat isolated words such as "balloon, balloon".

Secondly, the child may use bizarre speech, sounding like a word salad. The words may be understandable separately, but do not make any sense when used together. Examples are, "green rabbit", "Bufferin, one, two, three", or "happy puppy baby little". You may be able to think of it as sound input for the sake of the input, much like a verbal version of self-stimulation.

Finally, another class of behavior included in this category is immediate echolalia. For example, the adult may say "How are you, John?" and the child will answer "How are you, John?".

The adult may say, "Now it's time to draw." and the child will say, "draw".

Again, the key is depressed for the duration of the child's speech. You should not release the key after every word or phrase, but should release it if the child pauses or the adult speaks. You will not have to record "babbling". Also, do not record humming, grunting, squealing, or any sound which is unintelligible (including fussing or crying).

3. *Appropriate speech.* This behavior consists of intelligible, non-repetitive speech which is appropriate to the situation. If the child answers a question, makes a comment, labels an object, carries on a conversation, or reads aloud, you should record the behavior as appropriate speech. Depress the key when the child begins to speak, and release it when he finishes.

This category includes everything from the most primitive use of words, such as a simple "hi" or "go" (when the child is trying to leave), to the most complex conversation. The important element is that the child is using words correctly, meaningfully, and that he is readily understandable.

4. *Social non-verbal behavior.* There are two levels of this type of behavior. Level one describes certain kinds of interactions the child may display with the adult present. Included in this category are the simplest kinds of social relationships. Each party need only respond once. Thus, if the child makes a response, and the adult responds by completing the interaction, this is one response. No further response is necessary. There are no chains of response. Examples consist of two types.

- A. *Demand behavior.* The child grabs the adult's hand and tugs him toward the door.
- B. *Compliance.* In this case, the child simply complies to some request from the adult. The adult may say, "Sit down.", "Play ball.", "Put the block like this.", and the child does. You should briefly depress the button (for less than one second) when the child responds appropriately to a request. Also included in this category is simple imitation when it is not part of a game. The adult may say, "Jump, John." and then the adult may jump. If the child imitates the jumping, you should press the button.

All of these behaviors may or may not be accompanied by language. It makes no difference as long as the child is involved in some simple nonverbal response which depends upon the presence of the other person. Remember, the

important element is that the child would not be making the response if the adult were not there.

In the higher level of social nonverbal behavior, the interaction demands a variety and flexibility of response from both people. There is a longer interchange, in which the people must make several different responses to complete the interaction. The game of "Simon Says" is a good example of this kind of interaction. The child must watch, listen, and mimic or not mimic the adult depending on what the adult does. Games of pretending, playing ball, imitating drawing, follow the leader, and tag are also examples. Each person must watch and respond correctly and complete the game. Again, two people must be present for this type of behavior to take place.

Because of the instructions given to the adult with the child, in almost all cases the child will have to initiate this kind of interaction (the major exception being ball play). Therefore, you may see a considerable amount of appropriate verbal behavior simultaneously taking place. Remember to keep the key depressed throughout the entire interaction.

5. *Appropriate play.* There are also two levels of appropriate play. The lower level is defined as exploration and simple play. It means that the child is exploring or manipulating objects and that he shows interest but inability to use them properly, or that he has a lack of experience in their use. He may be using them for play (rather than self-stimulation, for instance) but is not, perhaps, using them as they were intended.

He may be stacking tiles or blocks, scattering things, putting crayons in boxes, handling and examining various toys, pouring water into containers, etc. Here you must make the distinction between handling or examining and staring or using objects in a bizarre, repetitive, or stereotyped manner.

He may pile up objects, fit tiles in a peg board, punch a bobo, ring the telephone, scribble with the crayons, pull the wagon, turn pages in a book, or make a rattle by placing small objects in a larger one. One common element here is that one response accomplishes as much as any series of responses to a given object. One response does not require another one, nor does it depend on a previous one. Stacking up one block does not require another be stacked to complete the stacking.

The higher level of appropriate play consists of the complex and appropriate use of objects, or participation in games in which there is a definite dependency of one response on another.

One response leads to or proceeds from another in the accomplishment of some project. In this category, a number of responses completes some whole which no response individually could complete. Examples include making a pattern or picture with tiles or crayons, building an object with blocks, reading, pulling the wagon to transport objects for a project, setting bowling pins up in the appropriate pattern and knocking them over, and completing a puzzle. Each response here adds something new to the ultimate goal of some project. The games listed under social nonverbal behavior have this same quality, interdependency of responses, and they should also be recorded (simultaneously). *Note.* There are several behaviors which may best be recorded by pressing and releasing a key immediately (a blip). This should be done in the case of social nonverbal behavior when the child obeys a command, or each time the child catches or throws a ball. It is not done each time the child stacks a block or fits a tile in appropriate play. Here you must use your own judgment. Do not record during pauses, but do not record a pause between every response. Are there any questions?

The reader may note that social non-verbal and appropriate play have been divided into two levels each in these instructions. This was done in an attempt to increase the discriminating power of these measures, and reflect a later development, not present in the recordings that we present in this paper.

The multiple-response measures do to some extent assess response generalization. That is, many of the behaviors we did score (particularly social non-verbal and play) were not specifically taught during treatment. But we had no way of knowing exactly how much of these behaviors were new and novel by the child, so that the recordings are not pure measures of response generalization. The measures do, however, lend themselves well to studies on stimulus generalization.

To assess stimulus generalization, the children were observed in a room separate from, and not associated with, the training situation and in the company of an unfamiliar adult. The room was equipped, like most playrooms, with the following toys: a wagon, paper and crayons, a bobo doll, a 9-in. rubber ball, three plastic

bowling pins, a plastic telephone, a magnetic board with numbers and letters that attach to it, 12 assorted wooden blocks, a 6-in. tom-tom drum, a hand puppet, and three simple wooden jigsaw puzzles. The child was observed in this room during sessions lasting 35 min. These sessions were divided into three conditions of 10, 10, and 15 min each. In the first condition (the Alone condition), the child was observed by himself in the playroom. In the second condition (the Attending condition), an unfamiliar adult was present and attended visually to the child, but made no comment, interfered in no way, and did not initiate any interaction with the child. If, on the other hand, the child initiated some activity that required the involvement of the adult, the adult performed those responses and made whatever comments necessary to complete the interaction. In the final condition (the Inviting condition), the adult encouraged the child to participate in several different kinds of activities. The adult invited the child to play with each of the 11 toys in the playroom in succession (1 min per toy), giving demonstrations of how to use the toy if the child appeared not to know how. The adult also attempted to initiate a simple game of "patticake" for 1 min. He also gave the child a 1-min series of simple commands that could be performed non-verbally, such as "Stand on one foot.", "Touch the floor.", and "Sit down.". Next, the adult asked a 1-min series of questions which could be answered either verbally or non-verbally. This series consisted of questions such as "Where is your nose?" or "Which block is bigger?". A final 1-min series of questions, which could only be answered verbally, was also asked. This series consisted of questions such as "How are you?" or "Where do you live?".

We have multiple-response measures on only 13 of the 20 children we have treated. This is so because we initially had considered these measures to be inappropriate for outpatients, since we had less control over their treatment. Since 1968, however, we have obtained multiple-response measures on the outpatients as well.

The first four children (Ricky, Pam, Billy, and Chuck) for whom we have multiple-response recordings received a "before" measure (in June, 1964) and recordings were then made on a monthly basis for the 14-month duration of their treatment. Pam and Ricky were discharged immediately to a local state hospital, while Billy and Chuck spent a short time (less than 6 months) with their families before being hospitalized in the same state hospital. Pam and Rick were returned to us for follow-up measures 2 yr later (1968). They were then briefly treated once more (24 hr for Ricky and one month for Pam), discharged to the state hospital again, and finally returned for a second follow-up 2 yr after that (1970). Pam and Rick received our treatment twice, interspersed by a period of no behavior therapy treatment; Billy and Chuck were treated once, but measured again 4 yr after discharge from our project (1970); they received an ABA design.

We replicated essentials of the treatment on a second group of children (Jose, Michael, and Taylor) who were hospitalized in 1965 and received 12 months of treatment, with multiple response measures before treatment and at three-month intervals during treatment. They were returned for follow-up measures 3 yr after treatment (in 1970).

The third group (Leslie, Tito, and Seth) to receive the multiple-response recordings were seen as outpatients. They were measured before treatment (1968) and after 1 yr of treatment, and received follow-up measurements 1 yr later (1970).

A fourth group (Kevin F., Ann, and James) to receive multiple-response recordings was also seen as outpatients. Measures were taken before treatment (1969), after 1 yr of treatment (1970), and with follow-up measures in 1972.

The first and second groups of children were inpatients. They received 8 hr of treatment per day, six to seven days a week. The parents of the first group were not involved in the treatment. With the second group, however, we began to train the parents in our treatment procedures.

The third and fourth groups were outpatients, and while we initiated training programs in the clinic, we otherwise served essentially as consultants (2 to 3 hr a week) to the parents, training them in shaping procedures.

Discharge procedures for these children differ for each individual case, depending on the rate of progress by the child, the skill of the mother as a therapist, and the prospects for enrolling the child in a special school. In general, our approach was gradually to phase the children out of the program. We decreased the number of sessions from three per week to once a month. After the child was officially discharged, a therapist visited the home several times during the first few months. Generally, by this time the parents had found a school placement for the child and our involvement became minimal. Parents were encouraged to call us when they encountered difficulties, and we spoke to them from time to time informally discussing the child's progress. Often, the therapist visited the school and discussed the child's case with the teacher, suggesting ways he might find effective in dealing with the child and encouraging the teacher to call on us if he encountered any difficulties.

The basic rationale for changing the treatment procedure from treating inpatients, with the parents as observers, to treating outpatients with the parents as therapists, became apparent from examination of the follow-up data.

Intelligence and social maturity. The Stanford Binet Intelligence Scale was administered before and after treatment either by an agency not associated with UCLA, or when this was not feasible, by a graduate student trainee in the UCLA Psychology Clinic. Nineteen of the 20 children received IQ testing. One child, Taylor, received the Merrill-Palmer Intelligence Test instead of the Stanford Binet. We will also present some data from the Vineland Social Maturity Scale, which was administered to the parents of the last 14 of the 20 children. The irregularities in the number of children who received the various tests does not reflect a systematic bias.

Rather, in the early phases of the program we did not consider generalization and follow-up data to be significant data for our study.

RESULTS

Multiple-response measures. Since the multiple response measures are the focus of this study, they are presented first. The results are presented as group averages, followed by discussions of changes in the individual groups and children. All the figures based on the multiple response measures have per cent occurrence of the behavior on the ordinate. This percentage

was obtained by calculating the duration of a behavior, to the nearest second, and dividing it by the duration of that condition (e.g., if the subject spent 200 sec in self-stimulatory behavior during the 10-min Alone condition, he would receive a measure of 33% self-stimulation at that time).

The first data, presented in Figure 1, give the before and after treatment scores for the various behaviors, averaged over all conditions for the four groups. The various behaviors are presented on the abscissa before (B) and after (A) treatment. Three groups are presented: T (total

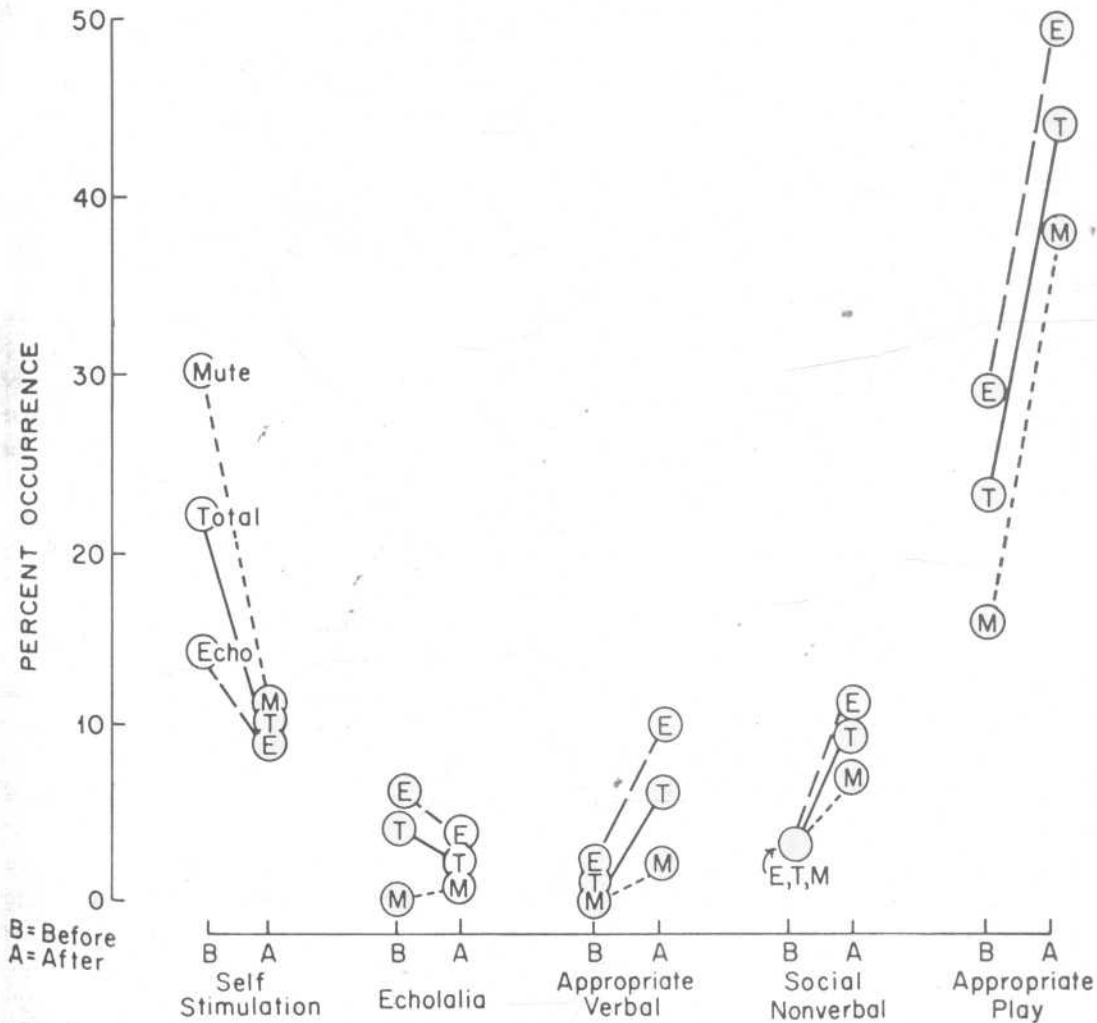


Fig. 1. Before (B) and After (A) multiple-response measures averaged over all conditions for the four groups. Per cent occurrence of each behavior is plotted on the ordinate. "E" refers to the average results for the echolalic children, "M" to the average results for the mute children, and "T" to the average results for the total group.

subjects); and the breakdown of that group into the children who were echolalic (E) and mute (M) before treatment. Looking first at the data from the total group, it is apparent that the inappropriate behaviors decreased while the appropriate behaviors increased. Specifically, self-stimulatory behavior was reduced to about one-half of its pre-treatment level. The amount of echolalic speech decreased only slightly when one considers the total group, but this is because the decrease in echolalic speech by the echolalic children was offset by the increase of echolalia in the mute children.

Turning to the appropriate behaviors, the children showed about four times as much appropriate verbal and social non-verbal behavior after treatment, and almost twice as much appropriate play. There were no exceptions to these changes; all the children improved.

The total group comprising Figure 1 consisted of five mute and eight echolalic children. If we examine the data on the mute children we can observe that, in addition to evidencing no speech, they showed more self-stimulation and less appropriate play. The mute children, in general, appear more behaviorally retarded than the echolalic children. The figure also suggests that the mute children show the largest gains in treatment. They show the largest proportionate reduction in self-stimulation and largest proportionate gains in the verbal behaviors. While this may be a correct inference, it must be remembered that our measuring system gives equal weights to all behaviors within the various categories. For example, while the mute children showed a proportionately greater increase in appropriate verbal behavior, the speech of the echolalic children seemed qualitatively superior to that of the mute children. More exact descriptions of the changes in speech are presented on film (Lovaas, 1969) and in a separate paper (Lovaas, *in preparation*). Perhaps it is sufficient to say that both mute and echolalic children improved with treatment, ignoring the more specific comparisons. The data are now presented separately for each group.

Group 1 (Rick, Pam, Billy, and Chuck) was measured on a monthly basis, enabling us to assess the rate at which the behavior changed. The data for these children are presented in Figure 2. Pam and Rick (both echolalic) are presented on the left side. Billy and Chuck (mutes) are presented on the right. The top part of the figure shows changes in verbal behavior, while the bottom part shows the non-verbal behaviors. For Rick and Pam, one can observe the gradual increase in appropriate speech. No trend is obvious for echolalic behavior. Billy and Chuck, who were initially mute, showed a rise in echolalic speech before it was replaced by appropriate language. Neither had appropriate speech before treatment; each had some appropriate speech afterwards. Inspection of changes in non-verbal behaviors shows a decrease in psychotic self-stimulation, and increases in appropriate play and social non-verbal behavior.

It is probably helpful to break the data down by conditions to demonstrate the degree to which the adult gained control over the child, and the extent to which the child initiated behavior independent of the adult's explicit direction. The reader is reminded that in the multiple-response measures, the "adult" was unfamiliar to the child, and during the Attending condition initiated no interaction with the child. Therefore, any social and language behavior during the Attending condition was an indication of spontaneous, "self-initiated" behavior. Figure 3 presents social non-verbal and verbal behavior separately for the Attending and Inviting conditions. Examining the data closely (top half of the figure), it is noteworthy that there was an absence of social non-verbal behavior in the Attending condition until about eight months of treatment. The appearance of this behavior signals the children's spontaneous initiation of behavior, a very important sign* of therapeutic progress.

The same spontaneous interaction was replicated in the case of appropriate verbal behavior (lower half of Figure 3). The data again indicate that the children began to initiate verbal contact

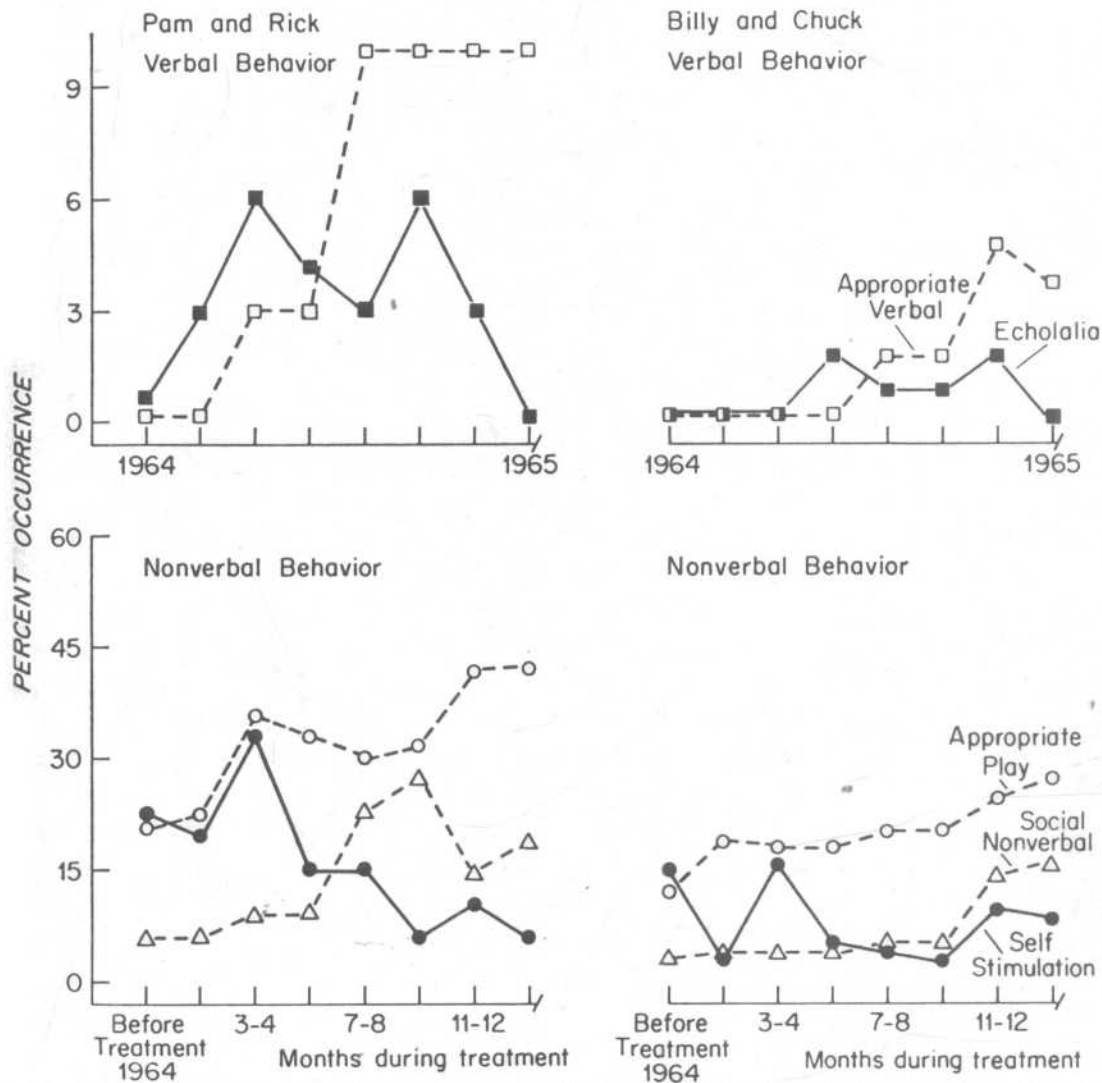


Fig. 2. Monthly multiple response measures for the first group. Rick and Pam's data are presented on the left, and Chuck and Billy's data are presented on the right. The top part of the figure shows changes in verbal behavior, and the bottom part shows changes in nonverbal behavior. Data are averaged over two-month periods.

with the attending adult after the eighth month of treatment. Predictably, both social non-verbal behavior and verbal behavior were higher during the Inviting than the Attending condition. In the Inviting conditions, the attending adult facilitated the children's social behavior by instigating numerous interactions. The facilitory effect increased as treatment progressed. It also seems reasonable to us that the children show more social non-verbal than language behavior because the latter is more difficult to build.

An important observation has to do with individual differences in the rate at which the children displayed these behaviors. Figure 4 shows the change in appropriate verbal behavior over the Attending *versus* the Inviting conditions for each of the first four children. As can be seen in the Attending condition (Figure 4), only Ricky and Billy progressed to the point where they came to initiate verbal behavior with the attending adult. However, all the children learned to interact when the adult initiated the

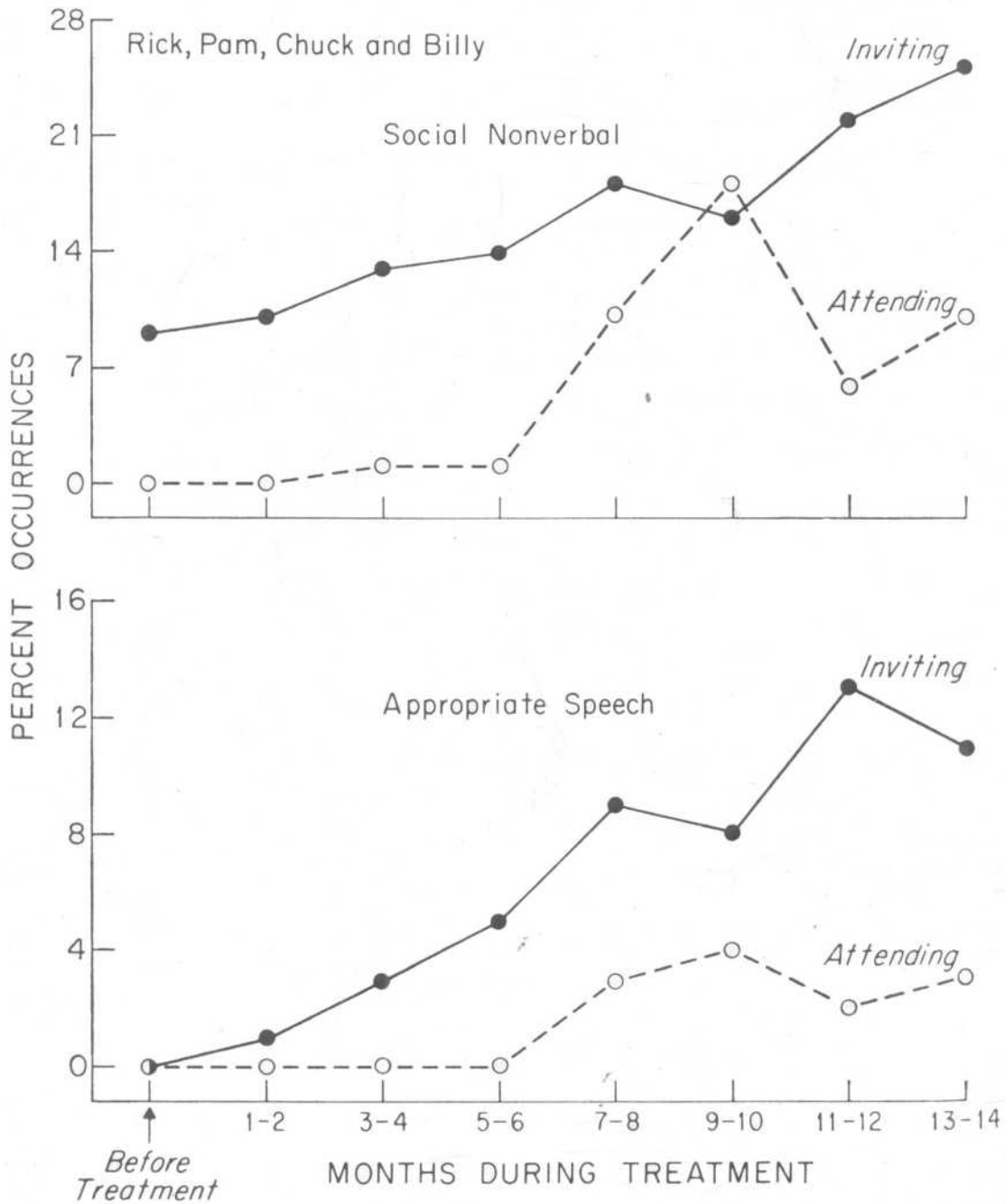


Fig. 3. Monthly recordings of the first group's social nonverbal and verbal behavior presented separately for the attending and inviting conditions. Data are averaged over two-month periods. Per cent occurrence of each behavior is presented on the ordinate.

conversation, as is indicated by data for the Inviting condition.

Group 2 (Taylor, Mike, and Jose) was treated similarly to Group 1; with two exceptions: first,

we employed no aversive stimulations (shock, spankings, etc.) for the first six months of treatment; and second, we initially planned a much less-demanding schedule for the children. That

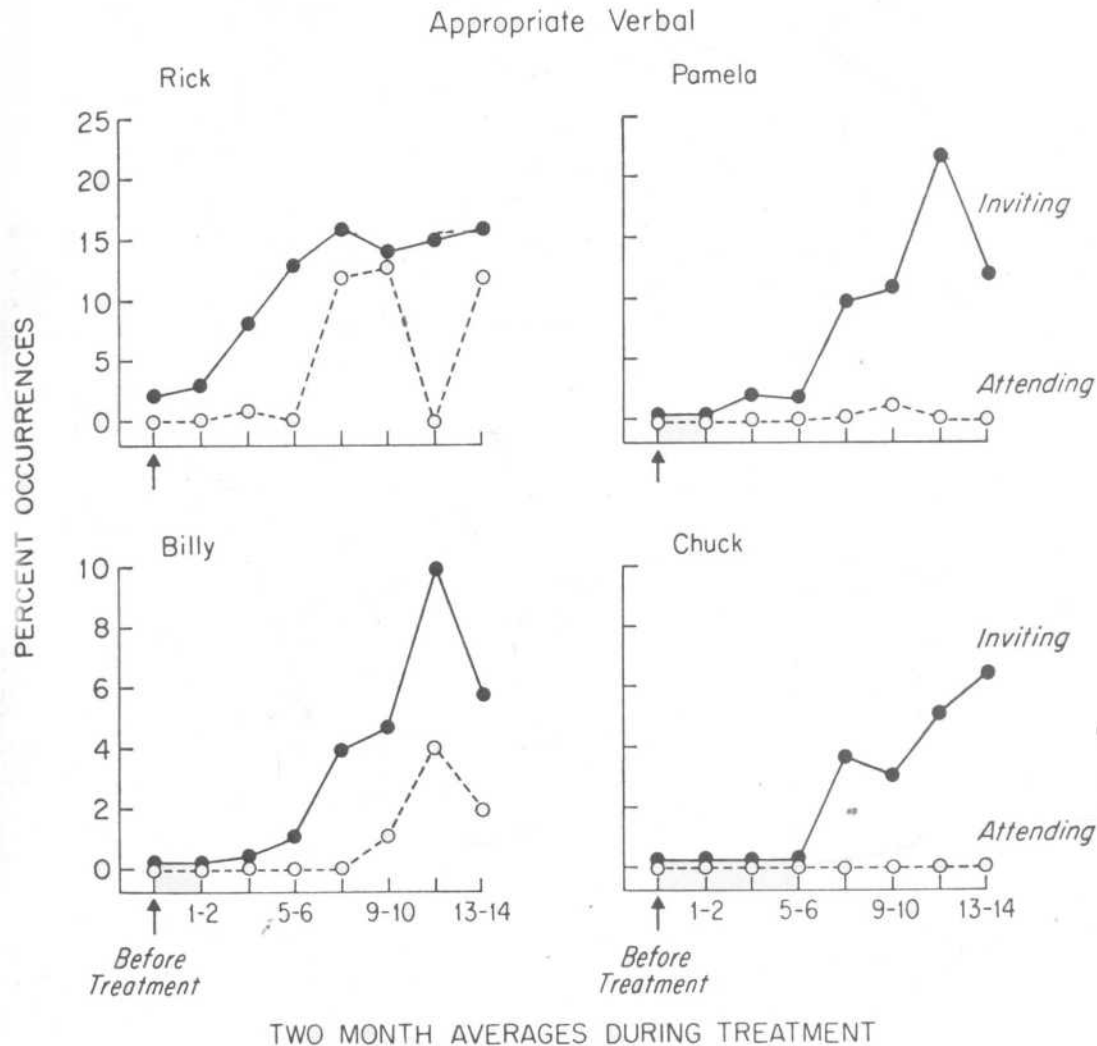


Fig. 4. Monthly recording of verbal behavior presented separately for the attending and inviting condition for each of the first four children. Per cent occurrence of the behavior is presented on the ordinate. Data are averaged over two-month periods.

is, we left a child at a certain level of mastery for a relatively long time before we introduced the next task. We also attempted some variation on imitation training by pairing food with the therapist's vocalizations, instead of demanding the difficult discriminations described earlier. We did not observe any particularly encouraging improvement in the children's behavior after six months of such treatment, so we returned to the more demanding treatment program the first group received. Essentially, then, Group 2 received the same treatment as the first, although it was somewhat less intensive. These

children also differed from those in Group 1 in that all three were mute.

The data on Group 2 are presented in Figure 5. The measures were taken every three months, as is noted on the abscissa. Results from Group 2 essentially replicate the results obtained from Group 1. There is a gradual replacement of inappropriate by appropriate behavior. We have not plotted changes in verbal behavior, because these were minimal, rising only to 1 or 2% after 12 months.

Group 3 (Leslie, Tito, and Seth) and Group 4 (Kevin F., Ann, and James) were all outpatients.

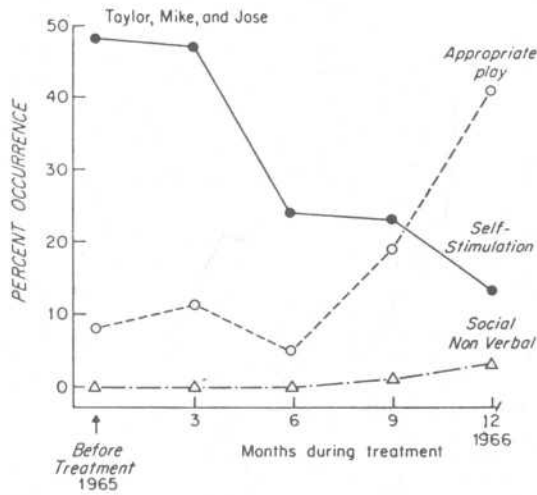


Fig. 5. Monthly multiple-response measures for the second group. Per cent occurrence of each behavior is plotted on the ordinate. Data are averaged over three-month periods.

For these children, we served more as consultants to the mothers, doing less direct therapeutic work with the children ourselves. James was essentially mute, the others were echolalic. Multiple response records for these children were made before and after 1 yr of treatment. Data from Group 3 (Leslie, Tito, and Seth) are shown on the left side of Figure 6, while the data from Group 4 are shown on the right side. The data from Groups 3 and 4 replicate the results from Groups 1 and 2: a decrease in psychotic behavior and an increase in normal behavior. Starting with the top line in each graph, one can observe a rise in Appropriate Play, Social Nonverbal, and Appropriate Verbal. Concurrently, there is a drop in Echolalia.

While the measures on Group 4 (on the right side of Figure 6) did not reflect greater improvement than with the other groups (this is most clear in the case of Appropriate Play), it seemed

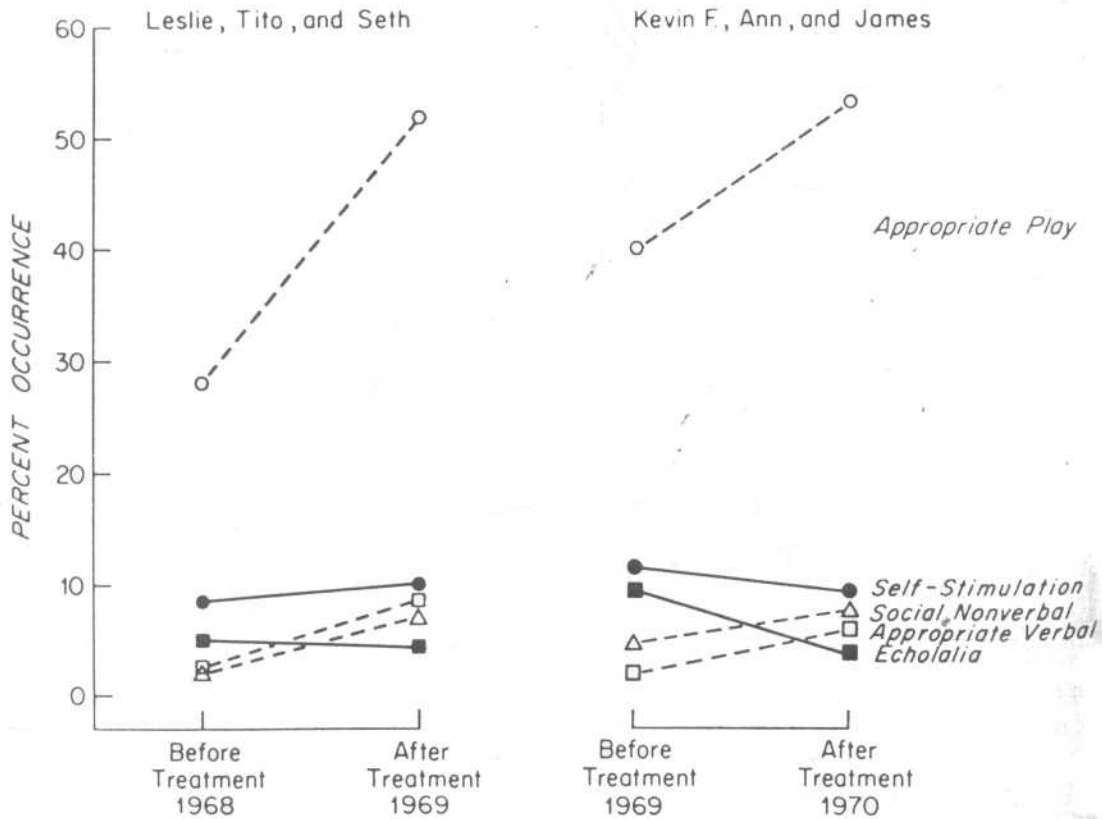


Fig. 6. Before and After multiple-response measures for Groups 3 and 4. Data for Group 3 are presented on the left, and for Group 4 on the right. Per cent occurrence of the behaviors is plotted on the ordinate.

clinically that the children in Group 4 made greater gains during treatment than the other children. The failure of the multiple-response measures to reflect this improvement may be based on the failure of those recordings to make discriminations beyond a certain level of behavioral complexity. We have previously (Lovaas, *et al.*, 1965a) pointed out that some of the behavioral categories failed to discriminate beyond certain ages for normal children. To overcome this difficulty we began to differentiate between different "levels" of social non-verbal and appropriate play (as was presented earlier in instructions for rater identification). We intended to improve the sensitivity of the recording procedures by making discriminations between certain behaviors; for example, differences in play behavior that involved "simple" acts like repetitively dropping beads into a jar (level 1), as compared to imaginative doll play (level 2). When the data are presented using these new categories, as is done in Table 1, it becomes apparent that Group 4 made most of its gain in the "higher" levels of social non-verbal and

Table 1

Before and after measures for groups three and four on levels I and II of social non-verbal behavior and appropriate play.

	Social Non-Verbal _r		Appropriate Play	
	I	II	I	II
<i>Group Three</i>				
Before Treatment	1	1	18	10
After Treatment	4	3	41	11
<i>Group Four</i>				
Before Treatment	4	1	32	8
After Treatment	5	3	37	16

appropriate play behavior, while Group 3 made most of the gains in the "lower" levels. If one plans to measure treatment effects on children who have a higher level of behavior development than the first three groups, then some attempts may have to be made at discriminating between "levels" of certain behaviors.

Follow-up measures. The four groups (13

children) have now been seen for follow-up data on the multiple-response measures. These measures were taken anywhere from 1 to 4 yr after termination of our treatment. The children may be divided into two groups, those who were discharged to a state hospital and those who remained with their parents. The overall data on the 13 children are presented in Figure 7. Percent occurrence of the various behaviors are plotted on the ordinate for before (B) and after (A) treatment, and the latest follow-up (F) measures are presented. "I" denotes average results for the four children who were institutionalized (discharged to a state hospital), and "P" denotes data for the nine children who have lived with their parents since their discharge from treatment. The trends in these data may be succinctly described. The children who were discharged to a state hospital lost what they had gained in treatment with us; their psychotic behavior increased in frequency (self-stimulation and echolalia). They appear to have lost all they gained of social non-verbal behavior, and they lost much of what they had gained in appropriate verbal and play behaviors. On the other hand, the children who stayed with their parents maintained their gains or improved further.

Let us examine these children more individually, discussing the follow-up data of Rick and Pam first. When we terminated Rick and Pam's treatment we decided to recommend to their parents that their children be institutionalized. We based this decision on two major considerations. First, we had made the mistake of isolating the parents from their children's treatment, such that they did not receive the training we did in handling their children. Secondly, these parents had other large commitments to their families or themselves. For example, Pam's mother had just given birth to a child with severe brain damage which required continuous care, she had several other children, and Pam was not an easy child for anybody to handle. Ricky's mother was divorced and needed full-time employment. There were other con-

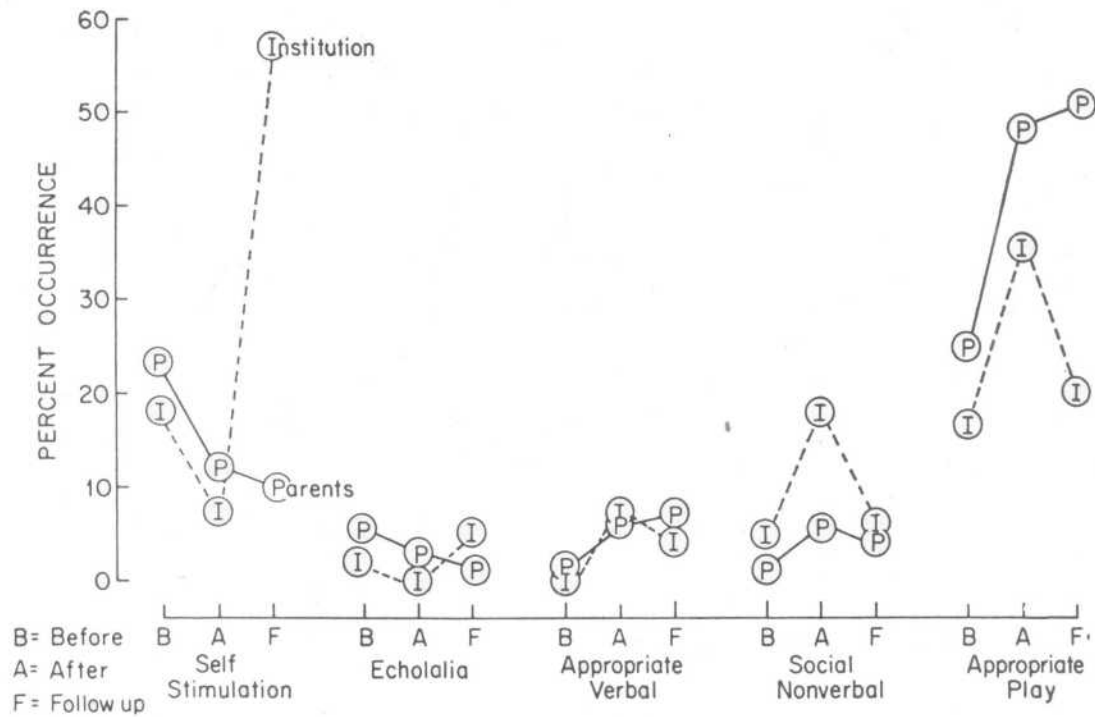


Fig. 7. Multiple response follow-up measures. Per cent occurrence of the various behaviors is plotted on the ordinate for Before (B) and After (A) treatment, and for the latest follow-up (F) measures. "I" refers to the average results for the four children who were institutionalized, and "P" refers to the average results for the nine children who were discharged to their parents' care. Per cent occurrence of the behaviors is presented on the ordinate.

siderations, involving the direction of effort on the research project. Provision for supervised foster home care, special schools, etc., were judged beyond our resources. In most treatment projects one is confronted with this option; either to invest one's time and resources into the treatment of a few children, or to concern oneself with replicability and generality of one's procedures for many children. When one runs a research project, one is fairly well restricted to the latter alternative.

Rick and Pam were discharged back to the state hospital they came from at the beginning of our treatment. It is difficult to specify the kind of environment a child enters when he becomes a patient in a large understaffed state hospital. Essentially they received custodial care. In any case, they did not receive behavior therapy; behavior therapy was new at that time, and considered harmful by most psychiatric

professionals. The emphasis in the state hospital was on "acceptance", which meant the children were encouraged to regress. The intensive demands we had made on them were not continued. For the most part, care consisted of attempts to make the children comfortable; they were allowed to self-stimulate and tantrum, they received some drug treatment, but the amount, kind, and duration of such treatment varied between the children, for any one child over time, etc. We now know (Lovaas and Simmons, 1969) that traditional interventions may worsen some psychotic behaviors. During this time, a foster home placement was attempted for Ricky, under the supervision of professionals with traditional orientations. Perhaps the turning point for the worse came for him when, after his school teacher reported that he was acting out in class, it was decided to remove him from school, instead of reprimanding him. And we shall show

that it would have been easy to prevent his subsequent relapse.

The 1968 follow-up measures on Pam and Rick are presented in Figure 8 (Follow-up 1, 1968). The figure also presents the before (1964) and after (1965) measures. It is apparent that in 1968, 3 yr after their first treatment, both children displayed a far lower frequency of

appropriate behaviors (speech, play, and social non-verbal are all down from 1965), and that their bizarre self-stimulatory behavior was significantly increased.

We decided to place them in the program a second time. The therapeutic effects of such an intervention would certainly provide a powerful demonstration of our treatment. It soon became

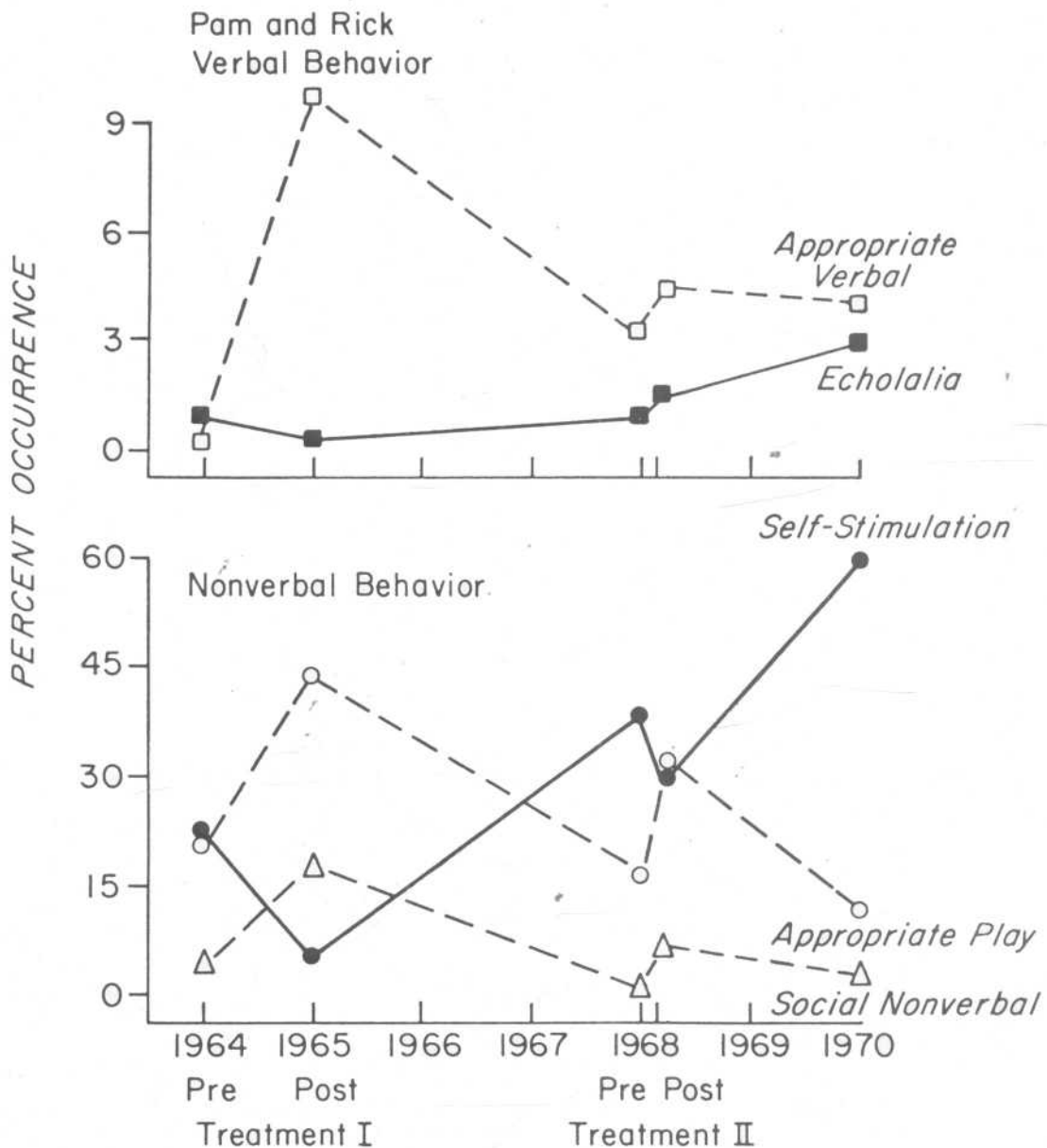


Fig. 8. Multiple-response measures for Rick and Pam presented Before (1964) and After (1965) first treatment, and for the first follow-up (1968), second treatment (1968), and second follow-up (1970). Percent occurrence of the behaviors is presented on the ordinate.

apparent that the children had not forgotten what we had taught them, but that their problem was essentially motivational. They were not afraid to behave inappropriately, neither did they behave appropriately in order to gain approval. The second treatment (Treatment II, 1968) consisted essentially of reinstating the contingencies they had experienced earlier. That is, if they behaved inappropriately (self-stimulated or were echolalic), they were punished. If they behaved appropriately, they were fed and approved of. The second treatment lasted 24 hr for Ricky and three weeks for Pam. The results for Treatment II, 1968 (Figure 8) indicate that this very short exposure to the program was reasonably effective. One can see a rise in the three appropriate behaviors (Appropriate Verbal, Appropriate Play, Social Non-Verbal) and a decrease in Self-Stimulation.

The children were institutionalized again and brought back for a final follow-up in 1970. As can be seen in Figure 8 (Follow-up II, 1970), they again regressed, as they had earlier. While Appropriate Verbal and Social Non-Verbal behaviors seemed to have remained stationary, one

can observe a loss in Appropriate Play, and a substantial increase in Self-Stimulation and Echolalia.

At the time that this report was written, Pam and Rick were both hospitalized. Clinically speaking, Pam showed few, if any, effects from the treatment she had received from us; she was extremely retarded in most areas of function. Rick had fared better. He was still definitely autistic, but the psychologist who examined him noted that he had received much training and had developed to a greater extent than many other autistic children. He was more verbal and rarely echolalic. He had generally flat affect although he could smile and showed some positive feeling and was definite about what he wanted—showing little ambivalence.

Billy and Chuck were discharged to the same institution under conditions similar to those of Pam and Rick. Their parents were essentially untrained and had other serious commitments and personal difficulties. At the time of Chuck and Billy's follow-up 4 yr later, they had retained most of their gains in Appropriate Play and Social Non-Verbal behaviors but they showed

Table 2

Group 2, 3, and 4 individual and averaged multiple-response measures per cent occurrence of each behavior is tabled for before (B), after (A), and latest follow-up (F) measures.

	<i>Self-Stimulation</i>			<i>Echolalia</i>			<i>Appropriate Verbal</i>			<i>Social Non-Verbal</i>			<i>Appropriate Play</i>		
	B	A	F	B	A	F	B	A	F	B	A	F	B	A	F
<i>Group 2</i>															
Taylor	74	32	11	00	01	00	00	00	00	00	00	03	01	05	38
Mike	18	06	27	00	01	00	00	00	00	00	06	02	17	56	39
Jose	53	01	04	00	01	00	00	00	00	01	02	03	07	62	65
Mean:	48	13	14	00	01	00	00	00	00	00	03	03	08	41	47
<i>Group 3</i>															
Leslie	12	04	19	10	11	01	03	12	16	03	10	11	47	58	43
Tito	02	15	08	01	02	01	01	05	18	00	01	02	07	42	51
Seth	12	12	18	04	00	00	05	11	02	03	08	04	31	56	36
Mean:	09	10	15	05	04	01	03	09	12	02	06	06	28	52	43
<i>Group 4</i>															
Kevin	03	02	03	25	11	00	02	16	07	02	02	02	51	79	68
Ann	25	19	02	03	02	05	03	05	13	02	21	02	26	25	66
James	07	08	00	00	00	00	00	01	06	11	02	01	44	55	49
Mean:	12	10	02	10	04	02	02	07	09	05	08	02	40	53	61

losses in Appropriate Verbal behavior, increased Echolalia, and showed a marked increase in Self-Stimulation. At the time of this report, Billy was in a foster home, where his adjustment was said to be marginal. Chucky's mother may be able to take Chuck back home.

The follow-up data on Groups 2, 3, and 4 are essentially presented in the group denoted by P (parent) in Figure 7. The children were at home with their parents, who had received some training in how to continue treatment with their children. Group 2 parents received less training than Group 3 and 4, and their children were treated on an inpatient basis. Group 3 and 4 children lived with their parents throughout treatment and their parents received extensive training by us. The children were evaluated from 2 to 4 yr after termination of treatment. The data on the individual children are presented in Table 2.

Considering the grouped data, we can see that the children whose parents were trained largely retained their gains or continued to improve. The gains the children made in appropriate play,

social non-verbal, and appropriate verbal behaviors were usually retained. The data indicate some increase in self-stimulatory behavior after treatment was ended.

Intelligence and social maturity. We have obtained IQ scores for 19 of the 20 children. Figure 9 shows the changes in these measures during the time the children were in treatment. IQ scores are plotted on the ordinate for before (B) and after (A) tests. The dotted lines indicate that the patient was untestable. Most of the children showed substantial changes with treatment, functioning in the mildly to moderately retarded level by the termination of treatment, while they were previously untestable. "Untestable" means that the children would not respond to the examiner's attempts to test them. For example, they would not sit in a chair if asked to do so, and they remained oblivious to the testing materials that were presented to them. After treatment, the children would cooperate; that is, they would respond to the examiner and engage in the behaviors he wanted (such as block building, etc.). Some of this change reflects extinction

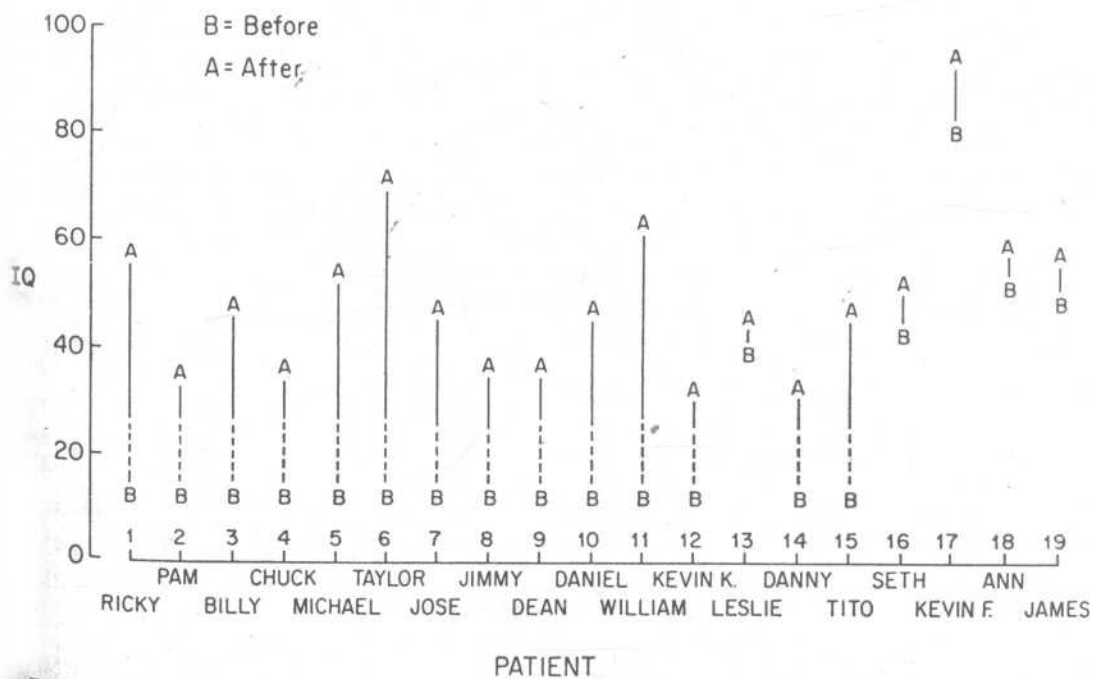


Fig. 9. IQ scores Before (B) and After (A) treatment. Dotted lines indicate the patient was untestable before treatment.

of interfering behaviors, while some reflects genuinely new acquisitions. It is an open question whether these changed IQ scores would be predictive of the children's future performances in school.

We obtained Vineland social quotients for the last 14 children we treated, and all of the children made large gains. The mean social quotient before treatment was 48, with a standard deviation of 20; the mean quotient after treatment was 71, with a standard deviation of 27. The changes that took place in social maturity are consistent with the IQ data in that the children showed large gains in their ability to look after their own practical needs. Much of this change was again due to a reduction in bizarre behavior, and the achievement of elementary social stimulus control. As with the other measures, there were no exceptions to the improvement; all 14 of the children had higher quotients after treatment than they did before treatment. We did not obtain follow-up data on these measures.

RELIABILITY

We attempted to solve the problems associated with reliability of the multiple-response recordings in two ways. First, we maintained, for the majority of recordings, at least two trained observers who were assigned in a non-systematic fashion to do the recordings. These observers changed over time, such that those who scored for the second half of a year were often different from those who scored the first half. This means they had different degrees of familiarity with the children. These steps, of rotating observers and bringing in new ones, probably helped to reduce observer bias in the recordings. Each new observer received about three to six 1-hr training sessions. The various behaviors were defined for him, he became familiar with the apparatus, and he worked with an experienced observer until the average difference between their scores over all behaviors was less than 20%. This was calculated by dividing the difference between given pairs of scores by the average of the two

scores, and then averaging these percentages for all behaviors. The reliability between observers was checked on a monthly basis, and if they exceeded the 20% given above, they were retrained. Table 3 presents pairs of observers' recordings randomly drawn from data on the first group of children. The table reflects high agreement between observers. The table also shows that during the first year we recorded physical contact, tantrums, and vocalizations, which we omitted in subsequent recordings. Physical contact and vocalizations were omitted because we felt they were not directly related to the child's chances for success in "normal" society. Tantrums were omitted because they were low-frequency behaviors that changed very quickly; therefore, no change over time was apparent in our measures.

The high degree of agreement between observers, and the use of non-systematic assignment across observations, supports the argument that the data can be replicated, and that they are not the product of a particular observer. However, observers' familiarity with the study, and knowledge that the children were in treatment may have increased the probability that ratings would reflect improvement. This seems unlikely when one considers the explicitness with which the behaviors were defined and the subtle changes that often appear in the data. These kinds of changes would be difficult to "fake" and hence appear to validate the children's improvement. Nevertheless, the study is greatly strengthened by the following investigation, which demonstrates that naive observers, scoring the sessions in a random order, produce data similar to that of our experienced observers.

Three observers, who were not familiar with the children and the purpose of the study, were introduced to the recording system and presented with video recordings (Sony EV 200 1-in) of the children displayed on an 8-in. television monitor. One tape was used for training purposes. The tape selected was a pre-treatment tape of a 7-year-old echolalic autistic child who displayed all of the behaviors the observers were required to

Table 3
Randomly Drawn Pairs of Observers' Recordings from the First Group of Children

<i>Child, date</i>	<i>RICKY—7/16/64</i>						<i>BILLY—6/25/64</i>						<i>PAM—4/1/65</i>					
	<i>Alone</i>		<i>Attending</i>		<i>Inviting</i>		<i>Alone</i>		<i>Attending</i>		<i>Inviting</i>		<i>Alone</i>		<i>Attending</i>		<i>Inviting</i>	
	<i>Joan</i>	<i>Gail</i>	<i>Joan</i>	<i>Gail</i>	<i>Joan</i>	<i>Gail</i>	<i>Gail</i>	<i>Irene</i>	<i>Gail</i>	<i>Irene</i>	<i>Gail</i>	<i>Irene</i>	<i>Diane</i>	<i>Joan</i>	<i>Diane</i>	<i>Joan</i>	<i>Diane</i>	<i>Joan</i>
Tantrums	0	0	2	2	1	2	0	6	0	0	1	0	0	0	0	0	3	0
Self-stimulation	25	25	7	7	15	14	3	5	2	1	2	1	30	27	16	11	6	1
Vocalization	0	0	5	4	1	1	11	14	6	6	8	9	0	4	0	0	1	1
Echolalia	0	0	0	0	1	1	0	0	0	0	0	0	8	3	1	0	7	9
Appropriate Verbal	-	-	0	0	2	2	-	-	0	0	0	0	-	-	3	1	11	13
Social Nonverbal	-	-	0	0	15	11	-	-	0	0	3	3	-	-	46	54	16	19
Physical Contact	-	-	0	0	0	0	-	-	0	5	1	1	-	-	7	7	1	6
Appropriate Play	18	20	28	31	45	45	7	7	6	2	22	22	14	12	73	63	31	36

record. Observers were given a total of nine 1-hr training sessions. In the first three sessions, observers viewed the tape to familiarize themselves with the behaviors to be scored, learned the position of the keys on the board, *etc.* These three informal sessions were followed by six sessions, after which observers' raw data were reduced. The observer was given feedback about whether his scores were high or low in relation to the mean score for a particular behavior over the three observers. The tape was then replayed to all observers, and the instances where they had disagreed were discussed. This kind of training was like that given to the observers who observed the children *in vivo*. The recordings were different in that the observers were required to divide social non-verbal and appropriate play each into two levels (see the Instructions for Observer Identification), which, with eye-to-face and self-stimulation, demanded the recording of eight behaviors. During the first three sessions, observers were required to record all eight behaviors simultaneously. This proved too difficult, so we then decided to simplify the task by playing the tape twice and asking the observer to record only four behaviors per observation session. Low agreement during Session 4 (Figure 10) reflects an adjustment to the new procedure to some degree. In addition, observers rarely disagreed about which category a particular behavior represented. Disagreements between observers were usually disagreements about the onset and termination of a particular behavioral segment. Much of this disagreement was basically a problem of skill, that is manual dexterity. One must observe, categorize, and record simultaneously. This means that the observer could not also look at the keyboard. The observer had to make decisions about which key to depress, at what time to depress it, and for how long. He had to observe, categorize, and record three other behaviors. It is a complex, on-going process and a mistake may be made at any of several points in the chain. The two most common errors were continuing to depress a key after a particular behavior had ceased and missing the precise

onset of a particular behavior because one or more of the other recordable behaviors were present.

The data showing the acquisition of agreement between these three observers over the last six sessions are given in Figure 10. In all cases, a mean figure for the three observers for any one behavior during a given condition was obtained. The data reduction was carried out as follows: the onset of the tape was clearly marked by the observer on the computer tape by the observer depressing all the keys for 5 sec. The end of the tape was similarly noted. Each tape was then divided into 10-sec segments. Agreement was recorded in an all-or-none fashion for each 10-sec segment. If two observers agreed, a "+1" was recorded. If they did not record the same behavior during that particular period, a "0" was recorded. Per cent deviation then refers to the per cent of these 10-sec segments during which a particular observer failed to record a behavior recorded by either of the other two observers or during which a particular observer recorded a behavior not simultaneously noted by either of the other two observers. Per cent deviation was averaged over all conditions and behaviors for a given observer for each session. O-7 (Observer 7), then, shows an average per cent deviation of 35% during the first session. The figure shows that observers eventually learned agreement.

After the completion of this reliability training, one of the observers (O-7 in Figure 9) scored the pre- and post-treatment tapes on three children (Michael, Ricky, and Jose) to assess whether a naive observer could record the treatment effects in agreement with our experienced observers. She scored the tapes in the following order: Mike's pre-treatment, Mike's post-treatment, Jose's post-treatment, Ricky's post-treatment, Ricky's pre-treatment and Jose's pre-treatment. All tapes were scored in order of condition Alone, Attending, and Inviting except Mike's, where the conditions were scored in order of Attending, Inviting, and Alone.

This is a direct test of the reliability of our

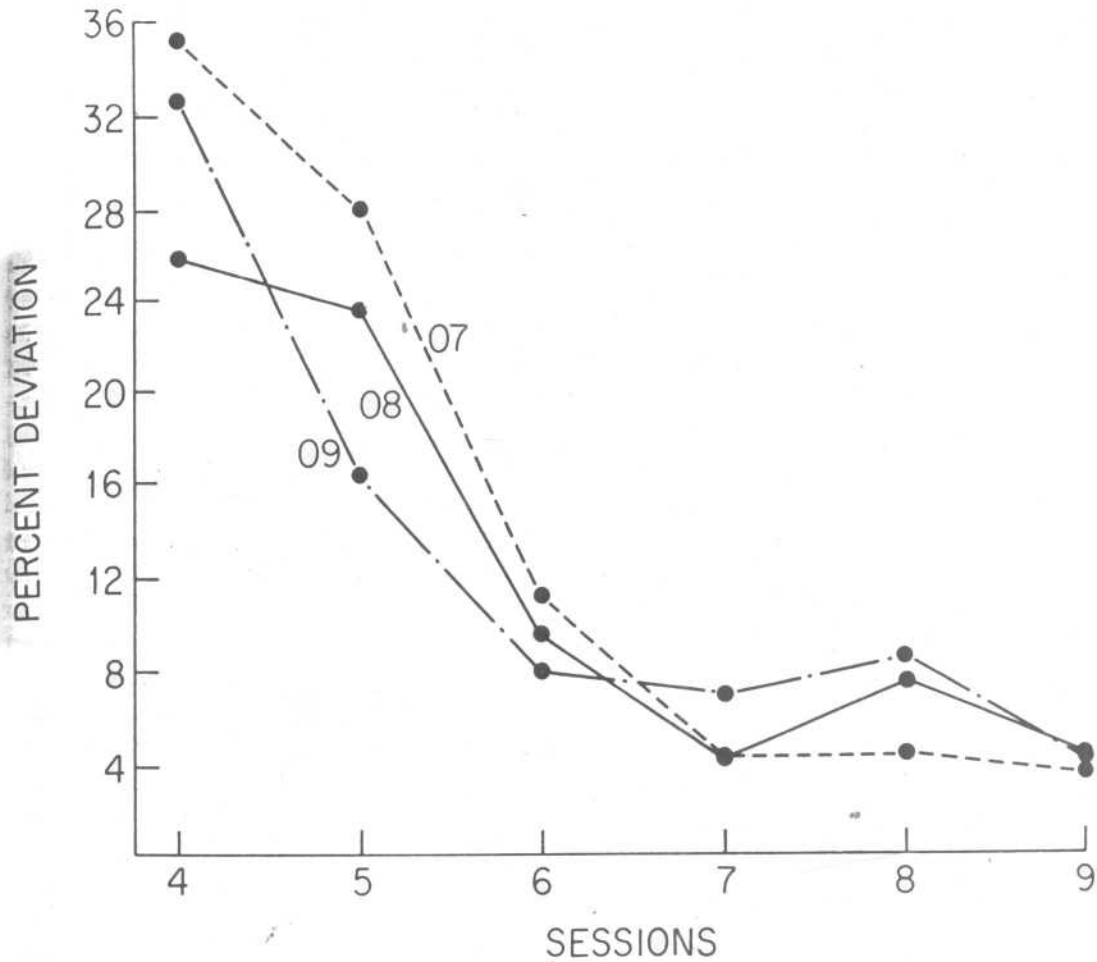


Fig. 10. The acquisition of agreement between observers 7, 8, and 9. Per cent deviation from the mean of the three observers' scores is plotted on the ordinate for the last six sessions. Per cent deviation is averaged over all conditions and behaviors for a given observer for each session.

Table 4
Tape and *In Vivo* Comparisons of "Before" and "After" Multiple-Response Measures

	JOSE		MICHAEL		RICKY							
	<i>Tape</i>		<i>In Vivo</i>		<i>Tape</i>		<i>In Vivo</i>					
	<i>Before</i>	<i>After</i>	<i>Before</i>	<i>After</i>	<i>Before</i>	<i>After</i>	<i>Before</i>	<i>After</i>				
Self-stimulation	33	0	53	1	11	1	18	6	10	15	25	13
Appropriate Verbal	0	0	0	0	0	1	0	0	1	2	0	3
Echolalia	0	2	0	1	3	6	0	1	1	3	1	2
Social Nonverbal	0	1	1	2	1	2	0	6	1	2	0	2
Appropriate Play	6	54	7	62	24	46	17	56	33	52	16	26

scoring procedures, as it removes potential effects for both observers' familiarity with the experiment, and the order in which they recorded the behaviors. Table 4 presents the comparisons between the scores obtained by an experienced observer recording the child *in vivo*, and a naive observer scoring off the tape of that same session. The absolute values between the two observers are different, probably because the video recording reduces fidelity (*i.e.*, there is particularly less fidelity with speech and facial expression). However, the naive observer recorded increased frequency of normal behaviors, and decreased frequency of pathological behaviors, similar to the changes recorded by the experienced observer. The only exception to this is Ricky's self-stimulation, which was observed to increase during the "after" measures when scored off the tape. Observers suggested that this may have been due to the large amount of Ricky's facial contortions, which were detected and scored as self-stimulation in the *in vivo* pre-treatment condition, but which could not be detected on the tape.

To summarize our findings concerning the reliability of our recordings: (1) there was a high degree of agreement for pairs of experienced observers who were randomly assigned to do *in vivo* recordings; (2) naive observers, unfamiliar with the study and children, could also be trained to show high agreement in their scoring of video recordings of the children's behavior; and (3) the direction of the behavioral change in treatment was scored essentially the same for a naive observer scoring pre- and post-treatment video tapes in a random order as it was for experienced observers scoring the sessions *in vivo*.

It may be appropriate to note here that the replicability of these results on reliability may be open to question, as the procedure requires that the trainer be very familiar with the types of errors naive observers are likely to make and to be able to note where such an error has occurred and provide appropriate feedback to the observer.

DISCUSSION

In summary, the major results of this study are that: (1) inappropriate behaviors (self-stimulation and echolalia) decreased during treatment, and appropriate behaviors (appropriate speech, appropriate play, and social non-verbal behaviors) increased; (2) spontaneous social interactions and the spontaneous use of language occurred about eight months into treatment for some of the children; (3) IQs and social quotients reflected improvement during treatment; (4) there were no exceptions to the improvement, however, some of the children improved more than others; (5) follow-up measures recorded 1 to 4 yr after treatment indicated that large differences between groups of children were related to the post-treatment environment (those groups whose parents were trained to carry out behavior therapy continued to improve, while children who were institutionalized regressed); (6) a brief reinstatement of behavior therapy could temporarily reestablish some of the original therapeutic gains made by the children who were subsequently institutionalized; and (7) the technique utilized for recording therapeutic change was reliable. Observers were trained to be able to recognize and record specific behaviors, the presence or absence of which may be considered diagnostic of autism.

Individual Differences

While the major findings listed above characterize each of the children in the group, there has been considerable heterogeneity with respect to the degree of improvement shown by each child. The delineation of "autism" is one area that will demand considerably more work. It has not been a particularly useful diagnosis. Few people agree on when to apply it. It is not a functional term in the sense that it is neither related to a particular etiology nor to a particular treatment outcome. Our children responded in vastly different ways to the treatment; Rick learned, in 1 hr, what Jose learned in 1 yr. Since there was such heterogeneity among the patients,

three fairly representative case descriptions will be presented in order to give a picture of the clinical implications of our findings: first, we present Scottie, who showed considerable gains in treatment; second, Tito, who showed moderate gains; and finally, Jose, whose progress was minimal.

Scottie, who was 4.5 yr of age at the start of treatment, spent much of his time staring into space and did not attempt to initiate interactions with people. If he was directly addressed, he would show a passive and friendly interest. When left to himself, he would self-stimulate; he was particularly attracted by spinning wheels. Scottie was echolalic, and he could label common objects, but he had very little communicative, and no spontaneous speech. He had to be washed and dressed by others, helped when he ate his meals, and he was not fully toilet trained. He was, however, relatively free from tantrums, and he could understand simple directions. Because of his social responsiveness when approached, and occasional appropriate play, he was considered less psychotic than most of the children we have seen. His social quotient was 68.

Initial treatment sessions took place in his home, lasting 2 to 4 hr each, several days a week. His treatment plan included programs designed to teach him communicative skills, as well as the behaviors necessary for him to take care of his own practical needs, and to take a more active part in his everyday life. He was taught common abstract terms, such as prepositions, words denoting temporal relationships as well as counting, singular *versus* plural, *etc.* The frequency of echolalia was decreased. Meals and almost all daily activities were made strictly contingent on verbal requests for them. At first he missed several meals. He was also taught to offer materials as well as to request them. Much emphasis was placed on conversational speech. He was asked a general question about something he had done, and then was asked progressively more specific questions. Any spontaneous responses he made were reinforced socially,

since he was responsive to social reinforcers. Again, the language program (Lovaas, *in preparation*) gives a rather complete discussion of these procedures. We focused our efforts on consultations with his grandmother, who was brought in to take special care of him, and who held him to a highly demanding schedule. She was different from most of the parents we had seen, since she did not tolerate withdrawal or other expressions of pathology.

Scottie is presently attending third grade in a normal elementary school. His social quotient is 100. He shows no trace of autism, and in all aspects must be considered a normal child.

Tito, at admission, was a hyperactive 5-yr-old boy who evidenced an extremely short attention span. Eye contact was absent. He had many compulsive rituals. For example, he would spend considerable time arranging objects in a straight line and would become very upset if the arrangement was disturbed. He "refused" to let his parents read the Sunday paper by becoming very upset when they removed the string that tied it in a bundle. He was untestable on intelligence tests, and he obtained a social quotient of 52. He was echolalic, but occasionally would use speech for communication. Tito's understanding of speech, however, was minimal. He resisted any involvement with people, and was unresponsive to displays of affection. He was extremely negative and clever at getting himself out of situations he did not like, often responding to the most elementary demands (*e.g.*, "sit in the chair") with extreme tantrums.

Tito was treated as an outpatient. He was seen for three sessions per week for 1 yr. We served primarily as consultants to his mother, who was very conscientious and warm. His treatment program included two main objectives. First, we tried to teach him to deal with frustrating situations more maturely rather than engaging in tantrums. This was an extremely demanding job, and he received many spankings. Secondly, he was taught those basic skills upon which he could build more complex behaviors, particularly in language. Included in this cate-

gory were pronoun usage, preposition usage, number concepts, relational concepts such as big and little, and social greetings. He was taught to comment upon his environment. He was also taught to draw, and to play more appropriately.

At discharge, Tito was observant and alert, but still appeared definitely educationally handicapped. He had made some progress in most areas, but his biggest gains were in language. His speech was quite spontaneous, and he could comment correctly on most social interactions. He obtained an IQ of 47 on the Stanford Binet and a social quotient of 63 on the Vineland. He now attends a school for retarded children 3 hr per day. His mother reports that he has continued to show improvement in most areas. He remains emotionally aloof with strangers, but is close to his mother. His future is uncertain. He may escape institutionalization if his progress continues.

Jose was 4 yr old at the start of treatment. His extreme negativism was reflected in tantrums, biting, and extreme stubbornness. He did not play with peers. He did not respond to his name or any commands. He had no speech, could not dress himself, was not toilet-trained, nor did he have any other self-help behaviors. Appropriate play was essentially absent. He was found to be untestable on intelligence tests. He had a social quotient of 59. In short, he was extremely behaviorally retarded.

He was treated as an inpatient at the UCLA Neuropsychiatric Institute for 1 yr. His mother was given some limited training in how to continue therapy with him as described above. His treatment was primarily designed to overcome his negativism and to build some basic language skills. The latter included simple labeling, color discrimination, response to simple commands, and form discriminations. Some work was also done on the reinforcement of spontaneous babbling.

We probably made slower progress with Jose than with any of the other children. At discharge, his gains in language were minimal. He would obey some commands; his vocabulary in-

cluded a number of common nouns, some names, and a few verbs. He would use these words to label objects or express a desire for something, but never for commenting. He would attempt to imitate new words spontaneously on occasion. His greatest improvements were social. His social quotient of 74 reflected increases in smiling, laughing, and self-help skills. He was partially toilet trained. He was testable on the Stanford Binet Intelligence Scale (IQ = 47).

At present, Jose uses only a few words spontaneously (*e.g.*, "car", "go to school", *etc.*), and what he has retained of his speech training is negligible. He can take care of himself at dinner, and is fully toilet trained. His greatest gains at home have been in his play, which has become elaborate and creative, enabling him to entertain himself. He appears indifferent to people. While at intake he appeared unaware of ("blind" and "deaf" to) social contact, it now looks more as if he does not care whether anyone is there or not. He has to be watched constantly. Otherwise, he runs away from home, going nowhere in particular. His parents fear he will be killed because he is unaware of many common dangers. His parents plan to place him in a nearby state hospital. He will be able to come home on weekends and short vacations. We concur in these plans.

It is important to note that given the considerable heterogeneity among patients diagnosed as autistic, it is not enough merely to say that one has treated autistic children. Considering that some children improve without treatment, and that these children are differentiated from those who do not by certain behaviors, a good diagnostician can select his patients so that the majority of the children would eventually improve independent of the treatment offered. No doubt such pre-selection of patients, which would yield a much more favorable base rate change than is true of a nonselected group, keeps many non-functional treatments alive.

We utilized several procedures that allow us to argue with some confidence that autistic chil-

children improve with behavior therapy. First, we have performed two within-subject replications (on Rick and Pam), and in both instances demonstrated that we could establish behavioral control at will over the course of time for these patients. Therefore, we can argue that their behavior change must have been due to our intervention. Secondly, we performed several between-group replications. All replications yielded similar results. Also, each group of children was treated independently of the others, demonstrating that we could replicate our treatment effects independent of any conditions that may have been specific to one group of patients. We attempted specifically to avoid pre-selecting patients who might provide a favorable base rate of change regardless of treatment. The majority of patients were selected specifically because they displayed behaviors (IQ less than 50, mutism or no appropriate play) that were considered to be poor prognostic indicators. In addition, we did not drop any patient from treatment once we began. Finally, we consider our measures to be socially meaningful, and independent of any given theory of autism.

Further evidence that directly supports our results has been provided by other behavior therapy programs, which also demonstrated improvement with autistic children (Hewett, 1965; Risley and Wolf, 1967; and Wolf, Risley, and Mees, 1964). Furthermore, Wolf, *et al.* (1967) provided some additional data that lend support to our follow-up results. Such data suggest that replication of the data presented here is practically feasible.

Reservations about the follow-up data. We have some confidence in the inferences we have drawn about the effectiveness of behavior therapy from the Before and After measures. The study was designed primarily to assess for change during treatment. But the reader should view the follow-up data with certain reservations; we did not initially design the project with follow-up assessment in mind and a number of variables were left uncontrolled. In retrospect:

- (1) The children should have been randomly as-

signed into the hospitalized *versus* parent-trained groups so as to minimize child characteristics associated with differential prognoses. Age of the child, testability on standard psychological tests, amount and kind of play at intake are some of the more obvious variables that should have been equated across groups. (2) We should have exercised more control over the post-treatment environments. For example, we were unable to assign children to special educational institutions, yet presence or absence of school is probably an important determinant of subsequent improvement or relapse. Similarly, future studies would do well in developing more objective assessment of the extent to which the parents are in fact continuing to implement the treatment procedures. Not all parents are equally good as behavior modifiers. The more successful parents seem to have the following features:

- (a) A willingness to use strong consequences, such as food and spankings, to be emotionally responsive, showing their anger as well as their love. Not all parents can do this; some people are more "gentle" or more permissive than others, preferring to let their children "grow" or "develop" while they, as parents assume a spectator role. Such parents do not do well with their autistic children.

- (b) The willingness to deny that their children are "ill". This means that they deny the child the "need" to be sick, and, instead, give him some responsibility.

- (c) The willingness to commit a major part of their lives to their children and to exercise some degree of contingent management throughout the day. This virtually rules out any professional or extensive social interests on the mother's part, requires a stable family structure, *etc.* Parental assistants, such as special tutors, can help out at times, but ultimately the parents must bear the major responsibility.

Despite these limitations, we have included the follow-up data because their validity seems strengthened when one considers the study as a whole. Basically, it is not surprising that the

children who were "discharged" to their parents tended to improve, because these children remained in a treatment environment.

Major strengths. Behavior therapy programs for autistic children help. That is their major asset. So far as we know, despite its limitations, it is the only intervention that is effective. Our program did not give everything to every child. Sometimes it gave very little to a particular child, but it did give something to each child we saw. The improvement was analogous to making from 10 to 20 steps on a 100-step ladder. Scotty probably started at 80 and gained 20; his treatment brought dramatic changes, he became normal and his change is irreversible. Jose, on the other hand, may have started at 10 and gained 10; the change was not all that dramatic.

We have been especially successful at suppressing self-destructive behavior. In minutes, we have been able to stop the self-destruction of children who have mutilated themselves for years. The suppression was highly discriminated (situation specific), but this merely meant that we had to apply the treatment in more than one environment.

We have also been successful at rearranging behavior. For example, if the child was not mute (if he already had psychotic speech), then we could help him make large strides in his language and intellectual behavior.

The gains the children made in treatment generalized. Our multiple-response measures clearly demonstrate that we obtained stimulus generalization. We realize that sometimes the generalization was not as broad as we might have suspected. The shift in our program to teach the parents to treat their own children is a direct attempt to build greater stimulus generalization.

We were interested in response generalization as well as stimulus generalization. We do have some data on the subject, but they are limited. Response generalization, like stimulus generalization, deals with efficiency. How much behavior can one get for free? We obtained

changes on the IQ tests for free; that is, we did not train the children on the test items, yet they improved. The data on response generalization are limited because psychology has little to say about it. We had no information about what changes to expect, which made it difficult to assess changes this first time around. From casual observation, during the treatment, the children looked much more "alert" and "aware" and became more effectively prosocial. It was particularly dramatic when, during the continuous demands of the therapy hour, a child would suddenly start sobbing and put his head in the therapist's lap; or when, after much hard work, a child appeared delighted over his new mastery. One does not often observe appropriate affect in autistic children.

Certain changes were difficult to assess because they seemed outside our behavioral framework. The children who were chronic toe-walkers (one of the soft neurological signs) began to walk normally after four or five months of treatment. Children who had never slept normally through the night began sleeping for 10 hr without interruption. Children who were chronically diarrhetic began having firm stools, etc. It is the search for this kind of generalized behavioral change that we feel will be particularly useful in future research.

Major weaknesses. There are many disappointments, and only a full appreciation of these will enable more realistic hopes now, and solutions in the future. We will discuss the major problems.

The most significant disappointment was the failure to isolate a "pivotal" response, or, as some might describe it, the failure to effect changes in certain key intervening variables. This means that in the beginning, we searched for one behavior which, when altered, would produce a profound "personality" change. We could not find it. We had once hoped, for example, that when a child was taught his name ("My name is Ricky") that his awareness of himself (or some such thing) would emerge. It did not. Similarly, the child who learned to

fixate visually on his therapist's face did not suddenly discover people. Our treatment was not a cure for autism. But we had to start somewhere. At least the child who learned his name was then in position to learn someone else's name. When he learned to fixate visually on his therapist's face, he could pay more attention to teaching cues.

The failure to isolate a pivotal response (or change a crucial intervening variable, or cure autism) can be discussed in two different veins. First, behavior therapy may not be the correct approach to treating autistic children. One could suggest in this regard that the "underlying pathology" (the intervening variable) is biochemical, and that the early detection and correction of this imbalance will enable the child to learn from his everyday environment, with little or no special educational remediation. If this should prove to be the case, our research has limited ultimate clinical value for autistic children. Behavior therapy alleviates some problems now, but the ultimate solution may involve correction of a biochemical imbalance. This is a viable alternative and intelligent people will pursue it.

Speaking of physiological variables, it is possible, of course, that the pathology is structural. It could be that something is non-functional, as is the case with the blind or deaf child. The repair of this structural deficit, an attempt to "connect up" the millions of neurons in order to correct or bypass the deficit, is beyond the limits of present medical technology. In this case, we must approach the problem as we approach blindness or deafness. In other words, it is sometimes the case that even where the underlying pathology is neurophysiological, the only feasible treatment may be essentially psychological. It is the early detection of blindness and deafness, and subsequent special remedial environments that allow blind and deaf children to develop normally. Without special consideration, they closely resemble the autistic children. The perceptual deficit that may underlie autism is more difficult to assess, hence it is more

difficult to remediate. We will speculate on some possible basis for such a perceptual deficiency later in this paper.

The second possibility is that behavior therapy is the correct approach, and that it is the problem that is erroneously conceptualized. A full discussion of this point involves evaluations of terms like "mental illness", "treatment", *etc.*, and that is beyond the scope of this paper. But it is important to bear in mind that "autism" is a hypothetical construct and a very shaky one at that. We have been expected, in a sense, to cure the children of someone else's inferences about them. There are no studies on "autism" that point to either a common etiology or a common response to treatment (or even a common response to experimental situations of much more limited scope). "Autism" was coined before a functional analysis of pathology. The public and emotional appeal associated with the term, not our scientific understanding of the "syndrome", helps the term survive.

There are other conceptual problems. Consider, for example, a behavior as easy to build as "looking at the therapist's face". In data language this behavior has a very limited meaning. However, the associated theoretical structure is very extensive, implying that the child is recognizing and evaluating another person. For autistic children, the behavior of looking or not looking at another person has acquired special significance on a purely conceptual level (*cf.* Hutt and Ounsted, 1966). According to some conceptions of the problem, one would expect major changes in the autistic child who started to look at others. In our research, we have found these changes to be of minor significance.

Research problems. We had anticipated that the children in the state institutions would regress. From the beginning, we have published studies (Lovaas, 1967; Wolf, *et al.*, 1964) that show that when the experimental reinforcers were withdrawn, behaviors weakened. Such extinction occurred whether we employed food or fear as reinforcers, and whether the behavior was physical contact, imitation, or abstract

speech. The shift from response to time-contingent delivery, the very procedure that we employ to demonstrate the effectiveness of our main treatment variable, also demonstrates our major weakness, the tenuousness with which the behavioral gains were maintained. Reversible baselines help one's research design, not one's patients.

The reversibility of the treatment effects was most dramatically observed with the first four children we treated: Rick, Pam, Chuck, and Billy. Although each one was making progress at the time of discharge, when we assessed them after 4 yr in the state hospital they appeared to have made no gains in appropriate behavior during that time, and showed large increases in self-stimulation. It is probably worth elaborating on this point, that they had learned nothing in the intervening 4 yr, but seemed to have stayed still. They gave the same smiles, the same facial expressions, the same words. This again emphasizes the point that without therapeutic, prescribed, contingent, functional, reinforcers, children like these do not improve or retain their improvement; and, since we are not yet in a position to help them acquire normal, social reinforcers, their post-treatment environment has to be controlled. In our philosophy, functional contingencies are reality; if removed, any child would fail to develop.

The reversibility of the treatment effects are not peculiar to autistic children. It has been observed in a large variety of behavior therapy programs, and has led Bandura (1969) to speak of the distinction between physical and psychological treatments. The work of others illustrates these problems. For example, Tharp and Wetzel (1969) attempted to avoid discrimination and reversal problems by placing the treatment in the hands of those persons who have control over most of the patient's reinforcers. Our failure to maintain the gains made by the first four patients underscores the need for that kind of intervention. Wahler's (1969) study provided a good illustration of the need for interventions across settings. When the chil-

dren he saw were treated in school, they did not necessarily change at home. When the contingencies were instated both in the home and at school, the behavior changed in both settings. Many therapists (*cf.*, Patterson and Bechtel, 1970) now argue that the child's parents are essential as mediators of treatment.

The extent of the reversibility of treatment effects will probably be some function of various patient characteristics. It seems that when the primary problem centers on the child's motivation, and when treatment relies on "artificial" or experimental reinforcers (stimuli that do not characteristically maintain the patient's behavior on the outside), then one invites certain problems. Food, slaps, and accentuated social reinforcement are not the reinforcers that maintain the daily behaviors of normal school-aged children. Our use of such reinforcers set up the exact conditions for the kind of discriminations and the kind of extinctions we did not want. The necessity for using primary reinforcers, rather than everyday, more natural ones, is a probable reason why the children regressed in the state hospital environment. The state hospital, since it did not prescribe contingent functional reinforcers, constituted an extinction run.

It is implied in the above discussion that we view the problem of maintaining the treatment gains (generalization over time) as a special case of stimulus generalization. When the child stayed home with his parents who had learned our techniques he did not regress (*i.e.*, extinguish) because the environments before and after discharge were similar (*i.e.*, we maintained stimulus generalization). However, the child who was discharged to a state hospital entered a new environment to the extent that it did not possess (or did not program) effective reinforcers. Remember that the children had not "lost" the behaviors we had given them (some "progressive disease" had not rotted their brains), they simply did not perform, they were unmotivated, unless we re-exposed them to the treatment contingencies. The point is that it is important for research in behavior therapy to be directed towards

ways of normalizing reinforcing functions so as to smooth the transition (prevent a discrimination) between the treatment and post-treatment environments.

The second major research problem centers on how to develop procedures for accelerating the acquisition of new behaviors. No doubt, the slow rate with which some behaviors were acquired was based on the children's inadequate motivation, as discussed above. However, it also seems to be the case that autistic children show deviations in attentional behaviors and that these deviations slow down their acquisition, particularly of that kind of learning that requires shifts in stimulus control. Autistic children appear to respond in an overselective manner to multiple cues. We referred to this problem as stimulus overselectivity (Lovaas, *et al.*, 1971). This finding has led us to consider redesigning many of our teaching procedures. For instance, we may well have to minimize our use of supporting prompts and prompt fading techniques, as these may provide interfering rather than helping stimuli in the learning process of these children.

Perhaps it is stimulus overselectivity that prevents or slows down the autistic child's acquisition of secondary reinforcers as well. If primary reinforcers had been used as consistently with normal children as we used them with our autistic patients, the associated environment would probably have acquired a larger range of reinforcing function. If secondary reinforcers are acquired classically (through the simultaneous presentation of neutral with already functional stimuli) then the autistic child, being overselective, may fail to respond to one of these inputs, and conditioning should then fail. Behavior therapy with the normal child, then, may not require detailed consideration of programs designed to build reinforcing function. To ensure long-lasting effects in the autistic child, the process of building secondary reinforcers would seem to require much more effort.

Finally, a major focus of future research should attempt more functional descriptions of

autistic children. As we have shown, the children responded in vastly different ways to the treatment we gave them. We paid scant attention to individual differences when we treated the first 20 children. In future, we will assess such individual differences, for example, by contrasting the effectiveness of behavior therapy for very young *versus* older autistic children. Presumably, the very young (before 2 yr of age) child should discriminate less well than the older child, hence generalize and maintain the therapeutic gains to a more optimal degree.

It is important to remember that behavior therapy is a treatment based on research rather than deduced from theory. It is a technology for producing behavioral change through environmental manipulations. Sensitivity to new findings produces constant change in treatment techniques. Only a short time ago, imitation training procedures were developed (*cf.*, Baer and Sherman, 1964). Our treatment changed and its effectiveness was greatly increased. Similar gains occurred when we developed procedures for building abstract speech. Just a short time ago, many argued that autistic children were unable to imitate, and that they were unable to form abstractions.

In closing, we should note that many of the procedures we have described are not new, but bear striking similarities to those described by Itard ("The Wild Boy of Aveyron") and by Sullivan (in Gibson's "The Miracle Worker") and recently by Clark ("The Siege"). We are especially struck by the similarity in their willingness to use functional consequences for the child's behaviors, the meticulous building of new behaviors in a piece-by-piece fashion, the intrusion of the education into all aspects of the child's life, the comprehensive, hour-by-hour, day-by-day commitment to the child by an adult, *etc.*

So the principles we employ are not new. Reinforcement, like gravity, is everywhere, and has been for a long time. The principles can be used to the child's advantage, or they can be used

against him. What is new in behavior therapy is the *systematic evaluation* of how these principles affect the child. It is not the content of behavior therapy that is new, but its research methodology. In that sense, we have an immense and often unappreciated advantage over those who preceded us; the methodology enables us to contribute in a *cumulative* manner to psychological treatment.

REFERENCES

- Baer, D. M. and Sherman, J. Reinforcement control of generalized imitation in young children. *Journal of Experimental Child Psychology*, 1964, **1**, 37-49.
- Baer, D. M., Wolf, M. M., and Risley, T. R. Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, 1968, **1**, 91-97.
- Bandura, A. *Principles of behavior modification*. New York: Holt, Rinehart and Winston, 1969.
- Brown, J. L. Prognosis from presenting symptoms of pre-school children with atypical development. *American Journal of Orthopsychiatry*, 1960, **30**, 382-390.
- Ferster, C. B. Positive reinforcement and behavioral deficits of autistic children. *Child Development*, 1961, **32**, 437-456.
- Ferster, C. B. and DeMyer, M. K. The development of performances in autistic children in an automatically controlled environment. *Journal of Chronic Diseases*, 1961, **13**, 312-345.
- Gibson, W. *The miracle worker*. New York: Bantam Books, 1960.
- Havelkova, M. Follow-up study of 71 children diagnosed as psychotic in pre-school age. *American Journal of Orthopsychiatry*, 1968, **38**, 846-857.
- Hewett, F. M. Teaching speech to an autistic child through operant conditioning. *American Journal of Orthopsychiatry*, 1965, **35**, 927-936.
- Hutt, C. and Ounsted, C. The biological significance of gaze aversion with particular reference to the syndrome of infantile autism. *Behavioral Science*, 1966, **11**, 346-356.
- Itard, J. M. G. *The wild boy of Aveyron*. New York: Appleton-Century-Crofts, 1932.
- Kanner, L. and Eisenberg, L. Notes on the follow-up studies of autistic children. In P. H. Hoch and J. Zubin (Eds.), *Psychopathology of childhood*. New York: Grune & Stratton, 1955. Pp. 227-239.
- Lovaas, O. I. *Behavior modification: teaching language to psychotic children*, instructional film, 45-min, 16 mm-sound, Appleton-Century-Crofts, New York, 1969.
- Lovaas, O. I., Berberich, J. P., Perloff, B. F., and Schaeffer, B. Acquisition of imitative speech in schizophrenic children. *Science*, 1966, **151**, 705-707. (a)
- Lovaas, O. I., Freitag, G., Gold, V. J., and Kassorla, I. C. Recording apparatus and procedure for observation of behaviors of children in free play settings. *Journal of Experimental Child Psychology*, 1965, **2**, 108-120. (a)
- Lovaas, O. I., Freitag, G., Kinder, M. I., Rubenstein, B. D., Schaeffer, B., and Simmons, J. Q. Establishment of social reinforcers in two schizophrenic children on the basis of food. *Journal of Experimental Child Psychology*, 1966, **4**, 109-125. (b)
- Lovaas, O. I., Freitag, L., Nelson, K., and Whalen, C. The establishment of imitation and its use for the establishment of complex behavior in schizophrenic children. *Behaviour Research and Therapy*, 1967, **5**, 171-181.
- Lovaas, O. I., Litrownik, A., and Mann, R. Response latencies to auditory stimuli in autistic children engaged in self-stimulatory behavior. *Journal of Abnormal Psychology*, 1971, **9**, 39-49.
- Lovaas, O. I., Schaeffer, B., and Simmons, J. A. Experimental studies in childhood schizophrenia: building social behaviors by use of electric shock. *Journal of Experimental Studies in Personality*, 1965, **1**, 99-109. (b)
- Lovaas, O. I., Schreibman, L., Koegel, R. L., and Rehm, R. Selective responding by autistic children to multiple sensory input. *Journal of Abnormal Psychology*, 1971, **77**, 211-222.
- Lovaas, O. I. and Simmons, J. Q. Manipulation of self-destruction in three retarded children. *Journal of Applied Behavior Analysis*, 1969, **2**, 143-157.
- Lovaas, O. I. Teaching language to psychotic children (*in preparation*).
- Metz, J. R. Conditioning generalized imitation in autistic children. *Journal of Experimental Child Psychology*, 1965, **2**, 389-399.
- Park, C. C. *The siege*. New York: Harcourt, Brace & World, Inc., 1967.
- Patterson, G. R. and Bechtel, G. G. Formulating the situational environment in relation to states and traits. Chapter in R. B. Cattell (Ed.), *Handbook of modern personality study*. Chicago: Aldine Publishing Company, 1970.
- Rimland, B. *Infantile autism*. New York: Appleton-Century-Crofts, 1964.
- Risley, T. and Wolf, M. Establishing functional speech in echolalic children. *Behaviour Research and Therapy*, 1967, **5**, 73-88. *
- Rutter, M. Prognosis: psychotic children in adolescence and early adult life. In J. K. Wing (Ed.), *Early childhood autism: clinical, educational and social aspects*. London: Pergamon Press, 1966.
- Skinner, B. F. *Verbal behavior*. New York: Appleton-Century-Crofts, 1957.
- Tharp, R. G. and Wetzel, R. J. *Behavior modifica-*

- tion in the natural environment.* New York: Academic Press, 1969.
- Wahler, R. G. Setting generality: some specific and general effects of child behavior therapy. *Journal of Applied Behavior Analysis*, 1969, 2, 239-246.
- Wolf, M., Risley, T., Johnston, M., Harris, F., and Allen, E. Application of operant conditioning procedures to the behavior problems of an autistic child: a follow-up and extension. *Behaviour Research and Therapy*, 1967, 5, 103-112.
- Wolf, M., Risley, T., and Mees, H. Application of operant conditioning procedures to the behavior problems of an autistic child. *Behaviour Research and Therapy*, 1964, 1, 305-312.

Received 16 June 1971.

(Final acceptance 7 September 1972.)

REVIEWERS' COMMENTS

The following comments have been abstracted and compiled from the more extensive comments of the three researchers who reviewed this manuscript.

Anyone familiar with the work of Lovaas and his associates would find this new generalization data most interesting and important. His previous work with some of the autistic children described in this manuscript satisfied the reliability and validity requirements of Applied Behavior Analysis and was clearly an important contribution to the understanding and treatment of childhood autism. New observational data, reflecting the generality of changes produced in these children (across time and situations) would be appropriate for a JABA audience because of these prior research considerations.

I think perhaps that this article presumes too much about a reader's familiarity with the authors' prior work. Their arguments, for example, that some of these subjects satisfied the requirements of a within-subjects' replication design are not at all convincing. To say that Group 1 "received an A B A design" is very misleading if we think of B as verified removal of the independent variable. As far as we can tell from this page, B only means that the children were not in the authors' Program. That kind of evidence is not very convincing if one is to argue that utility of a particular *kind* of treatment program (as the authors do). However, some of the kind of evidence needed to satisfy these design requirements is already published.

Next, I see some problems with the data re-

lating to the major focus of this manuscript: namely, generalization of the treatment effects over situations and time. The two-condition assessments of situational generalization (attending and inviting) do provide nice measures of the phenomenon. Since unfamiliar adults were present in both situations, the gradual increase in desirable behaviors is a good index of setting generality. However, the relevance of this generalization is, of course, dependent upon initial demonstrations that the behavior modification program was responsible for the improvement.

Continuing with the generalization data, I'm hesitant to make much of the follow-up data as it is described. The authors attempt to argue that the post-treatment environments (institution *versus* home) accounted for these data. However, as the authors point out, these groups' differences in follow-up could have been due to many factors. I think that the follow-up data should be plotted as a possible function of these environments, but the authors should not claim causal relationships between these environments and the follow-up measures.

The authors describe an elegant and rigorous analysis of observer behavior. They carefully studied the possibility of bias, habituation, and unreliability. The care with which they treated the problem of measurement reliability is very impressive and should serve as a model to other investigators. Nevertheless, I have a very serious question about one of their reliability studies. The authors describe an experiment involving three observers whose acquisition of agreement was studied. The results of this study indicated that after reading the definitions and practising informally with the apparatus for three sessions

the agreement between the observers was still very poor. A reliability shaping procedure was then begun. After two sessions where the observers received *feedback* about how they scored in comparison to the other naive observers and engaged in *discussion* of "instances where they had agreed and disagreed", there was much improved inter-observer agreement. This seems to mean that the written definitions were not sufficient to produce good agreements. It appears that it was necessary for the observers privately to work out definitions among themselves! Therefore, these private definitions were used to gather the data, and not the written definitions. It would thus seem that the reader should be alerted that the replicability of this measurement system is open to question. It is an open question whether another set of three naive observers either in the authors' laboratory or in a reader's laboratory would develop by feedback

and discussion methods the same informal and private response definitions. While this question about the replicability of that measurement system detracts from the confidence that one has in the replicability of that measurement system, it does nothing to alter the conclusions about the relationships that have been reported between the behaviors and the treatment variables. This should likewise be pointed out to the readers.

The manuscript by Lovaas, Koegel, Simmons, and Stevens is a major work by the most important group currently working in behavior modification research with autistic children. It has many outstanding aspects. The results are fascinating, the theoretical discussion section describes implications that have a significance much broader than autistic children. The experimental rigor and emphasis on reliable measurement are models to be carefully studied by other clinical researchers.