MANAGEMENT OF CHRONIC PSYCHIATRIC PATIENTS BY A TOKEN REINFORCEMENT SYSTEM[1]

R. C. WINKLER[2]

UNIVERSITY OF NEW SOUTH WALES

A token reinforcement system for the reduction of institutionalized behavior in a chronic psychiatric ward is described. Quantitative assessments were made of the effects of the system on five types of positively reinforced behavior, two types of unreinforced behavior, and two types of fined behavior. To assess whether these effects were a function of the token procedures, three experiments were carried out, the first removing tokens for a brief period, the second making tokens non-contingent on behavior, and the third removing fines in one area while maintaining them in another. The token procedures were found to be the source of the observed improvements.

Despite advances elsewhere, the management and treatment of chronic refractory patients has remained a major problem. The treatment approaches used, if any, have had only limited success and the behavior of chronic patients remains marked by absence of spontaneity, general inactivity, lack of self-care, and capricious tantrums. Coherent responses to the contemporary environment are rare and there is little interest in planning for the future.

Recently, a new type of ward program has been developed for these patients. The essential characteristic of the program is that explicit consequences are arranged for these patients' behavior. If a patient behaves well, his behavior is reinforced; if he behaves badly, he may suffer aversive consequences. The reinforcers are tokens, which can be exchanged for desired goods and activities. A number of studies has shown that these token reinforcement systems can within a short time markedly ameliorate the institutionalized behavior of chronic patients as a function of the token reinforcement procedures (Ayllon and Azrin,

1965; Schaefer and Martin, 1966; Atthowe and Krasner, 1968; Lloyd and Garlington, 1968).

In June, 1967, after six months of preparation, a token system was established at Gladesville Psychiatric Hospital, a major hospital in Sydney, Australia. The general aims of the program were (i) to carry out an experimental analysis of behavior within a token system, and (ii) to make the management of patients easier for the staff by moving patients towards rehabilitation; more specifically, to reduce institutionalized behavior, to increase active behavior, and to maintain that behavior over an extended period.

The Ward Population

The token system was established in a closed, refractory female ward with 66 beds. When the program began, 31 of the patients had been diagnosed as chronic schizophrenics, 18 as congenital mental defectives, and the remainder as having manic depressive psychoses or chronic brain syndromes. The mean age of the patients was 49 yr, ranging from 20 to 85 yr, and the average period of hospitalization was 12 yr, their hospitalization ranging from six months to 39 yr. The patients' behavior was characterized by excessive violence and screaming, as well as apathy and general lack of response to the ward environment. The staff allocation was not altered for the program, and the ward retained its normal function in the hospital throughout the program. No formal course of staff instruction could be given but as far as possible, each new nurse was in-

terviewed on her first day. The nursing staff
changed very rapidly according to an unpre-
dictable roster.

The Token System

The token system was similar to that de-
scribed by Ayllon and Azrin (1965). The pa-
tients earned tokens (small plastic discs) by
performing a wide range of behavior. Almost
all token-earning behavior was recorded daily
at least two weeks before the behavior began
to earn tokens and was recorded daily while it
was being reinforced. Each patient's purchases
at the ward shop were recorded and the num-
ber of tokens exchanged for every type of pur-
chase was also recorded daily.

Violence and unnecessary loud noise (tan-
trums, screaming, banging) were recorded
daily and fined by removing tokens from pa-
tients according to the principles of response
cost (Weiner, 1962). To ensure that patients
paid their fines, patients had to pay double the
usual prices when spending tokens for the
three days immediately following their refusal
to pay a fine. If during this double-prices pe-
riod, a patient paid her fine, her prices re-
turned to normal. This procedure also pro-
vided a technique for controlling violent or
noisy patients who at times had no tokens; if
they could not pay their fine, they were still
put on the aversive double prices.

Violence and noise were also recorded im-
mediately before the token system was intro-
duced and again immediately before the in-
troduction of fines, in order to assess any
possible changes in non-reinforced behavior.
Atthowe and Krasner (1968) showed that gen-
eralized changes in socialization occurred in a
token system, but it is not clear from their re-
port that socialization went completely un-
reinforced.

Stealing of tokens was prevented by giving
the patients involved uniquely colored tokens
and insisting they spend only that color. Spe-
cially designed reinforcement schedules were
used for individual patients who posed special
problems.

RESULTS

Despite the staff's fears that the patients
would not understand, the patients responded
quickly and most began to improve within
the first week of tokens. After six months, all

but one patient were spending tokens; there
were only six patients, of widely varying ages,
diagnoses, and lengths of hospitalization,
whose token-earning behavior had not im-
proved. Staff morale seemed to improve mark-
edly. Absenteeism, measured in terms of the
number of cases where a member of staff was
rostered for duty and did not appear, was ex-
amined in the four months before the pro-
gram began and the four months immediately
following. Absenteeism dropped 24% after the
program began, while in a comparable ward,
absenteeism over the same period dropped
only 3%.

Without exception, every type of behavior
that was reinforced, improved. Since the only
previous quantitative demonstration of im-
provements in a token system with adult psy-
chotics had been limited to two types of be-
havior (Atthowe and Krasner, 1968), almost
all reinforced behavior was recorded in the
present study. There was little judgment in-
volved in recording most of these types of be-
havior (*e.g.*, attendance in a room at a certain
time). Some, however, were more subject to
poor observer agreement and were made un-
der more pressing conditions. Early morning
behavior was one of the most difficult to re-
cord, since changes in the recording staff were
rapid, time was short, and criteria were more
complicated. The inter-observer agreement
on three aspects of early morning behavior was
therefore assessed.

Inter-Observer Agreement

Inter-observer agreement is usually assessed
by comparing the records made by two spe-
cially trained observers who are aware that
their observations are being compared. This
technique does not necessarily indicate the
agreement that might be achieved in practice,
particularly when the situation is complicated
by changing observers, time pressures, and
varying motivations. To obtain agreement
data more relevant to this situation, the nurses'
record books were, for a period of 11 days dur-
ing baseline recording, designed to allow the
possibility of a patient's behavior being re-
corded "in error" by an extra nurse. Neither
of the nurses involved in any double recording
knew that another parallel record was being
made in another record book. "Blind" agree-
ment was studied for three types of behavior:
getting up, dressing, and bed making. For each

type of behavior, 79 cases were obtained where two nurses had each recorded the same patient's behavior. These 79 cases were obtained from 36 different patients and were recorded by more than 25 different nurses. There was highly significant agreement, using Cohen's k (Cohen 1960), on dressing and bed making, and agreement approaching significance for getting up (getting up: 86.1% agreement, $Z = 1.67$, $p > 0.05$; dressing: 92.4% agreement, $Z = 3.29$, $p < 0.001$; bed making: 88.6% agreement, $Z = 5.71$, $p < 0.001$). The lower agreement for getting up was a function of the relatively disproportionate marginal totals for getting-up records, which limited the maximum value of k to 0.73. The absence of significant agreement for getting up, despite high percentage agreement, illustrates the inadequacy of using percentage agreement to assess rater agreement when marginal totals are disproportionate. Phi coefficient correlations between the first and second half of the 11-day period for 15 patients who had one record in each half ranged from 0.63 to 0.75 over the three behaviors. Assuming that the behaviors were relatively stable over time, this provides further evidence of rater reliability, as different sets of observers were involved in the two periods. This agreement was felt to be sufficient to discount the possibility that effects were measurement artifacts, particularly as the changes to be described were large and sustained over lengthy periods.

Positive Reinforcement

The effect of positive token reinforcement is illustrated by the following five responses:

1. *Attendance at morning exercises.* Figure 1 shows the mean number of patients starting the 25-min exercise period per five-day week before and after attendance earned tokens. The early improvements were maintained over the following 12 months, as shown in Fig. 1.

2. *Finishing morning exercises.* To reduce the number of patients wandering away from exercises (*i.e.*, to increase "concentration"), tokens were given at the end of the exercises to patients who had stayed. This criterion was subsequently changed so that patients not only had to stay but also had to perform the exercises well. Figure 1 shows the initial improvement in the mean number of patients staying per week. A year later, fewer patients met the

criterion, but criteria were very much more stringent. A further criterion, introduced after eight months, required 21 patients to attend exercises before tokens for starting and performing were given out. When there were fewer than 21 patients, 31% stayed and performed well, but when there were 21 or more patients, 69% stayed in the five weeks after the criterion was added. The extra criterion not only provided a continuous test of token control, but led to the patients bringing other patients to exercises.
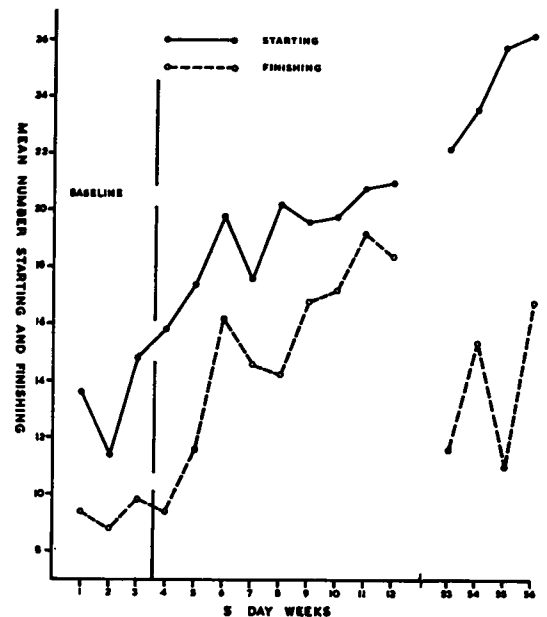

Fig. 1. Attendance at and number finishing morning exercises before and after positive token reinforcement.

3. *Getting up, dressing, and bed making.* Preparing patients for the day is one of the most difficult tasks for hospital staff (Gericke, 1966), particularly in chronic wards, where the patients' lack of self-help is a major problem. Getting up, dressing, and bed making were reinforced with the effects shown in Fig. 2. The percentage of patients not looking after themselves in the morning was markedly reduced, thereby giving the staff more time to spend on more profitable activities.

Active self-help was also increased by requiring that patients bring themselves to meals when the meal was announced, and locking the door seven min after the call. This procedure, used by Ayllon (1963) with an individual patient, was used for all patients, but
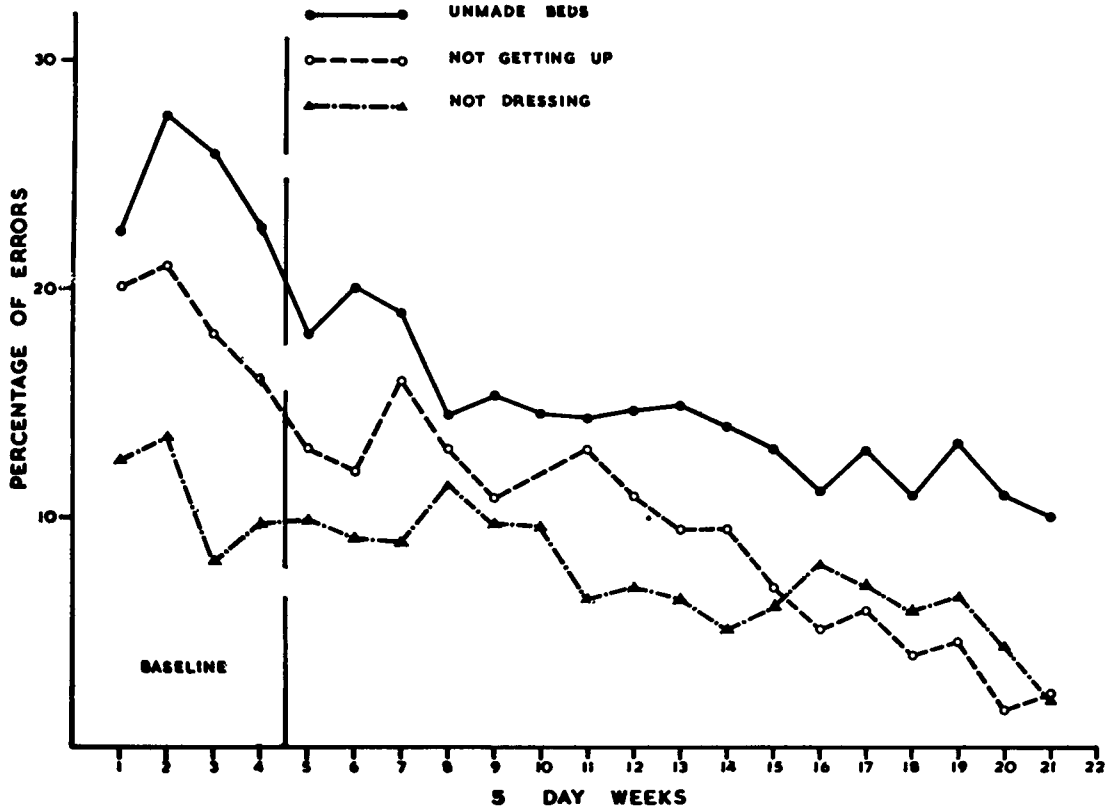
Fig. 2. Percentage of patients not getting up, not dressing, and not making their bed before and after positive token reinforcement.

particularly for four to six patients who refused to come to meals without being brought by the staff, saying they could not walk or their "voices" told them not to come. Whereas this behavior used to produce attention and a meal, it now resulted in no meal and being ignored by the staff. With the extinction of the old behavior and the reinforcement of self-help, there was a substantial reduction in the number of times patients failed to bring themselves to meals. Figure 3 shows the number of patients not at meals over all meals and at breakfast. The increase in the total number of missed meals in the first two weeks of the procedure reflects the behavior of patients who "tested the limits" of what the system allowed them to do. Unlike the other two meals, breakfast punctuality improved immediately, possibly because the time between breakfast and the preceding meal was longer than the corresponding intervals for the other meals.

These improvements, together with others reported elsewhere (Winkler, 1969), indicate the increase in patient activity, self-help, and

initiative that occurred with the introduction of the token system. Instead of lying or pacing around, the patients sought out activities without being asked. They saved, chose, and planned, showing a degree of self-control and understanding that was previously regarded by the staff as impossible for chronic patients.

Non-Reinforced Behavior

Behavior subject to reinforcement procedures was not the only behavior to improve. Both violence and noise were recorded daily for five weeks before the program was introduced and recorded again for five weeks, five weeks after the program had begun. During these periods, violence and noise were not involved in reinforcement procedures in any way. Violence was recorded by the five nurses rostered for the ward on their 12-hr shift and by the occupational therapist. Noise was recorded independently by the senior nursing staff and occupational therapist. Over the two recording periods, 59 and 11 different staff members recorded violence and noise respec-
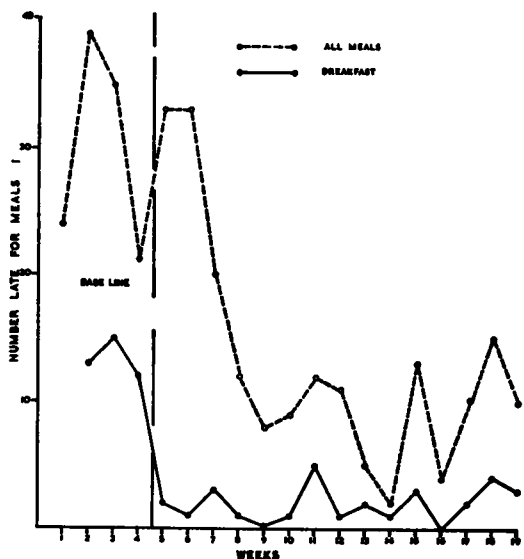
Fig. 3. Non-attendance at all meals and at breakfast before and after meals made contingent on punctuality.

Table 1

Non-Reinforced Behavior: Violence and Noise

Behavior	Physical Violence		Unnecessary Loud Noise
Recorders	All Recorders	Senior Sisters	Occupational Therapist
Mean episodes per week before tokens	44.4	52.0	54.2
Mean episodes per week after tokens	23.8*	24.6**	16.0**

*p < 0.05.
**p < 0.001.

tively. Table 1 shows that violence and noise were significantly reduced [Wilcoxon, T. (Wilcoxon, 1949) p < 0.05 and p < 0.001 respectively], without specific attempts to treat this behavior. Five weeks after the token system had begun, 47 of 57 patients were less violent, 56 of 57 were less noisy according to the senior sisters' records, and all recorded patients were less noisy according to the occupational therapist's records. The large number of recorders involved in recording violence, and the fact that different observers were involved in the two periods, support the reliability of the change in violence. The number of observers recording noise was smaller, but the records of the two independent observers both show substantial reductions in noise, thereby suggesting the reliability of this result. Atthowe and Krasner (1968) reported apparently unreinforced changes in social interaction and taking passes from the ward. However, a manual from their ward shows that taking passes was required for the maintenance of membership in a privileged group (Brockhoff 1966). Also, as attendance and performance at group activities earned tokens, at least some aspects of social interaction were probably reinforced. It is not, therefore, clear that the changes noted were actually changes in unreinforced behavior. Nevertheless, the present results demonstrate that such changes can occur. At-

thowe and Krasner interpreted their changes as indicating response generalization, but this is not an explanation. Such changes are more probably a function of improved staff morale. Violence and noise can be a function of staff conflict (Stanton and Schwartz, 1954), disinterest, or hostility towards the ward. A more enthusiastic and contented staff provokes fewer disturbances, and in turn, the reduction in disturbances improves morale.

Negative Reinforcement

Noise and violence were reduced further by the introduction of fines for this behavior. Fines were introduced 10 weeks after the program began, immediately after the second five-week period mentioned in the previous section. After initial disruptions, the number of episodes of noise per week dropped from pre-fine levels of 25 to 30 to 10 to 15 six weeks later. Violence dropped from 20 to 30 episodes per week before fines to 0 to 7 episodes per week six weeks later. Before the program, staff and patients were continually being attacked, sometimes so badly they required medical treatment, and by the end of the day, the nurses reported that they were emotionally exhausted by the persistent screaming and tantrums. After the program had been established, the nurses reported that the ward was quieter, less violent, and a better place in which to work. Fines used with individual problems confirmed the effectiveness of removing tokens as a procedure for reducing the

frequency of undesirable behavior. For example, Lil, a 56-yr-old mental defective schizophrenic, would change her clothing 7 to 10 times a day, according to five nurses' estimates, and had been doing so for many years. Lil was fined each time she unnecessarily put on a new dress. In the first week of fines, she changed unnecessarily only five times and over the next 12 weeks the number of unnecessary changes was reduced further to 1 to 2 a week.

Token Control

The improvements that coincided with the giving of tokens, and the sensitivity of performance at exercises to the availability of tokens suggested that it was the token procedures that were controlling the patients' behavior. However, the improving baselines and the improvements in unreinforced behavior suggested that other factors might also have produced the observed gains in reinforced behavior. It was therefore important to investigate to what extent tokens were actually controlling behavior. Three experiments were carried out.

In the first experiment, tokens for shoe cleaning were discontinued for three weeks and then reintroduced as before. The day before the experiment the following instructions were given:

> You know how you have been getting 1 token for every pair of shoes you clean at night. Well, from tonight on, you won't be getting any tokens for cleaning shoes for the next 3 weeks. (Repeated). We still want you to clean shoes as you have been doing. The night nurse will still come around and record whether you have cleaned your shoes just as she usually does. (Repeated). When the 3 weeks is up, we'll be going back to the tokens for shoes. Don't worry about the other tokens. You'll still be getting tokens for everything else just as usual. It's only shoe cleaning we're doing this for.

The reintroduction of tokens was also announced the day before it occurred. No explanation of the changes was given. Tokens were removed in the middle of the night nurses' roster, so one nurse was on duty both before and after the removal of tokens and another nurse before and after their reinstatement. Four different nurses recorded during the period of no tokens, thus limiting any systematic recording biases and other staff effects. Figure 4 shows the immediate decrease in the number of shoes cleaned when tokens were discontinued and the immediate increase on their reinstatement. All patients who had cleaned shoes in the baseline period showed a decline in performance during the experimental period. The slight increase in shoe cleaning in the third week of no tokens may have been due to some patients mistakenly believing tokens were due back one night, the first night of the week.
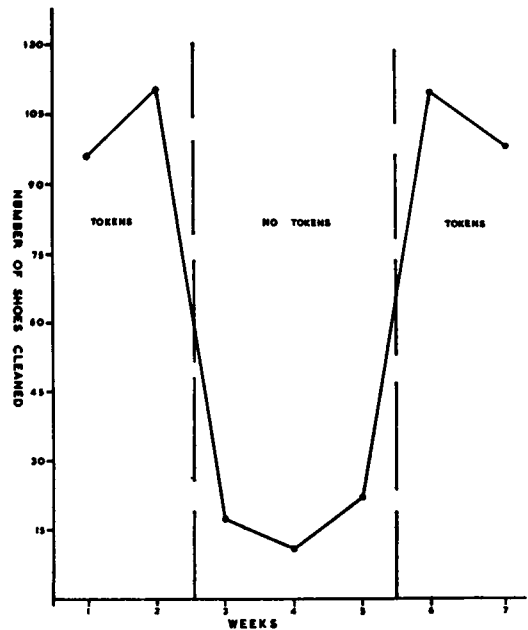

Fig. 4. Number of shoes cleaned as a function of the presence and absence of positive token reinforcement.

This experiment clearly indicated that tokens needed to be given out if the behavioral improvement was to be maintained. This result is consistent with Ayllon and Azrin (1965).

The second study was carried out at the hospital's Industrial Therapy Unit, where some of the ward's patients worked at various simple assembly type jobs. The number of items of work completed earned tokens on a fixed-ratio schedule, with the modification that tokens earned were given out at the end of the day to avoid the fixed-ratio pauses that otherwise occurred. For 10 days this contingent reinforcement was suspended and the 10 patients involved were given the same number of tokens each day regardless of their work output, *i.e.*, non-contingently. Each patient

was given the following instructions individually:

> You know you've been getting 1 token for every (30 or 70 or 100) bags you do at therapy; well, for the next 2 weeks, we're going to try something different; for the next 2 weeks, we'll be giving you the same number of tokens every day, regardless of how much you do; you usually earn . . . (*e.g.,* 6) tokens a day; sometimes a bit more, sometimes a bit less; well, we're going to give you 6 tokens *every* day for the next 2 weeks, regardless of how much work you do. OK? That starts on Monday.

The flat rate given at the end of each of the 10 days was set so that each patient received the same total number of tokens in the experimental period as in her baseline period. The work output of all patients on their different jobs dropped immediately with the introduction of non-contingent tokens. Figure 5 shows the decrease and subsequent increase in mean number of bags completed per week for the two patients whose job it was to make string handles for carry bags. Two patients had continued on the same job (stringing bags) from the baseline period before tokens were introduced. Under non-contingent tokens, one dropped from a mean of 151.2 to a mean of 87.0 bags and the other from 142.2 to 116.2. In the week before tokens were introduced, their mean outputs were 87.8 and 116.6 bags respectively. Under non-contingent tokens, pre-

vious gains disappeared and performance reverted to the level achieved before reinforcement began. The reduction in performance produced by giving tokens non-contingently is consistent with Ayllon and Azrin (1965), whose results however may have been, in part, a function of their instructions, which suggested patients take a vacation from work. Lloyd and Garlington (1968) found a similar effect.

The third experiment examined control of violence and noise by fines. Since noise and violence, particularly noise, could vary capriciously, a design was used that permitted control of extraneous factors such as staff changes, weather, and changed medications.

For 14 O.T. sessions, fines were discontinued in the O.T. room, while being maintained as usual in the ward. Since the O.T. room was separate from the ward, the fined and unfined areas were easily discriminated. The 20 to 25 patients who attended O.T. were in the ward at all other times, and thus moved from the fined to unfined area once or twice each weekday. Noise and violence, particularly noise, increased in the unfined area but not in the fined area, and when fines were reinstated in the O.T. room, noise and violence decreased, as shown in Figs. 6 and 7 respectively. (Re-
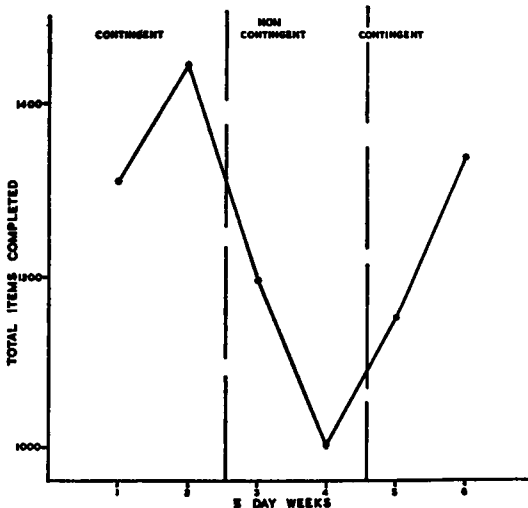
Fig. 5. Work output as a function of the contingency and non-contingency of tokens.
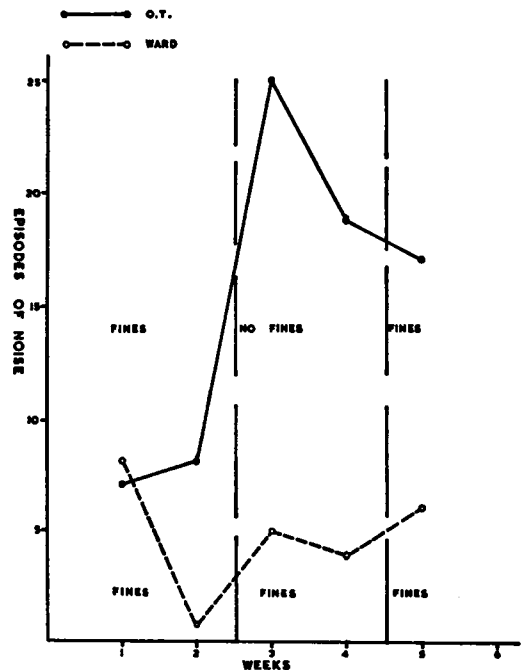
Fig. 6. Unnecessary loud noise as a function of the presence and absence of fines.

covery data were not complete for O.T. as the occupational therapist went on an unexpected holiday).

This third experiment showed that fines were necessary to maintain a low frequency of tantrums and screaming, and may have had a similar function with violence. The small change in violence with the removal of fines is difficult to explain. However, the increase in the unfined area corresponded with a decrease in the fined area, so that the relative change after removing fines is more marked. These results are consistent with Burchard (1967), who showed that the removal of tokens lost its controlling effect when made non-contingent.
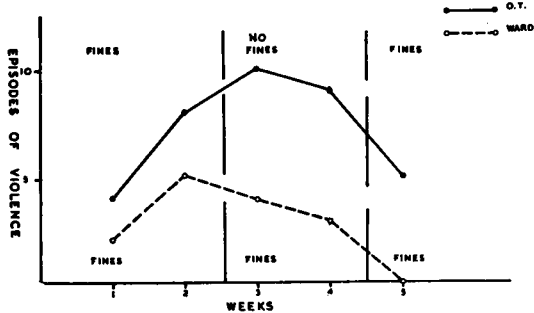

Fig. 7. Violence as a function of the presence and absence of fines.

DISCUSSION

Without exception, each type of behavior that was reinforced, improved. The patients learned to behave in ways the nursing staff reported that they previously felt were impossible. The staff reported, for example, that patients who previously refused to behave well came to demand opportunities to do so and their reinforced behavior was more consistent; more questions were asked and more activity occurred without prompting by the staff; generally, there was less institutionalized apathy to be seen around the ward; nurses did not have to badger uncooperative patients as frequently as before and at various points during their day, the time taken to complete jobs was shortened. The clinical success of the program could also be measured by the fact that the program, originally a research experiment, was continued by the hospital as part of its normal routine. Other wards were exhorted to begin similar programs and other hospitals in Australia began token systems.

The changes in patient behavior could not be explained solely by such factors as staff enthusiasm, extra luxuries, or increased expectations. The three experiments on token control strongly suggested that the patients' behavior was controlled by the token procedures; more specifically, through the contingent relationship between behavior and tokens. This, however, does not deny the importance of other factors, such as improved morale.

The degree of behavioral control found also indicates that, despite their usual appearance, chronic patients are particularly sensitive to their environment. Chronic institutionalized patients are often regarded as not being able to respond or to respond coherently to their environment. Superficial clinical experience strongly supports this argument. Yet it has been made clear that when an environment is created that reinforces active and coherent responses, such responses are forthcoming. It seems probable that much of the patients' initial institutionalized passivity developed through the hospital system reinforcing this passivity either by doing the patients' work for them or giving the passive patient more attention. Such conditions are not hard to find in psychiatric hospitals (Gelfand, Gelfand, and Dobson 1967). In a token system, patients are rewarded for activity, not for inactivity.

REFERENCES

Atthowe, J. M., Jr. and Krasner, L. Preliminary report on the application of contingent reinforcement procedures (token economy) on a 'chronic' psychiatric ward. *Journal of Abnormal Psychology*, 1968, **73**, 37-43.

Ayllon, T. Some behavioral problems associated with eating in chronic schizophrenic patients. In L. P. Ullman and L. Krasner (Eds.), *Case histories in behavior modification*. New York: Holt, Rinehart & Winston, 1965. Pp. 73-77.

Ayllon, T. and Azrin, N. H. The measurement and reinforcement of behavior of psychotics. *Journal of the Experimental Analysis of Behavior*, 1965, **8**, 357-383.

Brockhoff, L. A. *Ward 106 token program, incentives and costs—a manual for patients.* Unpublished manuscript. V.A. Hospital, Menlo Park Division, 1966.

Burchard, J. D. Systematic socialization: a programmed environment for the habilitation of antisocial retardates. *Psychological Record*, 1967, **17**, 461-476.

Cohen, J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 1960, **20**, 37-46.

Gelfand, Donna M., Gelfand, S., and Dobson, W. R. Unprogrammed reinforcement of patients' behavior in a mental hospital. *Behavior Research and Therapy*, 1967, 5, 201-7.

Gericke, J. R. Problems for behavioral therapy in mental hospitals. *Psychological Reports*, 1967, 20, 534.

Lloyd, K. E. and Garlington, W. K. Weekly variations in performance on a token economy psychiatric ward. *Behavior Research and Therapy*, 1968, 6, 407-410.

Schaefer, H. H. and Martin, P. L. Behavioral therapy for 'apathy' of hospitalized schizophrenics. *Psychological Reports*, 1966, 19, 1147-1158.

Stanton, A. H. and Schwartz, M. S. *The mental hospital*. New York: Basic Books, 1954.

Weiner, H. Some effects of response cost upon human operant behavior. *Journal of the Experimental Analysis of Behavior*, 1962, 5, 201-208.

Wilcoxon, F. *Some rapid approximate statistical procedures*. Stanford, Conn.: American Cyanamid Co., 1949.

Winkler, R. C. *Management of chronic psychiatric patients by a token reinforcement system*. Unpublished doctoral dissertation. School of Applied Psychology, University of New South Wales, 1969.