THE UNIVERSITY OF MANITOBA

A COMPARISON OF THE EFFECTS OF TWO PARADIGMS OF TIMEOUT UNDER TWO SCHEDULES OF POSITIVE REINFORCEMENT ON ANTONYM LEARNING BY SEVERELY RETARDED GIRLS

by

GARY REIGH HUGHES

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF PSYCHOLOGY

WINNIPEG, MANITOBA

February, 1973

ABSTRACT

A COMPARISON OF THE EFFECTS OF TWO PARADIGMS
OF TIMEOUT UNDER TWO SCHEDULES OF POSITIVE REINFORCEMENT
ON ANTONYM LEARNING BY SEVERELY RETARDED GIRLS

by

GARY REIGH HUGHES

Two paradigms of timeout (TO as punisher and TO as negative reinforcer) were employed under two schedules of reinforcement (CRF and VI) to determine effects on antonym learning by three severely retarded girls.

Although all conditions improved responding considerably, the highest rates of correct responding were obtained when subjects were on a VI - 30 sec. schedule of positive reinforcement with TO as a punisher for errors. However, accuracy of responding and achievement in terms of various criterion measures were not consistently related to the experimental conditions. TO appeared to have a positive effect on correct responding regardless of the paradigm.

The results are interpreted as favoring the view that TO may be used effectively in the training of retarded children.

ACKNOWLEDGEMENTS

I wish to express my sincere appreciation to Dr. Garry L. Martin for his advice, encouragement, aid, inspiration and all of the many other positive reinforcers that he so generously bestowed throughout the course of my program. My indebtedness to Dr. Joe Pear is also great since he first kindled my interest in operant behavior and provided guidance from initial training stage to final oral examination. Dr. M.D. LeBow is accorded my thanks for his part in my candidacy examination and dissertation proposal. Dr. G. Lowther, medical superintendent of the Manitoba School for Retardates has my sincere gratitude for his cooperation in permitting this and other research to be carried out at the school and for serving on my thesis examination committee. The comments and suggestions of Dr. R. Norton, the external examiner, were much appreciated. Thanks are also due to C.L. Bell, J. Anderson, Walter Mostoway, Judy Briggs and all the staff at Cedar Cottage for their help and friendly encouragement.

This research was supported in part by a grant number MA 3636 from the Medical Research Council of Canada and by Public Health Research Grant number 606-7-255 of the National Health Grants Program, Ottawa, Canada, and by research bursaries from the National Research Council of Canada and the National Institute on Mental Retardation.

TO

THE MEMORY OF

CLIFFORD J. ROBSON

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

One of the early landmarks in the history of the experimental analysis of behavior was the publication of Skinner's first book, The Behavior of Organisms, in 1938. The area has grown considerably since that time and, despite a slow start, it now occupies a substantial portion of the psychological literature. Two journals, the Journal of the Experimental Analysis of Behavior and the Journal of Applied Behavior Analysis, have been founded in order to disseminate the growing body of research that has employed this particular methodology. Within the last 15 years, a large number of texts have appeared in which the general procedures and concepts are dealt with at some length (e.g., Sidman, 1960; Honig, 1966; Ferster and Perrott, 1968; Millenson, 1967; Reynolds, 1968; Reese, 1966). In recent years, the application of these principles has led to the rapidly expanding field of applied behavior analysis (e.g., Ullmann and Krasner, 1965; Ulrich, Stachnik and Mabry, 1966, 1970; Bandura, 1969; Schaefer and Martin, 1969; Franks, 1969; Sloane and MacAulay, 1968).

Applied behavior analysis encompasses the full range of phenomena of operant conditioning. The scope of the present study was limited to an examination of only a few of these phenomena. More specifically, this study was an attempt to discover some of the variables that determine the effectiveness of timeout (TO) from positive reinforcement as a technique for producing more efficient performance.

Theoretical Background

While there are several ways in which TO may be arranged, TO from positive reinforcement as a punisher and TO from positive reinforcement as a negative reinforcer were the paradigms employed in the present study. TO may serve as a punisher when it is presented following the occurrence of a specified response. To be considered as an effective punisher, the presentation of TO must occasion a decrease in the frequency of that response. In two conditions in the present study, TO was used as response-contingent punishment to decrease errors. TO may serve as a negative reinforcer when a specified response of an organism terminates, delays or prevents the presentation of the TO. If TO is effective as a negative reinforcer, the effect of TO will be to increase the frequency of that specified response. In the two conditions in the present study where TO as a negative reinforcer was employed, TO was used to attempt to bring about an increase in correct responding. The end result of these two arrangements should be similar, that is, they should both serve to alter the ratio of correct to incorrect responses even though the problem is attacked from different directions. Thus, in the case of TO as a punisher, the ratio is altered by decreasing the frequency of errors. In the case of TO as a negative reinforcer, the ratio is altered by increasing the frequency of correct responses.

TO may be arranged in a variety of other ways. TO as a punisher, for example, may involve removing the subject from the situation in which reinforcement normally occurs or otherwise

preventing him from making the response. A typical extinction procedure, in which a previously reinforced response is no longer reinforced, may also be considered as a TO procedure. Ferster & Skinner (1957) suggested presentation of a stimulus (S^{Δ}) correlated with non-reinforcement and (for pigeons) turning out the lights of the experimental chamber as additional methods of imposing TO. Finally, TO may be arranged to serve as a positive reinforcer. In this situation, a given response of an organism must lead to the presentation of a period of TO, which in turn must result in an increase in the frequency of emission of that response. For example, a fixed-ratio (FR) schedule may become aversive when the ratio exceeds a certain level and subjects will voluntarily impose TO from such a schedule. In this situation, TO is considered to be a positive reinforcer of the response that led to the TO. A number of studies have demonstrated this phenomenon in basic research settings (Azrin, 1961; Appel, 1963) and it has recently been studied in an applied setting as well (Steeves, Martin and Pear, 1970).

Although the definition of TO as a period of time during which positive reinforcement is not available is quite well agreed upon (Verhave, 1966; Steeves et al., 1970; Martin, Moir and Skinner, in preparation, c), Leitenberg (1965) pointed out that there is no single set of operations which adequately defines TO from positive reinforcement. In reviewing various TO procedures used in studies with the retarded, Gardner (1969) also remarked that considerable variation exists among the specific operations involved. Further-

more, Gardner noted that TO punishment procedures usually include any of three elements, each of which could contribute to the deceleration effects frequently obtained. These three elements are: (1) the removal of some aspect of the stimulus environment which is identified either as a reinforcing stimulus or as a discriminative stimulus for reinforcing events, (2) the creation of a situation which may contain conditioned aversive properties, e.g., isolation may be aversive to some subjects, and (3) the operation of factors related to the reinforcing properties of the termination of TO. Therefore, a TO may be effective as a result of (a) loss of the possibility to obtain positive reinforcement, (b) the presentation of a conditioned aversive stimulus, (c) negative reinforcement following the termination of TO, or (d) any combination of these factors. Thus, in spite of the fact that TO has been the object of experimental study for well over a decade [compare early experiments by Ferster (1957, 1958) to the considerable research since that time (Leitenberg, 1965)], Azrin and Holz (1966) concluded that "the means where by a time-out stimulus acts as a punisher is not completely clear at the present time (p.390)." This fact not withstanding, Gardner (1969) noted that TO procedures still are very popular among those working with the mentally retarded.

Much of the work with TO has been predicated on the assumption that TO serves as a punisher, that is, it serves to lower the probability of the behavior upon which it is contingent. This assumption led Leitenberg (1965) to ask whether TO from positive

reinforcement was indeed an aversive event and to review the experimental evidence. On the basis of this review Leitenberg concluded that "TO satisfied sufficient criteria to conclude that it belongs to the class of stimuli called 'aversive' (1965, p. 428)." However, he suggested that this conclusion is only tentative in view of the fact that some differences between the effects of shock and TO were found and since some experiments allow interpretations other than that TO is aversive. Furthermore, he contended that much research remains to be done.

The main concern of the present study is not whether TO can be "officially" classed as a punisher, but whether it can be an effective tool for increasing "correct" responding on simple tasks performed by retarded children.

In the section to follow, the relevant basic and applied research has been reviewed. The distinction between basic and applied follows, in part, the suggestion by Baer, Wolf and Risley (1968). That is, a study was considered as applied research when the behavior under study was of importance to the individual being studied (e.g., learning to speak) and as basic research when it was not (e.g., lever pressing).

Recent Basic Research

Recently, Willoughby (1969) examined the effects of TO on a lever-press response of preschool children. In one experiment, a lever press was reinforced with tokens on a VI 10 sec. schedule with a 10 sec. TO (lever locked) administered on an FR 7. When

the punished response was the only one available, TO depressed response rate by approximately 25%. However, this suppression was of relatively brief duration. With the introduction of an unpunished alternative response (pressing a second lever), TO produced complete and lasting suppression. In a second experiment a response preference was established by reinforcing responses on one lever more frequently than those on another and then punishing responses on the preferred lever with a 10 sec. TO. Results showed that under punishment by TO the formerly unpreferred lever was chosen even though the frequency of reinforcement for the two levers was identical. Willoughby concluded that the practical considerations to be derived from this study are clear: if TO is used as a method of eliminating undesired behaviors in children, then the child should also be given the opportunity to emit a second unpunished behavior in addition to making TO a consequence of the undesired response. Willoughby (1970) performed another experiment with preschool children to discover whether a subject's preference for a response which was punished by TO would be in any way influenced by the specific consequences of the response serving as an alternative to TO. Two experimental conditions were used: in one the response alternative to TO was reinforced with the same frequency as the punished response; in the other, the response alternative to TO was not reinforced, i.e., extinguished. In the first condition, TO produced complete and lasting suppression

while in the second condition, the response that led to TO was significantly preferred over the response that was neither punished nor reinforced. That this should be so, however, appears to be quite obvious, i.e., to find that a child would prefer to work for reinforcement in the face of TO punishment (which is viewed as a rather mild form of punishment) than to work for nothing at all should surprise no one.

Martin (1970) performed a series of experiments to determine whether or not a maximization of the discriminative properties of TO was necessary to efficiently suppress behavior. Subjects in this study were retarded boys. Martin found that for TO to be maximally effective in suppressing behavior, stimuli correlated with these extinction periods must be consistently presented. Extinction periods or TOs which are non-signalled (i.e., which don't have consistent cues associated with them) are relatively ineffective in suppressing the behavior which they follow. Martin asserted that TOs must be correlated with specific and distinctive stimuli to effectively suppress behavior. TO must function as a discriminative stimulus for the absence of reinforcement. Unless there is a stimulus associated with TO, using TO to suppress certain behaviors may have the opposite result in that the unsignalled TO may actually increase the undesired behavior. Martin concluded that these findings should be helpful to those experimenters who seek to use TO to suppress

behaviors: to insure the effectiveness of their procedures, the discriminative properties of TO from positive reinforcement should be maximized. Unless this is done, TO may be inefficient, ineffective and may contribute undesirable results.

Carlson and Aroksaar (1970) performed an experiment designed to determine whether TO is aversive or whether the observed effects are artifacts attributable to the fact that behaviors which prevent or terminate TO typically also lead to increases in positive reinforcement. To test this the authors placed rats on a concurrent schedule with a TO being contingent upon one of two concurrent operants, but the availability of reinforcement for the alternative, unpunished operant was adjusted to equal lower levels than that for the punished operant. Responding on the punished lever was suppressed (punishment involved a TO by retraction of both levers) and rate of responding on the unpunished lever increased. The authors concluded that the results favored the view that TO functioned as an aversive stimulus in this context, lending further support to Leitenberg's (1965) tentative conclusion that TO is an aversive event.

No review of the area of the basic research in TO would be complete without at least a brief mention of matching-to-sample studies. Ferster and Appel (1961) found that when correct responses were reinforced on a continuous reinforcement schedule, TO following incorrect responses had little effect on the matching accuracy

of pigeons. On the other hand, when correct matching responses were reinforced on a VI 3 min. schedule, matching accuracy progressively improved as the TO duration for incorrect responses was raised from 1 to 10 to 30 to 60 seconds. Performance began to deteriorate when TO interval was 2 min. duration. Zimmerman and Ferster (1963) obtained generally similar results using different TO durations, that is, matching accuracy of pigeons increased as the frequency of TO increased. Optimal matching accuracy was obtained when every incorrect response was followed by a TO of intermediate duration.

Although TO has been found to be effective in a variety of situations, some investigators have found it to be an ineffective measure under certain circumstances. With mental patients as subjects, Holz, Azrin and Ayllon (1963) found that a 30 sec. TO was effective when positive reinforcement (cigarettes) were presented for plunger pulling on a VI 30 sec. schedule but ineffective when positive reinforcement was presented on a VI 90 sec. schedule. Baron, Kaufman and Rakauskas (1967) found TO to be ineffective in reducing rates of responding of college students' button pressing on an FR schedule for money reinforcements. These investigators suggested that the results of this study and those of Holz et al., (1963) could be attributed to the fact that reduction in rate of responding would have been accompanied by reduction in reinforcement frequencies as well.

The last study to be discussed in this section (Thomas, 1964),

is of particular interest. Thomas found that pigeons would perform a key-pecking response to avoid TO more frequently when positive reinforcement was on a VI 9 min. schedule than when it was on a VI 1 min. schedule. Thomas suggested that the lack of occurrence of avoidance behavior on the VI 1 min. schedule may have been due to the effects of that schedule's relatively high reinforcement frequency. This led to the inclusion of the means for determining the effects of reinforcement schedule in the present study.

The interest in schedule of reinforcement here is with the manner in which it effects the efficacy of TO, whether as punisher or negative reinforcer. While the literature related to schedule effects is prolific, the above-mentioned relation has not been extensively studied. In fact, very little research has been done in this area at all. The primary concern of this thesis was to look at some simple and easily applied ways of making TO a more effective tool for training the retarded in academic tasks. Therefore, since the variable of interest is TO, this literature is the subject of a fairly extensive review below. The literature on schedule effects, although peripheral here, has been the subject of extensive study previously, e.g., in Ferster & Skinner (1957), Morse (1966), Catania (1968).

Applied Research

Eliminating Undesirable Behavior

Considerable applied research has been done utilizing various

TO procedures to eliminate undesirable behavior in retarded children. Giles and Wolf (1966) used brief interruption of meals as an aversive event to eliminate behaviors of profoundly retarded subjects which interfered with the development of simple self-care skills. Peterson and Peterson (1967), in a program designed to eliminate the self-injurious behaviors of an eight year old severely retarded boy, successfully used TO from primary reinforcement (food) and TO from secondary reinforcement (social interaction). Hamilton and Allen (1967) used TO procedures plus response cost to control inappropriate mealtime behaviors of severely and profoundly retarded females. The basic procedure was one of removing the subject from the dining area without the rest of her meal. They obtained a rapid and permanent reduction in inappropriate dining room behavior from the daily average of 60 during baseline. A 12 month follow-up revealed a daily average of less than one inappropriate response. Hamilton and Stephens (1967) in work with a moderately retarded girl were successful in rapid elimination of floor-rolling and screaming using a procedure of TO plus physical restraint. It is interesting that success with a social isolation consequence was evident in this case even though the subject was described as being consistently socially isolated, withdrawn, and that she interacted with no one on the ward.

Tate and Baroff (1966) used a brief TO from physical contact

in an effort to eliminate undesirable self-injurious behavior in a nine-year-old blind boy. The results of this study indicated that the relatively simple procedure of immediate brief withdrawal of physical contact produced a dramatic reduction in the frequency of chronic self-injurious behavior. The usual extinction procedure of ignoring the behavior was not effective in reducing its frequency of occurrence.

Baer (1962) was able to suppress thunbsucking in three nursery school children who were persistent thumbsuckers, by showing them cartoon movies and programming interruptions of both picture and sound track. These interruptions were made contingent upon thumbsucking and the representation of the cartoon contingent upon the removal of the thumb from the mouth.

Bostow and Bailey (1969) used TO plus DRO (differential reinforcement of other behavior) to successfully reduce loud vocal responses in a female psychotic, and to effectively reduce aggressive behavior in a seven-year-old psychotic boy.

Barton, Guess, Garcia and Baer (1970) used TO to improve mealtime behaviors of retardates. In some cases the child was removed from the room and placed in a TO room for the remainder of the meal, in other cases, the subject had his tray removed for 15 sec. following undesirable behavior. The procedure led to a marked and useful reduction in the behavior.

Wahler (1969) found that parental use of a combined TO from

parental reinforcement and differential attention program effected dramatic and stable changes in the children's oppositional behavior. The author noted that TO was effective even though parental reinforcement value was initially quite low.

Wasik, Senn, Welch and Cooper (1969) used TO from social reinforcement contingent upon behavior classified as aggressive and resistive with culturally deprived school children. By the end of treatment, the desirable behavior had increased markedly and remained high through subsequent follow-up.

The literature on the application of TO procedures continues to grow. Most recent studies include the use of TO in rapid toilet training of the retarded (Azrin & Foxx, 1971), modifying enuresis (Nordquist, 1971), eliminating high-rate misbehavior (Pendergrass, 1972), modifying self-injurious behavior (Smolev, 1971), reducing stuttering (Martin & Haroldson, 1971; Adams and Popelka, 1971) and so forth. Even from this very brief review, it is obvious that TO has been used effectively to eliminate or modify a wide variety of undesirable behaviors.

Reducing Errors During Learning

In addition to its widespread use in the elimination of undesirable behavior, TO has also been employed as an aid to reducing errors or other incompatible behavior during learning. For example, TO has been used for increasing the efficiency of speech training (McReynolds, 1969). In two experiments with a brain-damaged five-

year-old boy, TO was used effectively to reduce undesired and increase desired responses. In Exp. I, the procedure consisted of temporarily halting language training (TO) following verbal jargon and resuming it (time-in) contingent upon the boy sitting quietly in his chair for approximately 30 sec. The jargon declined to almost zero for an extended period each time the procedure was employed. In Exp. II, the procedure consisted of halting language training (TO) after emission of undesired verbal responses which previously had been reinforced as the desired approximation of the target verbal behavior. Resumption of training (time-in) was made contingent upon the emission of the then-desired approximation of the target verbal behavior. In each experiment, the contingent TO and time-in of the language training that involved positive reinforcement effectively reduced the undesired and increased the desired responses.

Martin and Berndt (1970) used "time-out from speaking" to punish stuttering in a 12-year-old boy. They found that stuttering declined from a base rate of 28 per two min. to essentially zero following the introduction of a response contingent TO. Although the procedures initially resulted in a concomitant reduction in number of words read, by the last two or three TO sessions word output was back to near the base rate level.

Martin et al., (in preparation, c), in a comparison of two TO durations (7 1/2 and 17 1/2 seconds) as punishment for

errors during an academic task (learning antonym pairs) found that some subjects (two autistic boys) showed fewer errors, less inattentive behavior and better learning under 7 1/2 second TO condition while other subjects (one autistic boy and two retarded girls) showed little difference in performance under the two conditions. Martin, Hardy and McDonald (in preparation, a) in a comparison of several TO intervals as consequences for errors during a picture-naming task found that three out of four of the severely retarded girls who served as subjects in the experiment performed as well or better on a zero TO duration than 10, 20 or 30 sec. TO for errors. Martin, McDonald, Michalishyn and Hardy (in preparation, b) did a comparison of several TO intervals as consequences for errors during tracing training with autistic and retarded children. They found that three out of four subjects showed the highest percentage of errors under the no-TO condition but they suggested that the extra trials possible on a no-TO condition more than made up for the slightly higher percentage of errors under that condition. These results suggest the need for further careful research in the area to determine whether TO can be effective in such learning situations.

While it has been worthwhile to review some of the literature in which the TO procedure has been used, it must be pointed out that much of the work has had certain methodological deficiencies. Of the punishment studies reviewed by Gardner (1969), he noted

that only a few met the rigorous requirements of a "single organism, within subjects design." It is interesting that these methodological deficiencies should persist in the face of numerous exhortations to improve procedures (Gelfand and Hartmann, 1968; Gardner, 1969; Roos and Oliver, 1970; Pawlicki, 1970; Lovitt, 1970). As recently as 1968, Baer, Wolf and Risley noted that primary concern of investigators at this stage in the development of applied behavior analysis is usually with reliability rather than with parametric analysis or component analysis. But a mere two years later Lovitt (1970) was moved to remark that if behavior modification is to advance from a technology to a science, methodical parametric or component analysis must be undertaken. It is in the latter spirit that the present study was performed, within the limits of the applied research framework.

The Purpose of the Study

The TO duration studies performed under the direction of Martin (in preparation, a, b, c) have seriously questioned the justification for using TO during training programs since such procedures appeared to occasion inefficient use of training time. That is, the amount of time the child spent in TO might have been put to use more profitably with extra training trials. However, these studies have three things in common - first, positive reinforcement was presented on a continuous reinforcement

schedule (CRF); second, TO was administered in a punishment paradigm; and third, except for one study, there was no unpunished alternative response that could be emitted readily by the subject. In academic tasks, such as those studied in the Martin experiments, it is often difficult, if not impossible, to do anything about this third problem because it is the fundamental nature of the task. However, it is possible that by presenting positive reinforcement on an intermittent schedule, TO as a punisher might be more effective (e.g., Thomas, 1964). It is also possible that TO as a punisher is not effective in these situations. Perhaps a more fruitful approach would be that of McReynolds (1969) in which TO was used as a negative reinforcer to reduce errors.

Thus, the purpose of the study was to determine whether TO could be made more effective by altering the associated schedule of positive reinforcement, and whether one paradigm of TO might be more effective than another. In order to accomplish this, positive reinforcement was presented on either a VI or a CRF schedule and TO was administered either as a punisher or as a negative reinforcer. The resultant two by two arrangement yielded four distinct conditions which were then administered to each subject.

CHAPTER II

METHOD

Subjects

The subjects were chosen from the residents of Cedar Cottage, a residential treatment unit for 30 severely and profoundly retarded girls at the Manitoba Training School, Portage la Prairie, Manitoba. Since there are some important differences between the girls in terms of age and diagnosis, a brief description of each girl is given below.

Marion. Marion was born September 21, 1953. She was almost 18 years old at the time of the study. Marion was diagnosed as hydrocephalic, with encephalopathy due to unknown or uncertain cause. Recently, she obtained an IQ rating of 22 on the Stanford-Binet Intelligence Scale.

Valerie. Valerie was born September 15, 1952. She was almost 19 years old at the time of the study. Valerie was diagnosed as mongoloid. An IQ rating of 24 was obtained on the Stanford-Binet Intelligence Scale by Valerie recently.

Agnes. Agnes was born November 14, 1956. She was approximately 14 1/2 years old at the time of the study. Agnes was diagnosed as epileptic with encephalopathy due to unknown or uncertain cause. Recent testing indicated an IQ rating of 30 to 35 on the Stanford-Binet Intelligence Scale.

All three girls used in this study had been conditioned

previously to respond for token reinforcers, to sit quietly and make eye contact with the experimenter (attending behavior), and to follow simple commands. Although there was considerable variation in the extent and elegance of their behavioral repertoires, this was not crucial since the effect of the manipulated variables upon performance was assessed by intra-subject replication.

Apparatus

This research was conducted in one of the offices at Cedar Cottage. During the session, the child was seated in a tablet-arm chair, placed against the wall so that the child could not leave her seat. The experimenter was seated opposite the subject with a small table at his side containing the teaching materials, tokens (pegs in a pegboard), back-up reinforcers (various edibles such as Apple Jacks, Popcorn Curls and Smarties), stop-watch, cassette tape recorder, running time meters, light panel and data sheets. The teaching material consisted of pictures of objects depicting antonym pairs pasted on heavy paper.

The discriminative stimulus (S^D) light display consisted of a rectangular box divided into compartments with openings of approximately 4 in. by 5 in., each illuminated from the rear by means of a colored Xmas tree light (blue, green, red and gold) shining through a frosted plexiglass plate. A small 12 watt white light mounted on the side of this box served as the

signal for a reinforcement period. This seemed necessary in order to make the periods of TO, time-in (TI) and reinforcement as discriminable as possible. The S^D lights were operated through a multiple position switch which in turn was connected to a switch on a power control box. The same switch that operated the S^D lights also operated the Cramer running time meter on which TI was recorded. A second switch on the power control box operated the reinforcement light and a Grason-Stadler running time meter on which reinforcement time was recorded. A third switch operated the Lehigh Valley running time meter on which TO was recorded. The cassette recorder on which the VI-30 sec. schedule was recorded was operated through the time-in switch. Total session time was monitored with a stopwatch.

General Methodology

The research strategy common to the experimental analysis of behavior (Sidman, 1960), popularly known as the single-organism approach, was employed. To distribute sequential effects, the conditions were presented in a different order in each session with a brief rest period between each condition. Discriminative stimuli for the various conditions were maximally differentiated, this being a major requirement for the use of a multi-element baseline (Sidman, 1960).

Thus, the pairs of pictures depicting the various antonyms

were placed on a large, brightly coloured cardboard rectangle, the colour of this background varying from condition to condition. Antonym pairs to be learned under Condition I were always placed against a blue background, Condition II - green background, Condition III - red background and Condition IV - gold background. Similarly, a light panel situated on the experimenter's table and facing the subject was activated so that a blue light was on during Condition I and so forth. A description of the composition of the conditions is given in the section which follows.

The importance of clear discriminative stimuli signalling the onset and duration of the TO period as stressed by Martin (1970), was also taken into account. At the onset of TO, then, the experimenter turned away from the subject in an obvious manner, changed facial expression from pleasant to neutral, and turned off the discriminative stimulus light for that condition. When TO was terminated, the behavior involved in the above sequence was reversed.

General Procedure

To determine the effect of the schedule of positive reinforcement upon the effectiveness of TO, two schedules of positive reinforcement were used; namely, continuous reinforcement (a token and social reinforcement for every correct response) and a variable interval schedule (a token and social

reinforcement provided for the first response following expiration of a certain interval, namely, VI 30 sec.: range 15 sec. to 45 sec.). To determine the effect of TO as a punisher vs. TO as a negative reinforcer, two arrangements were used. In the first, TO followed an error and continued for the designated period (15 sec .) irrespective of whether the subject subsequently emitted the correct response. In the second, TO followed an error and continued until the subject emitted the correct response at which point TO was immediately terminated. (No reinforcement, other than that occasioned by termination of TO, was given for the correct response that terminated TO). The two levels of these two variables were paired with each other as follows:

Condition I - positive reinforcement on CRF, TO as punisher;

Condition II - positive reinforcement on CRF, TO as negative reinforcer;

Condition III-positive reinforcement on VI, TO as punisher;

Condition IV -positive reinforcement on VI, TO as negative reinforcer.

Each condition was presented in a separate 15 min. session with a 5 min. rest period between each session. The order of the sessions was alternated.

Specific Procedure

The procedure to be followed was similar to that of Martin

et al., (in preparation, c). The task was one in which the subject had to learn pairs of antonyms. For each pair of antonyms, five items were chosen and corresponding pictures were accumulated. For example, for the antonyms wet-dry, pictures of wet and dry land, faces, tin cans, hair, and cars were obtained.

Because the subjects had been previously taught to point to an item and say the name of the item when the experimenter said, "Show me the (name of the item)", a baseline of the subject's knowledge of the antonyms was obtained in the following way. The experimenter placed a pair of items (e.g., a picture of a wet face and a picture of a dry face) before the subject and said, "Show me the (name of one item, e.g., wet face)." The subject's pointing performance was recorded and immediately followed by another pair of pictures and another question. This procedure was continued without consequence for the subject's responses. Following every second question, the subject was asked to perform a previously acquired response unrelated to the experimental task for which he received a token. Tokens were exchanged for back-up reinforcers (edibles) on a 5/1 token/ reinforcer ratio as soon as five tokens were accumulated. The baseline was continued until each item of the antonym was tested on three separate trials, with the position of the antonyms alternated from trial to trial. From these data, probability

values for all items were calculated, that is, if the subject made the correct response three times out of six, the probability of getting that item correct was .50.

A period of preliminary shaping followed the baseline period. At this time, each subject was trained to make the alternative response when TO (as a negative reinforcer) was imposed following an error. On some initial occasions, this involved a physical prompt, that is, placing the subject's finger on the correct picture. Gradually these prompts were faded until the subject was making the alternative response upon presentation of the discriminative stimuli associated with TO alone.

Following preliminary shaping, the antonyms were placed in a pool, with a set of antonyms being withdrawn when required and used in one condition exclusively for each child. Higher probability sets were chosen first in order to help establish the contingency as firmly as possible. For each antonym set of five pairs of items, then, training was begun with the highest probability pair in the set and proceeded through to items of lower probability. The lowest probability item pair was used to test for generalization of the particular antonym concept. As previously mentioned, four daily sessions of 15 min. duration were run with each subject. Thus, each subject was exposed to each condition every day (five days a week), with the

sequence of conditions alternating from day to day. Within sessions, a trial was initiated when the subject made eye contact with the experimenter. A response was considered correct when the child pointed to the specified item within 10 sec. Any other response was considered incorrect.

In Condition I, in which positive reinforcement was on a CRF schedule, all correct responses were immediately reinforced with a verbal exclamation by the experimenter, "good girl!" and a token. Tokens were exchanged for back-up reinforcers on a 10/1 ratio as soon as 10 tokens were accumulated. The back-up reinforcer consisted of any one of a variety of edible reinforcers, such as Smarties, Apple Jacks, Corn Puffs, etc. Also in this condition, TO was used as a punisher for errors. Thus, all incorrect responses were followed by presentation of a 15 sec. period of TO, in which the experimenter turned away from the subject, changed his facial expression to neutral, and ignored her for that period. The TO period was further signalled by turning out the S^D light. It is to be noted here that no aversive event other than TO was used. That is, the experimenter did not simultaneously say "No!" or present any conditioned or unconditioned aversive stimulus. This was done so that the effects of TO could be assessed apart from any other means of decreasing behavior.

In Condition II, positive reinforcement for correct responses

was programmed as in Condition I. TO for errors, however, was programmed as a negative reinforcer rather than as a punisher. That is, when the subject made an error, the experimenter imposed the TO as described above with the exception that TO was not terminated until the subject emitted the correct response. No reinforcement, other than the termination of TO, was given for a correct response at this point to avoid contaminating the negatively reinforcing properties of termination of TO with positive reinforcement for making the alternative response.

In Condition III, in which positive reinforcement was on a VI schedule, the subject was reinforced as in Condition I following the emission of the first correct response that occurred after the expiration of the designated interval. A tape recorder on which brief tones had been recorded at appropriate intervals signalled the occasion for reinforcement to the experimenter through an earphone. The rest of the reinforcement procedure and the use of TO as a punisher were identical to Condition I.

In Condition IV, positive reinforcement for correct responses was programmed as in Condition III, while TO for incorrect responses was programmed as in Condition II.

Since multi-element baselines such as this require clear discriminative stimuli to differentiate conditions, care was

taken to ensure that each condition was marked by distinctive cues, as described previously.

The training procedure required the subject to correctly identify four items of an antonym pair before a second antonym pair was introduced. All antonyms were taught by the program described below for teaching the antonyms, wet-dry.

a) Two items (picture of a wet face and a dry face) were placed before the subject on a sheet of brightly colored cardboard and the corresponding S^D light was turned on. The experimenter said, "Show me the wet face. That's the wet face, (while pointing to the wet face). Show me the wet face." This was considered a prompt trial and was designated P1 (see the sample data sheet in Figure 1).

b) Following a correct response to a prompt trial the experimenter said, "Show me the wet face." This was considered a question trial and was designated in Figure 1 as Q1.

c) A prompt trial (P2) and a question trial (Q2) were presented for the item, dry face.

d) Question trials were then presented for the item wet face (Q1) and dry face (Q2) according to the sequence shown in Figure 1.

ANTONYM PAIR "Wet-Dry"

ITEM "Face"

SUBJECT _____

DATE _____

CONDITION _____

	✓ OR X	TOKEN	BACK-UP	TO DURATION		✓ OR X	TOKEN	BACK-UP	TO DURATION
P1					P1				
Q1					Q1				
P2					P2				
Q2					Q2				
Q1					Q1				
Q2					Q2				
Q2					Q1				
Q1					Q1				
Q2					Q2				
Q1					Q2				

Fig. 1. A sample data sheet showing the sequence of prompt trials (P) and question trials (Q) used to teach an item of an antonym concept. Trials marked "1" designate one member of a pair, e.g., "wet", while trials marked "2" designate the other member of the pair, e.g., "dry".

e) If the subject made a wrong response to a question trial, the appropriate time-out was given. Then the prompt trial for that item was presented again and then followed by the series of question trials shown below the double line in Figure 1.

f) When the subject responded correctly to the series of question trials for wet and dry face, then that item was considered to have reached a satisfactory criterion of learning within a session. The experimenter then taught wet-dry land, cars, and tin cans similarly. Each series of question trials followed a different sequence (see Figure 1) in order to control for the possibility of position learning by the subject.

g) At the start of the second and subsequent sessions, all items that reached criterion in the previous session were tested. If a subject responded incorrectly to an item on a test, then that item was retaught in the manner just described. Items that were responded to correctly were retested at the start of the following session. If an item was correct on its first presentation for three sessions in a row, it was considered as learned and dropped from subsequent testing. These criteria of learn-

ing, previously described by Risley and Wolf (1967), require proficiency both within a session and across three sessions.

h) When all items of a concept were learned, the fifth item was presented without consequences to test for generalization of the concept. The fifth item (e.g., wet-dry hair) was presented on ten occasions in order to evaluate guessing by the subject. If the fifth item was identified correctly on 80 percent of the trials, the concept was considered as generalized.

The subjects were run under each condition every day in the manner described above. An experimental probe (Sidman, 1960) was introduced at a stable point in the sessions, the nature of the probe being determined by performance under the various conditions. For Marion, TO as negative reinforcer (TONR) appeared to produce the highest percent correct performance. During the probe sessions, therefore, TONR was eliminated in order to determine whether this variable was instrumental in producing the better performance. For the same reason, TO as punisher (TOP) was removed during the probe sessions for Valerie and Agnes. Following the five sessions allotted to the probe, the contingency was reinstated and the subjects were run for an additional five sessions under the original conditions. The procedure during the

probe session involved the presentation of a prompt trial immediately upon the making of an error, rather than the imposition of TO.

The main dependent variables were percent correct responses and number of correct responses per minute. In addition, a number of other measures were recorded including number of words learned to the various criterion levels, number of generalization tests passed, number of tests failed, mean number of trials per session, mean total TO per session, mean number TOs per session, mean individual TO per session, mean total reinforcement time per session, mean number of reinforcements per session, mean individual reinforcement duration, mean total TI per session and ratio of usable to unusable time per session.

While it was originally planned to keep a record of wasted time, this plan was abandoned for two reasons. First, very little inattention was noted during the baseline phase. Second, time was limited since the design called for four periods of each condition within a session. Thus, if one were to permit wasted time, much valuable session time would be lost. Since the purpose of the study was primarily concerned with the effects of paradigm of TO under two schedules of reinforcement on the rate of correct responding, it was felt that it would be advisable to obtain as much data on this factor as possible, even at the expense of losing the data on wasted time.

Interobserver reliability checks were abandoned as well because the nature of the task allowed little room for error in judgement. That is, there was very little equivocation as to whether the subject pointed to one member of the set or the other.

CHAPTER III

RESULTS AND DISCUSSION

The results are presented and discussed separately for each subject first with some minor comparisons across subjects. In subsequent discussion, the similarities and differences across subjects are considered in greater detail.

Marion

Cumulative percent correct responses by conditions for Marion are shown in Figure 2. In spite of the fact that Condition II (CRF - TONR) appeared to produce consistently superior performance in terms of percent correct responses, it is questionable whether this difference is of any particular practical value. Table 1 appears to indicate that the maintenance of this superiority on the cumulative graph was due in part to high performance in the early sessions (Block 1). Table 1 also indicates that the difference over all trials between the condition producing the highest percent correct responses (Condition II) and the condition producing the lowest percent correct responses (Condition IV) is only slightly more than five percentage points. As Baer, Risley and Wolf (1968) pointed out, such a difference would hardly be a conspicuous example of applied behavior analysis. However, the dramatic positive difference between percent correct responses during the baseline period and during Block 1 is a clear illustration that each

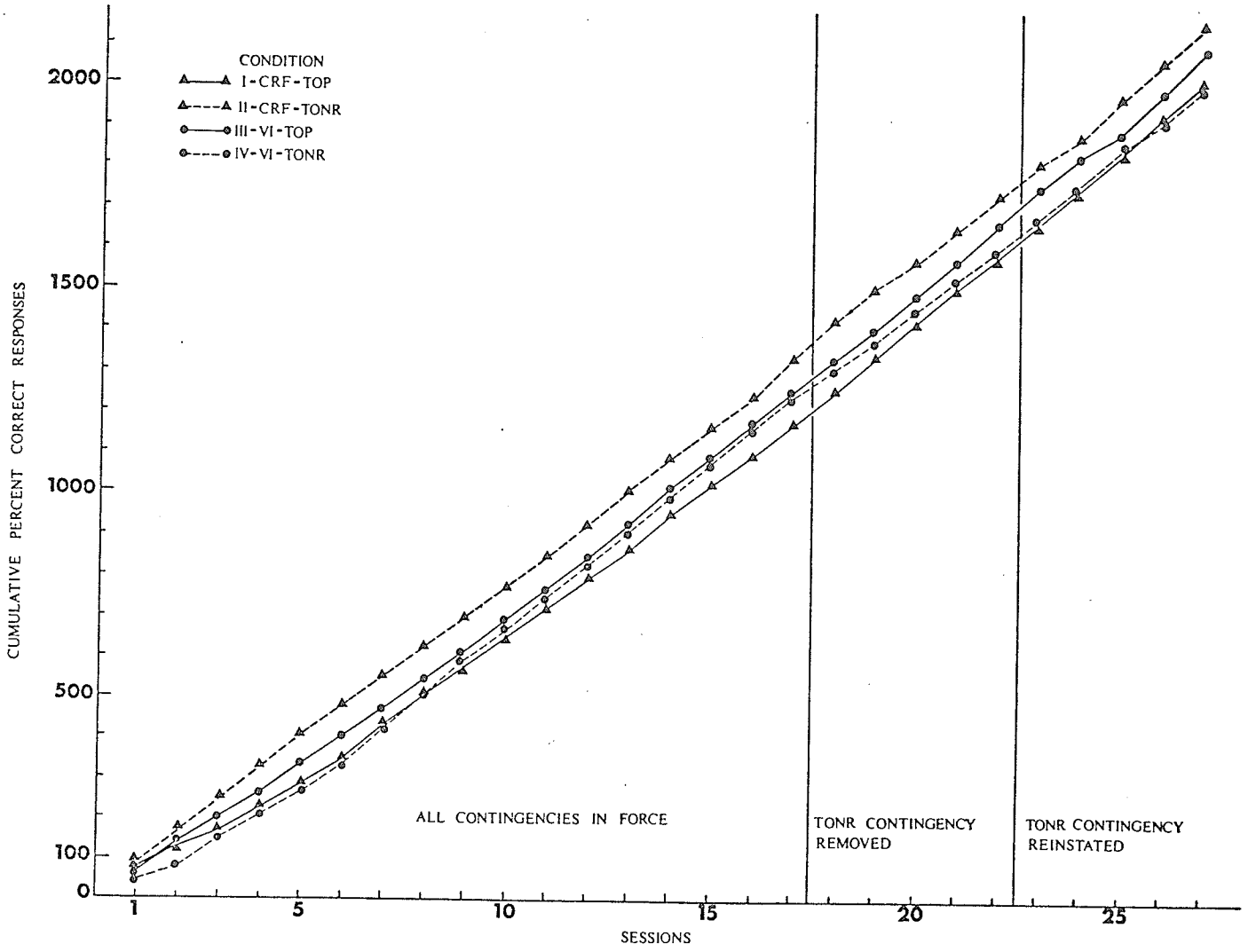

Fig. 2. Cumulative percent correct responses by sessions for Marion

TABLE 1

MEAN PERCENT CORRECT RESPONSES BY BLOCKS FOR MARION

		BLOCK					
CONDITION	BASELINE	1**	2*	3*	4*	5*	ALL TRIALS
I-CRF-TOP	54.44	62.04	72.95	75.39	84.01	84.79	74.81
II-CRF-TONR	50.83	79.47	75.20	81.85	79.42	85.23	80.18
III-VI-TOP	53.33	67.93	74.46	82.19	82.28	84.99	77.60
IV-VI-TONR	48.33	61.03	80.31	82.40	72.53	81.00	74.38

* 5 trials

** 7 trials

condition served to bring about a significant improvement in accuracy of performance.

It will also be noted in Figure 2 that the experimental probe (removal of the TONR contingency from Sessions 18 to 22 inclusive) had little apparent effect on cumulative percent correct responses in Condition II. However, it appeared to have the effect of decreasing cumulative percent correct responding in Condition IV. Although not apparent in Figure 2, this effect appeared to be reversible in that percent correct responses recovered almost to the previous level when the contingency was reinstated. This effect is quite clearly illustrated in Figure 3 and Table 1, which depict mean percent correct responses (noncumulative) by blocks[1]. The probe appeared to have little effect upon percent correct responding under Condition II but this effect was rather noticeable under Condition IV. This would seem to suggest that the schedule of positive reinforcement was the prepotent factor in determining performance in Condition II. That is, with positive reinforcement delivered under a CRF schedule there was little that TONR could add in such circumstances toward the improvement of performance. However, in Condition IV, when positive reinforcement was on a VI schedule, the TONR contingency appeared to be a contributing factor in the maintenance of performance.

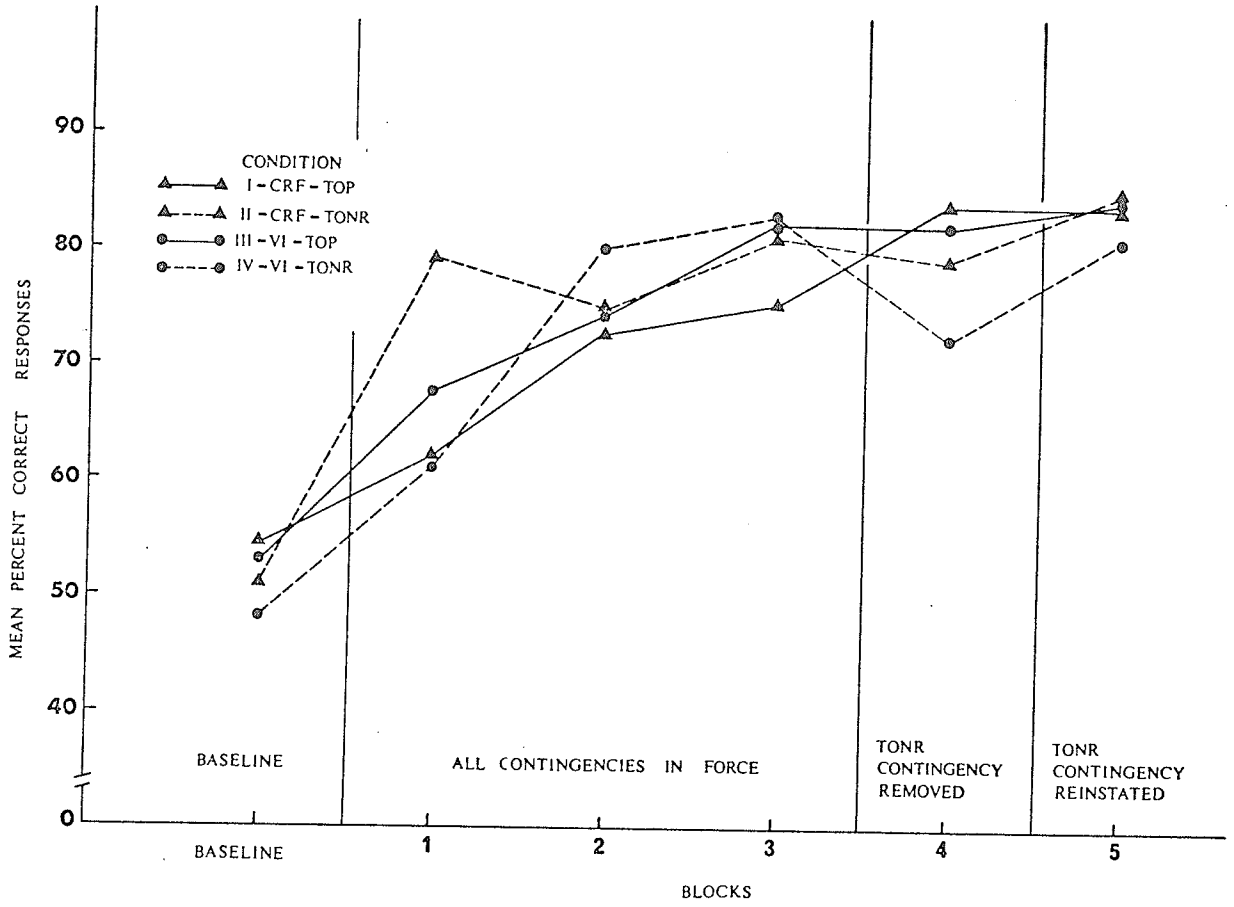

Fig. 3 Mean percent correct responses by blocks for Marion

As noted in Table 1 and Figure 3, when the TONR contingency for errors was eliminated in Condition IV, performance sharply declined (almost 10 percentage points) and dramatically improved again almost to the previous level when the TONR contingency was reinstated. When working on a CRF schedule, then, Marion appeared to require little else to produce and maintain good performance, although on a VI schedule, where reinforcements were relatively few and far between, Marion seemed to require the additional behavioral control occasioned by the TONR contingency.

During the probe period, number of reinforcements per session (reinforcement density) increased from 7.80 per session to 9.40 per session under Condition II and from 1.88 per session to 2.00 per session under Condition IV. This increase in reinforcement density was due to the fact that more reinforcement opportunities were available since TI was increased by the elimination of TO. However, reinforcement density appeared to have no positive effect on percent correct responding since the increase in reinforcement density under Conditions II and IV coincided with a decrease in percent correct responding.

A much more striking difference between conditions was found in correct responses per minute of TI as depicted in Figures 4 and 5 and Table 2. This rate represents responding as a function of available session time and excludes time

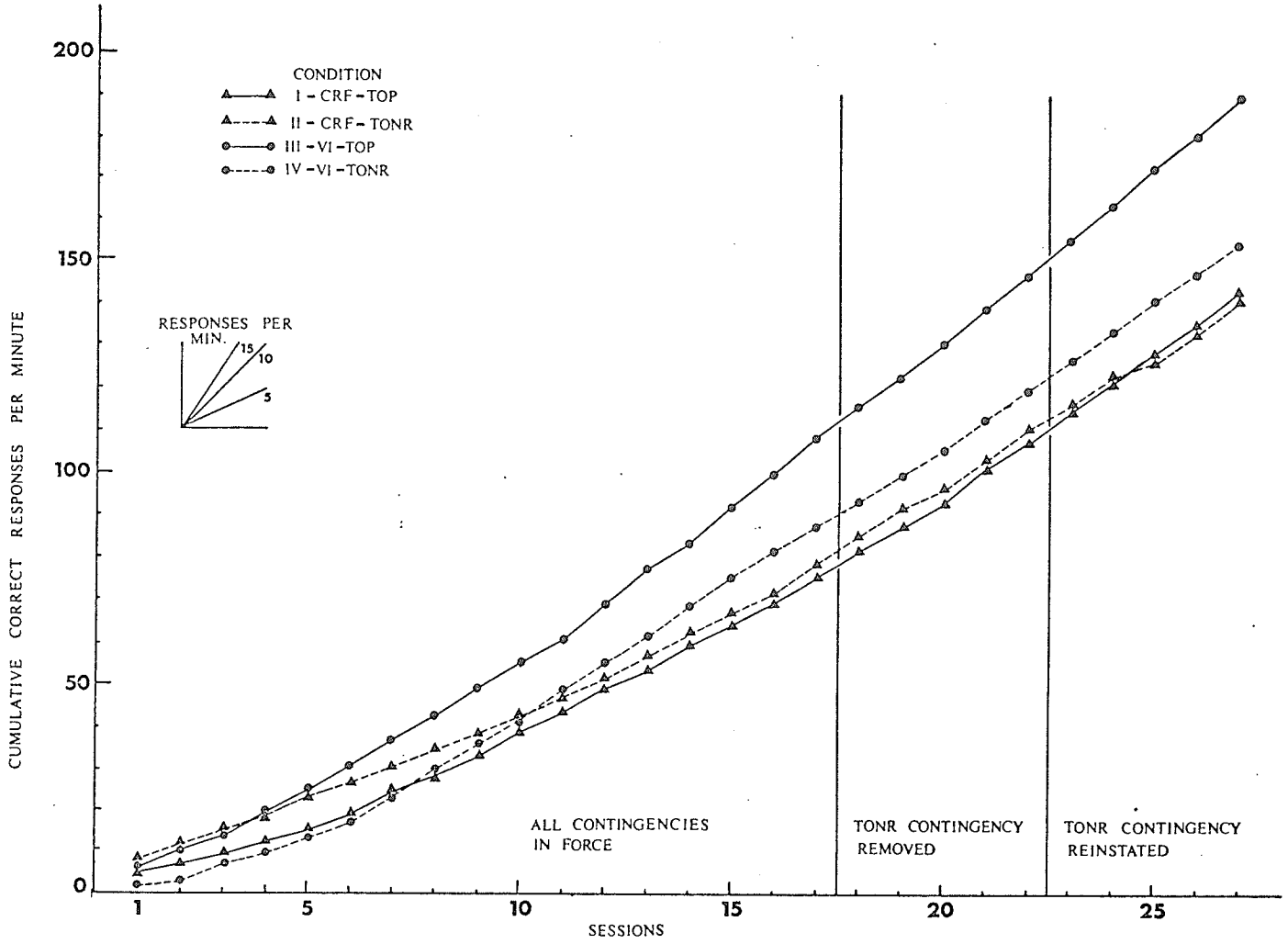

Fig.4. Cumulative correct responses per minute by sessions for Marion

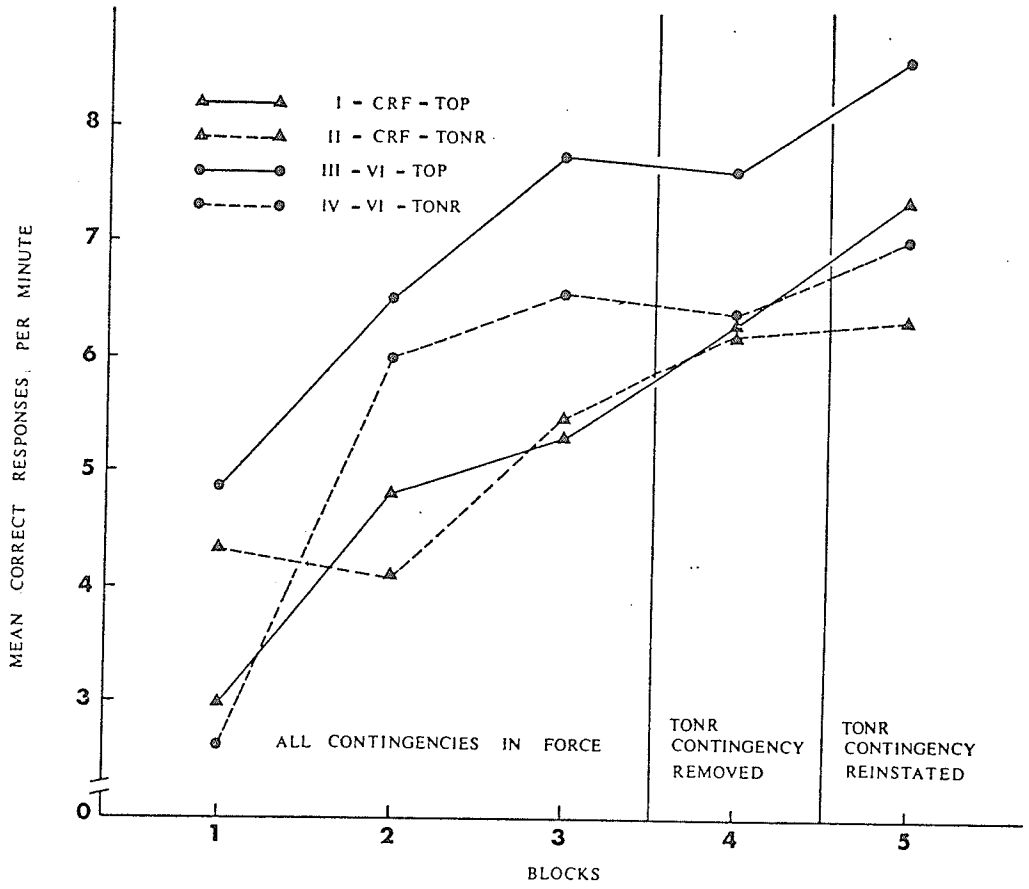

Fig. 5. Mean correct responses per minute by blocks for Marion

TABLE 2

MEAN CORRECT RESPONSES PER MINUTE BY BLOCKS FOR MARION

	BLOCK					
CONDITION	1*	2**	3*	4*	5*	ALL TRIALS
I-CRF-TOP	2.97	4.84	5.34	6.31	7.37	5.33
II-CRF-TONR	4.64	4.10	5.44	6.28	6.35	5.27
III-VI-TOP	4.86	6.50	7.74	7.64	8.64	7.03
IV-VI-TONR	2.61	6.01	6.56	6.39	7.10	5.75

* 5 trials

** 7 trials

spent in TO and reinforcement periods. According to this measure, Condition III produced rates that were from 23% to 33% higher than the other conditions. Condition II produced a high initial rate but this was not maintained in subsequent sessions. The fact that Conditions I & II produced generally lower rates than Conditions III & IV is consistent with the well-known finding that intermittent schedules produce higher rates of responding than continuous reinforcement schedules. In fact, close inspection of Figure 4 seemed to suggest that some combinations of reinforcement and TO were more effective during initial stages of performance, while the effectiveness of other conditions was not optimal until later stages.

The effect of the experimental probe (Trials 18 to 22 inclusive) upon the rate of correct responding is also seen in Table 2 and Figure 5. Rate of correct responding was not adversely affected by the probe under Condition II although a very slight decrease was seen under Condition IV (Table 2 and Figure 5). It might be concluded, therefore, that while the TONR contingency had some effect in maintaining percent correct responding (i.e., accuracy) under Condition IV for the reasons suggested above, it had little effect on the actual rate of correct responding (i.e., speed). The increased rate of correct responding under Condition II may have been due in part to the increase in reinforcement density under this condition.

Table 3 serves to shed some light on the reasons why there was some inconsistency between percent correct responding and rate of correct responding. While Condition II produced the highest percent correct responses it also produced the lowest rate of correct responding. This apparent paradox was due to the fact that the overall rate of responding (mean frequency of responses per session) was also relatively low. This means, of course, that rate of incorrect responding was low as well which accounts for the higher percent correct responding. Similarly, the high rate of correct responding under Condition III was not complemented by percent correct responding being correspondingly high because the error rate under this condition was also high. Similar explanations could be adduced for the other conditions.

While percent correct responding and rate of correct responding are important measures, certain other measures of the effectiveness of a procedure are relevant as well. In Table 4, Marion's performance in terms of words learned to various criterion levels is depicted. Here it will be noted that Conditions II and IV were equivalent with respect to words learned to first criterion and that this superior performance was fairly well maintained up to and including the number of generalization test words correct. That both Conditions II & IV should be generally equivalent in terms of criterion performance and that this performance should be high relative to the

TABLE 3

MEAN DURATIONS, FREQUENCIES AND RATIOS OF TI, TO & REINFORCEMENT FOR MARION

	MEAN DURATION PER SESSION IN SECONDS			MEAN FREQUENCY PER SESSION			MEAN INDIVIDUAL DURATION		PARADIGM EFFICIENCY
CONDITION	TI	TO	Reinf.	Total Resp.	TOs	Reinf.	TO	Reinf.	TI÷(TO+Reinf.)
I-CRF-TOP	507.47	217.56	163.26	54.22	14.43	6.61	15	25.34	1.33
II-CRF-TONR	674.56	14.48	189.68	67.37	14.22	7.94	1.63	25.89	3.30
III-VI-TOP	574.90	283.38	31.25	85.78	18.88	1.08	15	28.52	1.83
IV-VI-TONR	757.58	58.87	50.93	96.26	20.10	1.90	3.09	30.03	6.90

TABLE 4

CRITERION PERFORMANCE MEASURES FOR MARION

CONDITION	NUMBER OF ANTONYM PAIRS				NUMBER OF INSTANCES		
	LEARNED TO FIRST CRITERION	RETAINED ON FIRST ATTEMPT			SET CRITERION MET	GENERALIZATION TESTS CORRECT	CRITERION TESTS FAILED
		TO FIRST TEST	TO SECOND TEST	TO THIRD TEST			
I-CRF-TOP	12	10	7	4	1	0	13
II-CRF-TONR	32	27	19	16	6	2	20
III-VI-TOP	24	17	17	12	3	2	24
IV-VI-TONR	32	22	18	15	7	3	33

other conditions was somewhat puzzling at first. As previously noted, Condition II did show highest percent correct responses (Figure 2 & Table 1) but Condition IV was at the lowest level of this measure. The answer to this puzzle may lie in the fact that considerably more trials per session were possible under Condition IV than under Condition II (96.26 and 67.37 respectively) as shown in Table 3. This finding seems to support the contention of Martin et al., (in preparation, c) that the time spent in TO might be used more profitably to provide additional training trials. No apparent relationship was found between criterion performance and rate of correct responding in Conditions II & IV. Condition II, in fact, produced the lowest rate of correct responding while Condition IV produced an intermediate rate (Table 2). Even though the correct rate under Condition II was low, the overall rate and thus the incorrect rate were also low. Condition III, which produced the highest rate of correct responding and the intermediate level of percent correct responding would seem to be at approximately the appropriate relative position in terms of criterion performance. Similarly Condition I seemed to produce appropriate criterion performance relative to these other measures.

The fact that a higher number of criterion tests were failed under Condition IV than under Condition II even though general criterion performance was equivalent may also be attributed to the fact that more trials were presented and thus

more opportunities for failed tests occurred. The low and intermediate number of tests failed under Condition I & III respectively were proportional to other criterion performance measures and thus seem to require no further comment.

Table 3 shows mean durations and frequencies of TI, TO and reinforcement as well as an index of paradigm efficiency (ratio of usable to unusable time) for Marion. That the mean total TI should be greatest under Condition IV (see Table 3) and should decrease in the order shown (Condition II, III & I) is not surprising. Under Condition IV (VI - TONR), the least time was given over to other operations. Reinforcement occurred only intermittently and TO could be terminated by the subject. Similar explanations could follow down to the condition under which there was the least TI, Condition I. In this condition (CRF-TOP), the subject was reinforced following every response and punished with an automatic 15 second TO following every error. It might be expected that the conditions should take the same order in terms of number of responses per session as they did for amount of TI per session but this was not the case. Conditions II & III reversed the second and third positions from one of these measures to the other. That there should be less responses when there is more TI does not seem to follow. However, a possible explanation is that one schedule was less adient than the other and occasioned a lower rate of responding.

For Marion, it seemed that the CRF-TONR schedule was less response-evoking than the VI-TOP schedule. Since intermittent reinforcement schedules typically produce higher rates of responding, it seemed that the suppressive effect of TOP on overall responding (since it was designed to suppress incorrect responding) was not enough to override the effects of the positive reinforcement schedule. Similarly, the lower rate of responding occasioned by the CRF schedule was not offset by any increase in overall responding brought about by the TONR condition (which supposedly increases correct responding). Therefore, neither of the TO manipulations appears to have been sufficiently potent to override or noticeably modify the effects of the reinforcement schedule. On the other hand, when these pairings were reversed, that is, when the number of responses under CRF-TOP was compared with number of responses under VI-TONR, it was noted that the TO manipulations did appear to enhance the effects of the associated schedule of reinforcement. TOP appeared to drive the already low CRF response rate further down while TONR appeared to enhance the already high VI response rate.

It was to be expected that the greater amount of TO would be administered under Conditions I and III since the duration of TO was fixed. That the amount was greater under Condition III may have been due to the fact that more TI was available

and thus more opportunities for obtaining TO were presented than under Condition I. Another factor would seem to be involved in the case of the difference in TO per session between Conditions II and IV. While greater TI under Condition IV would account for the higher total TO and the greater number of TOs, this would not explain why the mean duration of an individual TO period was almost twice as high under Condition IV as under Condition II. Since it has been established that subjects will voluntarily impose TOs when they are on certain schedules (inferentially labeled aversive), it is possible that a subject will take a longer TO when a schedule is less reinforcing. This would seem to be the case with Condition IV where reinforcement was on a VI schedule. In other words, Marion seemed to be inclined to respond more slowly to get back into TI under the VI schedule than under the CRF schedule.

The results respecting reinforcement duration and frequency seem to require little explanation. Since Conditions I and II were under the CRF schedule, more reinforcements were available. Condition II was superior to Condition I because TI was greater thus providing more opportunities to obtain reinforcement. A similar explanation would account for the difference between Conditions III and IV.

The ratio of usable to unusable time [TI ÷ (TO + reinforcement)]

provides an index of paradigm efficiency. It will be noted that Condition IV showed the greatest efficiency by far according to this ratio. This is quite understandable since the least time is given over to reinforcement and to TO by the design of the condition. Of course, this fact alone would not be an especially powerful recommendation for a condition. However, when this fact is taken into consideration along with some measures of the effectiveness of the condition, the index does provide salient information. In comparing this index with the results shown in Table 4, it was noted that Condition IV was not only the most efficient use of time but it also produced the most effective performance, ranging from most words learned to first criterion to most generalization tests correct. Similarly for the other conditions: the more efficient the conditions were in terms of time usage, the more effective they were in terms of performance. Incidentally, this finding offers strong support for the contention of Martin et al., (in preparation, c) that it might be better to use the time spent administering TO by providing more trials.

Valerie

Cumulative percent correct responses for Valerie are shown in Figure 6. As was the case with Marion, cumulative correct responses revealed very little in the way of striking differences for Valerie. Such differences would probably be of little

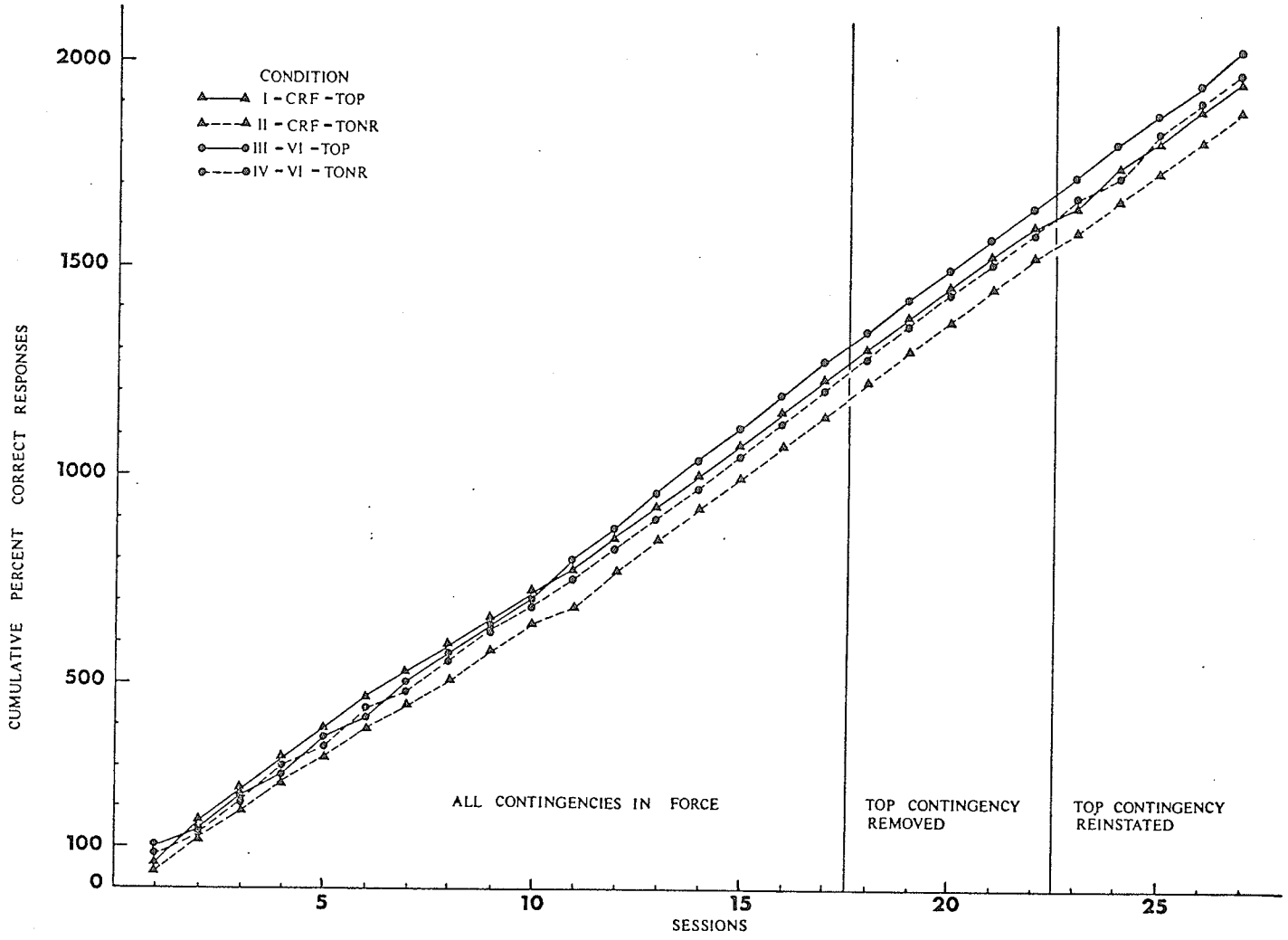

Fig.6. Cumulative percent correct responses by sessions for Valerie

practical value. The changes in position with regard to relative accuracy that were so apparent in the early sessions clearly points up the necessity for waiting until stable baselines are obtained before making any judgements about the relative utility of conditions.

Mean percent correct responding by blocks (noncumulative) is shown in Table 5 and Figure 7. Table 5 indicates that the difference between the highest and lowest condition in terms of accuracy was only slightly more than five percentage points and thus had little practical significance. However, the differences between percent correct responses during the baseline sessions and following the application of the contingencies were marked and consistent across all conditions. This is a clear indication that all manipulations produced a level of accuracy significantly above the base level.

The effect of the probe (in Valerie's case, removal of the TOP contingency from Trials 18 to 22 inclusive) is also seen in Table 5 and Figure 7. It was noted that both Conditions I and III showed a decrease in accuracy following this manipulation and that accuracy improved when the contingency was reinstated. However, while removal of the TOP contingency produced a decrease in accuracy under Condition I, this decrease was rather slight. The effect of this manipulation was much more obvious under Condition III. The fact that the probe brought

TABLE 5

MEAN PERCENT CORRECT RESPONSES BY BLOCKS FOR VALERIE

		BLOCK					
CONDITION	BASELINE	1**	2*	3*	4*	5*	ALL TRIALS
I-CRF-TOP	55.55	75.41	66.55	74.47	71.32	74.27	72.63
II-CRF-TONR	48.33	64.24	65.78	73.52	75.42	72.10	69.77
III-VI-TOP	51.11	71.88	73.65	81.06	73.79	76.84	75.18
IV-VI-TONR	53.33	70.22	68.09	75.76	75.15	76.88	73.00

* 5 trials

** 7 trials

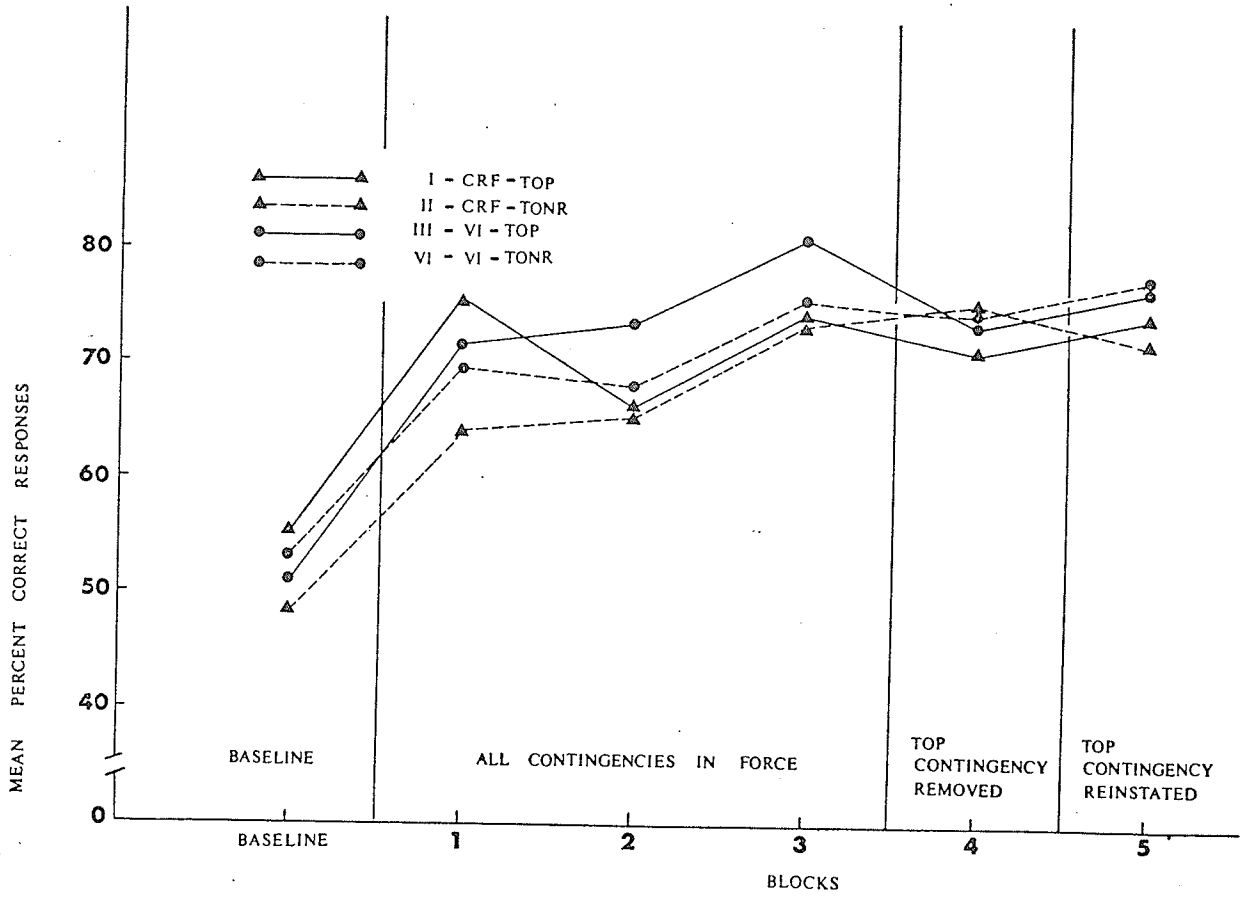

Fig. 7. Mean percent correct responses by blocks for Valerie

about a decrease followed by an increase after reinstatement of TOP under both conditions, coupled with the fact that similar changes were not observed in the case of the non-manipulated conditions, would seem to suggest that the contingency was having some effect in maintaining the level of accuracy.

During the period of the probe, reinforcement density increased from a prior level of 6.06 reinforcements per session to 8.80 reinforcements per session under Condition I and from 1.13 to 2.00 under Condition III. Reinforcement density, therefore, seemed not to be a particularly potent factor in determining accuracy. Although reinforcement increased during the probe, there was no corresponding increase in accuracy. In fact, accuracy decreased as noted above. However, this decrease in accuracy was less under Condition I than under Condition III suggesting a possible modifying effect since the increase in reinforcement density was greater under Condition I.

Rate of correct responding (cumulative correct responses per minute of TI) for Valerie is shown in Figure 8. Mean rates of correct responding by blocks are shown in Figure 9 and Table 6. Condition III was clearly superior to the others in terms of rate of correct responding. While overall rates (and thus, error rates) were higher for Condition III as well, this rate was not high enough to affect the relative position in terms of accuracy. The difference in relative positions of the other

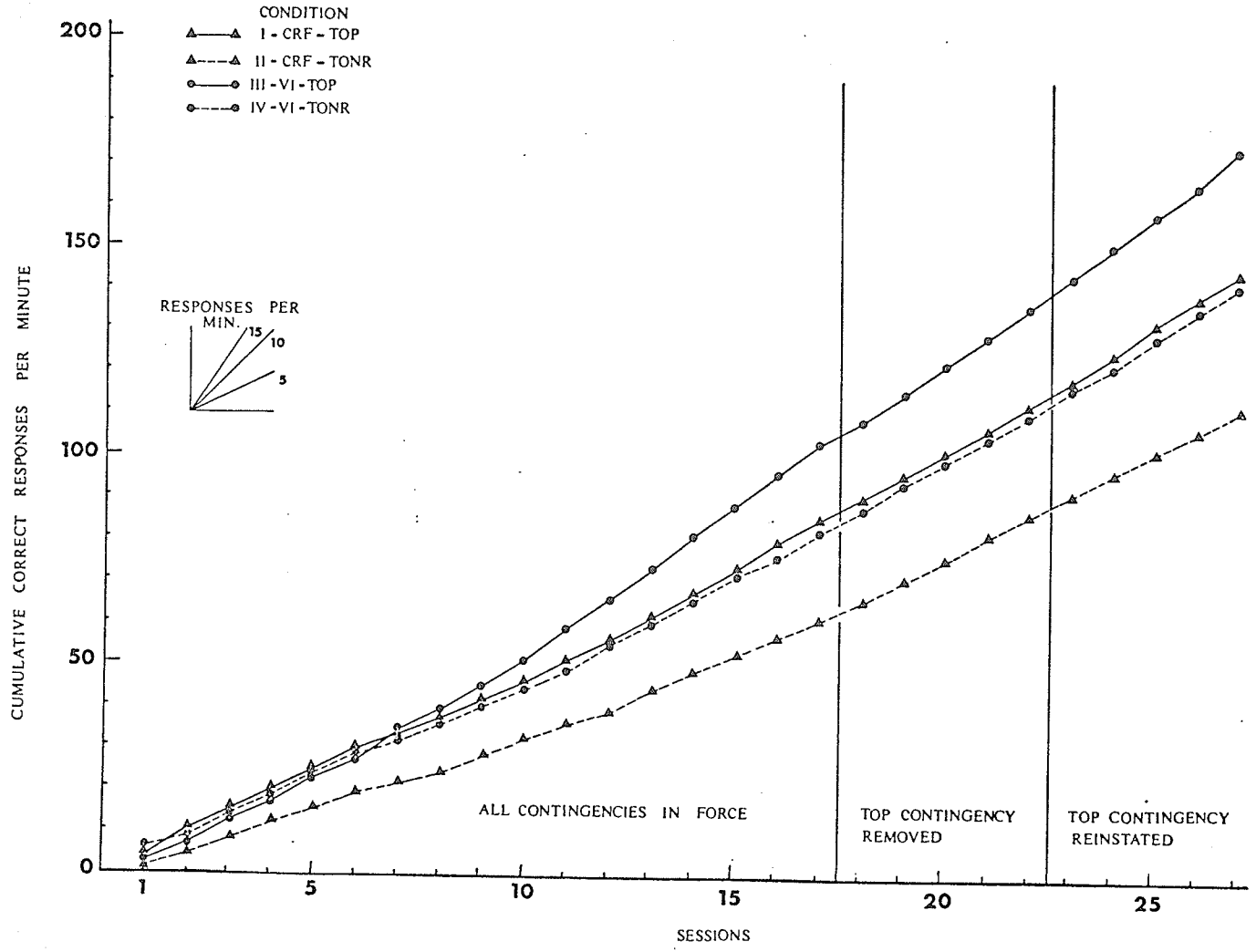

Fig. 8. Cumulative correct responses per minute by sessions for Valerie

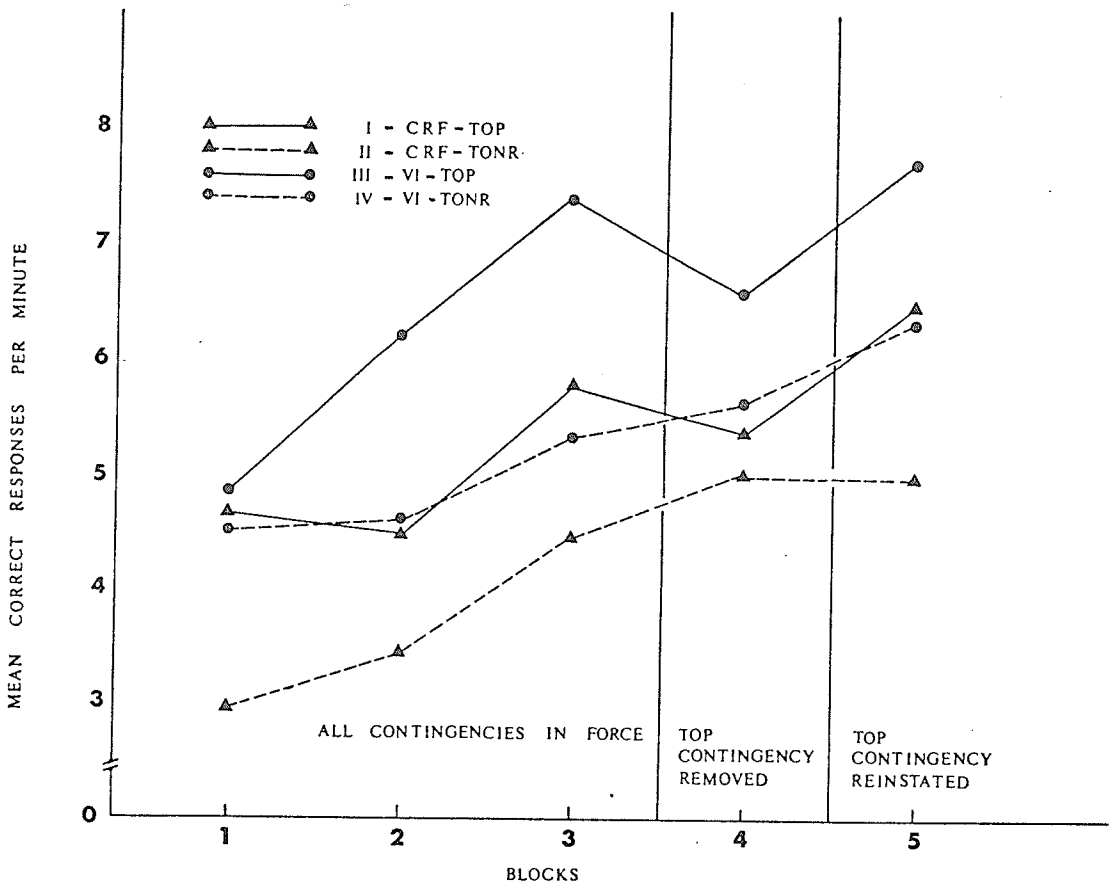

Fig. 9. Mean correct responses per minute by blocks for Valerie

TABLE 6

MEAN CORRECT RESPONSES PER MINUTE BY BLOCKS FOR VALERIE

	BLOCK					
CONDITION	1**	2*	3*	4*	5*	ALL TRIALS
I-CRF-TOP	4.68	4.53	5.80	5.41	6.48	5.33
II-CRF-TONR	2.98	3.46	4.48	5.06	4.96	4.10
III-VI-TOP	4.85	6.23	7.42	6.60	7.73	6.44
IV-VI-TONR	4.52	4.58	5.37	5.64	6.35	5.23

* 5 trials

** 7 trials

conditions on the correct rate and total rate measures might possibly account for the differences in relative positions in terms of correct rate and percent correct. Certainly, high overall rates in relation to correct rates accounts for the smaller differences between conditions on the percent correct responding measures.

The effect of the probe on Valerie's mean rate of correct responses was to bring about a decrease in both conditions affected. It will be noted that no corresponding decrease was seen in the nonmanipulated conditions. The decrease in rate of correct responding brought about by the probe was greater under Condition III than under Condition I. This could have been due to the possibility the CRF condition required very little additional impetus to produce its optimal rate, whereas the VI condition could be significantly enhanced by the addition of the punishment contingency. That the effect of the probe was a reliable one could be concluded from the fact that the probed conditions decreased during the probe and recovered afterwards while no such pattern is seen under the nonprobed conditions. As previously noted, reinforcement density increased somewhat during the probe period for each of the probed conditions.

If reinforcement density had been a potent factor in maintaining the rate of correct responding, one would have

expected this rate to increase during probe sessions under Conditions I and III since reinforcement density increased. Since the rates decreased, one is forced to conclude that either reinforcement density was not a factor or, that it was not powerful enough to override the effects of the TO manipulation. The latter interpretation seems preferable since the actual increases in reinforcement density were not great.

Criterion performance levels for Valerie are shown in Table 7. Relative standings on performance measures were ordered in a fashion that one might have predicted from the data on percent correct responses and rate of correct responses. Indeed, the condition at the highest level on these measures, Condition III, was highest on the first criterion measure. Conditions I and IV, which were virtually tied on the rate and percent measures, were tied on the first criterion measure. Condition II held its terminal position across these measures as well.

One interesting point in these data was that Condition III, which was the most successful condition overall, was also the condition under which most tests were failed. In fact, the number of tests failed under each condition coincided with the mean number of responses per sessions. Those conditions which produced a higher number of responses (and thus the opportunity

TABLE 7

CRITERION PERFORMANCE MEASURES FOR VALERIE

CONDITION	NUMBER OF ANTONYM PAIRS				NUMBER OF INSTANCES		
	LEARNED TO FIRST CRITERION	RETAINED ON FIRST ATTEMPT			SET CRITERION MET	GENERALIZATION TESTS CORRECT	CRITERION TESTS FAILED
		TO FIRST TEST	TO SECOND TEST	TO THIRD TEST			
I-CRF-TOP	10	8	7	7	2	1	9
II-CRF-TONR	4	1	0	0	0	0	11
III-VI-TOP	23	19	17	16	4	1	19
IV-VI-TONR	10	8	6	4	0	0	12

for more errors) also produced more failed tests.

Table 8 shows the mean durations and frequencies of TI, TO and reinforcement as well as the index of paradigm efficiency for Valerie. The fact that Condition III produced the highest number of responses per session is particularly interesting when it is noted that this condition had the second lowest amount of TI. What this means, of course, is that the overall rate of responding per minute of TI was far higher for Condition III than for the other conditions.

That the conditions which contained a VI component (Conditions III and IV) were higher in terms of mean duration of TO and frequency of TO than the conditions containing a CRF component (I and II) seems to be easily explicable. There is simply more response cost involved in failing to respond appropriately under a CRF than under VI schedule. A similar explanation seems appropriate for the difference between Conditions II and IV in terms of individual TO durations.

The reinforcement duration and frequency results for Valerie were identical to those of Marion and a similar explanation is offered.

A comparison of the paradigm efficiency ratios with the measures of paradigm effectiveness in Table 7 revealed an interesting inconsistency. While Condition III was clearly the most effective condition (i.e., highest number of words

TABLE 8

MEAN DURATIONS, FREQUENCIES AND RATIOS OF TI, TO & REINFORCEMENT FOR VALERIE

	MEAN DURATION PER SESSION IN SECONDS			MEAN FREQUENCY PER SESSION			MEAN INDIVIDUAL DURATION		PARADIGM EFFICIENCY
CONDITION	TI	TO	Reinf.	Total Resp.	TOs	Reinf.	TO	Reinf.	TI÷(TO+Reinf.)
I-CRF-TOP	454.87	219.04	210.09	57.89	13.96	6.18	15	34.37	1.06
II-CRF-TONR	613.49	31.50	240.96	59.37	17.69	7.31	2.07	33.09	2.25
III-VI-TOP	561.25	284.70	40.47	86.85	18.86	1.10	15	39.25	1.73
IV-VI-TONR	697.54	115.29	73.71	82.70	21.93	1.89	6.22	38.94	3.69

learned to first criterion), Condition IV was clearly the most efficient in terms of usable to unusable time. In fact, in Valerie's case no coherent relationship was found between these two criteria (unlike Marion's case, in which a symmetrical relationship was found). Thus Valerie's data did not support the contention of Martin et al., (in preparation, c) that the time spent administering TO would be better spent administering more trials.

Agnes

Cumulative percent correct responses for Agnes are shown in Figure 10. The results regarding cumulative percent correct responses for Agnes were similar to those of the other subjects in the sense that the differences between conditions were too slight to be of any particular practical significance.

Mean percent correct responses for Agnes are shown in Figure 11 and Table 9. Again, the improvement between baseline sessions and contingency sessions was obvious and dramatic. It will also be noted that this improvement was much greater under Conditions I and III than for the other conditions, especially Condition II. It is noteworthy that the two conditions associated with the TOP contingency showed the greatest increase from baseline to Block 1 in spite of being very different schedules of reinforcement. All the increases, of course, support the contention that the accuracy of responding was under

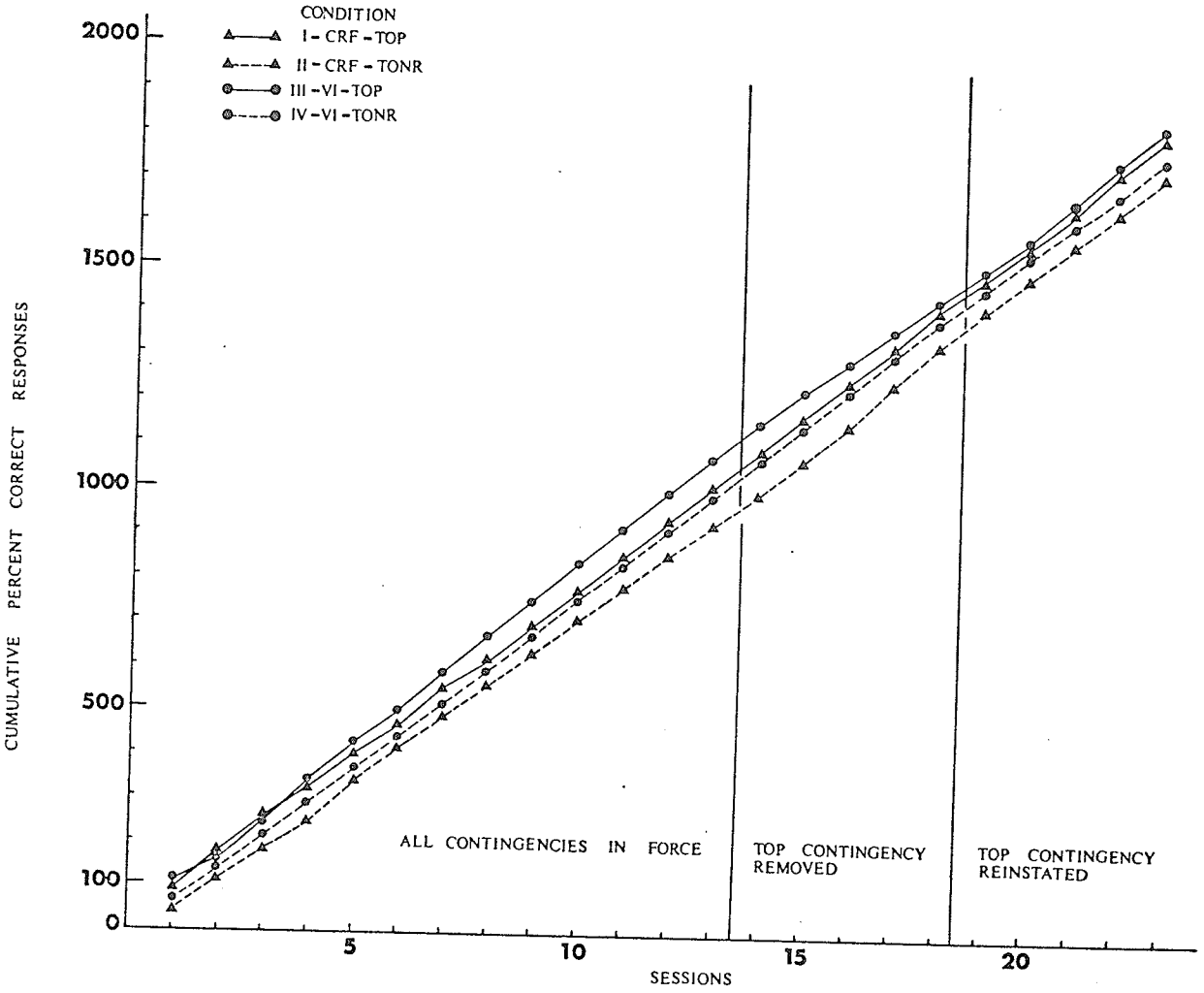

Fig. 10. Cumulative percent correct responses by session for Agnes

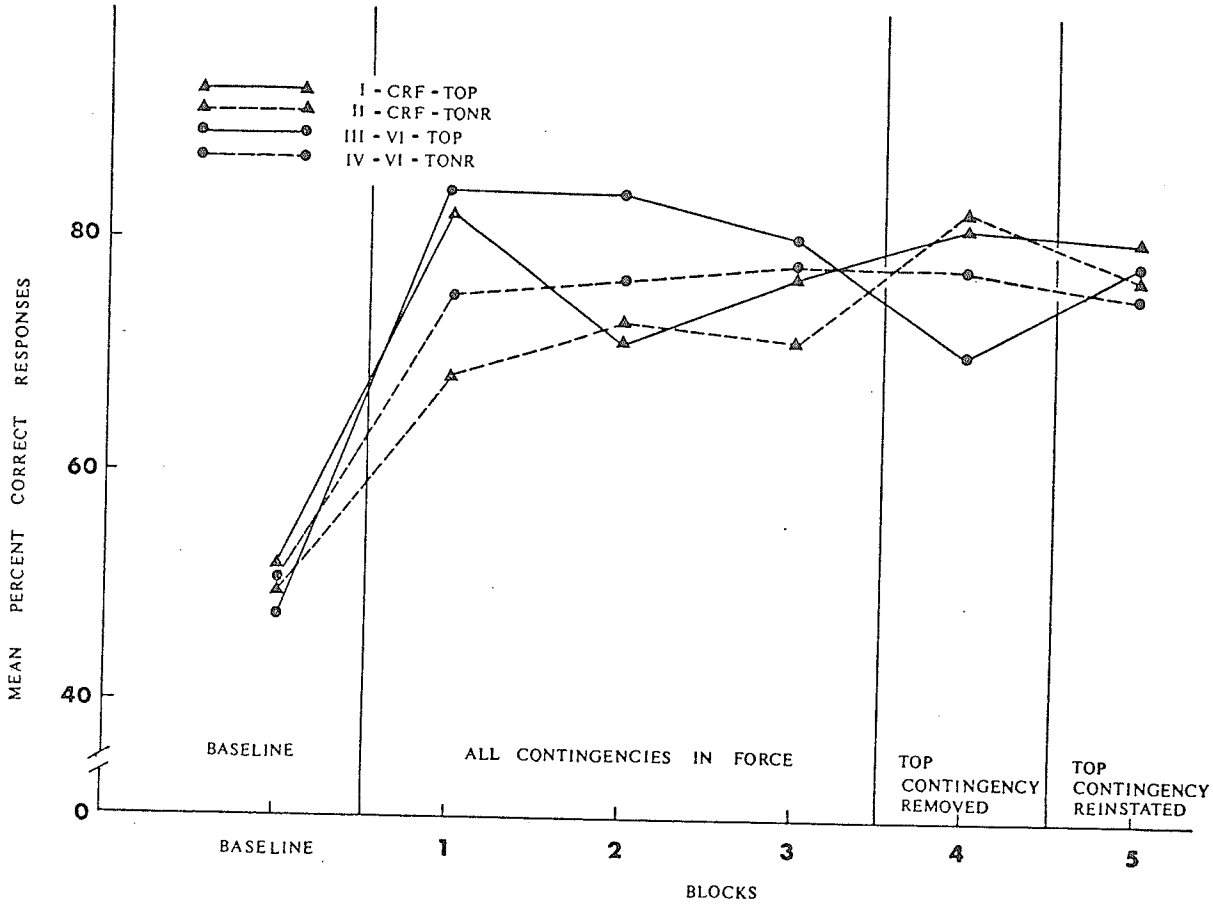

Fig. 11. Mean percent correct responses by blocks for Agnes

TABLE 9

MEAN PERCENT CORRECT RESPONSES BY BLOCKS FOR AGNES

		BLOCK					
CONDITION	BASELINE	1*	2*	3**	4**	5**	ALL TRIALS
I-CRF-TOP	51.67	82.98	71.81	77.22	81.74	80.49	78.98
II-CRF-TONR	50.00	68.40	73.14	71.67	82.03	77.81	74.94
III-VI-TOP	47.77	83.93	84.05	80.54	70.67	77.97	79.04
IV-VI-TONR	51.11	75.33	76.46	77.80	77.90	75.80	76.73

* 4 trials

** 5 trials

the control of the various contingencies.

Table 9 also illustrates that although there were some differences in the relative accuracy of responding, it will be noted that the overall difference between the most accurate (Condition III) and the least accurate condition (Condition II) was only slightly more than four percentage points.

The effect of the probe (removal of the TOP contingency from Trials 14 to 18 inclusive) was not consistent. Condition III showed a decrease in percent correct responding but Condition I appeared to be unaffected. The two conditions not manipulated during the probe showed no decrease. Reinforcement density increased markedly during the probe sessions under Condition I (from 3.92 per session to 8.00 per session) but not as much under Condition III (1.23 to 1.80). That the probe was successful in decreasing rate of responding under Condition III but not under Condition I could possibly be explained in terms of reinforcement density. It will be recalled that under Condition I reinforcement density more than doubled during the probe sessions. This factor may have been sufficiently potent to offset the effect of removal of the TOP contingency.

Cumulative rate of correct responding per minute of TI for Agnes is shown in Figure 12. Table 10 and Figure 13 show mean rates of responding by blocks. The mean rate of correct responding was highest under Condition III, with the superiority

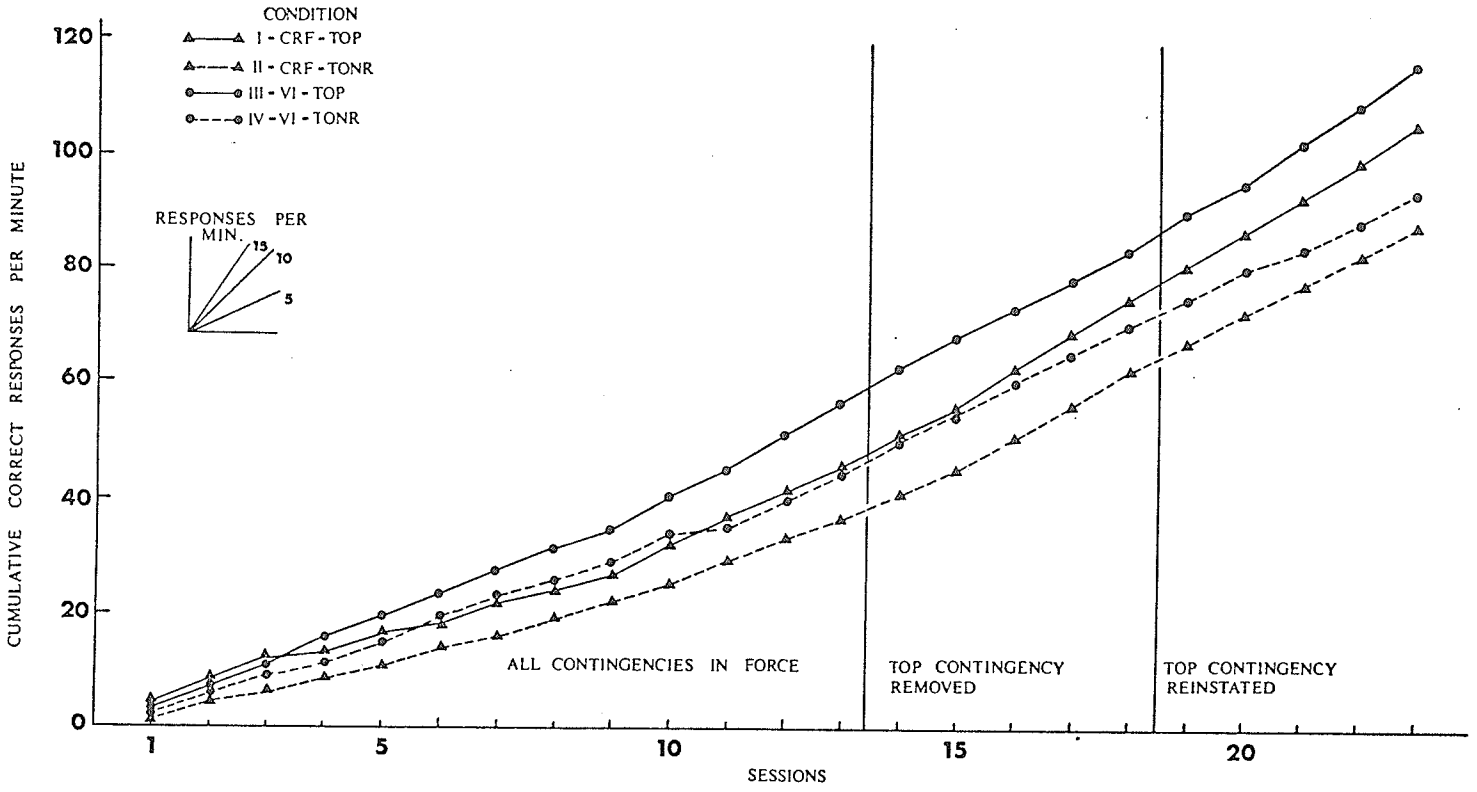

Fig. 12. Cumulative correct responses per minute by sessions for Agnes

TABLE 10

MEAN CORRECT RESPONSES PER MINUTE BY BLOCKS FOR AGNES

	BLOCK					
CONDITION	1*	2*	3**	4**	5**	ALL TRIALS
I-CRF-TOP	3.42	2.81	4.23	5.82	6.30	4.64
II-CRF-TONR	2.24	2.62	3.55	5.08	5.18	3.85
III-VI-TOP	4.00	3.96	5.07	5.24	6.69	5.08
IV-VI-TONR	3.27	3.18	3.93	4.98	4.84	4.11

* 4 trials

** 5 trials

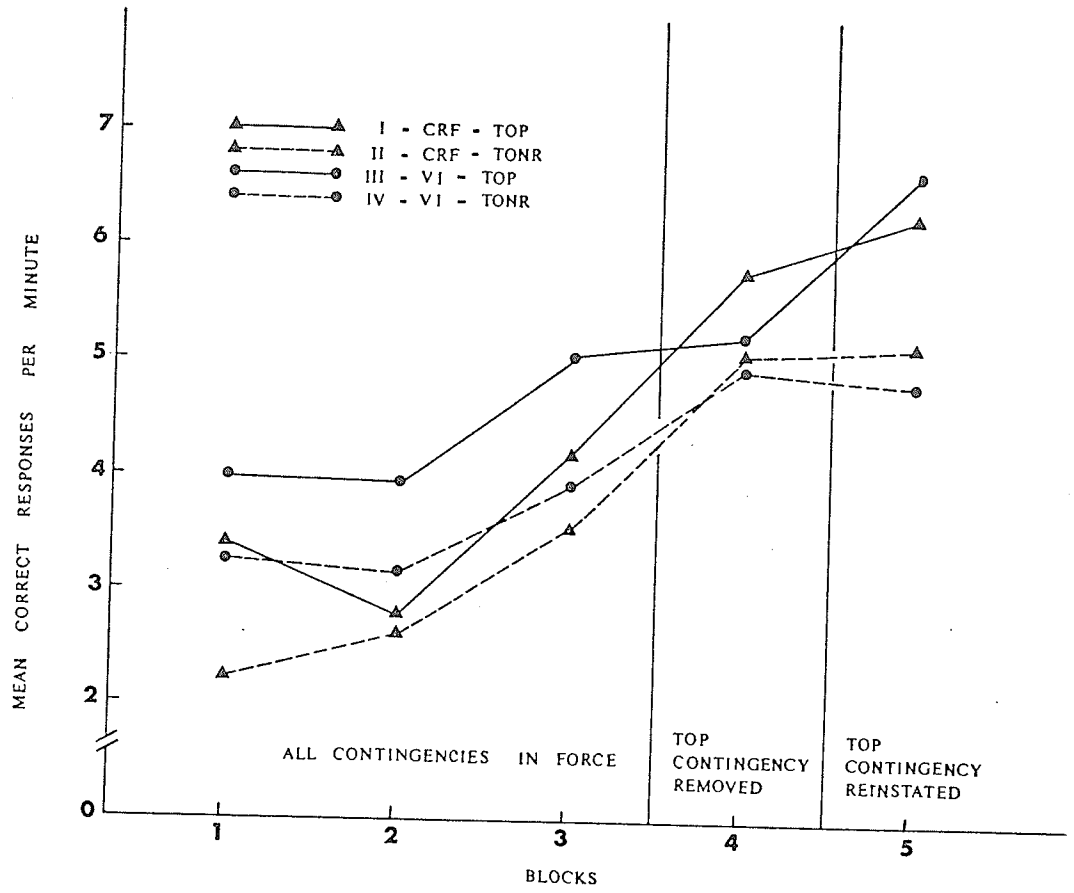

Fig. 13. Mean correct responses per minute by blocks for Agnes

over other conditions ranging from 9% to 32%. The differences between condition in terms of rate of correct responding were greater than the differences between conditions in terms of percent correct responding. Thus, higher rates were not necessarily associated with greater accuracy. A rather marked difference in rate over all conditions will be noted between Blocks 2 and 3. In fact, there is an abrupt change in rate between Sessions 9 and 10 with the mean correct rate during Session 9 equal to 2.92 correct responses per minute and the rate for Session 10 equal to 4.63 correct responses per minute, an increase of 65.41%. This change in rate coincided with a change of experimenters. Overall rate increased by 53.89% between these sessions indicating that error rate decreased, although percent correct responding increased by only 2.84%. The fact that the absolute rate of responding can be affected to such a great extent by the experimenter suggests a need for some caution in the choice of experimenters. If it is essential to make the most efficient use of session time, it might be desirable to "pre-test" experimenters for the rates that they induce in the subjects. Experimenter effects on percent correct responses, or accuracy, were negligible, however.

The fact that rate of correct responding increased during the probe sessions under both Conditions I and III would not have been predicted. Yet, when reinforcement density is taken into

account, a possible explanation for this seemingly aberrant result emerges. The increase in reinforcement density was much higher for Agnes than for the other subjects. Also the difference in reinforcement density between pre-probe and probe sessions under Condition I was much greater than under Condition III, which might account for the much greater increase in rate found under Condition I.

That the difference between Conditions I and III in terms of percent correct responding was so slight while the difference in terms of rate as shown in Table 10 was much greater might be explicable in terms of overall rate. While both correct rate and overall rate were high under Condition III (indicating a high error rate), the correct rate under Condition I was relatively high while the overall rate was relatively low (indicating a low error rate). This would seem to account for the negligible difference in percent correct responses between Conditions I and III.

Criterion performance measures for Agnes are shown in Table 11. These measures did not appear to be related to percent correct or to correct rate in Agnes's case. It will be recalled that the latter two measures were rank ordered Conditions III, I, IV and II while criterion performance was ranked Conditions II, I, III, and IV. The only condition retaining its relative position on all measures was Condition I.

TABLE 11

CRITERION PERFORMANCE MEASURES FOR AGNES

CONDITION	NUMBER OF ANTONYM PAIRS				NUMBER OF INSTANCES		
	LEARNED TO FIRST CRITERION	RETAINED ON FIRST ATTEMPT			SET CRITERION MET	GENERALIZATION TESTS CORRECT	CRITERION TESTS FAILED
		TO FIRST TEST	TO SECOND TEST	TO THIRD TEST			
I-CRF-TOP	12	7	6	5	1	1	13
II-CRF-TONR	13	11	10	10	2	2	6
III-VI-TOP	9	5	4	4	2	0	11
IV-VI-TONR	9	9	8	6	0	0	9

Nor did the number of tests failed appear to be related to overall rate of responding. In fact, the condition with the lowest overall rate (Condition I) produced the highest number of tests failed even though it was second highest in terms of overall criterion performance.

Mean durations and frequencies of TI, TO and reinforcement and the index of paradigm efficiency are shown in Table 12. It is interesting to note that although Condition IV had the greatest amount of TI by a considerable margin it produced only the second highest number of responses per session. It is even more striking that while Conditions II and III had an essentially equal amount of TI, there was a marked difference in overall rate of responding per session. It would seem to be necessary to appeal to some factor other than opportunity to make responses to explain this finding. Since Conditions III and IV were on the same reinforcement schedule, this could not be a factor. The only other alternative is the paradigm of TO but on this basis it seems more plausible to expect higher rates under TONR than under TOP since TOP is designed to suppress one component of the total response rate. To explain the discrepancy between Conditions II and III, it is necessary merely to refer to the finding that intermittent schedules produce higher response rates than CRF. The data on TO frequency and duration are in predictable order and seem to require no further comment. The difference in

TABLE 12

MEAN DURATIONS, FREQUENCIES AND RATIOS OF TI, TO & REINFORCEMENT FOR AGNES

	MEAN DURATION PER SESSIONS IN SECONDS			MEAN FREQUENCY PER SESSION			MEAN INDIVIDUAL DURATION		PARADIGM EFFICIENCY
CONDITION	TI	TO	Reinf.	Total Resp.	TOs	Reinf.	TO	Reinf.	TI÷(TO+Reinf.)
I-CRF-TOP	498.82	140.37	240.82	49.48	9.88	4.72	15	51.20	1.31
II-CRF-TONR	597.94	25.45	266.82	49.61	11.87	5.47	2.43	55.90	2.05
III-VI-TOP	606.59	177.91	82.80	70.78	12.65	1.16	15	70.07	2.33
IV-VI-TONR	723.83	40.37	109.95	65.39	15.13	1.74	3.72	71.73	4.82

individual TO durations was also in the expected direction. The frequency and duration of reinforcement periods was likewise in a predictable order. The fact that mean individual reinforcement durations were greater under the VI conditions (II and IV) than under the CRF conditions (I and II) is explicable in terms of the lower response cost involved in staying out of TI under a VI schedule. The expected order of the paradigm efficiency ratios was not found in Agnes's case, with Conditions II and III being in opposite order. This seems to have been the result of the combination of relatively high reinforcement time under Condition II and relatively low TO duration under Condition III. There seemed to be no coherent relationship between this efficiency index and relative standing on criterion measures, the former ranked Conditions IV, III, II and I and the latter ranked Conditions II, I, III and IV. This is the same sort of inconsistency that appeared in the case of Valerie, that is, no coherent relationship was found to exist between these two criteria. Thus, Agnes's data do not lend support to the contention of Martin et al., (in preparation, c) that the time spent administering TO would be more profitably spent administering more trials.

Comparisons Across Subjects

Figures 2, 6, and 10 depict the cumulative percent correct

responses for Marion, Valerie and Agnes respectively. It will be noted here that there was some inconsistency across subjects on this measure. For Marion, Condition II produced the highest percent correct responding, while for Valerie and Agnes, Condition III was superior. The positions of the intermediate conditions are even more inconsistent. For Valerie and Agnes, the poorest performance was obtained under Condition II which was the best condition for Marion on this measure.

Figures 3, 7, and 11 show mean percent correct responses by blocks. The major point of consistency here was the dramatic change in percent correct responses from baseline period to contingency sessions. The only other thing that occurred with even fair consistency was the decrease in correct responding in those conditions affected by the probe. This decrease appeared under one condition for Marion, under both for Valerie and one for Agnes. All conditions that showed the decrease during the probe also showed an increase in percent correct when the probe was eliminated, adding support to the view that the TO contingencies exerted a fairly reliable effect.

Figures 4, 8 and 12 illustrate an important consistency in terms of rate of correct responses. Here it will be noted that all three subjects exhibited the highest rate of correct responding under Condition III. Valerie and Agnes's data are consistent in terms of the ordering of the other conditions on

this measure.

Figures 5, 9 and 13 show mean correct responses per minute by blocks. Perhaps the first point of consistency that will be noted here is the general upward trend of the various rates. Again, the decrease in rate of correct responding appeared with fair consistency across the affected conditions. It was seen in one condition for Marion, in both for Valerie and in one for Agnes. Once more, those conditions that did show a decrease in rate during the probe increased in the subsequent sessions, lending support to the assertion of the reliability of the TO manipulations.

Since the entries in Tables 1, 2, 5, 6, 9 and 10 are simply the precise figures associated with the graphs cited above, it would be redundant to make comparisons on these.

Some of the consistencies across subjects shown in Tables 3, 8 and 12 are of minimal practical value since they were forced by composition of the condition. However, it is interesting to observe that a noticeably greater amount of time was spent in TO under Condition IV than under Condition II and that this was consistent over all three subjects. This has been explained previously in terms of the lower response cost involved in staying in TO when under a VI schedule than under a CRF schedule. One puzzling inconsistency involved mean individual reinforcement time. Here it will be seen that for Marion and Valerie these

times were fairly even over all conditions. However, in Agnes's case, much more time was spent consuming the reinforcer under the VI associated conditions than under the CRF associated conditions. This might also be explained in terms of the lower response cost involved in taking more time in reinforcement, although it is curious that it would not similarly affect the other subjects. The paradigm efficiency index produced a rank order of Conditions IV, II, III, I for Marion and Valerie, while the order for Agnes was IV, III, II, I. The extra time spent in reinforcement under Condition II was advanced as a possible reason for this interchange in Agnes's case.

Tables 4, 7 and 11 show the criterion performance measures. These tables illustrate the utter lack of uniformity across subjects in terms of such measures which is very difficult to account for.

General Discussion

TO seemed to be effective in contributing to higher rates of correct responding and higher percent correct responding as demonstrated by the effects of the experimental probe and the difference between baseline and contingency sessions. In a way, this supports Willoughby's (1969) conclusion that TO is effective when an alternative unpunished response is available. This statement is only tentative since there was no control condition to check the reliability of the phenomenon. Willoughby's

suggestion, in fact, was used simply in the hopes of enhancing the effects of the variables under study and it was not itself being tested. Similarly, the finding of Martin (1970), regarding the use of clear discriminative stimuli was employed although the validity of the suggestion was not tested here.

The results obtained in the matching-to-sample studies of Ferster and Appel (1961) were not supported by the present data. While these investigators found TO to be ineffective when correct responses were on a CRF schedule, the current study suggested that this was not always the case. The data also offer partial support for Zimmerman and Ferster's (1963) finding that optimal matching accuracy is obtained when every incorrect response is followed by TO. This was clearly demonstrated by the effects of the probe in the present study.

While Holz, Azrin and Ayllon (1963) and Baron, Kaufman and Rakauskas (1967) found TO to be ineffective under certain circumstances, this study appeared to suggest that TO was a fairly consistently effective variable in the conditions studied.

The present data lend support to McReynolds's (1969) finding that making TI contingent upon the emission of appropriate behavior increases the desired responses. This was true in at least some cases in the present study where the TONR contingency was more effective than the TOP contingency.

Martin and Berndt's (1970) results were also supported by

the present data. Clearly, TO reduces the frequency of behavior upon which it is contingent.

The series of studies performed by Martin et al., (in preparation; a, b, & c) suggested that TO was rather ineffective and concluded that the time spent administering TO might better be spent administering more trials. This notion was supported in the present study in one case where it appeared that simply having more trials produced better performance. However, the idea was not supported in the case of the other two subjects where the mere fact of having more trials was ineffective in improving performance.

Some attempt should be made to explain the inconsistencies in the effects of the various conditions across subjects. The reason for these inconsistencies is not directly known but there are some possibilities. First of all, there could be, and probably are, great differences in the behavioral histories of the subjects. Secondly, the genetic endowment most certainly differed. It will be recalled that one subject was hydrocephalic, one was mongoloid and one was epileptic. The nervous systems of the subjects have all apparently undergone damage the extent of which is unknown. All of these factors then, may have contributed to the observed differences. An interesting follow-up to this research might involve some sort of comparative study to determine whether the effectiveness of experimental manipulations is related in any way to clinical diagnosis.

CHAPTER IV

SUMMARY AND CONCLUSIONS

Many consistencies across subjects were observed in this study as well as some rather important inconsistencies. To take the consistencies first, it was clear from both the data on percent correct responding and the data on rate of correct responding that all experimental conditions had the effect of reliably enhancing both speed and accuracy of the antonym learning task. There is no equivocation in the data on this point. The effect of the experimental probe appeared with fairly high reliability. That is, percent correct responding and rate of correct responding usually decreased when the TO contingency was removed. Those occasions when responding did not decrease, it was found, were subject to an increase in reinforcement density as a result of removal of the TO contingency. Of the conditions that showed a decrease in responding during the probe, all showed an increase when the probe was terminated. There was also a general upward trend in rate of correct responding over sessions suggesting that the contingencies became more effective the longer they were in force. Time spent in TO under the TONR conditions was consistently greater under the VI schedule of reinforcement than under the CRF schedule. This would seem to suggest that the subjects discriminated the relative response

cost involved in remaining in TO under the two conditions. The index of paradigm efficiency was only fairly consistent with one subject showing a different pattern.

Turning now to the inconsistencies, the most efficient paradigm in terms of use of time was the most effective in terms of criterion performance in only one case out of three. Likewise, accuracy of responding and effectiveness of criterion performance were related in only one case as were accuracy of responding and speed of responding. Since the number of possible combinations of the factors of efficiency, effectiveness, accuracy and speed over three subjects is 18, to have three of these combinations turn up by chance alone would not be unusual. That there was only fair consistency across subjects regarding the conditions producing the most accurate responding is troublesome. However, this fact becomes less relevant when it is recalled that no consistent relationship was found between this measure and criterion performance. That is, when intra-session accuracy does not lead to better criterion performance anyway, it does not seem reasonable to lament the fact that this accuracy was not consistent across subjects. The lack of relationship between these two measures is puzzling nonetheless. There was no consistency between subjects on criterion performance measures either.

Several things might be concluded from the results of the

present study, some of these conclusions being rather more tentative than others.

First, it would seem to be clear that the sort of contingencies applied in this study are generally effective in increasing correct responding significantly above baseline levels.

Second, it would seem that the TO contingency is, in most cases, effective in reducing errors, except where the effect is overridden by the effects of reinforcement density.

Third, the evidence seems to support the conclusion that the highest rate of correct responding can be obtained with reinforcement on VI and TO employed as a punisher for errors. However, the superiority of this condition does not seem to be consistently related to superiority in terms of accuracy or criterion performance.

Fourth, it seems to be the case that the rate of correct responding continues to improve over sessions under all conditions, suggesting that perserverence in maintaining contingencies may be almost as important as the composition of the contingency itself.

Fifth, the fact that time spent in TO was consistently greater under VI than under CRF suggests that the organism is capable of discriminating differences in the response cost values of various behaviors, and that it behaves accordingly.

Sixth, the efficiency of the paradigm in terms of use of time is not related in any consistent way to accuracy of responding or criterion performance.

Seventh, it seems that accuracy of responding in terms of percent correct is not related to the conditions in any consistent fashion. However, the differences obtained between conditions on this measure were too small to be of practical significance at any rate.

Eighth, criterion performance results were not related in any coherent fashion to the conditions, nor was this measure consistently related to any other measure.

In summary, then, it would appear that the research into the use of TO in the training of the retarded has revealed many puzzling and seemingly contradictory facts. But the research up to the present time has barely scratched the surface in this area. Many more carefully-designed studies need to be undertaken before the variables determining the effectiveness of this tool can be established with any degree of confidence.

REFERENCES

Adams, M.R. and Popelka, G. The influence of "time-out" on stutterers and their dysfluency. Behavior Therapy, 1971, 2(3), 334-339.

Appel, J.B. Aversive aspects of a schedule of positive reinforcement. Journal of the Experimental Analysis of Behavior, 1963, 6, 423-428.

Azrin, N.H. Time out from positive reinforcement. Science, 1961, 133, 382-383.

Azrin, N.H. and Foxx, R.M. A rapid method of toilet training the institutionalized retarded. Journal of Applied Behavior Analysis, 1971, 4, 89-99.

Azrin, N.H. and Holz, W.C. Punishment. In W.K. Honig (Ed.) Operant behavior: Areas of research and application. New York: Appleton Century-Crofts, 1966, pp. 380-447.

Baer, D.M. Laboratory control of thumbsucking by withdrawal and representation of reinforcement. Journal of the Experimental Analysis of Behavior, 1962, 5, 525-528.

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.

Baron, A., Kaufman, A., and Rakauskas, I. Ineffectiveness of "time out" punishment in suppressing human operant behavior. Psychonomic Science, 8, 329-330.

Barton, E.S., Guess, D., Garcia, E., and Baer, D.M. Improvement of retardates' mealtime behaviors by timeout procedures using multiple baseline techniques. Journal of Applied Behavior Analysis, 1970, 3, 77-84.

Bijou, S.W., Peterson, R.F., Harris, F.R., Allen, K.E., and Johnston, M.S. Methodology for experimental studies of young children in natural settings. Psychological Record, 1969, 19, 177-210.

Bostow, D.E., and Bailey, J.B. Modification of severe disruptive and aggressive behavior using brief timeout and reinforcement procedures. Journal of Applied Behavior Analysis, 1969, 2, 31-37.

Carlson, J.G., and Aroksaar, R.E. Effects of time-out upon concurrent operant responding. The Psychological Record, 1970, 20, 365-371.

Catania, A.C. (Ed.) Contemporary research in operant behavior. Glenview, Illinois: Scott-Foresman, 1968.

Ferster, C.B. Withdrawal of positive reinforcement as punishment. Science, 1957, 126, 509.

Ferster, C.B. Intermittent reinforcement of a complex response in a chimpanzee. Journal of the Experimental Analysis of Behavior, 1958, 1, 163-165.

Ferster, C.B., and Appel, J.B. Punishment of S responding in matching to sample by time-out from positive reinforcement. Journal of the Experimental Analysis of Behavior, 1961, 4, 45-56.

Ferster, C.B., and Perrott, M.C. Behavior principles. New York: Appleton-Century-Crofts, 1968.

Ferster, C.B., and Skinner, B.F. Schedules of reinforcement. New York: Appleton-Century-Crofts, 1957.

Franks, C.M. (Ed.) Behavior therapy: Appraisal and status. New York: McGraw-Hill, 1969.

Gardner, J.M. Behavior modification research in mental retardation: Search for an adequate paradigm. American Journal of Mental Deficiency, 1969, 73(3), 844-851.

Gardner, W.I. Use of punishment procedures with the severely retarded: A review. American Journal of Mental Deficiency, 1969, 74, 86-103.

Gelfand, D.M., and Hartmann, D.P. Behavior therapy with children: A review and evaluation of research methodology. Psychological Bulletin, 1968, 69(3), 204-215.

Giles, D.K., and Wolf, M.M. Toilet training institutionalized, severe retardates: An application of operant behavior modification procedures. American Journal of Mental Deficiency, 1966, 70, 765-780.

Hamilton, J., and Allen, P. Ward programming for severely retarded institutionalized residents. Mental Retardation, 1967, 5, 22-24.

Hamilton, J., and Stephens, L. Reinstating speech in an emotionally disturbed mentally retarded young woman. Unpublished paper, Gracewood State School, Georgia, 1967. Cited by W.I. Gardner in Use of punishment procedures with the severely retarded: A review. American Journal of Mental Deficiency, 1969, 74, 86-103.

Holz, W.C., Azrin, N.H., and Ayllon, T. Elimination of behavior of mental patients by response-produced extinction. Journal of the Experimental Analysis of Behavior, 1963, 6, 407-412.

Honig, W.K. Operant behavior: Areas of research and application. New York: Appleton-Century-Crofts, 1966.

Leitenberg, H. Is time-out from positive reinforcement an aversive event? A review. Psychological Bulletin, 1965, 64, 428-441.

Lovitt, T. Behavior modification: Where do we go from here? Exceptional Children, 1970, 37, 157-167.

Martin, G.L., Hardy, L.M., and McDonald, S. A comparison of several time-out intervals as consequences for errors during picture-naming training with retardates. In preparation, (a).

Martin, G.L., McDonald, S., Michalishyn, M., and Hardy, L. A comparison of several time-out intervals as consequences for errors during tracing training with autistic and retarded children. In preparation, (b).

Martin, G.L., Moir, J., and Skinner, M. A comparison of two time-out intervals as consequences for errors during a learning task with autistic and retarded children. In preparation, (c).

Martin, J.A. The discriminative properties of timeout from positive reinforcement. Paper presented at the meeting of the Eastern Psychological Association, Atlantic City, New Jersey, April, 1970.

Martin, R., and Berndt, L.A. The effects of time-out on stuttering in a 12 year old boy. Exceptional Children, 1970, 37, 303-304.

Martin, R.R. and Haroldson, S.K. Time-out as a punishment for stuttering during conversation. Journal of Communication Disorders, 1971, 4(1), 15-19.

McReynolds, L.V. Application of timeout from positive reinforcement for increasing the efficiency of speech training. Journal of Applied Behavior Analysis, 1969, 2, 199-205.

Millenson, J.R. Principles of behavioral analysis. New York: MacMillan, 1967.

Morse, W.H. Intermittent reinforcement. In W.K. Honig (Ed.), Operant behavior: Areas of research and application. New York: Appleton-Century, 1966, Pp. 52-108.

Nordquist, V.M. The modification of a child's enuresis: some response-response relationships. Journal of Applied Behavior Analysis, 1971, 4, 241-247.

Pawlicki, R. Behavior-therapy research with children: A critical review. Canadian Journal of Behavioral Science, 1970, 2, 163-173.

Pendergrass, V.E. Timeout from positive reinforcement following persistent, high-rate behavior in retardates. Journal of Applied Behavior Analysis, 1972, 5, 85-91.

Peterson, R.F., and Peterson, L.R. Mark and his blanket: A study of self-destructive behavior in a retarded boy. Paper presented at a meeting of Society for Research in Child Development, New York, 1967. Cited by W.I. Gardner in Use of punishment procedures with the severely retarded: A review. American Journal of Mental Deficiency, 1969, 74, 86-103.

Reese, E.P. The analysis of human operant behavior. Dubuque, Iowa: Wm.C. Brown, 1966.

Reynolds, G.S. A primer of operant conditioning. Glenview, Illinois: Scott, Foresman, 1968.

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

Roos, P., and Oliver, M. Evaluation of operant conditioning with institutionalized retarded children. American Journal of Mental Deficiency, 1969, 74(3), 325-330.

Schaefer, H., and Martin, P. Behavioral therapy. New York: McGraw-Hill, 1969.

Sidman, M. Tactics of scientific research. New York: Basic Books, 1960.

Skinner, B.F. The behavior of organisms: An experimental analysis. New York: Appleton-Century-Crofts, 1938.

Sloane, H.N., and MacAulay, B.D. (Eds.) Operant procedures in remedial speech and language training. New York: Houghton Mifflin, 1968.

Smolev, S.R. Use of operant techniques for the modification of self-injurious behavior. American Journal of Mental Deficiency, 1971, 76(3), 295-305.

Steeves, J.M., Martin, G.L., and Pear, J.J. Self-imposed time-out by autistic children during an operant training program. Behavior Therapy, 1970, 1, 371-381.

Tate, B.G., and Baroff, G.S. Aversive control of self-injurious behavior in a psychotic boy. Behavior Research and Therapy, 1966, 4, 281-287.

Thomas, J.R. Avoidance of time-out from two VI schedules of positive reinforcement. Journal of the Experimental Analysis of Behavior, 1964, 7, 168.

Ullmann, L.P., and Krasner, L.(Eds.) Case studies in behavior modification. New York: Holt, Rinehart and Winston, 1965.

Ulrich, R., Stachnik, T., and Mabry, J. (Eds.) Control of human behavior. Chicago: Scott, Foresman, 1966.

Ulrich, R., Stachnik, T., and Mabry, J. (Eds.) Control of human behavior. Vol. II: From cure to prevention. Chicago: Scott, Foresman, 1970.

Verhave, T. The experimental analysis of behavior. New York: Appleton-Century-Crofts, 1966.

Wahler, R.G. Oppositional children: A quest for parental reinforcement control. Journal of Applied Behavior Analysis, 1969, 2, 159-170.

Wasik, B.H., Senn, K., Welch, R.H., and Cooper, B.R. Behavior modification with culturally deprived school children: Two case studies. Journal of Applied Behavior Analysis, 1969, 2, 181-194.

Willoughby, R.H. The effects of time-out from positive reinforcement on the operant behavior of preschool children. Journal of Experimental Child Psychology, 1969, 7, 299-313.

Willoughby, R.H. The influence of different response consequences on children's preference for time-out. Journal of Experimental Child Psychology, 1970, 9, 133-141.

Zimmerman, J., and Ferster, C.B. Intermittent punishment of S responding in matching to sample. Journal of the Experimental Analysis of Behavior, 1963, 6, 349-356.

FOOTNOTES

1. Blocks were used in order to illustrate more clearly the changes that took place at various stages of the study. The blocks before, during, and after the probe were of equal size. The size of the other two blocks was determined by the observed transition points in the data. This applies to all three of the subjects.

2. In Tables 3 & 4, 7, 8, 11 & 12 certain sessions (where overall time deviated from 900 seconds by plus or minus 10%) were omitted from calculation of the means where this might have materially and spuriously affected the result. Those sessions affected by the probe were also omitted.