ENGINEERING THE BEHAVIORAL ENGINEER¹

ROLF LOEBER

KINGSTON PSYCHIATRIC HOSPITAL, KINGSTON, ONTARIO

Twenty-eight nursing staff members treated, by operant methods, a simulated headbanging patient. They were randomly assigned to four experimental conditions, involving improvement *versus* non-improvement of the "patient" and promise of reward *versus* no promise of reward. The promise of reward significantly improved accuracy of treatment. Improvement of the patient, however, did not improve accuracy.

A few years ago there was a widely circulated joke among psychologists about the experimental rat in the Skinner box bragging of his success in conditioning the experimenter. Before this came the research by Rosenthal and Fode (1961) on experimenter bias in students experimenting with rats. These things both have in common a change of focus from the subject to the experimenter as the target of interest.

The rat's remarks point out a generally neglected area of study. In what ways are experimenters, who apply conditioning principles to their subjects, themselves susceptible to those principles? What sort of reinforcements and schedules are most influential for their performance as conditioners? Very little research seems to have been done to answer these questions. Schaeffer and Martin (1966) illustrated how easily an observer may be extinguished by a non-speaking patient. Panyan, Boozer, and Morris (1970) studied the reinforcing value of the number of operant training sessions on non-professional personnel, although they did not consider the negative or positive results of the sessions. Furthermore, a very interesting, but rather poorly controlled study was done by Bricker, Morgan, and Grabowski (1968) in which attendants were rewarded with tokens.

The problem of producing efficient behavioral engineers is of particular relevance for token economies. The ill fate of many of these projects may be partly explicable by the changing reinforcers experienced by the staff. Initially, such positive reinforcers as novelty, prestige, *etc.* may be predominant, while later in the program such things as boredom with

routine, failure of patients to improve, *etc.* become more prominent.

The present study examined in a limited way some aspects of operant control over the behavioral engineer.

Procedure

Nurses and attendants on the day or evening shift of four psychiatric wards were individually asked to volunteer for a project "to find out better ways to cure patients with specific problems in a shorter time". In addition to the verbal information, a supporting letter was also sent to them, in which it was stated that a boy was to be treated in order to reduce his persistent headbanging and that their cooperation would be appreciated.

The experiment itself was conducted in the psychological laboratory at the hospital, located in a different building. Upon entering the corridor in the laboratory, and passing the room from which the "patient's" headbanging was sounding, the following was said to the subject:

"In this room is the boy with the headbanging behavior which you will attempt to reduce. Since he is afraid of strangers you will not be able to see him because this would upset him even more. You will work from this other room. He is pro-

¹I am grateful for the help of Dr. A. McGhie and R. Van Allen, without whom this study would have never materialized. Furthermore, I acknowledge the kind assistance of the Nursing Staff, W. Leckett, and Mrs. M. Loeber. Reprints may be obtained from the author, Kingston Psychiatric Hospital, Kingston, Ontario, Canada.

tected from hurting himself by wearing a helmet”.

In fact, this headbanging was produced by playing a tape loop before the experimental tape, so that there was no artificial break from the time the subject entered until he left the laboratory. In the meantime, the subject was led into the experimental room, where he was asked to sit behind a table. Then the following instructions were given.

“We are trying to find out better ways to cure patients, in a shorter time. You can assist us in this by participating in treating this boy. You must realize that this is not in any way a job evaluation. This is just an opportunity to let you practice a promising technique. Nobody will be told about your level of performance.

The boy you are going to treat is an out-patient, ten years old, with a long history of headbanging. In this respect he is a considerable problem for his parents. The aim of the procedure is for you to reduce the headbanging of the boy through use of these buttons which administer or withdraw tokens. It has been demonstrated that his headbanging can be reduced temporarily, but not yet permanently, depending for a great part on the skill of the therapist. The boy has learned that he can exchange the tokens later for his favourite candies or toys. He will be able to see how many tokens he receives from your pressing the button A. In the same way he can see the tokens being withdrawn by your pressing button B. The tokens will be visible to him in an enclosed dispensing container so he cannot obtain them until after the training procedure.

Button A, when pressed twice, releases two tokens in the container and Button B, when pressed once, removes one token from this container.

Your task will be to press button A *twice* in succession *immediately* when the boy's headbanging ceases. This will reward him for stopping his headbanging. In addition, you are to press button B *once immediately* whenever the boy's headbanging begins. This deprives him of a token because he has started headbanging again.

It is of utmost importance to apply the button in the way we told you or else you might increase the headbanging instead of decreasing it. You may refer to this instruction sheet beside you to help remind you of the procedure”.

At this point the instructions were interrupted by the second experimenter entering the room, asking whether he would go and see if the boy was alright. The second experimenter then left open the door of the experimenter room and went to the boy's room, addressing the latter while opening his door, all this designed to be overheard by the subject. Later the second experimenter returned to report that the “boy” was okay and couldn't hurt himself.

The next step was to have the subject practise at the buttons, while the following instruction was given:

“In order for you to become acquainted with the use of these buttons, let us listen to the headbanging. Press button A two times when he stops for awhile” (practice). “Now press button B, but only when the headbanging begins again. Press button B only one time” (practice). “Now apply button A two times whenever he stops headbanging and apply button B whenever he starts again” (practice). “Do you have any questions about performing this task or what is required?”

Half of the subjects were told that they could earn a reward:

“Depending upon your performance in being able to administer tokens with as few mistakes as possible, you may receive an amount of money in the sum of \$4.00. It is very important that you follow the instructions, and if your performance is accurate you will obtain the \$4.00 later.”

Finally, all the subjects were told:

“Now that you are familiar with what is required, I will proceed to the adjoining room where I must turn on the apparatus which controls the token dispenser. You will work here alone, until I finally return to tell you to stop. This will be in approxi-

mately 20 minutes. Is everything clear to you? You can start now."

After this the subjects were left alone, and the experimental tape was started. The subjects could check the procedures on a paper near the buttons, which again emphasized the necessity immediately to react with the presses on the buttons at the beginning and end of each of the volleys of headbanging.

At the completion of the experimental tape, one of the experimenters would instruct the subject with the questionnaire, while the other experimenter would play a similar game for the subject of checking if the "patient" was alright. Finally, the confidentiality of the experiment was emphasized to the subject.

The task schedule of two presses to present and one press to withdraw a token, was chosen, after a pilot experiment, as being at an appropriate difficulty level for unexperienced operant conditioners. One of the reasons for choosing this task schedule was that under all circumstances tokens could be withdrawn.²

The effects of two independent variables in the subjects' accuracy were investigated: monetary reward and feedback from the "patient". Half the subjects were promised \$4.00 if they performed sufficiently accurately; the other half were told nothing about money. Assignment to the two reinforcement conditions was random.³ Half the subjects in each condition were exposed to an experimental headbanging tape that presented volleys of bangs of constant average frequency throughout (the "non-improvement" condition). The other half of the subjects in each condition were exposed to an experimental headbanging tape that presented volleys of bangs of gradually decreasing frequency (the "improvement" condition). Assignment to the improvement-non-improvement conditions was also random. In the "non-improvement" condition, volleys consisted of three to seven bangs separated by 1 to 2-sec intervals, with intervals between volleys varying between 4 to 7 sec. In the "improvement" condition, the intervals between volleys gradually increased from 4 to 76 sec and the number

of headbangings per volley decreased from seven to one.

Errors in performance were defined in advance as: (a) pressing Button A or B more or fewer times than instructed; (b) applying Button A or B too early; (c) pressing Button A more than 3 sec after the last headbang of a volley; (d) pressing Button B during or after the second headbang of a volley. Three channels of a polygraph were used to record separately the responses on Button A, the responses on Button B, and the individual headbangs. This permitted objective scoring of errors.

An open-ended questionnaire was developed (see Figure 1) to assess the subjects' suspiciousness of the experimental procedure. This questionnaire was administered at the end of the session. "Suspiciousness" was defined in advance as: the inference that subject suspects or is sure that (a) there is no real patient, or (b) that the headbanging is simulated.

RESULTS

The questionnaires of two subjects (both male) showed suspiciousness as noted by two independent judges. None of the other subjects showed any suspicion. Two further subjects did not keep the appointment, so that

1. Do you think you were successful in applying the therapy: Yes___ No___
2. Could you detect any improvement in the patient's behavior: Yes___ No___
If yes: What do you think is the reason:
If No: What do you think is the reason:
3. Did you feel at ease in using this type of therapy: Yes___ No___
If Yes: Explain:
If No: Explain:
4. Do you think such therapy might be applicable for other purposes: Yes___ No___
If Yes: Explain:
If No: Explain:
5. Would you mind doing the same for another patient at another time: Yes___ No___
If Yes: Why:
If No: Why:
6. Do you think that the fact that you could not see the patient interfered with your therapy: Yes___ No___
If Yes: Why:
If No: Why:
7. Do you have any comments?

Fig. 1. Suspicion Questionnaire

²It is perhaps worth noting that this reinforcement schedule does not necessarily represent an efficacious one from the "patient's" point of view.

³However, after all the subjects had participated, each of them was paid \$4.

data are presented for 28 of the original 32 subjects, seven subjects from each of the four experimental conditions.

Since the maximum possible number of errors was less in the improvement than in the non-improvement condition, a correction was applied to the former scores, to make them comparable. The correction was as follows:

Corrected "improvement" error score = Initial score $\times \Sigma NI \text{ max} / \Sigma I \text{ max}$, where $\Sigma NI \text{ max}$ is the maximum number of errors possible in the non-improvement condition, and $\Sigma I \text{ max}$ is the maximum number of errors possible

in the improvement condition. Table 1 shows the corrected error score per type of error for each subject, as well as the total corrected error score, and the averages per experimental condition.

Table 2 presents the results of an analysis of variance of the total error scores.

Only the reward condition achieved significance. The improvement or non-improvement of the patient appears to have had no influence on the number of errors made.

Analyses of variance were also computed for the different types of errors. Only in the case

Table 1
Distribution of Corrected Types of Error Per Subject and Their Averages

		<i>Pressing button A or B more or less times than required</i>	<i>Pressing button A or B too early</i>	<i>Pressing button A more than 3 sec too late</i>	<i>Pressing button B during or after the second headbang of a volley</i>	<i>Total errors</i>
reward- improvement group	A	0	10.65	0	0	10.65
	B	0	14.91	8.52	2.13	25.56
	C	2.13	10.65	2.13	0	14.91
	D	2.13	4.26	6.39	0	12.78
	E	2.13	4.26	0	0	6.39
	F	4.26	17.04	2.13	0	23.43
	G	0	4.26	12.78	0	17.04
		$\bar{x} = 1.53$	$\bar{x} = 9.43$	$\bar{x} = 6.36$	$\bar{x} = 0.30$	$\bar{x} = 15.82$
reward- non-improve- ment group	H	0	3	4	2	9
	I	0	7	2	2	11
	J	0	5	3	0	8
	K	1	3	6	0	10
	L	2	3	3	0	8
	M	1	4	1	1	7
	N	1	4	3	0	8
		$\bar{x} = 0.71$	$\bar{x} = 4.14$	$\bar{x} = 3.14$	$\bar{x} = 0.71$	$\bar{x} = 8.71$
no reward improvement group	O	0	10	3	0	13
	P	2	20	15	5	42
	Q	2.46*	3.69*	7.38*	8.62*	22.15*
	R	16	19	9	9	53
	S	1	4	26	2	33
	T	0	9	2	3	14
	U	2	8	2	0	12
		$\bar{x} = 3.35$	$\bar{x} = 10.53$	$\bar{x} = 9.20$	$\bar{x} = 3.94$	$\bar{x} = 25.02$
no reward non-improve- ment group	V	0	8.73	2.91	0	11.64
	W	0	12.78	0	4.26	17.04
	X	4.26	6.39	4.26	2.13	17.04
	Y	4.26	21.30	0	6.39	31.95
	Z	4.26	0	23.43	8.52	36.21
	AA	2.13	4.26	38.34	4.26	48.99
	BB	8.52	4.26	0	0	12.78
		$\bar{x} = 3.35$	$\bar{x} = 8.25$	$\bar{x} = 9.85$	$\bar{x} = 3.65$	$\bar{x} = 25.07$

*The error scores of Subject Q were also corrected, because he misunderstood the instructions and had to be reminded to start responding, while part of the experimental tape already had been played.

Table 2
Analysis of Variance of Total Error Scores

Source	Sum of squares	d.f.	Mean square	F	p
Reward Condition	1144.84	1	1144.84	7.67	.025
Improvement Condition	86.66	1	86.66	0.58	n.s.
Reward $\times$ Improvement	90.21	1	90.21	0.61	n.s.
Error	3578.40	24	149.10	—	—
TOTAL	4900.11	27	—	—	—

of error type d. (pressing Button B during or after the second headbang of a volley) did the reward condition achieve significance ($F = 11.77$, $p = < 0.005$). However, the other types of error (see averages in Table 1) show also a tendency for the reward condition to be different, although not different to a significant degree.

Correlations were computed between total error scores and age, sex, last grade obtained at school, position at hospital, and years of nursing experience. No significant relationships were found between the total error score and any of the variables.

All the 14 subjects in the improvement condition reported in the questionnaire that they detected improvement of the patient (none of this group denied this). Of the subjects in the non-improvement group, five reported improvement, while eight others could not detect any improvement (furthermore, one subject did not fill in this part of the questionnaire). Fisher's exact probability test showed that $p = 0.00058$, while the phi coefficient was -0.64 . The conclusion seems to be warranted that subjects in the improvement condition reported improvement of the "patient" significantly more than did subjects in the non-improvement condition.

Similarly, all 14 subjects in the improvement condition responded positively on the question concerning whether they were successful in applying the therapy (none of this group denied this). However, eight subjects in the non-improvement group also reported that they were successful with the therapy, while five subjects of the same group thought the opposite. Fisher's exact probability was 0.0159 ; the phi coefficient was -0.44 , so that it can be concluded that subjects in the improvement condition tended to report more success in therapy than subjects in the non-improvement condition.

DISCUSSION

It was expected that both promise of reward and improvement of the "patient" would increase accuracy of the subjects in applying the "therapy". This expectation was borne out in the case of promise of reward, but not for the improvement condition. Also, there was no significant interaction between the reward and improvement conditions.

The design of this experiment has several drawbacks, such as the fact that the subjects were not required to meet any specified criteria of performance during the practise period. Since the experiments required 10 days, some interaction between the subjects probably occurred. For this reason, the confidentiality of the experiment was emphasized to each participant, but this, as later appeared, did not prevent some discussions among the subjects about the experiment.

That only two out of 30 subjects were judged to be suspicious on the questionnaire, does not exclude the possibility that the subjects thought they were expected to play the role of "good subject", or that the questionnaire was not sensitive enough to pick up suspicions. Future research in this field should take more account of the hypotheses subjects have about the purpose of the experiment.

The fact that optimal contingencies were not applied, may have influenced the performance of the inexperienced subjects. Experiments with realistic contingencies would clarify this issue.

The clear-cut effect of a promise of monetary reward is interesting, and has obvious implications for behavior modification programs regarding training and maintenance of an adequate performance level by nursing staff members who administer them. Provision of monetary reinforcers may often be difficult in practical settings, but there is a multiplicity of

other potentially useful reinforcers that could be applied, such as upgrading, social reinforcers, selection of preferred days off, *etc.* It would certainly seem justified to search for and make use of reinforcements for nursing staff members, in a planned and contingent manner.

The same may be true for other kinds of behavioral engineers, such as psychology students, parents, school teachers, and patients working with other patients (for the latter, see, *e.g.*, Whalen and Henker, 1969, and Ludwig, Marx, and Hill, 1971). All too often it appears that the required behavior of behavioral engineers is hard to teach, or that "anti-therapeutic" behavior is emitted (see *e.g.*, Ayllon and Michael, 1959; Azrin, Holz, Ulrich, and Goldiamond, 1961; Ayllon and Haughton, 1964), and that the rate of desired responding of behavioral engineers decreases over time. Strangely enough, most studies in which behavioral modifiers are used provide very limited information about their training and ways to maintain their responding.

It seems necessary to break down the behavior of behavioral engineers into component responses. Some of these responses are listed in a rating scale to measure the skill in applying behavior modification techniques developed by Gardner, Brust, and Watson (1970), *e.g.*, "gets the subjects' attention", "gives the reward quickly", "properly ignores 'bad' behavior", *etc.* Listing these responses, however, is not sufficient; it is also necessary to identify the stimuli that control them, and to clarify the contingencies of reinforcement.

The field of operant conditioning can benefit from more detailed experimentation (which is now underway), especially by varying the task schedules over longer periods of time, introducing concomitant tasks (to imitate the complexity of the usual ward work, or school teaching), examining the effect of different types of reinforcements as well as different schedules of reinforcement of the behavioral modifier, and varying his expectations about the result of his intervention.

Training of operant conditioners might be feasible in the future by using multi-channel video-tapes in which desired and undesired

behaviors are recorded on different channels. These behaviors can then be programmed individually to appear on the screen at the appropriate time. The programming equipment might also check whether the behavioral engineer is responding adequately at the right time, and would give him feedback about his performance, if necessary, immediately, *e.g.*, in the form of token reinforcers.

Only a few of the many aspects of engineering the behavioral engineer have been touched upon here. Concluding, it can be stated that this field is still in an embryonic stage of scientific development.

REFERENCES

- Ayllon, T. and Michael, J. The psychiatric nurse as a behavioral engineer, *Journal of the Experimental Analysis of Behavior*, 1959, 2, 323-334.
- Ayllon, T. and Haughton, E. Modification of symptomatic verbal behavior of mental patients, *Behavior Research and Therapy*, 1964, 2, 87-97.
- Ayllon, T. and Azrin, N. The token economy, Appleton-Century-Crofts, New York, 1968.
- Azrin, N. H., Holz, W., Ulrich, R., and Goldiamond, I. The control of the content of conversation through reinforcement, *Journal of Experimental Analysis of Behavior*, 1960, 4, 25-30.
- Bricker, W. A., Morgan, D. G., and Grabowski, J. G. Development and maintenance of a behavioral modification repertoire for attendants working with behaviorally retarded children, *George Peabody College*, 1968, unpublished.
- Gardner, J. M., Brust, D. J., and Watson, L. S. A scale to measure skill in applying behavior modification techniques to the mentally retarded, *American Journal of Mental Deficiency*, 1970, 74(5), 633-636.
- Ludwig, A. M., Marx, A. J., and Hill, P. A. Chronic schizophrenics as behavioral engineers, *Journal of Nervous and Mental Disease*, 1971, 152(1), 31-44.
- Panyan, M., Boozer, H., and Morris, N. Feedback to attendants as a reinforcer for applying operant techniques, *Journal of Applied Behavior Analysis*, 1970, 3, 1-4.
- Rosenthal, R. and Fode, K. L. The effect of experimenter bias on the performance of the albino rat, *Rapport University North Dakota*, 1961.
- Schaefer, H. H. and Martin, P. L. Behavioral therapy for "apathy" of hospitalized schizophrenics, *Psychological Reports*, 1966, 19, 1147-1158.
- Whalen, C. K. and Henker, B. A. Creating therapeutic pyramids using mentally retarded patients, *American Journal of Mental Deficiency*, 1969, 74(3), 331-337.

Received 30 November 1970.

(Revised 18 June 1971.)