

TECHNICAL REPORT

No. 65-1

REINFORCING VALUE OF INFORMATION

Prepared by:

Derek P. Hendry

This investigation was supported by Grant No. Nsg 189-61 between National Aeronautics and Space Administration and Space Research Laboratory, Department of Psychology, University of Maryland, College Park, Maryland.

In psychological work with human beings it is taken for granted that information about favorable or unfavorable states of the environment is reinforcing. That is, people will perform some task just to find out whether some other reinforcing event is likely to occur or likely not to occur. It is obvious that knowledge of potential hazards and benefits in the environment has great survival value. It is curious that the reinforcing value of information in infra-human organisms has not been generally accepted by psychologists, though the point of view has been cogently presented (Berlyne, 1960).

In traditional psychological theories there is only one mechanism for the acquisition of reinforcing value: for a stimulus to become a conditioned reinforcer, it must be temporally paired with a reinforcer (Hull, 1943, 1952; Kimble, 1961). A variant on this theory is generally called the discriminative stimulus hypothesis. According to this hypothesis a stimulus becomes a conditioned reinforcer if the stimulus is the occasion for a reinforced response (Dinsmoor, 1950; Keller & Schonfeld, 1950; Skinner, 1938). Such a stimulus is called a (positive) discriminative stimulus. The evidence is almost unanimous that a discriminative stimulus is a conditioned reinforcer (Kelleher & Bellub, 1962). A discriminative stimulus is informative in the loose sense that it predicts the occurrence of reinforcement. We may formally (but still tentatively) define an informative stimulus as one which bears a fixed relation to reinforcement. All discriminative stimuli are then, by definition, informative stimuli. The hypothesis that all informative stimuli are reinforcing (the "information hypothesis") implies that all positive discriminative stimuli are reinforcing.

However, the information hypothesis is more general than the discriminative stimulus hypothesis since informative stimuli are not restricted to positive discriminative stimuli. An important class of stimuli which are informative but not positive discriminative stimuli are negative discriminative stimuli. (The usual symbol for a positive discriminative stimulus is S^D , and for a negative discriminative stimulus S^Δ . When we wish to refer to discriminative stimuli without regard to their sign we shall use the symbol $S^{\Delta D}$). It is widely agreed that S^Δ s, far from being reinforcing, are actually aversive. There is considerable evidence that S^Δ , or time-out (TO) from positive reinforcement, is a condition from which animals will learn to escape (e.g. Ferster, 1958). This result seems at first sight to require rejection or modification of the information hypothesis. However, another possible conclusion is that a stimulus may have both positive and negative reinforcing properties. This is an appealing line of argument. A stimulus is negatively reinforcing if the animal learns to escape it (i.e. terminate it). Therefore, the defining operations imply tests in which the stimulus is turned on by the experimenter and turned off by the animal. There is no logical reason why both these operations with a particular stimulus should not be reinforcing. It has been assumed, not because of any substantial evidence, but probably on the basis of verbal conventions, that a positive reinforcer is ipso facto, not a negative reinforcer, and vice versa.

It is therefore the conclusion of this train of argument that the fact that S^Δ s may be negatively reinforcing does not prevent them from being positively reinforcing as well, as demanded by the information

hypothesis. More concretely, an animal may learn some behavior to produce S^Δ , and also learn other behavior to terminate S^Δ .

That S^Δ s are both positively and negatively reinforcing is a conclusion which is quite congenial with our everyday experience. People in a situation which may or may not lead to some benefit are generally eager to know as soon as possible whether the benefit will materialize. To be informed that the benefit will not materialize is generally regarded as aversive in the sense of producing an emotional state, suppressing behavior, or even escaping the information by forgetting.

The experiment described in this report was designed to show whether information was reinforcing and whether S^Δ was both positively and negatively reinforcing.

METHOD

An adolescent female chimpanzee served in the experiment over a period of about 10 months. She had a previous history of reinforcement on ratio and multiple schedules. The chimp lived in the experimental space, which was a stainless steel, temperature controlled, air-conditioned chamber 5 ft x 5 ft x 7 ft high. On one wall, 2 ft above the floor, was a platform, above which was a 30 in. square aluminum work panel containing lights and levers (see Fig. 1). Water was continuously available, but the daily ration of food had to be obtained by working on the experiment 7-10 hrs daily. The experimental diet of monkey pellets was supplemented at regular intervals with fruit, vegetables, and vitamins.

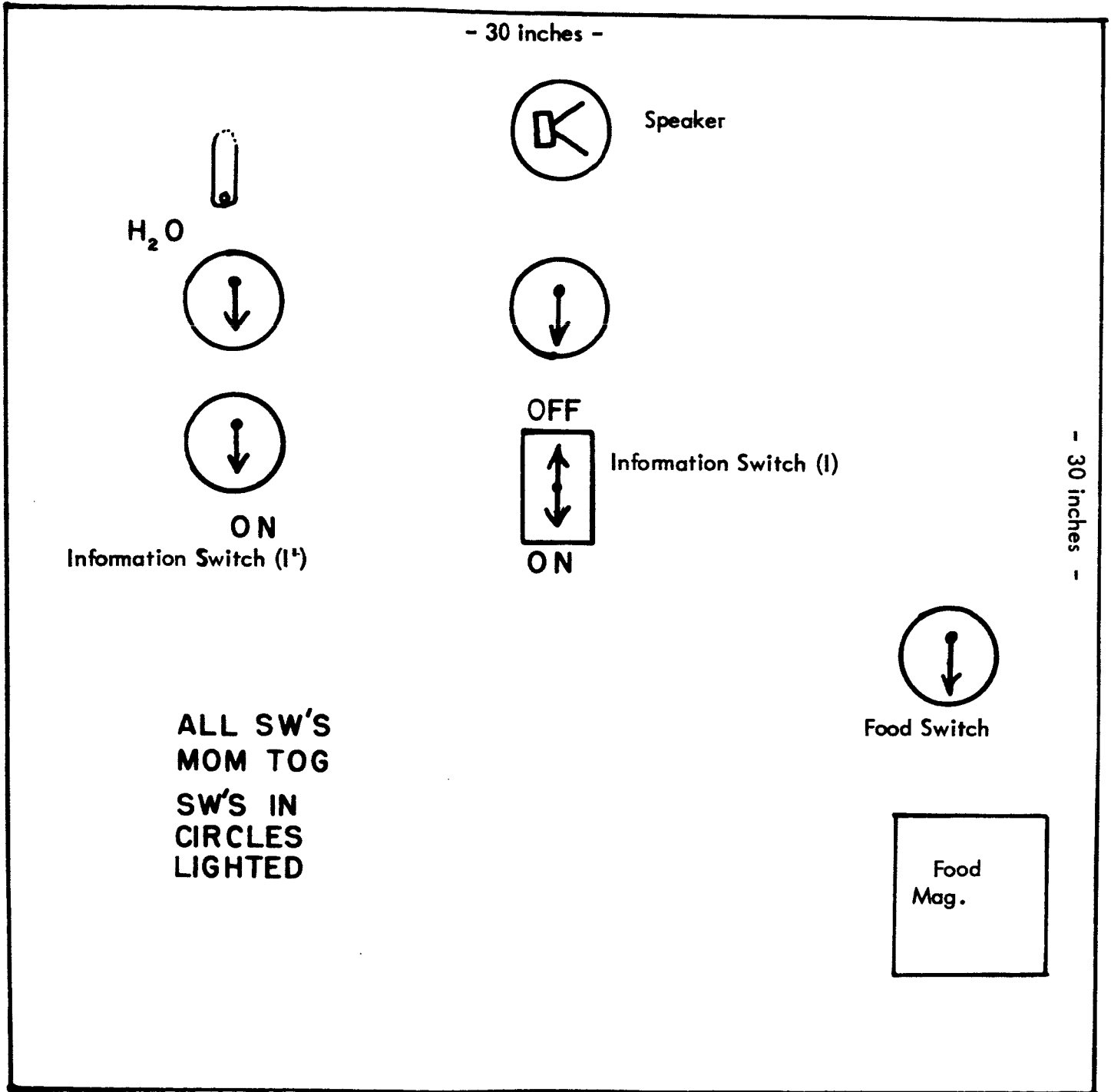


Fig. 1 The work panel; showing the position and function of the operanda. The information ON switch was changed from I to I' half way through the experiment. See text.

The experimental procedure was designed to produce observing behavior. The basic schedule was a multiple (Tandem VR). In the presence of a blue stimulus light, a Tandem FI 2 FI 2 FI 2 FI 2 was scheduled on the food switch. That is, at least one response after the end of 2 min followed by at least one response at the end of the next 2 min, followed by at least one response at the end of the next 2 min, followed by at least one response at the end of the next 2 min, produced food. After food was given at the end of the Tandem schedule, the stimulus color was red. Now responses were reinforced on a VR schedule - every Nth response on the average was reinforced. The actual value of N varied probabilistically according to the following scheme. The actual schedule was either VR 100 or Extinction (Ext). The choice of VR 100 or Ext was determined probabilistically every 30 sec. If there were an equal chance of VR and Ext then the effective schedule would average out at VR 200. The VR part of the multiple schedule terminated after 8 min, provided at least one reinforcement had been obtained in it. Each 8 min period of VR or Tandem will be referred to as a "trip".

One tone was associated with each of the components of the Tandem schedule. The tones ranged from 100 cps to 4000 cps and were adjusted to give phenomenally equal intervals to the human observer. The lowest to highest tones were associated respectively with the first to last components of the Tandem schedule. The tones were presented only after responses on the center (information) switch. In the VR component of the multiple schedule 2 other informative stimuli were available for responses on the information switch. They were loud clicks at either

1/sec or 10/sec. The 10/sec click indicated that VR 100 was in effect, and the 1/sec click indicated that Ext was in effect.

The schedule was chosen because it had been shown that animals would learn to produce (a) informative stimuli which distinguished between VR schedules and Ext (Kelleher, 1958), (b) stimuli associated with the components of Tandem FI 1 FI 1 FI 1 (Dillow, 1964).

One more important feature has to be described. The informative stimuli were produced by downward pressure on the information switch. The stimuli could be terminated by upward pressure on the same switch. At a later stage of the experiment stimuli could be obtained only by depressing one of the side switches.

It is important to note that responding on the information switch had no effect on the scheduling of reinforcements. However, when the schedule changed so that the stimulus that was on was no longer appropriate, the stimulus terminated.

During the course of the experiment changes were made in the number of food pellets given at the end of the Tandem schedule, the response requirement to turn the stimuli on and off, the duration of the stimuli, the time between changes from VR 100 to Ext, and the probability that VR 100 or Ext was in effect. The reason for and effects of these changes will be described in the results.

RESULTS

The contingencies of reinforcement quickly took effect, so that a high rate on the food switch occurred during red, and a lower rate occurred during blue. An appreciable response rate also occurred

initially on the information switch. Fig. 2 shows the response rate over the first 18 sessions. Responding for the clicks in VR Ext increased but responding for the Tandem stimuli gradually declined; in sessions after those shown in Fig. 2, response rate on the information switch during Tandem remained virtually at zero while an appreciable rate was maintained during VR Ext.

The lack of observing behavior in Tandem was thought possibly to be due to the low incentive value of the reinforcement, which was initially 2 pellets. Therefore, after session 30 the number of pellets was increased to 10, for the duration of the experiment. Number of pellets did increase the overall rate in Tandem, but the observing response rate did not go up also.

Another explanation for the disappearance of observing behavior in Tandem is that the stimuli became redundant either because the slight noise produced by the programming equipment at the end of each 2 min was discriminated or because of the chimp's accurate self-generated timing stimuli. No other attempts were made to restore observing behavior during Tandem since the continued absence of observing behavior during Tandem showed that the observing behavior generated in VR Ext was itself under discriminative control. The lack of observing responses in Tandem therefore served as a useful control datum throughout the experiment.

For the first part of the experiment the probability of VR and Ext was approximately 0.5 and the only 1 response was required to turn the stimulus on (for 20 sec) or to turn it off. Under these conditions the chimp learned to turn on S^{Δ} and turn off S^{Δ} . When performance

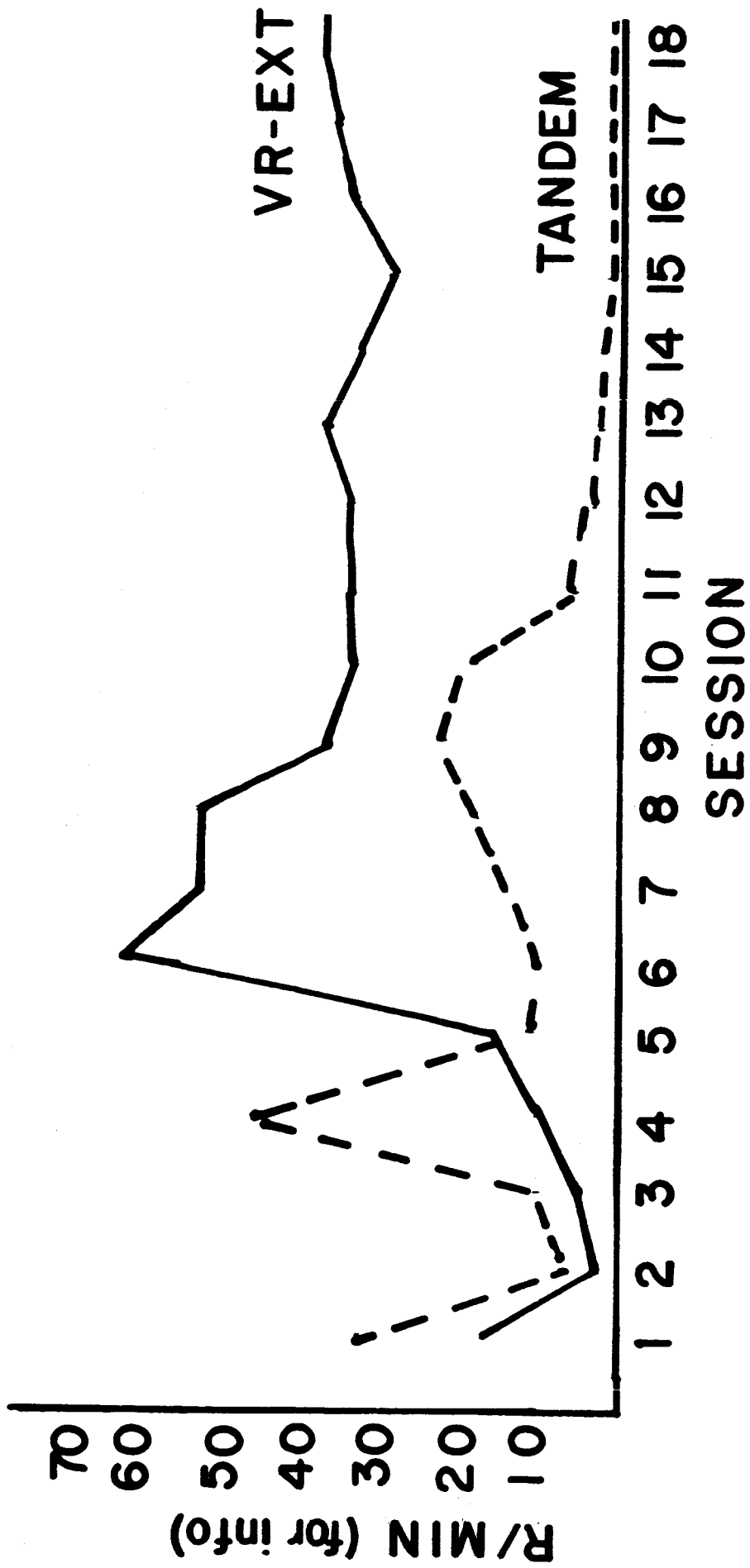


Fig. 2 Response rate on the information switch to turn on one set of stimuli associated with the components of Tandem FI 2 FI 2 FI 2 and another set associated with the components of Mix VR Ext.

was stable the number of responses to turn $S^{D\Delta}$ on or off was raised to 2 then 5, then 10, then 15, then 20.

Figures 3 and 4 illustrate the effect of these schedules. In Fig. 3 line 2 shows that the informative stimuli were very frequently obtained during the Mix schedule, but not during the Tandem schedule. In the Mix schedule, appropriate performance on the food switch is controlled by S^D and S^Δ . Comparison of line 3 and line 4 shows that the informative stimulus was frequently turned off when it was S^Δ , but not when it was S^D . A higher overall rate of information responses was obtained when S^Δ was available than when S^D was available. This effect is clearly shown in the first trip on line 3, where the occurrence of pips happens to indicate that S^Δ was available; where pips occur the slope of the record is greater than where they do not. Close examination of line 3 shows that when S^D was available, information responses exhibited an even spacing, corresponding to the duration of $S^{D\Delta}$. Thus, one response turned on S^D , and the next response was not made until S^D went off, usually 30 sec later. The rate of informative responses is higher when S^Δ is available because S^Δ was not allowed to remain on for its timed maximum duration.

Fig. 4 shows typical performance when the schedule on the informative switch was FR 20. In line 3 the pens remained deflected while S^Δ was on, and responses to escape S^Δ therefore show cumulatively with the pen deflected. The record clearly shows that responses to turn off S^Δ were emitted at a lower and more irregular rate than responses to turn on $S^{D\Delta}$. Moreover, escape responding during S^Δ is sometimes not sustained, as at A. Instead of escaping S^Δ , the chimp sometimes

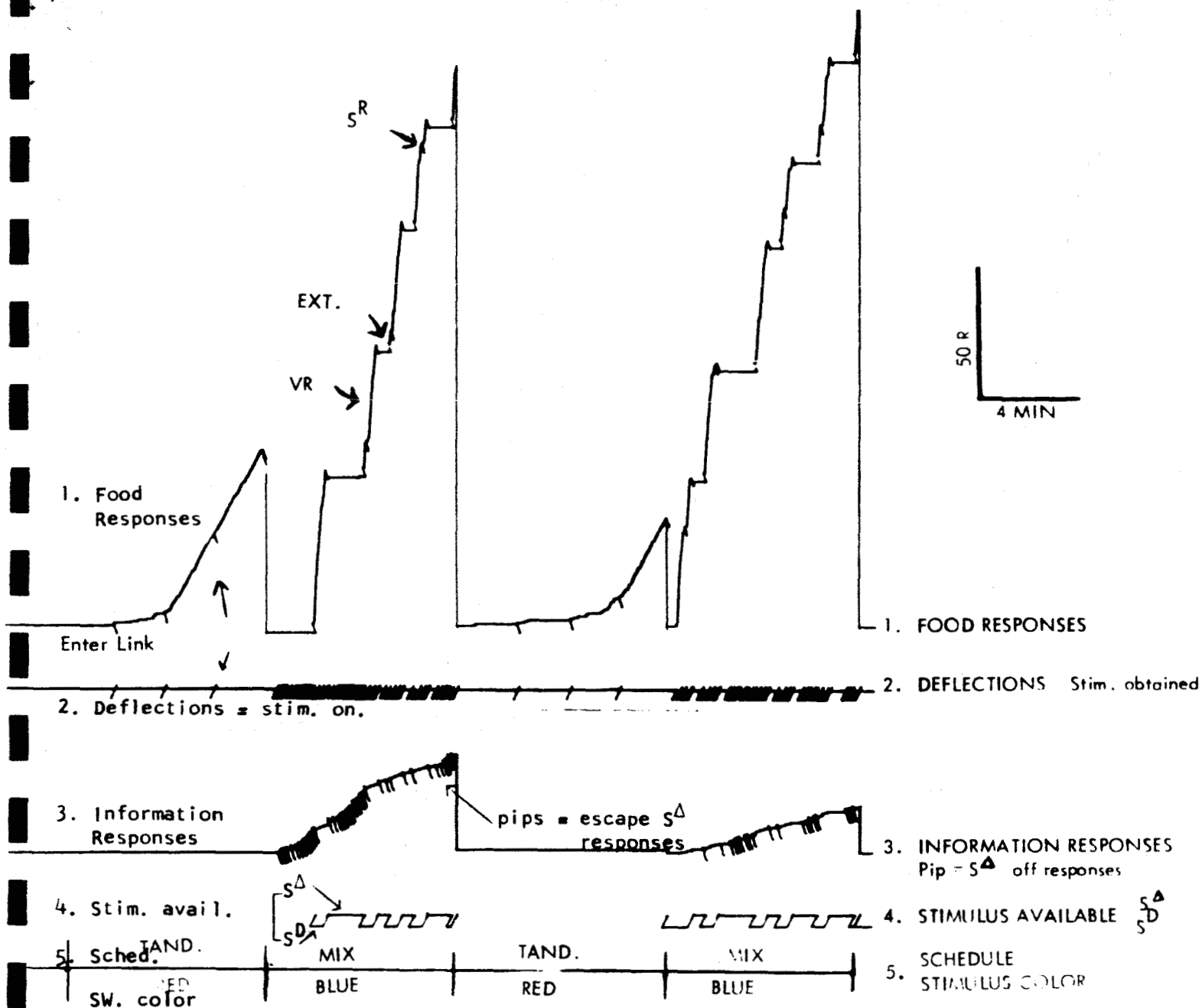


Fig. 3

Typical performance in Mult (Tandem FI 2 F I 2 FI 2 FI 2 Mix VR 100 Ext) With six different stimuli available, each correlated with a different component. One response on the information switch was required to turn on the informative stimulus, and another, different, response was required to turn it off. Lines 1 and 3 were made to vent to the baseline simultaneously.

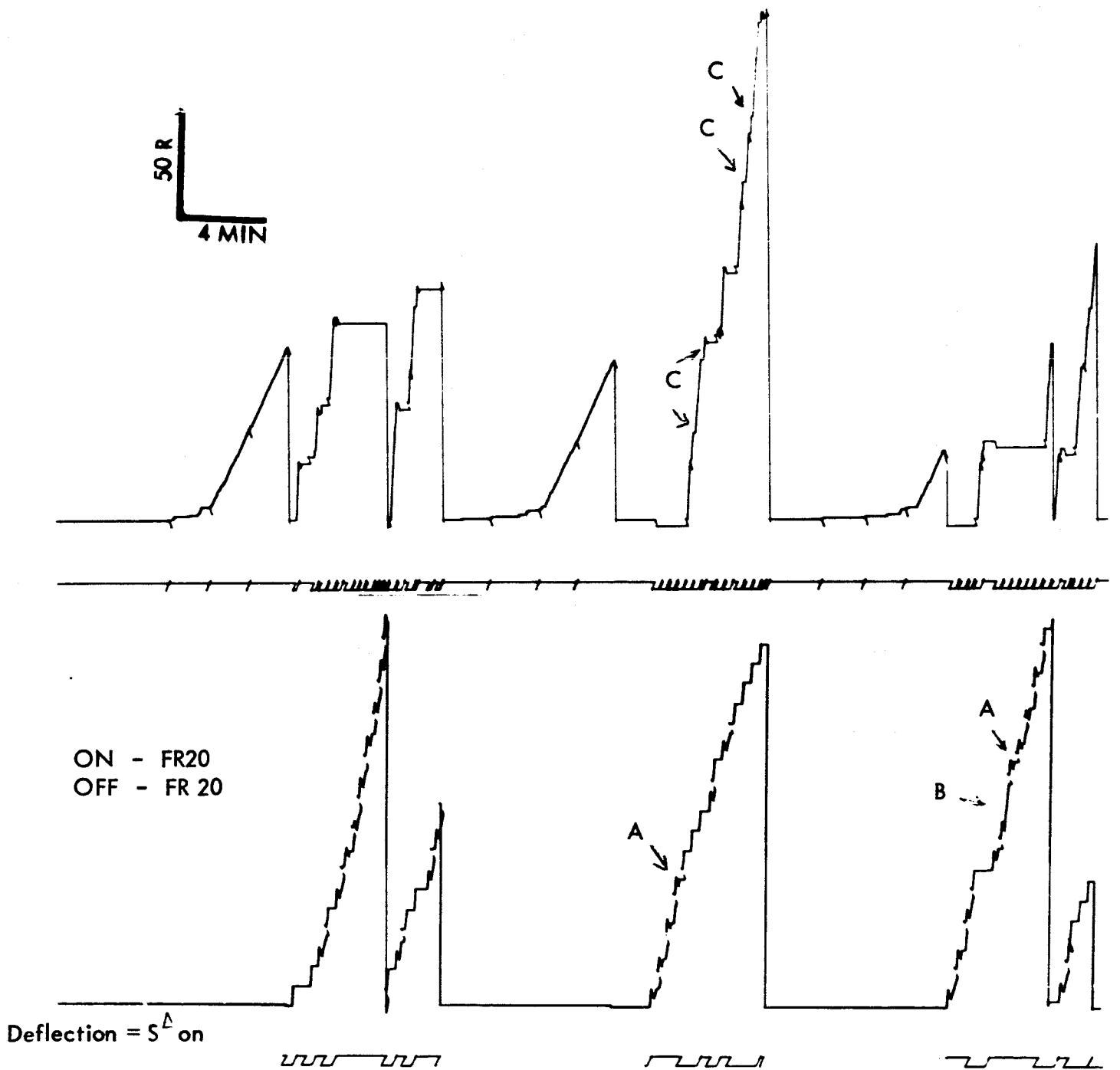


Fig. 4

See Fig. 3 and text. Twenty responses were required to turn on and to turn off the informative stimuli. The pen recording information responses remained deflected downwards for the duration of S^{Δ} . All information responses, those that turned on the stimuli, and those that turned them off, are recorded on line 3.

vigorously operated the information switch (on), which, of course, was without effect. An instance of this behavior appears at B. That the responses at B were ON responses rather than escape responses is shown by the fact that they occurred at a high rate, and that they were not effective in terminating S^{Δ} ; When S^D terminated, the chimp stopped responding on the food switch until S^{Δ} was again obtained. This behavior accounts for the pausing in the food switch responding, examples of which are shown at C.

To summarize: In the Mix schedule there is either a high rate or a zero rate and virtually continuous observation of S^{Δ} . After S^D comes on, no observing responses occur until it goes off 30 sec later. After S^{Δ} comes on, it is usually, but not always, turned off. After S^{Δ} is turned off, S^{Δ} is usually turned on again immediately. Note that S^D is never turned off, a result consistently obtained throughout the experiment. The observing behavior serves to keep the chimp more or less continuously informed of whether S^D or S^{Δ} is in effect, and the performance on the food switch is usually appropriate. With FR 20 S^{Δ} is turned on, but S^{Δ} is not so readily or so promptly turned off. We can make the straight forward inference that the positive reinforcing value of S^{Δ} is greater than the negative reinforcing value of S^{Δ} ; that is, elimination of the absence of S^D maintains more behavior than the elimination of S^{Δ} .

In order to test whether responding on the information switch was genuine observing behavior in the sense of being maintained specifically by the production of S^{Δ} , the information switch was made inoperative, and its function was transferred to another switch (I' in Fig. 1).

(The stimulus-termination function of the center switch was not transferred.) In a single session the chimp abandoned the original information switch in favor of the new one. She continued, however, to turn off S^{Δ} by pressing up the original information switch. The result confirms that the behavior was maintained by the production of the informative stimuli.

It is known that, except for extreme values, longer S^{Δ} s are more punishing (Ferster & Appel, 1961). On the other hand shorter S^D s are probably not less reinforcing. Accordingly, we predicted that shortening the duration of S^{Δ} would eliminate escape behavior but not observing behavior. We therefore changed the duration of S^{Δ} to 5 sec and required 5 responses to turn the informative stimuli on or off, which still allowed the stimulus to be terminated before it went off automatically. Fig. 5 shows a result representative of those obtained over several sessions. In Fig. 5 it can be seen that escape behavior has been eliminated, while observing behavior remains unimpaired.

Compare the performance on the information switch when S^D and S^{Δ} are available. Both S^D and S^{Δ} are produced repeatedly; as soon as the stimulus goes off it is turned on again. However, the slope of the cumulative record is the same whether S^D or S^{Δ} is obtained. If S^{Δ} had been turned off (5 responses) after it was obtained, the additional responses would appear on the record and cause the slope of the line to be steeper during S^{Δ} periods than during S^D periods. Instead the animal maintains S^{Δ} , works on the food switch so long as S^D is on, and, if S^{Δ} comes on, waits until it goes off before turning on S^{Δ} again.

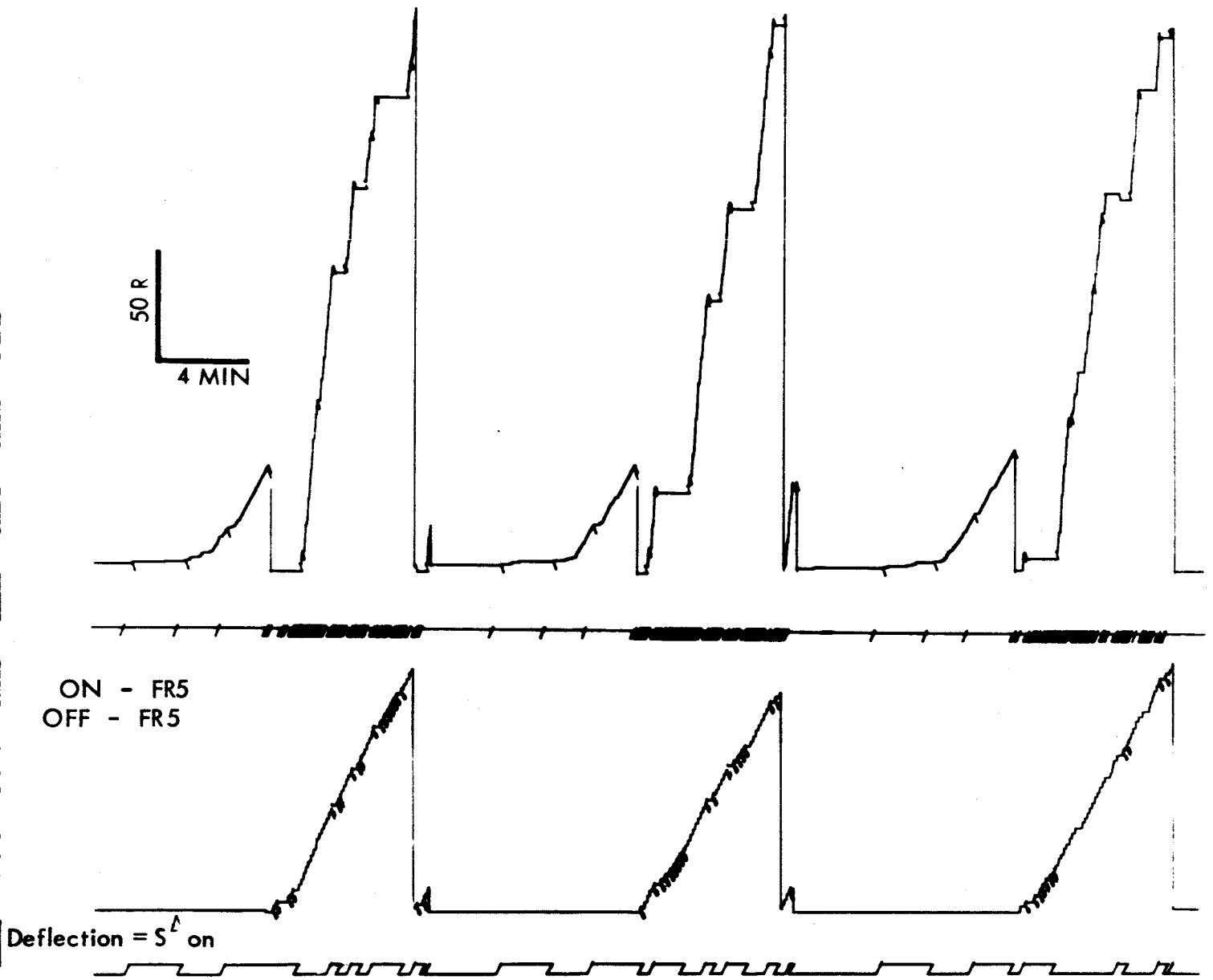


Fig. 5 See Fig. 4 and text. Five responses were required to turn on and to turn off the informative stimuli. The timed maximum duration of the informative stimuli was 5 sec.

Those committed to traditional theories of conditioned reinforcement could argue that observing behavior, though producing both S^D and S^Δ , is actually reinforced only by S^Δ . The arbitrariness of this view is obvious when we consider that the operations for establishing S^D are identically those for establishing S^Δ . There is no way, in the training situation, to distinguish reinforcing effects of S^D and S^Δ . Special tests for reinforcing value must be made. It is customary to test the reinforcing value of S^D , but not that of S^Δ , since it is assumed that S^Δ is simply aversive. Evidently, the best way to test for reinforcing value of S^D or S^Δ is to determine whether a new response can be learned using S^D or S^Δ as the reinforcer. However, this method, though the best available, has the drawback that extinction of conditioned reinforcement rapidly occurs, presumably because what the animal is learning is that the stimuli are not informative in the new situation. That is, the stimuli are not correlated differentially with reinforcement. The very conditions of the test preclude the possibility that the stimulus is informative since only one stimulus is used.

In the present experiments we sought to show that it was the differential relation of S^D and S^Δ to reinforcement that was the source of conditioned reinforcement. Rather than trying to show that traditional theories are wrong in their implication that only S^D is reinforcing, we tried to show that the claim is unnecessary and unenlightening compared to the insights provided by the information hypothesis.

Performance was stabilized under the following conditions: five responses were required to turn on $S^{D\Delta}$ and five responses were required to turn off $S^{D\Delta}$. The duration of $S^{D\Delta}$ was 20 sec. $S^{D\Delta}$ was changed with

a given probability (P) every 30 sec. Then P was systematically changed from 0.1 to 0.9, where the numbers refer to the probability of obtaining S^D . Five to ten successive sessions were given at all the selected values of P . Thus, when P was 0.1 the VR 100 schedule was in effect on the average only for one 30 sec period in every ten. When P was 0.9 VR 100 was in effect for nine 30 sec periods in every 10. $0.00 < P < 1.00$ defines the condition for obtaining both S^D and S^Δ , $P=0$ defines extinction and $P=1$ defines exclusive positive reinforcement (no Ext. periods). The outcome of the situation just described is not unequivocally predicted by any theory. However, if all the burden of maintaining performance is placed on S^D , it is fairly clear that stronger observing behavior should be obtained as P approaches 1.00. The information hypothesis suggests a different possibility. It is clear that as P approaches either 1.00 or 0.00 S^{Δ} becomes less significant to the animal in terms of reinforcement frequency. Moreover, by the usual measures, the amount of information contained in a 2-event system decreases with the rarity of one of the events. Therefore, it seems safe to say that the information hypothesis suggests that the conditioned reinforcing value of S^{Δ} will be greatest with P at middle values and least with P at extreme values.

Figure 6 shows typical performance with P at moderate values. The middle trace shows the maintenance of information responses $[R(I)]$ and escape responses. There are also several instances of $R(I)$ occurring during S^Δ , as at A. At B, for example, S^Δ terminated automatically and was rapidly turned on again by the continuing $R(I)$. This process might have been superstitiously maintaining the inappropriate $R(I)$.

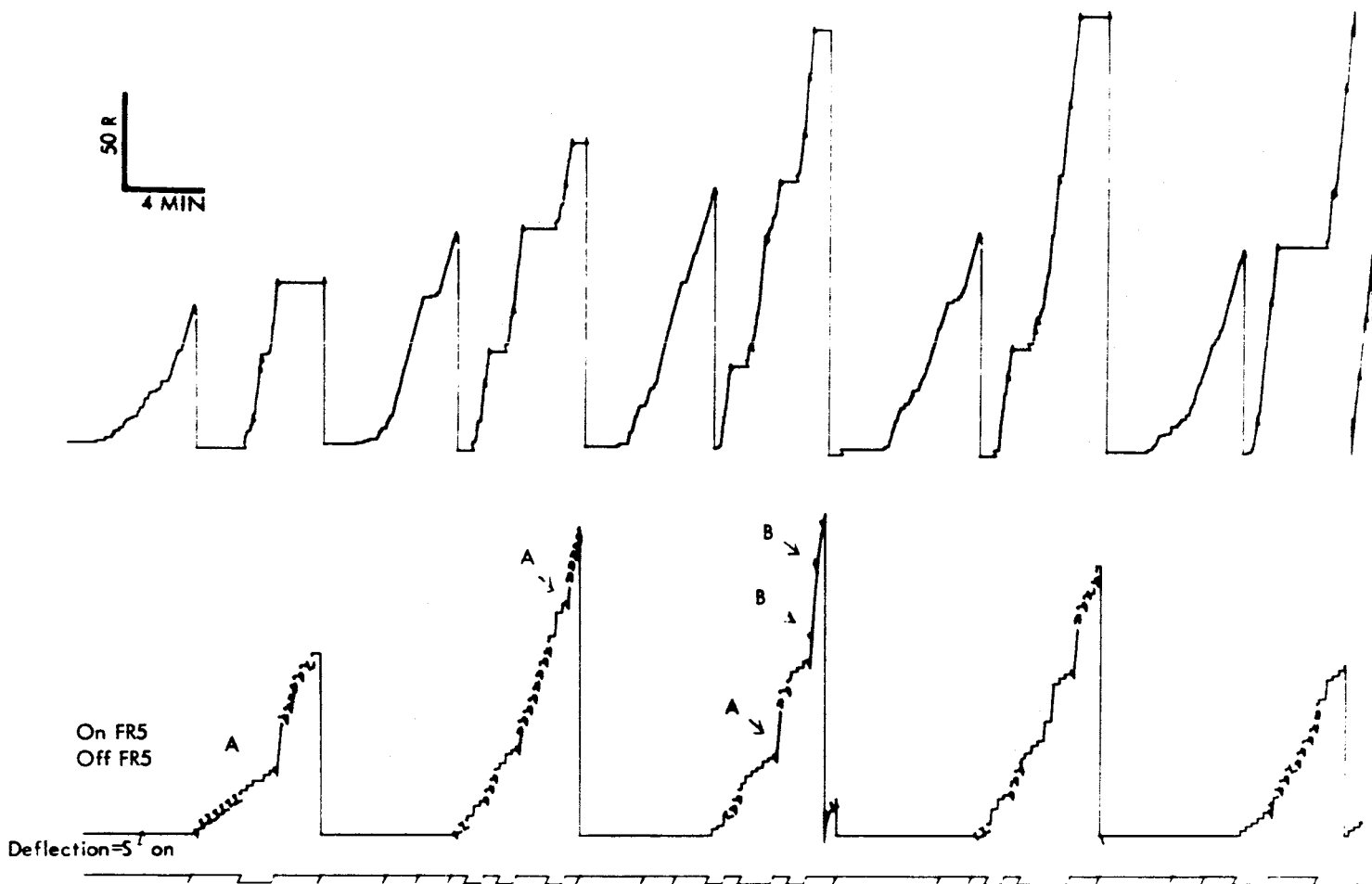


Fig. 6 See Fig. 4 and text. Five responses were required to turn on and to turn off the informative stimuli. The probability of VR 100 (rather than Ext) was about 0.5. The maximum duration of $S^{D/}$ was 20 sec.

Figure 7 shows the number of S^{DA} obtained per trip as a function of the relative frequency of S^D . The relative frequency of S^D varied continuously, so the median of the means over class intervals of 0.1 or 0.05 are plotted in Fig. 7. Had S^{DA} been on continuously, the results would have shown a straight line at 24 since an 8 min trip can contain 24 - 20 sec periods. However, the number of S^{DA} per trip can exceed 24 since S^A (and also S^D) could be turned off and on again. The steep increase in number of S^{DA} obtained at low values of P is due to the steep increase in durations of trips at low P values as a consequence of trips being prolonged until one reinforcement occurred. It should be noted that the large number of S^{DA} obtained at low P values were actually mostly S^A . When this is taken into account, the number of S^{DA} per min is roughly constant, and its absolute value shows that the animal more or less continuously maintained the informative stimuli.

A measure of response rate to turn on the stimuli was obtained by dividing the mean number of $R(I)$ per trip by the mean duration of absence of S^{DA} per trip. The medians of the mean numbers of $R(I)$ are shown in Fig. 8. The highest rates of responding occurred at intermediate values of P , not at the higher values. Since the stimuli were off for such a relatively small proportion of each trip, variations in response rate to turn the stimuli on had no detectable effect on the number of S^{DA} actually obtained (Fig. 7).

The aversiveness of S^A might be expected to vary according to its relative frequency. Fig. 9 shows the relative frequency of escaping S^A as a function of the relative frequency of obtaining S^A . The chimp was more likely to escape S^A when the probability of obtaining S^A was at intermediate values.

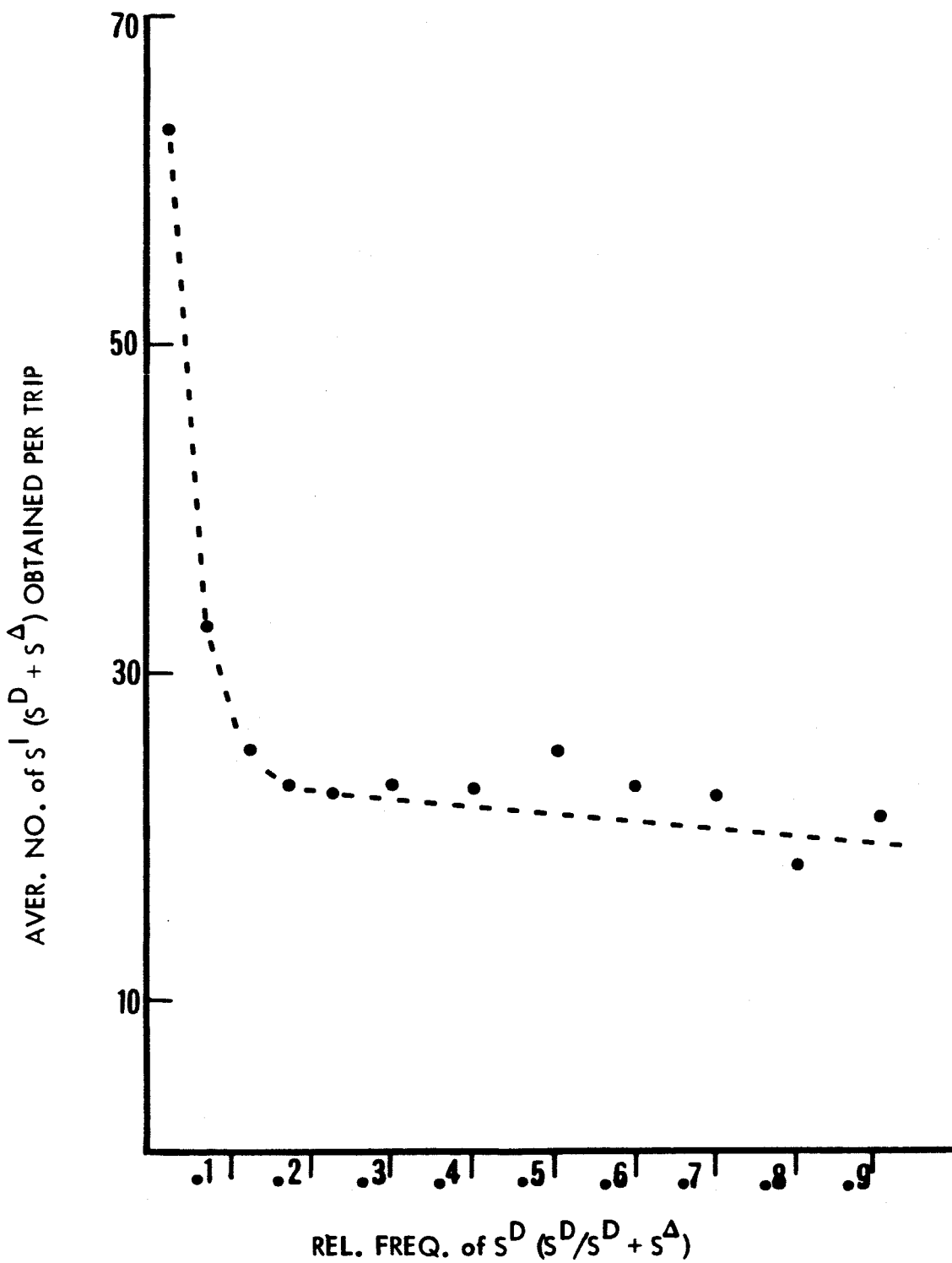


Fig. 7 Average number of information stimuli (S^D or S^A) obtained per trip as a function of the likelihood of obtaining S^D .

The relations shown in Fig. 6-9 represent rather complex behavior which is, nevertheless, quite orderly. We may summarize the observations as follows: when the blue light came on (Mix VR Ext), the animal first worked on the information switch to obtain S^{DA} . If she obtained S^D , she worked on the food switch. When S^D went off, she stopped working on the food switch and returned to the information switch. If she obtained S^Δ , she did not work on the food switch but often worked on the escape switch. Briefly, if S^D were on, she worked for food; if S^Δ were on, she worked to terminate it; if neither S^D and S^Δ were on, she worked to obtain them. She more readily turned on S^D (Fig. 8) and escaped S^Δ (Fig. 9) when the situation contained most uncertainty. She continued to produce S^Δ when S^D was very rare, and in these conditions she rarely escaped S^Δ .

The results provide some evidence that the reinforcing value of informative stimuli is positively related to the amount of uncertainty in their absence. Since the informative stimuli used in this experiment reduced the uncertainty to zero, the results may be described as showing that conditioned reinforcement is a direct function of amount of uncertainty-reduction.

Similarly, S^Δ is most aversive when it is obtained in a situation of most uncertainty. This is a surprising result, since escape of S^Δ puts the animal back into the situation of most uncertainty. One might expect that when S^Δ is very rare, the animal would have the strongest tendency to escape it, since subsequent responding for information would be more likely to produce S^D .

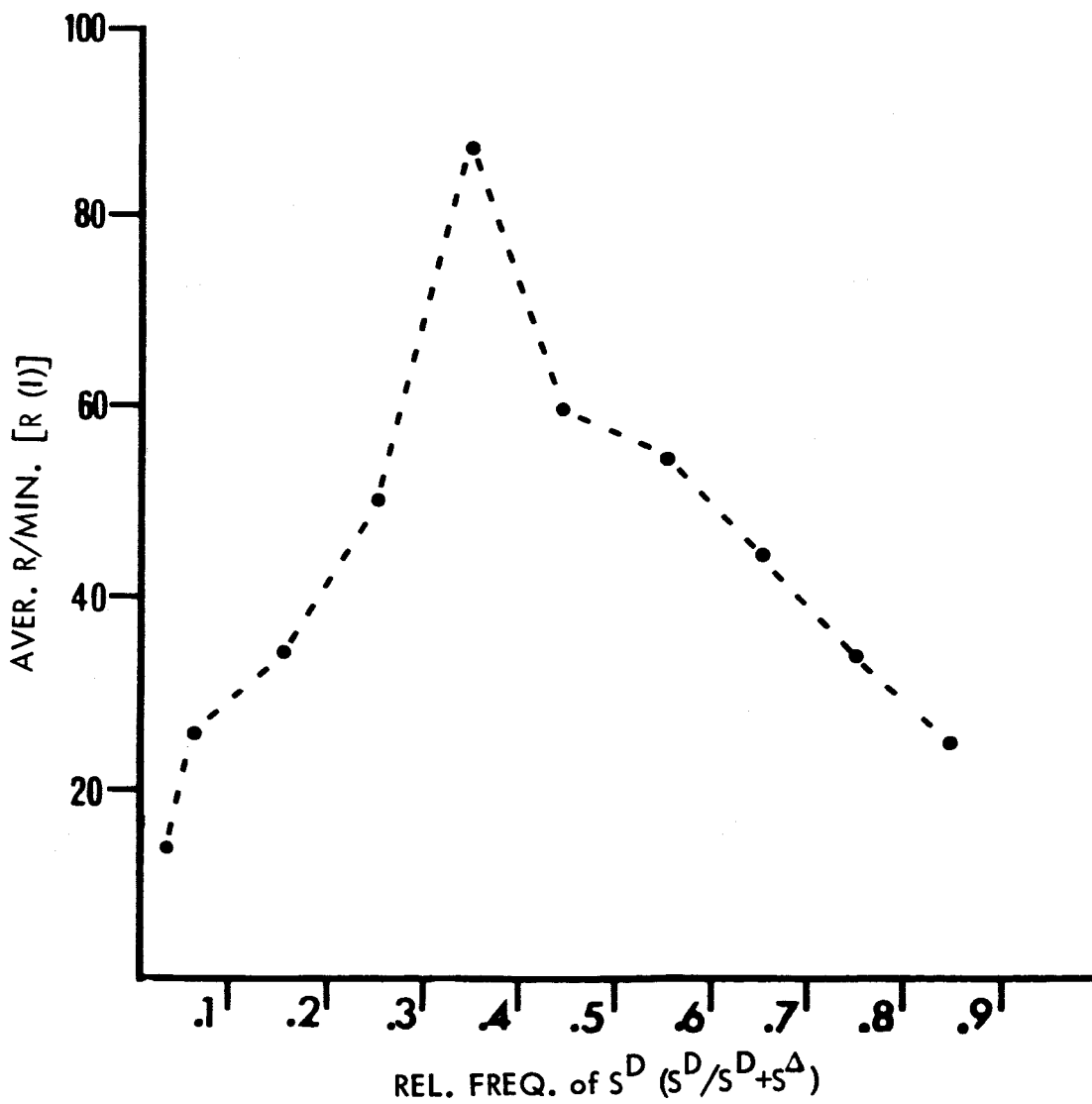


Fig. 8 Average number of information responses as a function of the likelihood of obtaining S^D .

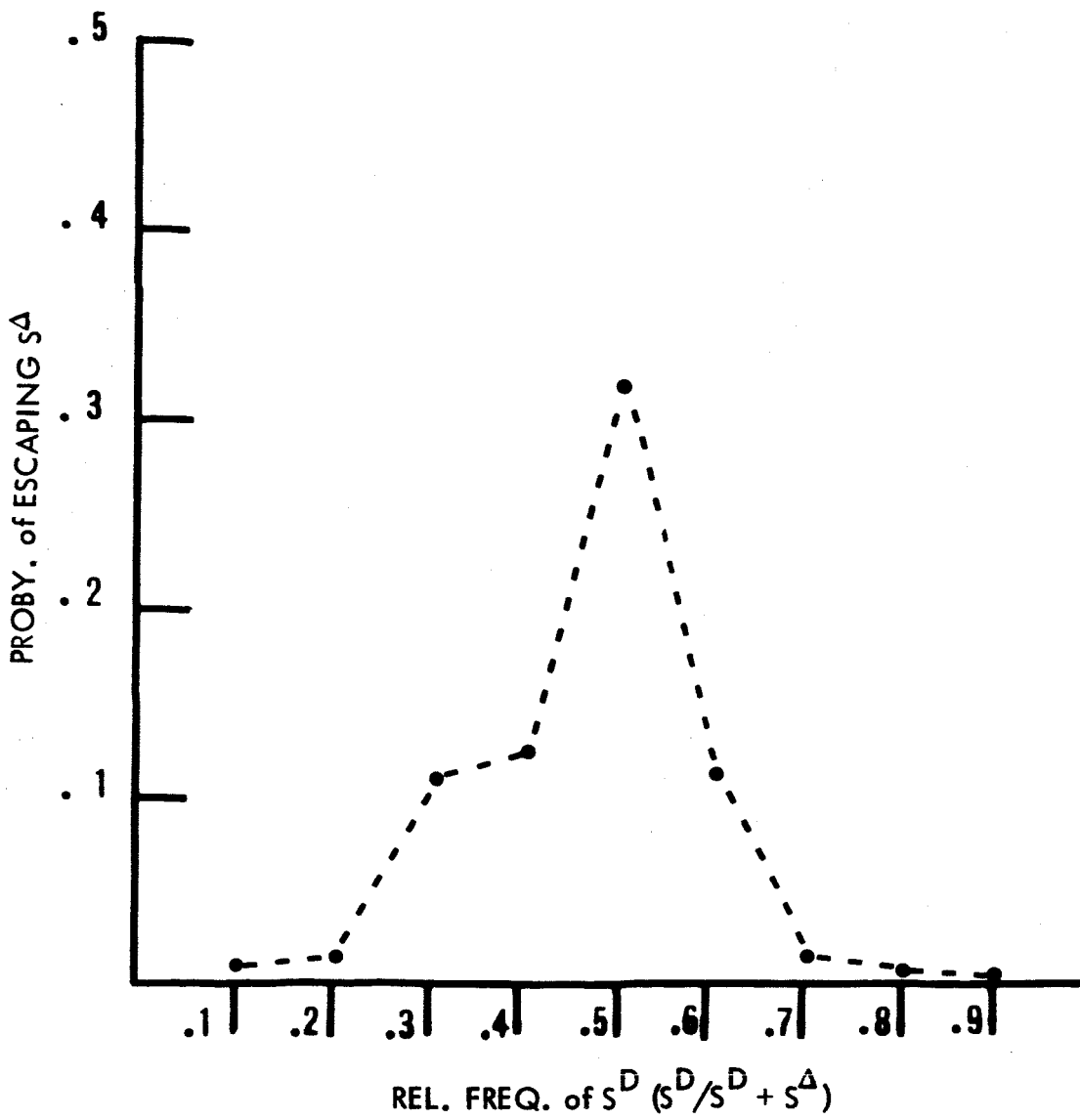


Fig. 9 Probability of escaping S^A as a function of the likelihood of obtaining S^D .

The systematic performance generated by the procedures used in this experiment and the relations between variables which it revealed encourage further exploration of the guiding hypothesis. While some of the results might be explained by traditional theories of conditioned reinforcement, the results as a whole could not be explained without considerable strain and the invention of several ad hoc principles.

References

1. Perlyne, D.E. Conflict, arousal and curiosity. McGraw Hill: New York, 1960.
2. Dillow, P.V. Reinforcing and aversive effects of stimuli in extended chained fixed-interval schedules. Unpublished dissertation, Durham University, 1964.
3. Ferster, C.B. Control of behavior in chimpanzees and pigeons by time-out from positive reinforcement. Psych. Monogr., 1958, 72, No. 8, Whole No. 461.
4. Ferster, C.B. & Appel, J.B. Punishment of S^{Δ} responding in matching-to-sample by time-out from positive reinforcement. J. exp. Anal. Behav., 1961, 4, 45.
5. Hull, C.L. Principles of behavior. Appleton-Century: New York, 1943.
6. Hull, C.L. A behavior system. Yale University: New Haven, 1952.
7. Kelleher, R.T. Stimulus-producing responses in chimpanzees. J. exp. Anal. Behav., 1958, 1, 87.
8. Kelleher, R.T. & Gollub, L.R. A review of positive conditioned reinforcement. J. exp. Anal. Behav., 1962, 5, 543.
9. Kimble, G.A. Conditioning and learning. Appleton-Century-Crofts: New York, 1961.
10. Skinner, B.F. The behavior of organisms. Appleton-Century: New York, 1938.