PRONOUN TRAINING IN AN ECHOLALIC AUTISTIC CHILD

by

Geoffrey Hogenson Peck

A Thesis Submitted to the Faculty of the

DEPARTMENT OF PSYCHOLOGY

In Partial Fulfillment of the Requirements
For the Degree of

MASTER OF ARTS

In the Graduate College

THE UNIVERSITY OF ARIZONA

1 9 7 6

STATEMENT BY AUTHOR

This thesis has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Geoffrey A. Parker

APPROVAL BY THESIS DIRECTOR

This thesis has been approved on the date shown below:

Wayne R. Carroll

WAYNE R. CARROLL
Associate Professor of Psychology

April 29, 1976

Date

TABLE OF CONTENTS

TABLE OF CONTENTS--Continued

LIST OF ILLUSTRATIONS

LIST OF TABLES

ABSTRACT

A program for training production of three personal pronouns was developed for the purpose of assessing the extent of extraverbal generalization to two types of stimuli, "reversed role" and "new persons." Although the program was successful, the results of the generalization test were equivocal concerning the level of generalization obtained for the two types of stimulus. While it was expected that neither type of generalization would occur above chance level, this expectation was not confirmed for "new persons" generalization. The results tended to support the conclusions of previous research that generalization of trained responses to persons not involved in the original training situation is not an automatic outcome of operant training procedures. A special supplemental program for training generalization of correct responding to "reversed role" stimuli, based on intermixing of those stimuli with original training stimuli, was successful. Several problems related to the nature of the experimental design were discussed, in particular those of overgeneralization, loss of control over echolalic responding, and the biasing effects of intermixed generalization probes. Finally, some limitations of the multiple baseline approach as an evaluative tool for language training programs were discussed in the light of new developments in this area.

INTRODUCTION

The establishment of functional speech in language deficient children has received much experimental attention in the last few years. Early studies, such as those by Lovaas (1968) and Risley and Wolf (1967), demonstrated the efficacy of imitation, fading and differential reinforcement procedures in the establishment of specific functional speech repertoires. Unfortunately, these studies generally failed to provide any experimental evaluation of the generalization of their trained speech skills (Garcia 1974). Recently, however, the issue of generalization of trained speech skills has been much emphasized in the experimental literature. The reason for this new emphasis upon generalization appears to lie in the supposed equivalence between the linguistic concept of "generative" language and the behavioristic notions of generalization of response and "functional response class." The term "generative" language refers to the fact that the language repertoire of a child contains novel responses, which have not been modeled or directly trained, that are often related to other responses in that repertoire (Lutzker and Sherman 1974). The term "functional response class" concerns instances in which an entire group of responses shows the effects of an experimental manipulation which was carried out on only a few members of that group (Wheeler and Sulzer 1970).

Since the discovery of its existence in the speech repertoires of normal children by Brown and Bellugi (1964), nativists such as

Lenneberg (1969) and McNeill (1966), as well as behaviorists such as Staats (1968) and Sherman (1971), have stressed the importance of the development of generative language in their theoretical accounts of language acquisition. The issue of generalization is therefore theoretically significant in that experimental demonstrations of generalization as a result of imitation and reinforcement procedures provide evidence that similar factors may be relevant to the normal process of language acquisition. Furthermore, the issue of generalization has obvious applied significance, since the establishment of broad and novel repertoires of functional speech in language deficient subjects is usually very time consuming and requires procedures which are maximally efficient.

Studies concerned with the training of generative language have covered a broad range of specific skills. Lovaas (1968) used imitation and reinforcement procedures for training psychotic, language deficient children in the production of object labels, prepositions, personal and possessive pronouns, and verbs. Similarly, Guess et al. (1968) and Guess (1969) trained retarded children in the generative usage of noun plural endings. Such procedures have also been effective in the training of generative language in the area of syntax. For example, Fygetakis and Gray (1970) increased the generative usage of the auxiliary verb "is" in aphasic children; Schumaker and Sherman (1970) obtained generative production of verbs in both past and present tenses in retarded children; and generative production of sentences with subject-verb agreement was trained in both young retardates and normal infants

by Lutzker and Sherman (1974). Also, Wheeler and Sulzer (1970) produced generative usage of complete sentences in a child who originally used "telegraphic" phrases.

Actually, the issue of generalization in speech production is more complex than it appears to be in the literature. Several distinct types of generalization may occur as the result of a given experimental manipulation. For example, production of correct verb tense endings may generalize from trained verb stems to untrained verb stems in the same class (Schumaker and Sherman 1970). If such generalization occurred, it would demonstrate the existence of a functional response class. This was actually a case of intraverbal stimulus generalization, in that one component of the response, the verb ending, generalized to new verb stems, which acted as descriminative stimuli for the various verb endings. On the other hand, another form of stimulus generalization may occur, in which a particular response may generalize to new extralinguistic stimuli. Garcia (1974) and Kale et al. (1968) obtained generalization of trained verbal responses to persons other than those involved in initial training. Although this latter form of stimulus generalization may be less interesting from a linguistic point of view, it is still an important and desirable result of language training procedures, since a subject's speech skills cannot be truly functional unless they can be used in a much greater number of extralinguistic contexts than those in which they were originally trained.

The nature of generalization is probably influenced by the nature of the response class whose members are being trained, as well as the

complexity of the discrimination involved in the training. A discrimination task may be either unidimensional or multidimensional. Given a discrimination task involving four blocks, a large blue one, a large red one, a small blue one, and a small red one, the discrimination would be unidimensional if the criterion for correct responding involved the size dimension alone. On the other hand, if both the size and color dimensions became relevant, then the discrimination would be multidimensional. Production of the plural ending in response to objects varying in number involves a simple unidimensional discrimination. Actually, the subject is required to attend to both the type of object and the number, but there is only one dimension per response component. On the other hand, production of simple statements involving a pronoun and a verb phrase in response to a particular person performing some action, as was trained by Lovaas (1968), involves both a unidimensional discrimination for the action and a multidimensional discrimination for the pronoun.

In pronoun production, the correct pronoun to be used with regard to the referent person of the sentence is determined by a hierarchy of semantic and syntactic discrimination dimensions (Waryas 1973). The first levels of this hierarchy are semantic in nature and deal directly with the referent person. The first dimension, as shown in Figure 1, concerns whether or not the referent person is the speaker of the sentence. If this is the case, then a first person pronoun is appropriate; if not, then the next dimension becomes relevant. The second dimension concerns whether or not the referent person is the listener, or addressee, of the sentence. If this is the case, then a second person pronoun is

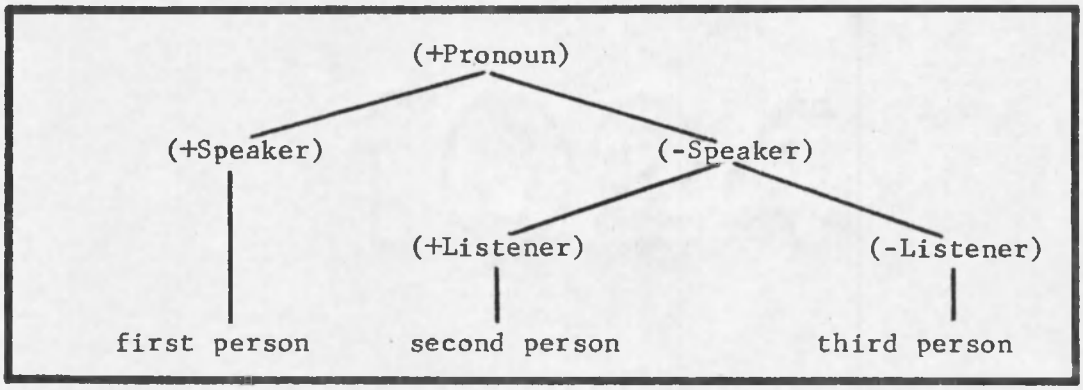

Figure 1. The first steps in the pronoun discrimination dimension hierarchy.

appropriate; if not, then a third person pronoun should be used. Of course, the first person pronoun will always be used by the speaker for himself only, as though it were an alternative for his name. On the other hand, choice between use of the second and third person pronoun for the referent person, when relevant, is determined by the direction of the sentence, rather than by any intrinsic characteristics of the referent person. This fact distinguishes the pronoun system from many other elements of language, such as noun and verb stems, which do depend only upon intrinsic characteristics for proper discrimination.

Two distinct types of extralinguistic stimulus generalization may be obtained as a result of training a subject in the production of first, second and third person pronouns. Lovaas (1968) trained subjects to produce simple sentences including the second person pronoun "You" in response to a question by the trainer concerning what action he was performing, and the third person pronoun "He" in response to a question by

the trainer concerning what action was being performed by a third person. It is possible that the response "You" may have generalized to the person who originally acted as the third person if he were to act as the trainer and perform some action. Similarly, if the original trainer were to perform an action, and the original third person were to act as the trainer and give the question, it is possible that the response "He" may have generalized to the original trainer. This type of generalization may be termed "reversed role" (RR) generalization, since it involves the trainer and third person in opposite roles as discriminative stimuli for production of the two pronouns. It is also possible that the responses "You" and "He" may have generalized to two persons acting as trainer and third person who had never been involved in the training of those pronouns. This type of generalization may be termed "new persons" (NP) generalization for both second and third person pronouns. Thus, for each pronoun, two separate types of extralinguistic stimulus generalization are possible. Unfortunately, Lovaas (1968) only evaluated the existence of intraverbal generalization of the trained pronoun responses to novel action labels.

There is reason to believe that neither type of extralinguistic generalization will occur at a significant level as an automatic result of training the second and third person pronouns to only one individual in each case. A problem frequently encountered in working with autistic subjects is that of "stimulus overselectivity," a tendency to focus on some dimensions of a multiple stimulus array while ignoring the rest (Lovaas et al. 1971). It is possible that, because of stimulus

overselectivity, a subject might attend to the person who was performing the action but not to the issue of whether or not that person was the listener. If this were the case, the subject's pronoun responses would become "stimulus bound"; that is, they would simply be a result of association with the specific individuals involved in the original training. In such a case, the occurrence of either "reversed roles" or "new persons" generalization at significant levels would be unlikely. A similar problem was found by Sailor and Taman (1972) in training autistic, language deficient children to produce the prepositions "in" and "on." Each preposition was trained with separate stimulus objects during the initial training condition. Apparently, the type of stimulus object, rather than its prepositional relationship with the referent object, became the cue on which the subjects based their preposition responses. In a second condition, where all stimulus objects were associated with both preposition responses, the subjects showed a lower proportion of correct responses, indicating that they had failed to generalize their preposition responses to those objects that were not associated with them during the initial training condition.

The purpose of this study was to experimentally evaluate the efficacy of imitation and differential reinforcement procedures in producing both "reversed role" and "new persons" generalization of the second and third person pronoun responses in a language deficient autistic child. The complexity involved in proper pronoun production offers the field of language training a special test of the strength and flexibility of these procedures in obtaining appropriate generalization

effect. Since generalization often does not occur as an automatic result of the application of training procedures, the need for effective techniques for programming it has been increasingly emphasized recently (Baer, Wolf and Risley 1968; Lovaas et al. 1973; Garcia 1974; Stokes, Baer and Jackson 1974). A test for the two types of extralinguistic generalization of the second and third person pronoun responses was administered in order to test the hypothesis that appropriate responding would not occur above the level of chance for either type of generalization. At the end of the study, a special program for increasing the level of correct responding to "reversed role" generalization stimuli from chance level to a criterion level of 80 percent was administered.

METHOD

Subject

The subject of this study was a four-and-one-half-year-old male. At the time of the study, he had exhibited behavior patterns characteristic of the autism syndrome. His repertoire of language skills was deficient, but included a small but increasing number of spontaneous appropriate verbalizations. However, the majority of his utterances were echolalic in nature. Since January 1974, the subject had been participating with his family in a general program of applied behavior analysis for the development of appropriate social, language and self-help behaviors at the Child Psychology Laboratory at The University of Arizona. Verbal training had produced a significant increase in his repertoire of object and action labels, but no systematic attempt had been made to train appropriate usage of pronouns. During the course of this study, the subject continued to participate in those programs being administered at the Child Psychology Laboratory.

Experimental Design

The experimental design involved in this study generally followed that used by Lovaas (1968). Lovaas first trained production of the second and then the first person pronoun separately. Trials for these two were then intermixed to insure proper discrimination. Following attainment of criterion performance on the discrimination trials,

production of the third person pronoun was trained, and then trials for the first, second and third person pronouns were intermixed until criterion performance was reached. Thus, the second person pronoun played a pivotal role in the training. It was first discriminated from the first person pronoun, thereby involving the (± Speaker) dimension, and then discriminated from the third person pronoun, involving the (± Listener) dimension. The format of the present study for training the production of each of the three pronouns and proper discrimination between them consisted of (1) a baseline condition (condition B), in which pronoun responses to the three pronoun training stimuli plus the generalization probe stimuli were tested; (2) training of the second person pronoun (condition P1); (3) training of the first person pronoun (condition P2); (4) discrimination training for the first and second person pronouns (condition D1); (5) training of the third person pronoun (condition P3); (6) discrimination training for the second and third person pronouns (condition D2); and (7) discrimination training for the first, second and third person pronouns (condition D3). The sequential training of the three pronoun responses allowed for the application of a "sequential analysis," or multiple baseline design, across the production training conditions as a means of evaluating the efficacy of the general training procedures in establishing pronoun production in response to the training stimuli (Baer et al. 1968, Baer and Sherman 1970).

After completion of condition D3, a generalization test, consisting of the initial training stimuli used for the second and third person

pronouns randomly intermixed with the generalization probe stimuli, was administered. Finally, one of the two types of generalization, "reversed role" generalization, was trained for both the second and third person pronouns through the use of a special discrimination training procedure.

Training and Generalization Probe Stimuli

The stimuli to which the subject gave his pronoun responses consisted of the performance of an action by either the subject or one of the persons directly involved in the training conditions or the generalization test. The context in which all training and generalization probe stimuli were presented did not vary. The trainer always stood directly in front of the subject. Imitative prompts, stimulus questions, commands and reinforcements were administered exclusively by the trainer, so that the subject became oriented toward the trainer and directed his pronoun responses to that person. The third person stood to one side, either right or left, of the trainer. During all training conditions and the generalization test, the same stimulus question, "What is happening?" was used. The stimulus question served simply as a cue for eliciting the subject's pronoun response. The fact that the nature of the stimulus question did not vary as a function of the type of stimulus presented, contrary to the Lovaas (1968) procedure, prevented the subject from relying upon its content as a discriminative stimulus for correct pronoun production during the various discrimination training conditions and the generalization test.

As discussed earlier, production of the correct pronoun is determined by two dimensions of discrimination: (1) whether or not the referent person of the sentence is the speaker, and (2) whether or not the referent person of the sentence is the listener. The referent person of the sentence was always the person who performed the action. In the stimulus used for training the first person pronoun, "I," the subject performed the action. In the initial training and two generalization probe stimuli for the second person pronoun, "You," the trainer performed the action. For those stimuli involving the third person pronoun, "She," the third person performed the action. All three initial training stimuli used the same two persons, person I and person II, for the trainer and third person respectively. For "You" and "She" "reversed role" generalization probe stimuli, person II acted as the trainer, while person I acted as the third person. The "new persons" generalization probe stimuli for both "You" and "She" involved two new persons, person III as the trainer and person IV as the third person.

General Procedure

Pretraining

Before the baseline measurement was taken, modeling and differential reinforcement procedures were used to train the subject to produce three different action labels, "pointing," "waving," and "clapping," in response to the performance of those actions by persons I, II, III, and IV. Training continued until the subject produced ten consecutive correct action labels in response to randomly intermixed presentation of

each of the three actions. The subject was then trained to respond to the command, "(subject's name), (imperative verb form)!" for each of the three actions until he had reached a criterion performance of ten consecutive compliant responses to randomly intermixed action commands.

Baseline

After completion of pretraining sessions, a baseline measure was taken for the level of pronoun production in response to the three initial training stimuli and the four generalization probe stimuli (condition B). The method for presentation of the baseline stimuli was identical to that used in the production and discrimination training conditions except that no imitative prompts or reinforcers were given. On both attempts of each trial for all stimuli, the action was performed and was followed approximately two seconds later by the stimulus question. All seven different stimuli were randomly intermixed, with the type of action also varying randomly. Baseline measures were taken on three separate days. An independent observer was used on one session in order to assess the reliability of the response measures.

Production Training

Following condition B, production training and discrimination training for each of the three pronoun responses was carried out. The general procedure for production training of the pronoun responses was an adaptation of the procedures used by Lovaas (1968) for training production of pronoun plus action label responses, and by Risley and Wolf (1967) for dealing with echolalic responding. The procedure used in the

present study for training production of the pronoun responses was identical for the three initial training stimuli, except that in the presentation of "I" stimuli, the trainer first elicited performance of the action by the subject with an action command. Training trials were administered in 20 trial blocks. Each training trial consisted of a "first try" and an optional "second try." On the "first try," initiation of the action by the referent person was followed approximately two seconds later by either an imitative prompt (i.e., a model) of the correct pronoun plus action label response, no verbal stimulus, or the stimulus question from the trainer, depending upon what phase of production training the subject was in. The two-second delay between onset of the action and presentation of one of the above stimuli by the trainer reduced the danger of the subject's attending to the verbal stimulus without attending to the referent person's performance of the action (Lovaas et al. 1971). If the subject responded with the correct pronoun within approximately five seconds, he was rewarded by the trainer, and the trial was terminated without use of the "second try." The subject readily produced the correct action labels throughout the training sessions. Therefore, it was not necessary to make reward contingent upon inclusion of the correct action label in the response. If the subject did not produce the correct pronoun within the alloted time, or if he responded with an incorrect pronoun, the "second try" was used. The "second try" was identical to the "first try," except that it always contained an imitative prompt of the correct pronoun plus action label response. During the final phase of pronoun production training, the

"second try" also included presentation of the stimulus question just before the imitative prompt. Use of the optional "second try" during pronoun production training was based on the fact that the subject was familiar with this procedure through his participation in training programs at the Child Psychology Laboratory and responded well to its use.

At the beginning of production training for each of the three pronouns, the trainer used an imitative prompt containing the correct pronoun plus the correct action label on the "first try" of each trial. Once the subject reached a criterion of correct pronoun production in response to imitative prompts on the "first try" for ten consecutive trials, the imitative prompts were faded out by (1) dropping out the action label, and then (2) reducing the volume of the prompt. After the subject had begun to give the correct pronoun response on the "first try" without the imitative prompt, the stimulus question was faded in on both tries by increasing its volume from whisper level to normal voice level. Production training for each pronoun was terminated when the subject had reached a criterion performance of 80 percent correct pronoun production in response to the stimulus question on the "first try" in a 20-trial block.

Discrimination Training

The procedure for administration of trials during pronoun discrimination training, with the exception of "You"-"I" discrimination (D1), was identical to that used at the end of each condition of production training. On the "first try" of each trial, only the stimulus question was presented with the performance of the action. Discrimination

training was carried out in 20-trial blocks with ten stimuli for each pronoun randomly intermixed, except for "I"-"You"-"She" discrimination training (D3), in which ten trials for each pronoun were randomly intermixed in 30-trial blocks. Training for each of the three discriminations was continued until the subject had reached a criterion performance of 80 percent correct pronoun production for each of the two or three different stimuli in a 20- or 30-trial block. In the case of "You"-"I" discrimination training, correct responding to both stimuli had deteriorated significantly. Several procedures for increasing correct pronoun production for "You"-"I" discrimination were administered: (1) non-random presentation of "You" and "I" trials (trials 80-100 and 120-140); (2) use of "Both hands down" command for inappropriate imitation of the trainer's performance of the action on "You" trials (trials 160-300); (3) use of imitative prompts on the "first try" (trials 180-240); and (4) use of immediate verbal praise after production of the correct pronoun on the "first try," or immediate use of an imitative prompt after the production of an incorrect pronoun.

Previous research on verbal training (Risley and Wolf 1967), as well as recent results in working with the subject, indicated that a combination of verbal praise plus food reward was more effective as a reinforcer than was verbal praise alone. Therefore, during the pronoun production and discrimination training conditions, a continuous schedule of verbal praise ("Good boy!," "That's right!," "Good, (subject's name)!") plus food reward (juice) was employed. No verbal punishment was ever given by the trainer for production of an incorrect pronoun.

Intertrial intervals varied in duration from two to five seconds. There were often two or three minute-long rest periods placed between five-trial blocks during training sessions. The trainer dealt with minor cases of inappropriate and disruptive behavior by the subject by ignoring them. In the case of screaming, hitting or self-directed aggression, the trainer physically restrained the subject until he became calm. Whenever the subject exhibited off-task behaviors, such as looking away from the experimental setting or self-stimulation, the trainer responded with the command, "Look at me!," and, if necessary, physically oriented the subject's head toward the stimulus being presented.

Generalization Test

The test for generalization was carried out in one session. Ten presentations of each of the four generalization probe stimuli, "You" (RR), "She" (RR), "You" (NP), and "She" (NP), were randomly intermixed with 20 presentations each of the initial training stimuli for "You" and "She." Thus, for the 80 trials, there was an equal number of initial training and generalization probe stimuli. The six different stimuli were intermixed in such a way that there were no obvious sequential dependencies between any of them. No stimulus was presented on two consecutive trials. Furthermore, no more than two generalization probe stimuli were presented consecutively in any five-trial block. Consequently, the opportunities for the subject to guess whether or not a particular trial was a generalization probe trial were reduced. Originally, a special "You"-"She" discrimination training session was planned for insertion after condition D3, in which a VR 2 schedule of

reinforcement would have been used. Use of such a schedule would have allowed the subject to adapt to the similar schedule used in the generalization test (Guess 1969, Garcia 1974). Unfortunately, time limitations prevented the use of this special training session. Presentation of the initial training stimuli was carried out in the same fashion as at the end of "You"-"She" discrimination training. For the four different generalization probe stimuli, the method of presentation was almost the same as that used for the baseline measurement. Although no reinforcements or imitative prompts were given, the procedure was modified such that the "second try" was not administered if the correct pronoun was produced on the "first try." The purpose of this modification was to reduce discriminability between the initial training and generalization probe stimuli. The generalization test was divided into 16 five-trial blocks, with short rest periods between each block.

Generalization Training

Because of time limitations, only one of the two types of generalization, "reversed role" generalization, was trained. In the case of "reversed role" generalization, each pronoun response must generalize to the person other than the one who was involved in the initial training stimulus for that response. The subject was required to discriminate between the initial training and "reversed role" generalization probe stimuli for each pronoun in order to produce the correct pronoun in response to the generalization probe stimuli. The method for training correct discrimination of the two "reversed role" stimuli from their initial training counterparts was based on the

principle of matching of opposites. During the first 20 trials, each of the initial training stimuli was matched with its "reversed role" counterpart such that "You" (initial) was followed by "You" (RR), then by "She" (initial), and then by "She" (RR), each being presented five times. At first, an imitative prompt was presented after the stimulus question on the "first try," but was later faded out. The four different stimuli were then randomly intermixed with the same procedure as was used in the initial "You"-"She" discrimination training condition. Trials continued until the subject had reached a criterion performance of 80 percent correct pronoun production in response to the stimulus question on the "first try" for each of the four different stimuli in a 40-trial block.

Experimental Setting

As with the baseline measurement, all pronoun production and discrimination training sessions, as well as the generalization test and generalization training sessions, were carried out in the subject's home. The setting for these sessions was the living room. Training sessions contained either one or two 20- or 30-trial blocks, and ranged in duration from 20 to 30 minutes. These sessions were usually given four days per week. Prior to the beginning of the sessions, all toys and other distracting items were removed, and the subject was situated in a chair. Other than the subject, the trainer, and the third person, there were two additional persons in the room, the author and an independent observer. Both were situated off to one side of the chair so as not to distract the subject. The author, who acted as a data recorder,

also supervised the trainer and third person concerning the type of stimulus to be used on a given trial, the kind of verbal stimulus to be given by the trainer, and means of controlling inappropriate or disruptive behaviors. Recent research (Goldstein and Lanyon 1971, Kozloff 1973) has emphasized the value of the use of parents and siblings as active agents of behavior change in the home setting. The use of such persons in a training program allows for long-term maintenance of target behaviors after they have been trained. Therefore, the subject's mother and his two sisters were used as persons I, II, and III respectively. A fourth person, who also acted as an independent observer during the generalization test, acted as person IV.

Scoring of Responses and Reliability

The method for scoring the subject's responses was the same for all seven different stimuli throughout the baseline, production training, discrimination training, generalization test and generalization training conditions. If, on any given "first" or "second try" of a trial, the subject responded within the five-second response interval by using only the pronoun which was appropriate for the stimulus, then that response was scored as being correct. A pronoun response was scored as incorrect if it contained a pronoun which was inappropriate for the stimulus, or if it contained a combination of the appropriate pronoun and an inappropriate pronoun. Echolalic repetitions of either the stimulus question or action command on either the "first" or "second try" of a given trial were marked as such. The occurrence of irrelevant verbalizations by the

subject were recorded in the same manner. Also, if none of the above responses occurred on a given try, then that try was marked as having no response. Echolalic and irrelevant responses, as well as non-responding, were recorded by the author during all conditions except for the generalization test.

An independent observer was present during the baseline measurement. Another was present for a majority of the "You"-"I" discrimination training sessions and also for one "reversed role" generalization training session. A third person acted as an independent observer during the generalization test session. The independent observers recorded only the occurrences of correct and incorrect pronoun production. Reliability between the author and the independent observers was based on whether or not, for each try scored, the scores were the same.

RESULTS

Reliability

Reliability of pronoun response scoring between the author and the independent observers ranged from 84 to 99 percent. The average level of interobserver agreement was 88 percent. The levels of agreement on the generalization test and generalization training period were 93 and 99 percent respectively. These levels were computed on the basis of the ratio of the number of tries which were scored the same to the total number of tries scored. Agreement was based only on the score given for the pronoun response. The other measures were not considered in computing reliability.

Production Training

As can be seen in Figures 2, 3, and 4, production training for each pronoun (conditions P1, P2, and P3) was rapid and free from error responses. During "You" production training, production of the pronoun dropped significantly after the imitative prompt had been faded out. However, it did return to criterion level within three 20-trial blocks. On the last block, when the stimulus question was faded in, production of the pronoun had reached the 100 percent level. The course of "I" production training was very similar, with the level of production dropping after the imitative prompt had been faded out and then increasing up to criterion level. Production training for "She" was very rapid

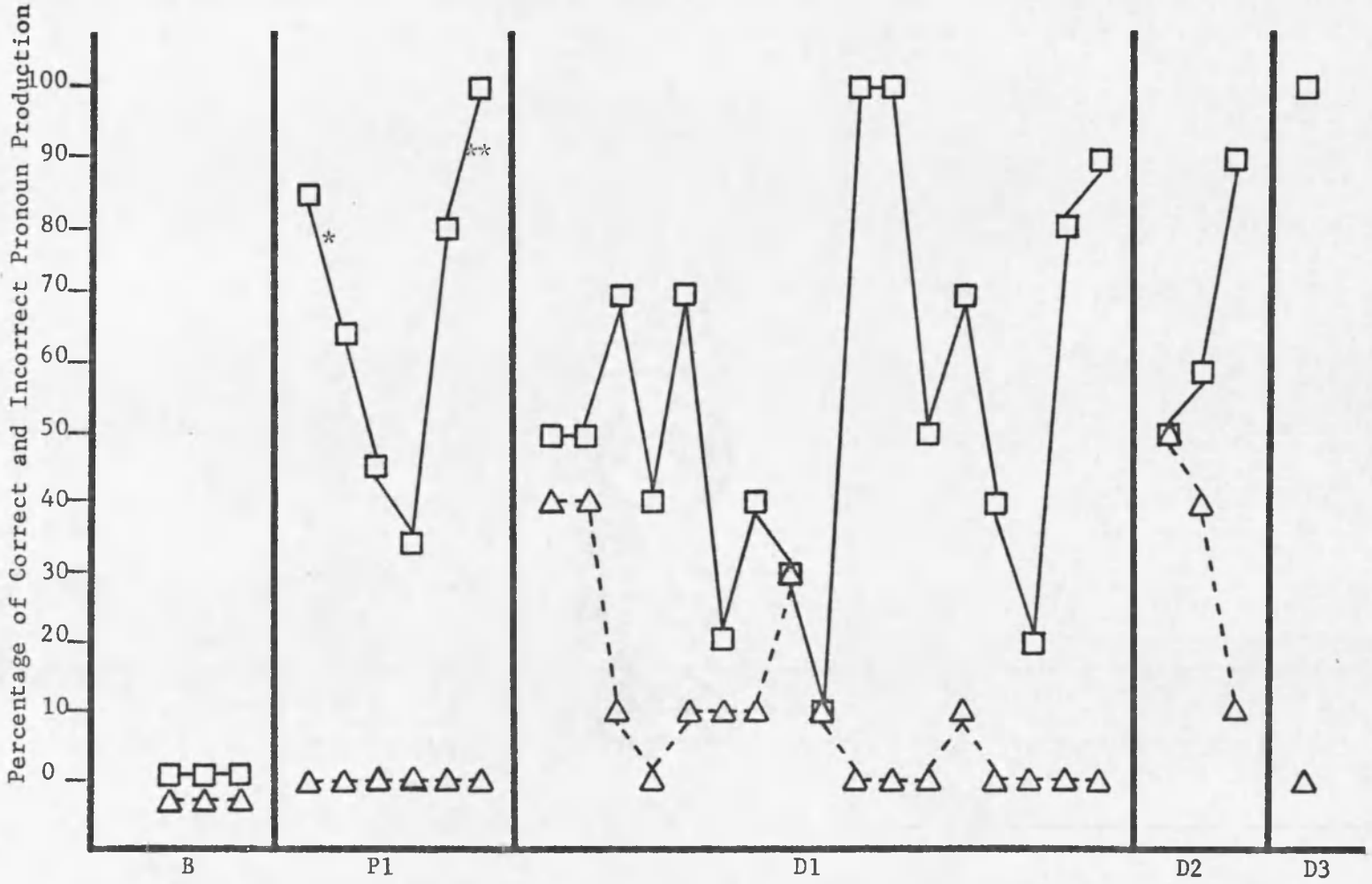

Figure 2. Percentage of correct and incorrect pronoun production on the "first try" for "You" trials per 20-trial block during production and discrimination training conditions. -- □—□ = correct pronoun production; △ -△ = incorrect pronoun production. A single asterisk (*) = fading out of imitative prompt; a double asterisk (**) = fading in of stimulus question.

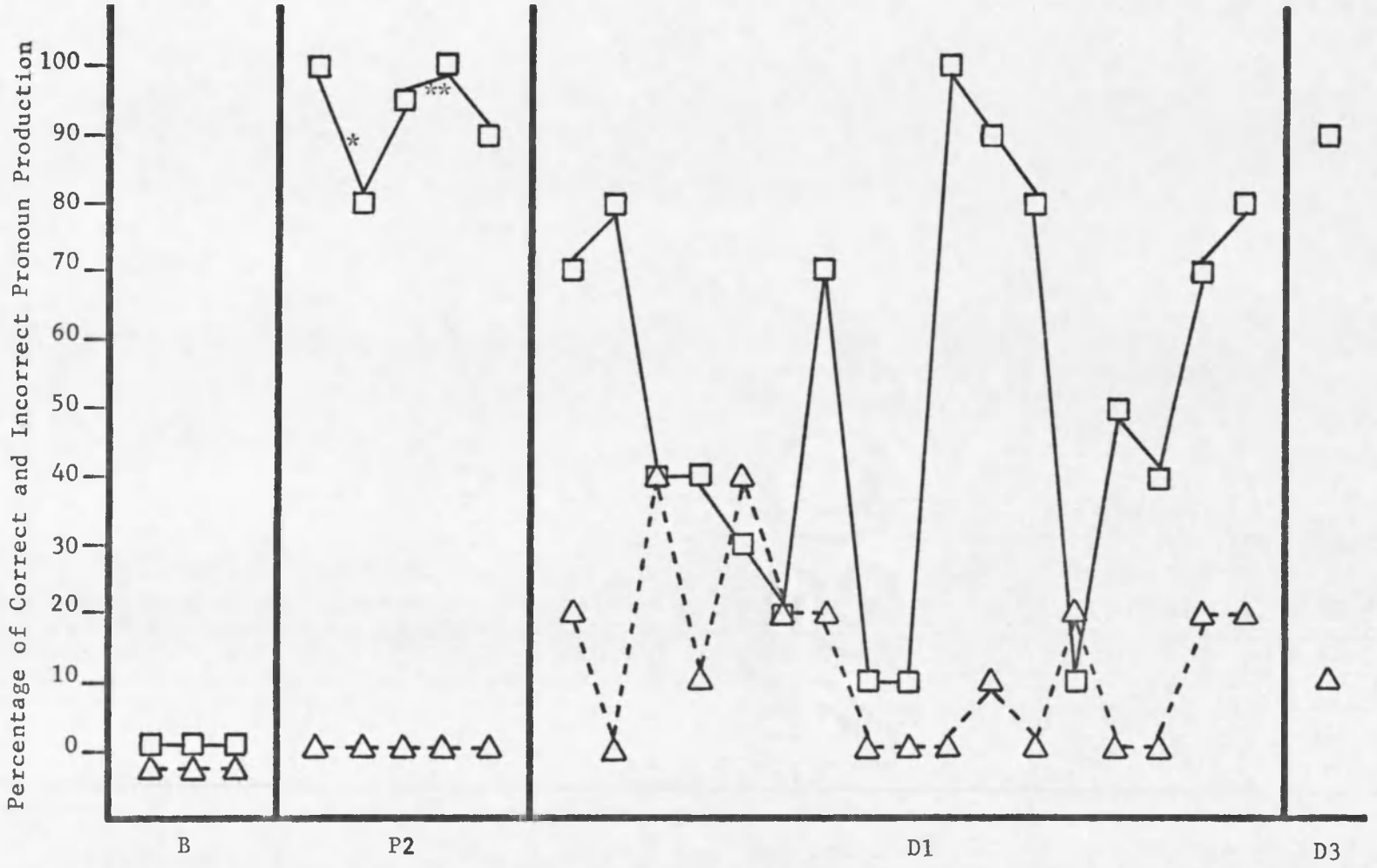

Figure 3. Percentage of correct and incorrect pronoun production for "I" trials. -- □—□ = correct pronoun production; △- -△ = incorrect production. A single asterisk (*) = fading out of imitative prompt; a double asterisk (**) = fading in of stimulus question.

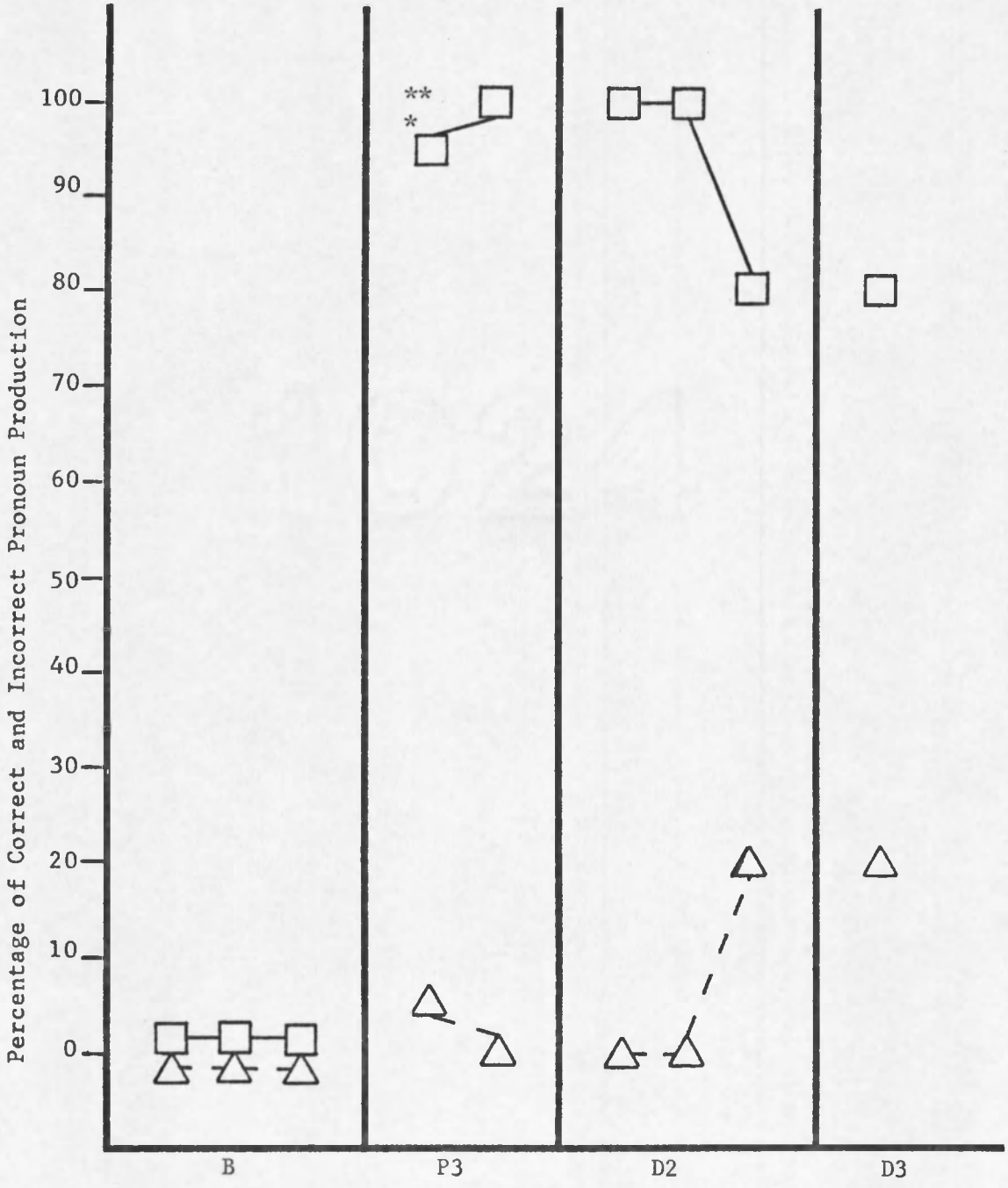

Figure 4. Percentage of correct and incorrect pronoun production for "She" trials. -- □—□ = correct pronoun production; △ △ = incorrect production. A single asterisk (*) = fading out of imitative prompt; a double asterisk (**) = fading in of stimulus question.

in comparison to the training of the other two pronouns, but it was the only condition of the three in which any errors were made. There was a trend across conditions P1, P2 and P3 in that as each successive pronoun was trained, there were fewer trials required to reach criterion performance. One possible reason for this is that during the first few weeks of training, the subject exhibited many disruptive behaviors, especially tantrumming, at the beginning of the training session, whereas these disruptive behaviors had diminished in frequency by the time that "She" production training had begun.

Discrimination Training

Figures 2 and 3 show the course of discrimination training (condition D1) for "You" and "I" respectively. It can be seen that the level of correct production of each pronoun on the "first try" had deteriorated significantly as training blocks progressed. By the ninth 20-trial block (trials 160-180), correct production of both pronouns was down to ten percent. The rapid rise in correct production in the next 20-trial block was a result of the reinstatement of imitative prompts on the "first try." However, when the prompts were faded out during the twelfth 20-trial block (trials 220-240), production again dropped to below criterion level. The use of non-randomized presentation of the two stimuli, used earlier in training, did not appear to facilitate correct pronoun production. During the early discrimination training sessions, the subject began to periodically imitate the trainer's performance of the action on "You" trials. As a means of preventing confusion between "You" and "I" stimuli, the "Both hands down" command was used

during trials 160-300 whenever the subject began to imitate the trainer's action. This procedure did not seem to have any facilitative effect on pronoun production and had two harmful side effects. During its use, the frequency of the subject's tantrumming and other disruptive behaviors, including irrelevant verbal responses, increased significantly, while the frequency of pronoun production dropped after removal of the imitative prompts. When use of the "Both hands down" command was discontinued, the frequency of disruptive behaviors decreased, and pronoun production soon reached criterion level. The use of immediate feedback for pronoun responses on the "first try" during trials 260-340 may have facilitated attainment of criterion performance. However, there is no means of experimentally evaluating its effect.

Interestingly, in spite of the deterioration in pronoun production, there was almost never a problem in discrimination between the two pronouns. The level of incorrect pronoun production for both pronouns was almost always below that for correct pronoun production. For both pronouns, there were only two blocks in which the level of incorrect pronoun production was equal to that for correct pronoun production and only two blocks, both for "I," in which it was higher. Generally, the subject tended to produce incorrect pronouns more frequently on "I" trials than on "You" trials, indicating that he was biased towards production of "You," the pronoun with which he was most familiar.

In contrast to D1, attainment of criterion performance on "You"-"She" (D2) and "I"-"You"-"She" (D3) discrimination training was much more rapid. During the first block of D2, imitative prompts on the

"first try" were used and then faded out. In the next training session, which contained only 30 trials, the trials were divided such that the first ten trials constituted one block and the last 20 constituted the final block. In general, the subject tended to produce incorrect pronouns more frequently on "You" trials than on "She" trials, indicating that he was biased towards production of "She," which had been trained just prior to D2. In D3, attainment of criterion performance required only one 30-trial block. As in P3, conditions D2 and D3 were relatively free of disruptive behaviors and irrelevant verbal responding.

Generalization Test

Table 1 shows the results of the generalization test. Correct pronoun production on the "first try" in response to the "reversed role" generalization probe stimuli for both pronouns was not above chance level. In the case of "She" (RR) probe stimuli, the subject produced three times as many incorrect pronouns as he did correct ones, whereas correct and incorrect pronoun production were equally frequent for "You" (RR) probe stimuli. On the other hand, the frequency of correct pronoun production in response to the two "new persons" generalization probe stimuli was above chance level, although it was not as high as the level of correct responding to the two initial training stimuli. In general, production of incorrect pronouns was more frequent for "She" (RR) and (NP) than for "You" (RR) and (NP) stimuli, indicating that the subject was biased towards production of "You," although the bias was greater for the latter type of generalization than for the former. The subject also made slightly fewer errors in responding to the two stimuli in

Table 1. Generalization test results.

Stimulus	Percent Correct		Percent Incorrect	
	"first try"	"second try"	"first try"	"second try"
You (initial)	85	00	00	00
She (initial)	80	75	15	25
You (RR)	50	40	50	40
She (RR)	20	00	60	75
You (NP)	60	50	10	00
She (NP)	60	50	40	25

which person I, the trainer for both initial training stimuli, performed the action than to the other two stimuli. Correct pronoun production on the "second try" in response to the generalization probe stimuli was similar in nature to those on the "first try."

Generalization Training

Figure 5 shows the results of the special "reversed role" generalization training period. Criterion performance was reached as rapidly as in the initial "You"-"She" discrimination training condition. During the course of generalization training, correct responding to "You" (initial) stimuli remained at the 100 percent level, while correct responding to "She" (initial) stimuli, after suffering a drop below criterion level in the beginning, quickly rose to above criterion level.

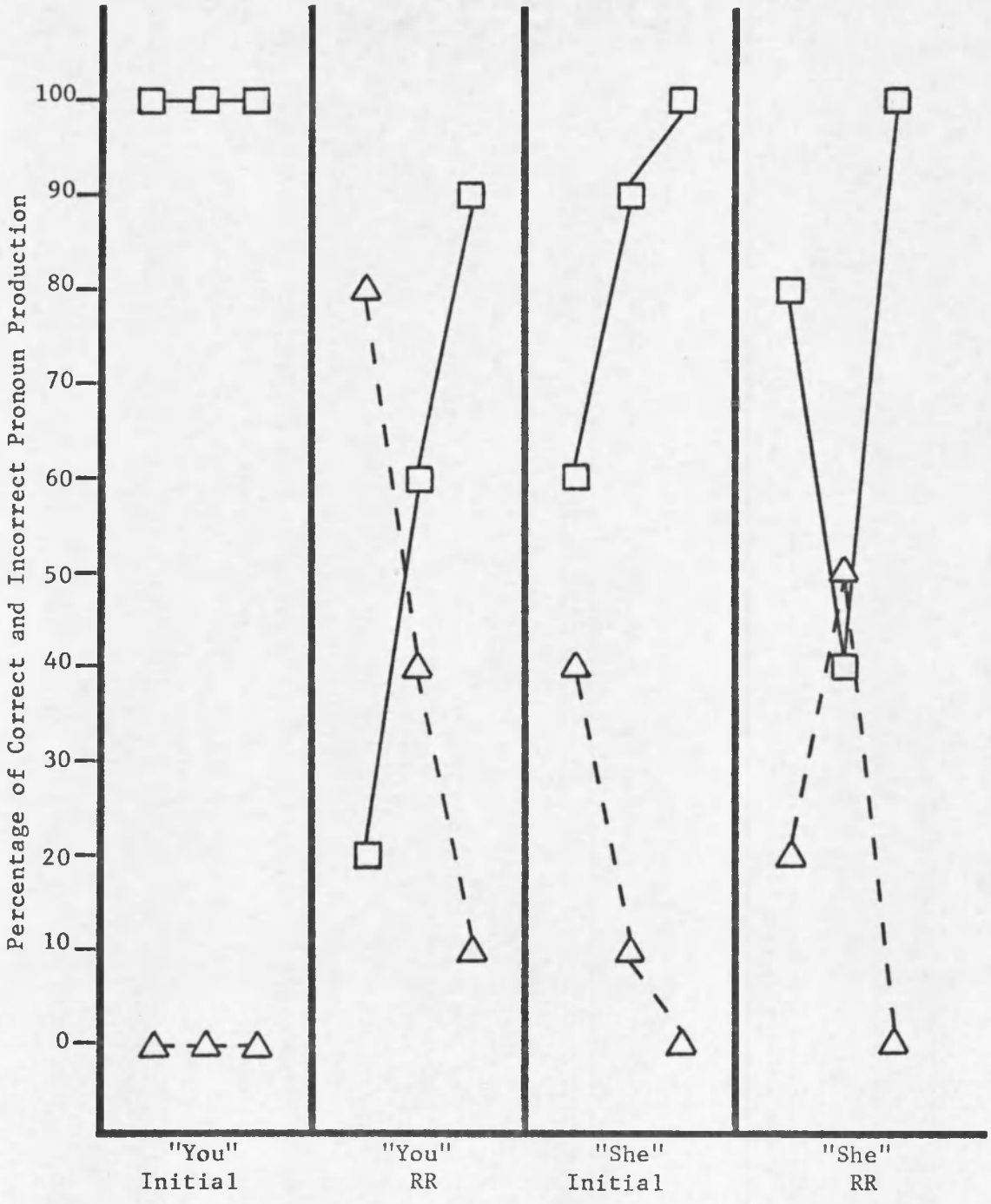

Figure 5. Generalization training: Percentage of correct and incorrect pronoun production on the "first try" per trial block. -- □—□ = correct pronoun production; △-△= incorrect pronoun production.

In the case of the two generalization stimuli, this relationship was reversed. The subject generally produced the correct pronoun on the "first try" more frequently in response to "She" (RR) than to "You" (RR) stimuli. As opposed to the results of the generalization test, correct pronoun production in response to "She" (initial) and (RR) stimuli was only slightly less frequent than in response to "You" (initial) and (RR) stimuli. The highest frequency of incorrect pronoun production was associated with "You" (RR) stimuli, followed by "She" (RR), "She" (initial), and "You" (initial) stimuli respectively. The above relationships would indicate that (1) as in the generalization test, discrimination between the two initial stimuli was more accurate than between the two (RR) stimuli; (2) the subject was biased towards production of "You" in response to initial training stimuli and "She" in response to (RR) stimuli; and (3) as in the generalization test, incorrect pronoun production was generally less frequent in response to the two stimuli in which person I, the trainer for both initial stimuli, performed the action than in response to the other two stimuli. Responses on the "second try" were almost always correct.

Echolalia, Irrelevant Responding and Non-responding

As shown in Table 2, the average percentages of the other response measures, echolalia, irrelevant responses and non-responding, varied widely across the production and discrimination training conditions. During the baseline measure, echolalic repetition of the stimulus question was common, as was to a lesser extent non-responding.

Table 2. Average percentages of occurrence of echolalia, irrelevant responses and non-responding for "You," "I" and "She" stimuli per trial block in each training condition.

	Echolalia				Irrelevant Responses				Non-responding			
	"You"		"I"		"You"		"I"		"You"		"I"	
	1st try	2nd try	1st try	2nd try	1st try	2nd try	1st try	2nd try	1st try	2nd try	1st try	2nd try
B1	45.0	0.0			20.8	05.8			14.2	0.8		
B2			93.0	02.0			09.0	0.0			02.0	0.0
C1	54.1	25.7	67.1	23.2	29.4	06.9	20.0	12.4	04.1	01.9	02.4	03.4
C3	0.0	0.0	80.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
			"She"				"She"				"She"	
B3			08.0	0.0			0.0	0.0			0.0	0.0
C2	30.0	0.0	10.0	0.0	03.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
C3			10.0	0.0			0.0	0.0			0.0	0.0

Fading-in of the stimulus question occurred during the last 20 trials of "You" production training (P1) and the last 30 trials of "She" production training (P3). Echolalic responding to the stimulus question in P1 was frequent at first, but then dropped out after presentation of the eighth trial, while it only occurred twice in P3. Echolalic responding to either the action command or the stimulus question during P2 occurred frequently, but it also dropped out at the end of training. The frequency of irrelevant responses and non-responding was generally 20 percent or lower for all conditions. As a rule, all three response measures were at their highest frequency in condition D1. Echolalia for "You" trials declined steadily from an average of 54 percent during D1 to zero percent in D3. For "She" trials, the average frequency of echolalia was always ten percent or less. On the other hand, the average level of echolalia for "I" trials during conditions P2, D1 and D3 was much higher than it was for "You" and "She" trials. In general, there was a trend towards decreasing levels of all three response measures after condition D1. During the generalization training period, these responses had dropped out completely.

Table 3 shows the average percentage of trials in which echolalia co-occurred with correct pronoun production, incorrect pronoun production, and non-pronoun production for each pronoun throughout the production and discrimination training conditions. Interestingly, on the "first try" in all cases, except "She" trials during P3 and D2, the over-all percentage of co-occurrence between echolalia and correct pronoun production was significantly higher than that for co-occurrence between

Table 3. Average percentage of trials in which echolalia co-occurred with correct pronoun production, incorrect pronoun production and non-pronoun production for "You," "I" and "She" in each training condition.

	"You"						"I"					
	Correct		Incorrect		Non-pro.		Correct		Incorrect		Non-pro.	
	1st try	2nd try	1st try	2nd try	1st try	2nd try	1st try	2nd try	1st try	2nd try	1st try	2nd try
B1	45.0	0.0	0.0	0.0	0.0	0.0						
B2							04.6	13.6	0.0	0.0	04.0	0.0
C1	32.0	25.5	02.9	0.0	21.2	0.0	34.1	22.7	07.6	0.0	24.7	01.0
C3	0.0	0.0	0.0	0.0	0.0	0.0	70.0	0.0	10.0	0.0	0.0	0.0
							"She"					
B3							03.3	0.0	03.3	0.0	0.0	0.0
C2	30.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	06.7	0.0	0.0	0.0
C3							10.0	0.0	0.0	0.0	0.0	0.0

echolalia and incorrect pronoun production. The level of co-occurrence of echolalia with both incorrect pronoun production and non-pronoun production was generally below ten percent. Only during "You"-"I" discrimination training (D1) was the level of co-occurrence between echolalia and non-pronoun production higher than this. The percentage of co-occurrence of echolalia with incorrect pronoun production was never much higher than it was with correct pronoun production or non-pronoun production. On the "second try," co-occurrence between echolalia and correct pronoun production was, except in the case of P2, less frequent than it was on the "first try," while co-occurrence of echolalia with incorrect pronoun production and non-pronoun production was almost nonexistent.

DISCUSSION

Efficacy of the General Procedure

Sequential training of "You," "I" and "She" pronoun responses to criterion level was successful. However, a basic flaw in the implementation of the experimental design makes it difficult to infer with certainty that the imitation and differential reinforcement procedures used in production training of the three pronoun responses were actually responsible for the change in pronoun production. Although baseline measures of responding to each of the three initial training stimuli were obtained in condition B, no additional probes for responding to "I" and "She" stimuli were given after production training for "You" was completed. Thus, evidence that the level of correct pronoun production in response to "I" and "She" stimuli did not change significantly until imitation and differential reinforcement was applied in those training conditions is not available.

"Reversed Role" and "New Persons" Generalization

The results of the generalization test were equivocal concerning the hypothesis that neither form of generalization would occur at a significant level. As was expected, the more difficult of the two forms, "reversed role" generalization, did not occur at a satisfactory level. But the level of performance in response to "new persons" generalization

probe stimuli was higher than was expected. Of course, it is possible that the degree of generalization obtained by this test was somewhat inflated because of the administration of response contingent "second tries" on generalization probe trials. However, examination of the generalization test data does not reveal any trend towards a greater degree of correct pronoun production as trials progressed for either form of generalization. Furthermore, it is important to note that both forms of generalization were tested in the context of a complex discrimination task. That is, the subject was required to match each pronoun response with three of six discriminative stimuli, which differed from each other in terms of both the identity of the persons involved and their relationship to the subject. Had the discrimination task been less complex, the level of generalization might have been greater. Nonetheless, the results of the generalization test are valid in the sense that the discrimination task with which the subject was confronted roughly approximated the kind of situation which he must face in his natural environment.

Taken as a whole, the results of the generalization test tend to support the conclusions of Kale et al. (1968), Garcia (1974), and Stokes et al. (1974) that generalization of trained verbal responses to persons not involved in the original training situation is not always an automatic result of simple application of training procedures. In the case of "new persons" generalization, the level of correct responding was just above chance, while it was definitely below chance in the case of "reversed role" generalization. The cause or causes for this deficiency in generalization are difficult to determine directly, but the previously

discussed trends in incorrect pronoun production provide some clues. These biases in responding may have been a result of a failure by the subject to regularly attend to the relationship between himself and the performer of the action. The results of the generalization test also suggest that responding on the basis of the relational aspect of the stimuli was not what was trained during the initial training conditions. It is possible, as was suggested by Lovaas et al. (1971) concerning "stimulus overselectivity" in autistic children, that the subject's responding was affected by his tendency to focus on a limited portion of the total stimulus array, the identity of the performer of the action, during both the initial training conditions and the generalization test. Although the results of the generalization test may be explainable in other terms, they are generally congruent with the findings of Lovaas et al. (1971) and Sailor and Taman (1972), both of which tend to support the above conclusion.

The choice of "reversed role" generalization over "new persons" generalization for further training was based on the fact that the former was not obtained as a satisfactory level. Furthermore, "reversed role" generalization requires the subject to respond in a manner counter to what was trained in the initial training conditions. The subject's responses must have been based only on the relational aspect of both the initial training and "reversed role" stimuli in order to be correct. The "matching of opposites" principle was employed as a means of demonstrating to the subject that the correct response for each "reversed role" stimulus was the same as that for its initial training counterpart,

even though the identity of the performer of the action was changed. The results of generalization training indicate that the use of this procedure in conjunction with imitative prompts and differential reinforcement at the beginning of training had the desired effect. There is no way of determining, however, if this "matching of opposites" principle was an essential feature of the over-all procedure, since design limitations, to be discussed later, prevented a comparison between it and simple random presentation of the different stimuli.

Limitations of the Experimental Design

One other explanation for the subject's deficiency in generalization is that it was a product of features inherent in the design of the experiment rather than in the operant procedures involved in its implementation. This explanation is not in conflict with the one previously offered concerning the effects of "stimulus overselectivity," since it is probable that all experimental results are a product of interaction between characteristics of the subject and those of the training procedure. The subject's failure to discriminate adequately between the two "reversed role" stimuli during the generalization test can be viewed as an example of "overgeneralization," in which trained responding to the initial training stimuli overgeneralized to the two "reversed role" stimuli on the basis of the identity of the performer of the action, causing the majority of responses to these stimuli to be incorrect. Overgeneralization of grammatical rules, for example, the use of "foots" instead of "feet" (Brown and Bellugi 1964), is a fairly common phenomenon in normal language acquisition and is usually rectified without special

training. On the other hand, overgeneralization of a trained response in a language deficient child (Guess et al. 1968, Guess 1969, Schumaker and Sherman 1970) is much less likely to follow such a course. As Ruder and Smith (1974) have pointed out, overgeneralization may be useful as a proof of the power of some training procedure, but it is of questionable desirability as a product of a language training program.

Ruder and Smith (1974) recommend that language training programs be designed to minimize the degree of overgeneralization in order to maximize their efficiency. In many cases, the language responses chosen for training are linguistically interdependent. That is, they can be classed together as surface manifestations of some underlying linguistic element, as in the case of plural noun endings or of pronouns. Ruder and Smith suggest that linguistically interdependent responses should be trained concurrently, that is simultaneously within the same modality (Schroeder and Baer 1972), rather than sequentially in order to prevent the occurrence of overgeneralization. Concurrent training is preferable to sequential training presumably because for any set of stimulus-response combinations, simultaneous use of differential reinforcement procedures for all combinations will train the subject to use each response only with the stimulus for which it is appropriate before those responses have a chance to generalize to other stimuli. Another reason for the use of concurrent instead of sequential training is that the former is more congruent with the nature of the subject's linguistic environment. As in the case of the generalization test, a person's linquistic environment can be viewed as a complex discrimination task which requires that

he choose from sets of simultaneously available response alternatives in order to make correct responses.

In the case of the present study, "reversed role" and "new persons" stimuli can be seen as semantically equivalent variants of the initial training stimuli for each pronoun response. Thus, both the stimuli and the responses are linguistically interdependent. The present study embodied a sequential training format for the initial and "reversed role" stimuli. Because of the interdependent nature of these two sets of stimuli, concurrent training might be more appropriate.

Recently, an interesting finding was obtained by Garcia (1974) concerning the effects of two different forms of generalization probe upon subjects' generalization of trained responses. After training of several conversational response forms to picture stimuli in the presence of two trainers was accomplished, two forms of probe session for generalization to untrained pictures were administered. One form, "general" probe sessions, involved presentation of just untrained pictures, with reinforcement administered on a noncontingent basis. The other form, "intermixed" probe sessions, involved intermixed presentation of both trained and untrained pictures, with reinforcement administered on a variable schedule for correct responding to the trained pictures. Through the use of a multiple baseline design across trainers, it was demonstrated that while generalization of responding to the untrained pictures did not occur for either of the trainers during general probe sessions, it did occur for each of them in sequence as soon as they participated in intermixed probe sessions. After the subjects were exposed

to intermixed probe sessions with each of the two trainers, a third trainer, who had never been involved in the original training sessions, was able to obtain generalization of responding to untrained pictures. This last result can be viewed as both further support for the efficacy of the procedure and as an indication of its potency and potential as a training tool.

It could be argued that the differences between the effects of the two forms of probe are due to their use of different forms of reinforcement rather than to the difference in the type of stimuli used. This is a weak argument, since the difference between partial reinforcement and noncontingent reinforcement is probably a subtle one for most subjects, especially under probe session conditions, which are relatively short in duration. Furthermore, there is evidence that reinforcement is not really an essential factor for the obtaining of generalization. McReynolds and Engman (1974) administered generalization probe stimuli intermixed with trained stimuli without using any reinforcement and were still able to obtain a high level of generalization. Perhaps what is essential is that the subject experience some association between already trained stimuli and unfamiliar stimuli, with reinforcement acting only as a facilitator of general responding. Interestingly, McReynolds and Engman found that as the number of trained stimuli increased from one to three, the level of generalization to untrained stimuli went from a low level of 15 percent of those presented to a level of over 80 percent. This finding suggests that a subject's tendency to generalize to untrained stimuli may be strengthened simply by adding more trained

stimuli to the generalization probe. However, since no controls were provided, it is not possible to overrule the possibility that the rise in the level of generalization was merely due to repetition of the probe sessions, as would be predicted from the Garcia results.

The results of Garcia's (1974) study indicate that the intermixing of trained with untrained stimuli is in some way essential for obtaining generalization. Apparently, a subject's ability to generalize to untrained stimuli is at least partly determined by the type of generalization test used. If this is true, then studies which have intermixed trained with untrained stimuli in their generalization probes, such as Guess (1969), Schumaker and Sherman (1970), Wheeler and Sulzer (1970), and Lutzker and Sherman (1974), may not have obtained pure measures of the efficacy of their training procedures in producing generalization. It is possible that the generalization obtained in these studies was due to the use of intermixed probes rather than to any feature of the actual training procedures used. Unfortunately, any attempt to answer this question by presenting untrained stimuli in isolation from trained stimuli may obtain the same results as did Garcia in his use of general probe sessions, namely little or no generalization. Perhaps the best way to resolve this dilemma would be to consider the intermixed probe as a special procedure for obtaining generalization, as Garcia has suggested, rather than as a test procedure per se. At the same time, a pure measure of generalization, using the procedure for general probe sessions, may yield positive results if it were administered after generalization to another set of untrained stimuli had been

established through the use of intermixed probes. In any case, instances of isolated "follow-up" generalization test are either extremely rare or non-existent in the realm of language training experiments. The use of such tests would be a worthwhile step toward increasing the scope of the functional analysis of generalization.

The above criticism of intermixed probes as a generalization test procedure may also be applicable to some extent to the procedure used in the present study. However, since the subject was required to respond in a complex discrimination task on the basis of relational aspects of the untrained stimuli, the effects of association between trained and untrained stimuli were probably either irrelevant or counter-productive to correct responding. Unfortunately, the extent of the subject's generalization to stimuli other than those involved in training and testing was not determined because of time limitations.

Another problem which may be attributable to the experimental design was the resurgence of echolalia during the discrimination training conditions. Echolalia had been present to some degree during the three production training conditions, but was under control by the end of training through the use of the fading-in procedure. What may have caused this resurgence is difficult to determine. One possibility is that the return of echolalia was a result of an increase in incorrect pronoun production and a consequent rise in the frequency of negative feedback during discrimination training. However, this is not a plausible explanation, since Table 3 shows that the subject was echolalic in conjunction with correct pronoun production generally much more

frequently than in conjunction with incorrect pronoun production. An alternative explanation would be that the return of echolalia was simply a result of the fact that the subject was required to discriminate between different stimuli and was consequently under a greater level of stress. There was a significant drop in the average level of echolalia from D1 to D2, while it was fairly low in the three-way discrimination in D3 and nonexistent during the generalization training condition. It would seem from this trend that echolalia dropped out as a function of the subject's increased proficiency and familiarity with the discrimination tasks. Further support for this conclusion comes from the fact that there were similar trends in the levels of irrelevant responses and non-responding as well. Of course, it is also quite likely that the reduction in the level of echolalia as discrimination training conditions progressed was at least in part a case of extinction due to lack of immediate reinforcement.

Several implications may be drawn from the above discussion. Echolalia is considered to be one of the defining characteristics of childhood autism and has been regarded by many to be pathological in nature (Risley and Wolf 1967, Lovaas et al. 1973). But, as Risley and Wolf have shown, it can also be used productively for the establishment of language skills and effectively controlled through fading procedures. The results of the present study suggest that the traditional assumptions concerning the controllability of echolalia as well as its pathological nature may not be entirely accurate. Although echolalic responding can be controlled by fading procedures in situations involving

non-discriminative training of responses to verbal cues, control will not necessarily be maintained after the subject is required to use those responses in a discrimination context. Furthermore, although echolalia may not actually be facilitative toward correct responding in a discrimination task, it would seem that it is at least not always detrimental. There are also several practical implications. If echolalic responding to a verbal cue reappears in a discrimination task after being controlled in a previous production training condition, one way of dealing with the problem would be to follow the Risley and Wolf procedure by eliminating the verbal cue and fading it in again. However, such an approach could be quite time consuming, especially if the frequency of echolalic responding is in fact a function of the subject's proficiency and familiarity with the discrimination task. Two alternative approaches are available. Perhaps the simplest method would be, as in the present study, to refrain from dealing directly with echolalia while continuing discrimination training until the problem goes away. Support for this approach comes from evidence that echolalia in a sample population of autistic children declined concurrently with increases in various appropriate behaviors, possibly as a product of differential reinforcement procedures (Lovaas et al. 1973). The other alternative would be to establish non-echolalic responding to the non-verbal stimuli during production and discrimination training conditions and wait until criterion performance has been attained in the discrimination training condition before fading in the verbal cue. The question of which of the three alternatives is the most practical in terms of time consumption and resistance to loss of

control cannot be answered by the present study. However, it is open to experimental evaluation and is worth answering.

Limitations of the Multiple Baseline Approach

Within recent years, the multiple baseline approach has enjoyed increasing popularity as an evaluative tool for language training procedures. The multiple baseline approach has several features which make it preferable to other approaches such as the ABAB design (Baer et al. 1968, Baer and Sherman 1970, McReynolds 1974). Perhaps what most recommends the use of the multiple baseline design for clinical applications is its flexibility. It can be used across responses (Schumaker and Sherman 1970, Lutzker and Sherman 1974), across subjects (Stokes et al. 1974; Guess, Sailor and Baer 1974), across settings, and even across trainers (Garcia 1974). In spite of this flexibility, there are several difficulties which require caution in the use of this approach to language training programs. These difficulties are not intrinsic to its methodology, but rather pertain to the nature of linguistic behavior as well as to recent developments in the area of language training programming.

Since the early use of operant techniques for language training in language deficient children, research has been focused mainly on developing and refining procedures which are both practical to apply and optimally functional for the child (Guess et al. 1974). Typically, these studies have taken some limited portion of the total range of language behavior as the subject of their functional analysis. Only recently has

there been any attempt to develop comprehensive programs for dealing with children who suffer from varying degrees of language deficiency (Guess et al. 1974, Stremel and Waryas 1974). These latter approaches differ from the former studies not only in scope but in complexity as well. In the program developed by Stremel and Waryas, target responses are arranged in a hierarchical structure starting with the most basic responses and progressing towards more and more behaviorally complex ones. While a multiple baseline analysis may be applicable to the training of a simple set of equally complex responses such as prepositions (McReynolds 1974), the use of such an analysis cannot be applied to any series of responses in which there is more than one level of complexity. For example, a sequential analysis of the training of construction (a), (noun) + (adjective), followed by the training of construction (b), (noun) + is + (adjective), would be subject to the criticism that prior acquisition of construction (a) contributed more to the learning of construction (b) than did the training procedure in question. This criticism would still hold even if probes administered after training of (a) showed no change in the level of (b). In general, the choice of related responses such as those above is in violation of one of the basic rules for using the multiple baseline design, namely that the target responses be functionally independent of each other (Baer and Sherman 1970). Thus, the multiple baseline approach is not suited to demonstrating the efficacy of program for training a series of responses which are hierarchically arranged.

Another area in which programming considerations may conflict with assumptions governing the use of the multiple baseline design is that of procedural variations. It may be desirable when training a set of responses to make minor changes in the training procedure according to the nature of the responses or the stimuli involved in their training. In the case of the present study, the same general procedure was used throughout, but the procedure used for training correct responding to "reversed role" stimuli was not an exact copy of that used during the initial training period. It involved the intermixing of untrained stimuli with previously trained stimuli. While such variations may not directly prohibit the use of a multiple baseline design, they make interpretation of results difficult. This is because there are no objective criteria for determining when procedural variations are within the bounds of some basic procedure or when they actually constitute different procedures. If the variations are great enough to constitute separate procedures, then the usefulness of the repetitive aspect of the multiple design as a means of discounting the possibility of coincidence is lost. On the other hand, keeping the same procedure for responses which could be trained more effectively through the use of minor variations merely in order to allow the use of a multiple baseline design may have long term negative effects for a language training program. Similarly, the requirement that training of some responses be delayed so that a sequential analysis can be carried out stands in opposition to the previously discussed need for concurrent training of linguistically interdependent responses. Here too adherence to the rules of the

multiple baseline approach may produce negative byproducts, such as overgeneralization or lack of generalization, which reduce the efficiency of the over-all training program.

One conclusion which can be drawn from the above discussion is that Sherman and Baer (1969) were somewhat one-sided in their claim that subservience of experimental design to a therapeutic goal may result in an antitherapeutic outcome. The obverse is true as well. This does not mean that the use of the multiple baseline approach is entirely unsuitable for the analysis of language training programs. Rather, it means that its use must be restricted in some cases and that greater use of its inherent flexibility should be made. While there appear to be several reasons for rejecting the use of multiple baseline designs across responses as a tool for evaluating training programs, the same does not apply to the use of other dimensions of sequential analysis such as subjects, trainers and stimuli. The use of a multiple baseline design across subjects has several advantages over the other alternatives. First, the use of a wide variety of subjects in a sequential analysis of some procedure can provide evidence that the procedure is useful for a large proportion of the population of language deficient children. The same can be done for programmed sequences of procedures as well. On the other hand, this approach can also be used to identify portions of the general population for which the procedure will not produce the desired outcome. In general, the use of the multiple baseline approach across subjects is well suited for the experimental analysis of training

programs in their entirety, while the other dimensions may be better suited for the analysis of limited segments or features of such programs.

As Guess et al. (1974) have pointed out, the history of the field of functional analysis of language behavior has reached a crossroad. Research has sufficiently demonstrated that the general procedures of operant training are powerful tools for modifying a wide variety of language responses. However, other issues remain to be dealt with. Perhaps the two most important issues for the field at this time are (1) the comparison of alternative procedures for training the same set of responses, and (2) the evaluation of the extent of generalization of trained language behaviors to stimuli outside of the initial training context. Previous forms of experimental analysis have been focused on demonstrating that a given procedure was in fact responsible for the obtained behavior change. While no comprehensive procedure for comparing the efficacy of various training procedures has yet been developed, several issues stand out as being useful criteria for comparison of training procedures. Questions which relate to the efficacy of a given procedure concern (1) how much training was required, (2) whether or not control over previously operating undesirable behaviors was achieved and maintained, (3) whether or not any undesirable behaviors were produced through the use of the procedure, (4) to what extent were any such behaviors produced, and (5) to what extent was generalization produced. The last criterion can be divided into three sub-areas: (1) the extent of intraverbal generalization, (2) the extent of extraverbal generalization, and the extent of cross-modality generalization. Cross-modality

generalization refers to generalization of expressive language skills to their receptive (comprehension) counterparts, or vice versa. Although there is some evidence that training of skills in one modality facilitates learning in the other modality (Asher 1972), the majority of studies in this area indicate that cross-modality generalization is not an automatic outcome of training in just one modality (Ruder and Smith 1974; Ruder, Smith and Herman 1974; Guess 1969; Guess and Baer 1973).

As discussed earlier, the use of a multiple baseline design across responses was not suitable for demonstrating the efficacy of a set of procedural variations. However, the use of such a method might be suitable for the comparison of two or more different procedures. Garcia's (1974) use of the multiple baseline approach in his study can be seen as the application of an AB design, with replication across trainers, in which the general probe procedure acted as the baseline condition and the intermixed probe procedure acted as the treatment condition. In comparison to condition A, which did not produce the desired results, condition B did so in several cases. There is a serious problem with this method, however, in that it provides no way of controlling for the possibility that the subject's experience with condition A facilitated the behavior change obtained by condition B. Of course, this does not mean that Garcia's conclusions were invalid, since there was already ample evidence from other studies that use of intermixed probes without prior exposure to general probes produced generalization. On the other hand, when two alternative procedures of roughly equal efficacy are to be compared, the problem of possible positive or negative effects

of one upon the other due to sequential application makes interpretation difficult. However, this does not mean that use of the multiple baseline approach is incompatible with comparison of alternative procedures. Guess et al. (1974) are currently engaged in a research program which involves simultaneous use of a multiple baseline design across subjects and comparison of alternative procedures at various steps in their training program. While experimental analysis of a given procedure can be made either across responses or across subjects, comparison between procedures can be made by use of a group design, in which each procedure is assigned to a separate group of subjects, thus eliminating the need for sequential application. Such a method appears to be the best solution to this problem.

SUMMARY

A program for training use of the first, second and third person pronouns, based on the work of Lovaas (1968) and Risley and Wolf (1967), was developed for the purpose of assessing the extent of extra-verbal generalization to two types of stimuli, "reversed role" and "new persons." Training of the three pronoun responses to criterion level was achieved, although it was not clearly demonstrated that the training procedures involved were responsible for the results. The results of the generalization test were equivocal with regards to the expectations concerning the level of generalization which would be obtained for each type of stimulus. As was expected, the level of generalization to "reversed role" stimuli was below chance; however, generalization to "new persons" stimuli was slightly above chance level. In general, the results of the generalization test indicated that the subject was responding to all stimuli on the basis of the identity of the performer of the action rather than his relation to the subject. The results also tended to support the conclusion of Garcia (1974) and others that generalization of trained responses to persons not involved in the original training situation is not an automatic outcome of operant training procedures. A special supplemental program for training generalization of correct responding to "reversed role" stimuli, based on intermixing of the two types of stimuli, was successful.

Several negative byproducts which may be attributable to the nature of the experimental design were discussed. Of main importance was the problem of overgeneralization. Recent work in the field of language programming suggests that overgeneralization may be prevented through the use of concurrent rather than sequential training of linguistically interdependent responses. Another problem was that of loss of control over echolalic responding. Several alternative means of dealing with this problem were discussed. Thirdly, a discussion of the traditional procedure for administering generalization probes pointed to the possibility that this procedure is not an accurate measure of the extent of generalization but may be in fact a powerful procedure for training generalization.

The final section of this paper consisted of a discussion of some of the limitations of the multiple baseline approach as an evaluative tool for language training programs. It was suggested that the use of multiple baseline designs across responses involves methodological assumptions which are at odds with current trends in language training, which include the use of hierarchical training sequences, concurrent training of linguistically interdependent responses, and procedural variations. It was stressed that other dimensions of analysis, such as subjects, settings, stimuli and trainers be employed more frequently as alternatives. Finally, the issue of comparison of alternative training procedures was discussed. A list of criteria upon which to base comparisons was suggested. A discussion of the applicability of the multiple baseline approach to comparing different procedures indicated that it

would not be as suitable as a group design, although its use would not be incompatible with comparing procedures.

REFERENCES

Asher, J. Children's first language as a model for second language training. Modern Language Journal, 1972, 56, 133-138.

Baer, D. M. and J. A. Sherman. Behavior modification: Clinical and educational applications. In H. W. Reese and L. P. Lipsitt (eds.) Experimental child psychology. New York: Academic Press, 1970, pp. 643-672.

Baer, D. M., M. M. Wolf and T. R. Risley. Some current dimensions of applied behavior analysis. Journal of Applied Behavior Analysis, 1968, 1, 91-97.

Brown, R. and U. Bellugi. Three processes in the child's acquisition of syntax. In E. Lenneberg (ed.) New directions in the study of language. Cambridge: The M.I.T. Press, 1964, pp. 131-161.

Fygetakis, L. and B. B. Gray. Programmed conditioning of linguistic competence. Behavior Research and Therapy, 1970, 8, 153-163.

Garcia, E. The training and generalization of a conversational speech form in nonverbal retardates. Journal of Applied Behavior Analysis, 1974, 7, 137-149.

Goldstein, S. and R. Lanyon. Parent clinicians in the language training of an autistic child. Journal of Speech and Hearing Disorders, 1971, 36, 552-560.

Guess, D. A functional analysis of receptive language and productive speech: Acquisition of the plural morpheme. Journal of Applied Behavior Analysis, 1969, 2, 55-64.

Guess, D. and D. M. Baer. An analysis of individual differences in generalization between receptive and productive language in retarded children. Journal of Applied Behavior Analysis, 1973, 6, 311-329.

Guess, D., W. Sailor and D. M. Baer. To teach language to retarded children. In R. L. Schiefelbusch and L. L. Lloyd (eds.) Language perspectives--acquisition, retardation, and intervention. Baltimore: University Park Press, 1974, pp. 529-563.

Guess, D., W. Sailor, G. Rutherford and D. M. Baer. An experimental analysis of linguistic development: The productive use of the plural morpheme. Journal of Applied Behavior Analysis, 1968, 1, 297-306.

Kale, R. J., J. H. Kaye, P. A. Whelan and B. L. Hopkins. The effects of reinforcement on the modification, maintenance, and generalization of social responses of mental patients. Journal of Applied Behavior Analysis, 1968, 1, 307-314.

Kozloff, M. A. Reaching the autistic child: A parent training program. Champaign, Illinois: Research Press, 1973.

Lenneberg, E. H. On explaining language. Science, 1969, 164, 635-643.

Lovaas, O. I. A program for the establishment of speech in psychotic children. In H. N. Sloane and B. D. MacAulay (eds.) Operant procedures in remedial speech and language training. Boston: Houghton Mifflin Co., 1968, pp. 125-156.

Lovaas, O. I., R. Koegel, J. Q. Simmons and J. S. Long. Some generalization and follow-up measures on autistic children in behavior therapy. Journal of Applied Behavior Analysis, 1973, 6, 131-166.

Lovaas, O. I., L. Schreibman, R. Koegel and R. Rehm. Selective responding by autistic children to multiple sensory input. Journal of Abnormal Psychology, 1971, 77, 211-222.

Lutzker, J. and J. A. Sherman. Producing generative sentence usage by imitation and reinforcement procedures. Journal of Applied Behavior Analysis, 1974, 7, 447-460.

McNeill, D. Developmental psycholinguistics. In F. Smith and G. A. Miller (eds.) The genesis of language: A psycholinguistic approach. Cambridge: The M.I.T. Press, 1966, pp. 15-84.

McReynolds, L. V. Application of systematic procedures in clinical environments. In L. V. McReynolds (ed.) Developing systematic procedures for training children's language, American Speech and Hearing Association Monograph, #18. Danville, Illinois: Interstate Press, 1974, pp. 131-140.

McReynolds, L. V. and D. L. Engman. An experimental analysis of the relationship of subject and object noun phrases. In L. V. McReynolds (ed.) Developing systematic procedures for training children's language, American Speech and Hearing Association Monograph, #18. Danville, Illinois: Interstate Press, 1974, pp. 30-46.

Risley, T. R. and M. M. Wolf. Establishing functional speech in echolalic children. Behavior Research and Therapy, 1967, 5, 73-88.

Ruder, K. F. and M. D. Smith. Issues in language training. In R. L. Schiefelbusch and L. L. Lloyd (eds.) Language perspectives--acquisition, retardation, and intervention. Baltimore: University Park Press, 1974, pp. 565-605.

Ruder, K. F., M. D. Smith and P. Hermann. Effect of verbal imitation and comprehension on verbal production of lexical items. In L. V. McReynolds (ed.) Developing systematic procedures for training children's language, American Speech and Hearing Association Monograph, #18. Danville, Illinois: Interstate Press, 1974, pp. 15-29.

Sailor, W. and T. Taman. Stimulus factors in the training of prepositional usage in three autistic children. Journal of Applied Behavior Analysis, 1972, 5, 183-190.

Schroeder, G. and D. M. Baer. Effects of concurrent and serial training on generalized vocal imitation in retarded children. Developmental Psychology, 1972, 6, 293-301.

Schumaker, J. and J. A. Sherman. Training generative verb usage by imitation and reinforcement procedures. Journal of Applied Behavior Analysis, 1970, 3, 273-287.

Sherman, J. A. Imitation and language development. In H. W. Reese and L. P. Lipsitt (eds.) Advances in child development and behavior. New York: Academic Press, 1971, pp. 239-272.

Sherman, J. A. and D. M. Baer. Appraisal of operant techniques with children and adults. In C. M. Franks (ed.) Assessment and status of the behavior therapies and associated developments. New York: McGraw-Hill, 1969, pp. 192-219.

Staats, A. W. Learning, language and cognition. New York: Holt, Rinehart and Winston, 1968.

Stokes, T. F., D. M. Baer and R. L. Jackson. Programming the generalization of a greeting response in four retarded children. Journal of Applied Behavior Analysis, 1974, 7, 599-610.

Stremel, K. and C. Waryas. A behavioral-psycholinguistic approach to language training. In L. V. McReynolds (ed.) Developing systematic procedures for training children's language, American Speech and Hearing Association Monograph, #18. Danville, Illinois: Interstate Press, 1974, pp. 96-130.

Waryas, C. Psycholinguistic research in language intervention programming: The pronoun system. Journal of Psycholinguistic Research, 1973, 2, 221-237.

Wheeler, A. and B. Sulzer. Operant training and generalization of a verbal response form in a speech-deficient child. Journal of Applied Behavior Analysis, 1970, 3, 139-147.