EFFECTS OF RANDOM REINFORCEMENT SEQUENCES[1]

M. J. MORGAN

UNIVERSITY OF CAMBRIDGE

Rats were exposed to a random sequence of reinforcement on two levers, such that there was no way to predict from the previous reinforcement which lever would deliver reinforcement next. The rats showed a tendency to repeat the choice that had just produced reinforcement, despite the absence of an overall contingency that differentially reinforced such repetition. However, this tendency decreased with continued exposure to the schedule. Runs of successive reinforcements on a lever increased the probability of pressing that lever, but only slightly, and only in the earlier phases of training. The more quickly a press was made after reinforcement the more likely it was to be on the lever that had delivered that reinforcement. Repetition of choice followed by reinforcement should be viewed as a naturally occurring behavior in the rat, but not necessarily as a behavior that will continue without differential reinforcement of repetition.

In a choice situation there are two or more behavioral alternatives, one of which usually receives differential reinforcement with respect to the others. What is the behavioral unit controlled by reinforcement in these circumstances? The simplest view is that it is the topographically defined operant, for example, a peck on the left key. But it is a truism that other descriptions of a particular "left-key response" are possible, in principle an infinite number. It could, for instance, be described as a peck on the key that gave the preceding reinforcement. To find out which of the potential descriptions are valuable in the analysis of behavior, one must try them out and see which are productive (Skinner, 1938). The present study concentrated upon descriptions of choice responding in terms of response repetition tendencies.

Shimp (1966) studied the repetition behavior of pigeons with a schedule that reinforced left-key pecks with a higher probability than right-key pecks. He reported a tendency of the birds to peck on the just-reinforced key, even though there was no differential reinforcement for such repetition behavior. Williams (1972) examined behavior of pigeons with a schedule in which repetition did receive differential reinforcement. The optimal strategy of pecking on the just-reinforced key was learned by the birds when the reinforcement probability for repetition was 0.80, but not when it was 0.65. The reverse of this contingency occurs on concurrent-interval schedules where, as Shimp (1966) pointed out, reinforcement probability on a key increases with time since a peck on that key was reinforced. This may explain why Killeen (1970) found the reverse of repetition (response alternation) in a concurrent interval schedule. Presumably, in such a schedule the contingency favoring alternation acts in opposition to the repetition tendency described by Shimp. Within the error of an experiment the two effects could conceivably cancel out, which may explain why Nevin (1969) found neither repetition nor alternation in a concurrent-interval schedule.

Interpretation of experiments in which repetition is reinforced would be easier if one knew more about the strength of the repetition tendency in the "baseline" condition, when repetition is neither encouraged nor discouraged by the schedule. This was the case in Shimp's experiment, but his data are complicated by the fact that the different key pecks were reinforced with different probability. A simpler case would seem to be one in which

the overall probability of reinforcement for the two operants is the same, as in Williams' study, and in which the probability of reinforcement on a key is independent of what happened on the last reinforcement, as in Shimp's study. This was the aim of the present experiment. Rats were trained in a two-lever situation with a random sequence of reinforcement distributed between the two levers. There was thus no differential reinforcement for repetition of just-reinforced choices, or for the reverse. To prevent "absorption", that is, an exclusive preference for one or other of the levers, the rat had to collect a reinforcement made available on one lever before proceeding to the next reinforcement. This is similar to the technique described by Menlove, Moffitt, and Shimp (1973), except that they used a "sampling without replacement" procedure to ensure that equal numbers of reinforcements were given on the two keys within each session. Since this procedure involves potentially undesirable nonrandom sequences as each session draws to a close, the present experiment used a "nonreplacement" procedure. In other words, the sequence of reinforcements was entirely unpredictable, without artificial constraints.

Each reinforcement was delivered after a random number of presses on the lever that had been selected to give the next reinforcement, a press on that lever having a stationary probability of success (random-ratio schedule). Presses on the other lever were ineffective. Once reinforcement had been obtained, the next lever to deliver reinforcement was randomly determined by a computer routine.

METHOD

Subjects

Ten experimentally naive male hooded rats were obtained from Animal Suppliers Ltd. and caged in groups of four and three.

Apparatus

The experimental chamber was a Campden Instruments CI-410 two-lever rat station fitted with a magazine (CI-442) delivering 45-mg Noyes Standard Diet food pellets. In this chamber, pellets are delivered into a recess between the two levers, in front of which is a panel (about 5 cm wide), hinged at the top, that the rat must open to collect pellets. The levers (each 4 cm wide, placed with centers 13 cm apart), which required just over 10 g (0.10 N) for operation, operated sealed-reed switches; operation of the panel activated a microswitch. The food recess could be illuminated by a 2.8-W bulb; general illumination of the compartment was provided by a 2.8-W bulb on the ceiling. The chamber was enclosed in a sound-resistant housing (CI-412) provided with an extractor fan. Scheduling apparatus was located in a separate room. White masking noise was played inside the experimental room.

Scheduling and data collection were carried out by a CTL Modular 1 computer, using the on-line language ONLI developed by C. Crook and S. E. G. Lea in the Cambridge Psychological Laboratory. Data were subsequently analyzed by the Cambridge University "Titan" computer.

Terminology

Each of the variable number of lever presses made between reinforcements can occur either on the lever that delivered the last reinforcement, or on the other lever. Unfortunately, there is no existing terminology for these two classes of response, the terms "stays" and "shifts" having been pre-empted to mean repetitions and changeovers of the preceding response, not of the preceding *reinforced* response. (This distinction is vital for present purposes.) In the present account, a repetition of the response that was reinforced on the preceding trial is called a REPR; repetition of the response not reinforced on the preceding trial is called a REPU. A repetition of the immediately preceding response is a STAY; a changeover is a SHIFT.

As an example, consider the following sequence of left-(A) and right-(B) lever responses and reinforcements (R); AR AABBR ABBR. In this sequence, there are four REPRs and three REPUs (the first response is not assigned to either class); the overall REPR probability is 0.57. There are four STAYS and three SHIFTS. The position of a response after reinforcement is called its ordinal position (N). Thus, in this example the REPR probability in the first ordinal position (N = 1) is found to be 0.5; in the second position (N = 2) it is unity.

The number of successive reinforcements (M) for pressing a lever refers to a run of reinforcements of length M occurring on one lever without intervening reinforcements on the other lever. In this experiment, the probability of a run of length M is twice as great as that of a run of length M + 1.

Procedure

After magazine training with the levers absent from the compartment, the rats were trained to press the left lever on a continuous schedule of reinforcement (a pellet for every press), and subsequently the right lever (with only one lever present at a time). Both levers were then introduced, with the following schedule in operation from the outset of training. After each reinforcement, a random-number routine in the computer determined, without constraint upon the number of successive reinforcements on one lever, which lever would deliver the next reinforcement. Left and right levers were selected with equal probability. Presses on the selected lever delivered a reinforcement with probability of 0.5 (random ratio with mean of two responses per reinforcement: RR 2) without constraint upon the number of presses made before reinforcement. The overall probability of reinforcement for a lever press on this schedule, given random distribution of lever presses between the two levers, is 0.25. After each reinforcement the food recess light came on, and the houselight went off. This stimulus condition remained in operation until the rat had pushed open the panel in front of the food recess and subsequently had withdrawn its head, thus returning the microswitch attached to the panel to its closed position. When the houselight was off, lever presses were not recorded; thus, the post-reinforcement pause, or time elapsing between a reinforced press and the succeeding press, necessarily includes the time during which the rat has its head in the food recess.

In Sessions 1 to 18 on this schedule, the rats were on a free-feeding diet in their home cages; in Sessions 19 to 32 they were fed for 2 hr daily at the end of each 20-min testing session. In block 33 to 46 this deprivation condition continued, and the probability of reinforcement on the correct lever was lowered from 0.5 to 0.25 (thus, the schedule became RR 4).

RESULTS

The tendency to repeat responses that had been associated with the preceding reinforcement (REPRs) and the tendency to repeat the last response made (STAYs) are shown session-by-session in Figure 1. The figure reveals that the probability of both STAYs and REPRs was considerably greater than 0.5 in early sessions, and that it declined as testing proceeded to an asymptote that was still consistently greater than 0.5. The probability of STAYs, but not REPRs, increased slightly in the second block of sessions, when deprivation was in effect, and increased still further in the third block when the reinforcement probability was lowered.

Thus, the analysis reveals that the rats did indeed tend to press the lever associated with
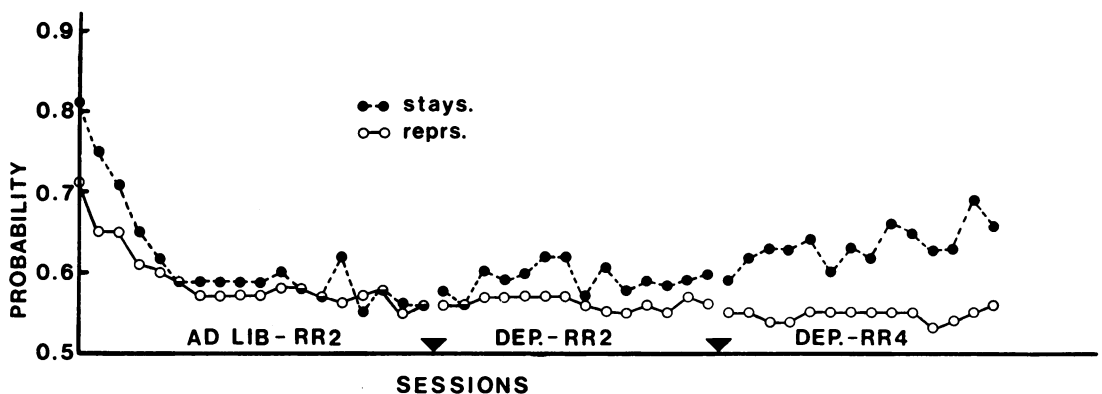

Fig. 1. Session-by-session changes in repetition of the preceding lever press (STAYs) and repetition of the just-reinforced lever press (REPRs) expressed as a fraction of all lever presses. Data are means over 10 rats. Changes in conditions, as labelled on the abscissa, occur at the vertically pointing arrows.

the preceding reinforcement, in that the probability of REPRs was greater than 0.5. This was despite the absence of a differential reinforcement contingency favoring this behavior. However, since the rats also showed a tendency to repeat all responses, as demonstrated by the data for STAYs, the question arises whether reinforcement in fact had any selective effect. A little thought will show that this reduces to the question: was the probability of a STAY greater when the preceding response was reinforced than when it was unreinforced? In other words, to answer the question we need to look for changes in the probability of a STAY with ordinal position of the response after reinforcement. This analysis is presented in Figure 2.

The most important fact seen in Figure 2 is that the probability of a STAY is higher for the first response after reinforcement than for any others. This decisively demonstrates the temporary "strengthening" effect of reinforcement; nor is this a trivial eliciting effect of reinforcement, for it will be noted that the effect of reinforcement in immediate terms is to remove the rat from the vicinity of the bar to that of the food magazine. So much for the first ordinal position. In the second and third positions, on the contrary, it will be seen that the probability of a repetition was lower than at any other time. What this means is that if

the first response after reinforcement is not reinforced, the rat will tend to SHIFT to the other lever; this shifting tendency is greater than at later ordinal positions.

The data for REPRs, shown in the right-hand side of Figure 2 are in broad agreement with these conclusions. The fall in probability of a REPR below 0.5 in the few ordinal positions after the first is to be noted and compared to the dip in the curve for STAYs. The approximation of the curves to an asymptote of 0.5 in later ordinal positions is clearly to be expected if the effects of the preceding reinforcement determine behavior to a diminishing degree, as the number of intervening unreinforced responses increases.

Concerning differences between blocks of sessions, the clearest fact evidenced by Figure 2 is that the probability of STAYs was greater in the final block of sessions than in the previous two, this being true at all ordinal positions after reinforcement. Further light on this fact is shed by Figure 3, which presents the session-by-session changes in the probability of a response repetition immediately after reinforcement, and the postreinforcement pauses associated with repetitions (REPRs) and nonrepetitions (REPUs). It will be seen that deprivation, as would be expected, decreased the postreinforcement pause, as did the lowering of reinforcement probability in
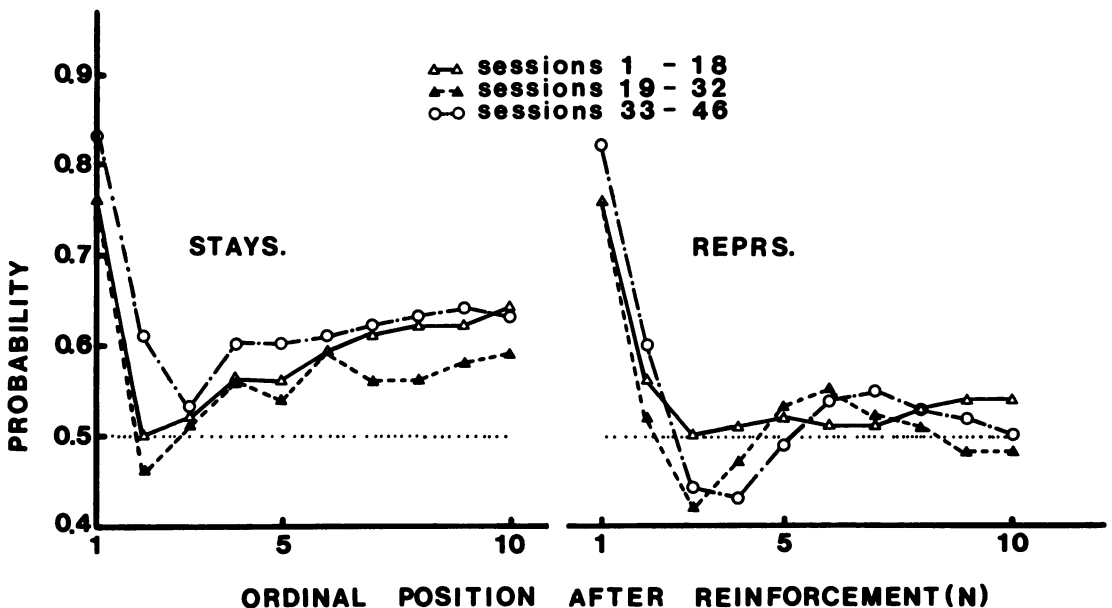

Fig. 2. The probabilities of STAYs (left-hand figure) and of REPRs (right-hand figure) at different ordinal positions after reinforcement, plotted separately over three blocks of sessions. Results are the mean for 10 rats.

the third block of sessions (33 to 46); also, both during deprivation and as a result of lowered reinforcement probability, the repetition tendency in the first ordinal position increased. This suggests an association between repetition and fast responding after reinforcement, and the existence of such an association is further suggested by the fact, also apparent from Figure 3, that the postreinforcement pause of REPRs tended to be smaller than that of REPUs.

The association of repetitions with shorter pauses is further illustrated in the individual data (Figure 4). This shows, for the final block of sessions only, the temporal distribution of pauses terminated by a REPR, separately from the distribution of pauses terminated by a REPU. Each point represents the percentage probability (abscissa) that the length of postreinforcement pause was smaller than or equal to the value on the ordinate. It will be apparent that in general, the white points (REPUs) lie above the black (REPRs); therefore it follows that pauses terminated by REPRs tended to be shorter.

The same result held inside other blocks as well. Thus, the number of rats showing a smaller median pause for REPRs than for REPUs was nine (out of 10) in the first block, and again nine in the second block.

The next question is why the probability of STAYs and REPRs declined during early exposure to the schedule (Figure 1). Was this decline due to an unintended but nevertheless real differential reinforcement contingency operating in favor of SHIFTs and REPUs? It will be recalled that the intention was to eliminate all differential reinforcement from the situation, but it can be argued that this aim was not completely achieved, as the following analysis shows.

When a lever is designated "incorrect", it remains so until at least the next reinforcement. Thus, to continue pressing on a lever in the absence of reinforcement is an inefficient strategy. Every time a lever is pressed and does not deliver reinforcement may be considered "information" that the lever is incorrect this trial. The best strategy is to press the lever that has been pressed least since the preceding reinforcement, in other words, to alternate responses until reinforcement is obtained (Lea and Morgan, 1972). The effect will be greater the higher the probability of reinforcement on
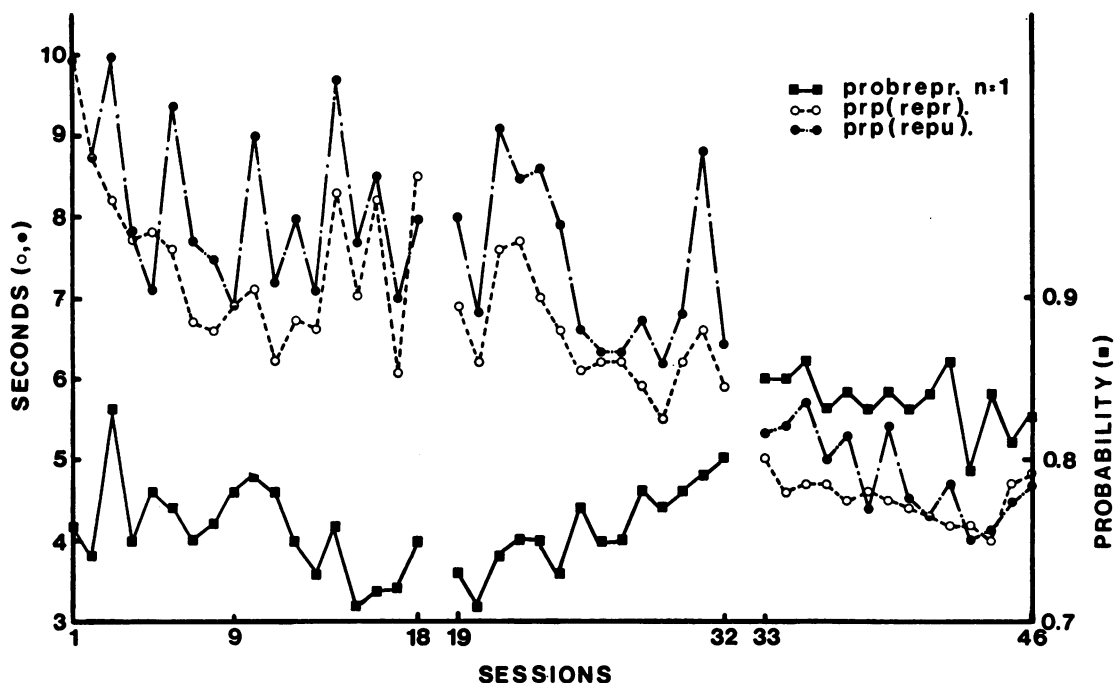

Fig. 3. Session-by-session changes in the probability of a REPR in the first ordinal position, and the median postreinforcement pauses (PRP) of REPRs and REPUs. Changes in conditions occur at the breaks in the curves. The probabilities are means of 10 animals; the PRP's are medians of the combined distributions of the 10 rats.
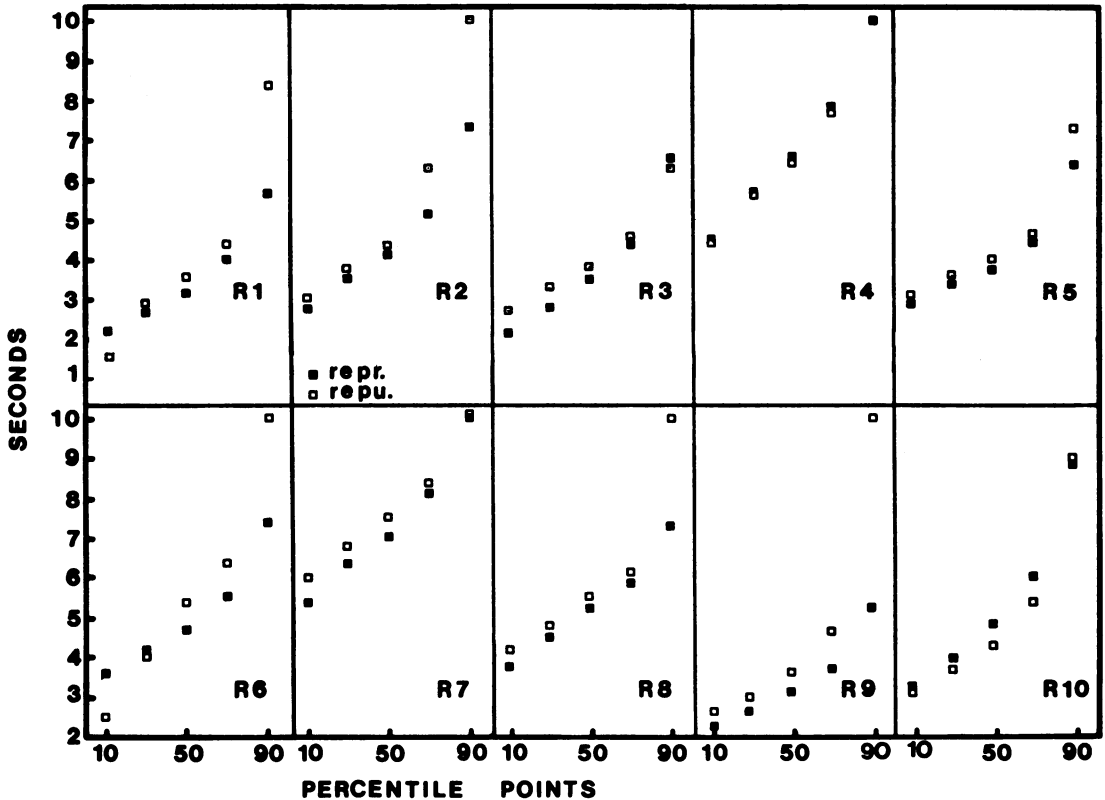
Fig. 4. Distribution of postreinforcement pauses in Sessions 33 to 46, shown separately for lever presses that were the same as the just-reinforced lever press (REPR) and those that were different (REPU). Data are shown separately for each of the 10 rats. The data points represent the percentage of times (abscissa) that were smaller than or equal to the time on the ordinate. Note that with the exception of Rat 4 and Rat 10, pauses for repetitions (REPRs) are smaller than for nonrepetitions (REPUs), particularly when longer pauses are considered.

the "correct" lever. If the probability is unity, it is inefficient to make more than one response without reinforcement. The "payoff" for exact alternation may be calculated as follows. If p is the (stationary) probability of reinforcement on the "correct" lever, then the probability of reinforcement for the first of two different responses is $0.5p$; the probability for the second response is also $0.5p$; thus, the probability of at least one of the responses being reinforced is p. Thus, the mean number of *pairs* of responses before reinforcement is $1/p$, and the mean number of responses is $2/p$. However, in half the cases on the average, a "pair" finishes after a single response, so the mean responses per reinforcement is actually $2/p-0.5$, which is 0.5 less than the mean expected number given purely random choice. For example, on the RR 2 schedule, the mean expected number given random choice is 4.0, and given alternation 3.5.

The relevant data are shown in Figures 5 and 6. First of all, it will be seen from the session-by-session data on response/reinforcement ratios in Figure 5, that performance was inefficient in the first few sessions; the animals obtained on average fewer than one reinforcement for every four presses. As testing proceeded, the ratio came nearer to, or even slightly improved upon, that expected from random choice. In the third block of sessions, when reinforcement probability was decreased by putting RR 4 into effect, the number of reinforcements obtained was slightly lower than that expected from random choice.

Further information on this is given in Figure 6, which presents the individual data relating the probability of a STAY to the ordinal position of response after reinforcement. Printed under the number identifying each rat are figures showing for that rat the mean number of responses per reinforcement during Ses-
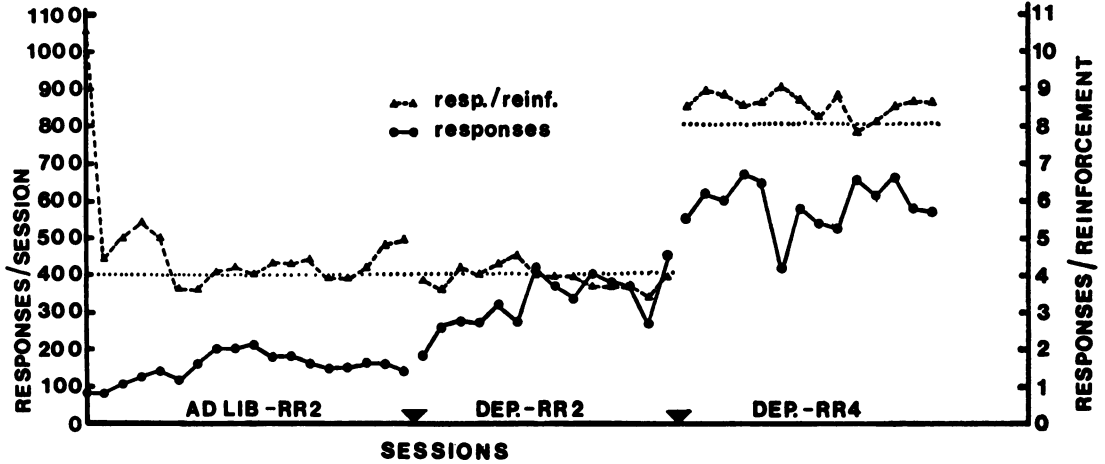
Fig. 5. Session-by-session changes in response rate (solid line and left-hand ordinate) and in the ratio of responses to reinforcements (dotted line and right-hand ordinate). Data are means over 10 rats.
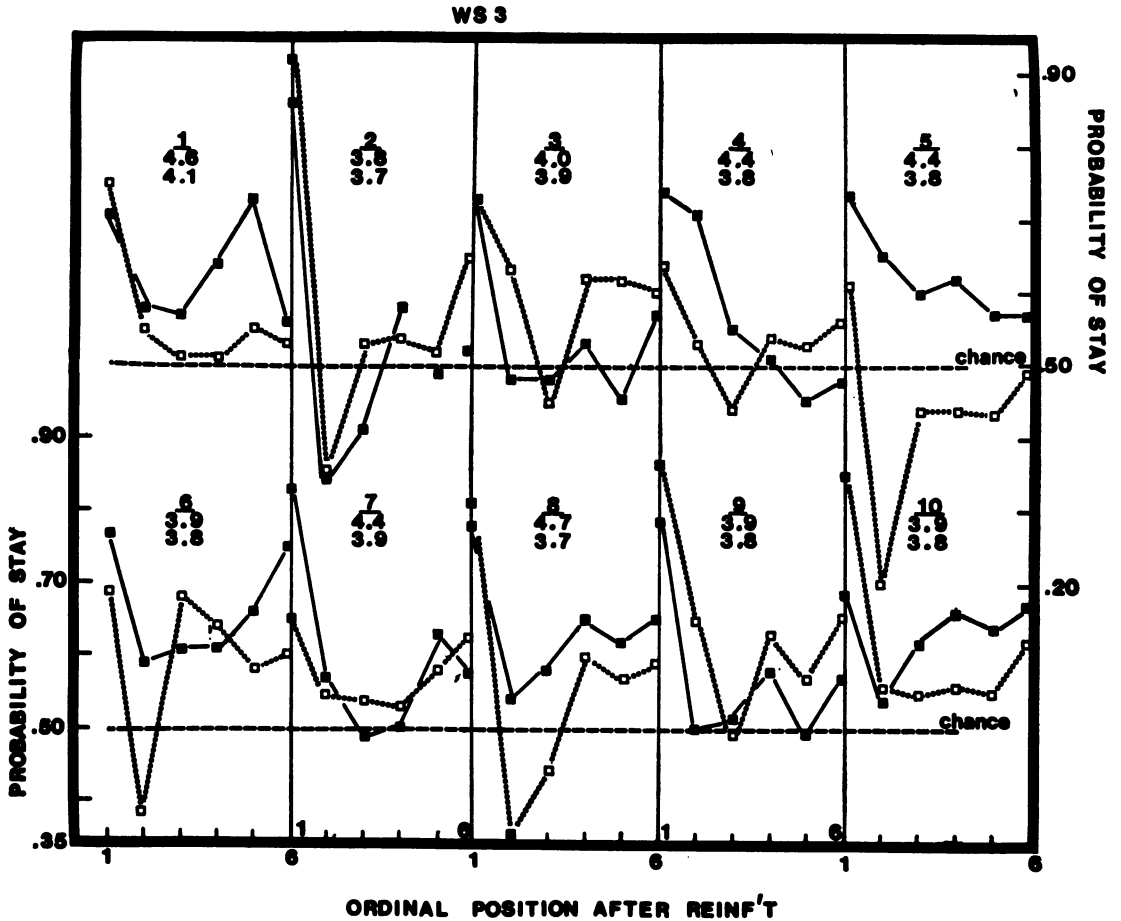
Fig. 6. Individual data showing the probability of a STAY at each ordinal position after reinforcement. Each pair of curves is from one rat, identified by the underlined numbers above the curves. The solid lines represent data from Sessions 1 to 10; the broken lines from Sessions 19 to 32. Below the number identifying each rat are two figures, indicating the average ratio of responses to reinforcements in Sessions 1 to 10 and 19 to 32. The right-hand ordinate refers to Rats 1 to 5 (top row), the left-hand ordinate to Rats 6 to 10 (bottom row).

sions 1 to 10 (top figure) and the mean number during Sessions 19 to 32 (bottom figure). Every rat shows a lower figure, that is, a more favorable ratio, in the later block of sessions. Figure 6 contains the data that allow one to test the hypothesis that improvements in the ratio are accompanied by decreases in the probability of "staying". If the argument presented above is correct, we expect better-than-chance performance to be correlated with an alternation tendency. That is, an increase in reinforcements per lever press should be associated with a decrease in the probability of a STAY in the second and subsequent ordinal positions. In Figure 6, the improvement in responses per reinforcement for each rat can be read from the numbers above the curves, and the changes in probability of a STAY are seen in the curves themselves. It will be apparent that there is in fact an association between responses per reinforcement and an alternation tendency. In particular, the two rats that achieved the best payoff (3.7 responses per reinforcement) both showed above-chance alternation at the second ordinal position. Rat 1, with the worst payoff, does not alternate at any ordinal position. However, there are some puzzling features of the data, particularly for Rat 10 (R10), which show little evidence for alternation, but above-chance payoffs. Perhaps this kind of variability in the data is the penalty one pays for deliberately surrendering control to random-number generators.

It may seem a problem that the rats achieve better-than-chance reinforcement payoffs, even though the overall probability of a STAY was greater than 0.5. The reason is that overall STAY probability is less important than the probability in the second and third positions, where the probability of a STAY was often lower than 0.5. These positions are especially important because the probability of a STAY in the first position is irrelevant, and the later positions exert a progressively diminishing influence because the rat is less likely to reach them. Thus, the critical fact in interpreting the better-than-chance reinforcement payoff is that in Sessions 19 to 32 there was significant alternation in eight of the rats at either the second or third ordinal positions.

In block 33 to 46 the rats obtained slightly fewer than the chance level of reinforcements. The difference from the RR 2 condition probably arises as follows: (1) the probability of

STAYs was greater in the RR 4 condition, and this as we have seen adversely affects payoff; (2) the relative numbers of responses at later ordinal positions rises on the "leaner" schedule. And since the probability of a STAY tends to be greater in these positions than in earlier ones (except for the first, which is irrelevant) the effect will be to increase the adverse effects of STAYING behavior.

In the random schedule used here, it will happen sometimes by chance that a long string of successive reinforcements will occur on the same lever. Since it has been shown that a single reinforcement strengthens the repetition tendency, it seems natural to suppose that runs of reinforcements will increase that tendency still further. Figure 7 shows, however, that the effect is surprisingly small, and in certain rats nonexistent. There is a suggestion that the effect is more pronounced in early sessions (1 to 10) than after exposure to the schedule (Sessions 11 to 19).

DISCUSSION

The data may be summarized as follows. Given a choice between the two levers with randomly distributed reinforcements, the rats showed a significant tendency to repeat previous responses (STAYs), and this tendency was greater if the previous response was reinforced. The probability of repeating a reinforced response (REPR) decreased with time since the reinforcement. Both the overall tendency towards response repetition, and the increase in this tendency caused by reinforcement, were increased by food deprivation. They were also increased by reducing the probability of reinforcement. The reinforcement "payoff" (number of responses per reinforcement) was poorer than would be expected by chance in early sessions, and slightly better than chance after exposure to the schedule. Runs of reinforcement on the same lever had little effect on REPR probability, and such effect as was seen was more pronounced in early sessions of training.

The decline in STAYs with exposure to the schedule can be explained by the contingency favoring alternation, which had been arranged to prevent absorption on one or other lever. This interpretation is in agreement with the observed changes in reinforcement "payoff".
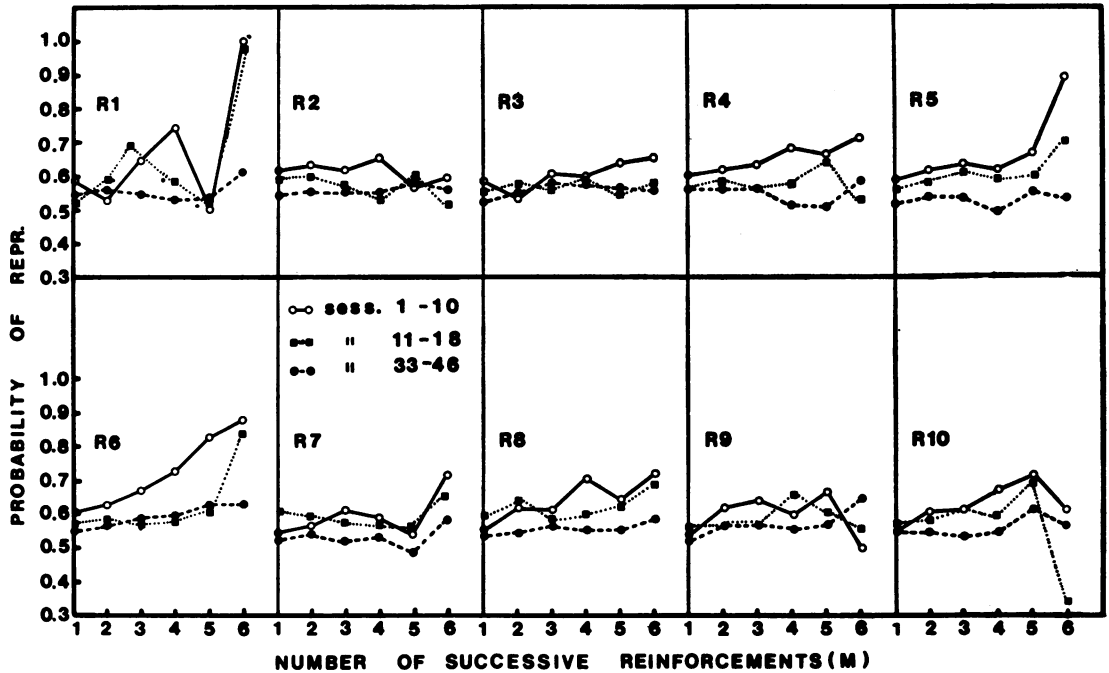
Fig. 7. Individual data showing the probability of a press on the lever associated with the preceding reinforcement, as a function of M, the number of successive reinforcements that have occurred on that lever. Data are shown separately for each of three blocks of sessions. Rises in REPR probability with M are most obvious in the early sessions.

Changes in REPR probability with variation in deprivation and scheduled reinforcement probability can be explained by the influence of these variables upon response rate. The fact that increased REPR probability at the first ordinal position after reinforcement was associated with increases in response rate can further be related to the finding that the more quickly a response is made after reinforcement, the more likely it is to be a REPR. A possible suggestion from these data is that the "strengthening" effects of reinforcement in some sense decay with time since the reinforcement. The word "decay" should be thought of here only in a descriptive sense, not as implicating a model such as memory. The data do not demonstrate that the passage of time is a causal factor in controlling REPR probability, only that there is an association of postreinforcement pause and REPR probability, possibly with a third variable unidentified in this experiment. To show a causal relation one would have to manipulate time since reinforcement as an independent variable, perhaps by imposing a variable timeout period after reinforcements.

The main question posed by these data is why there was a "strengthening" effect at all in the first ordinal position. It must be stressed that nothing in the animal's choice behavior at the first ordinal position can affect reinforcement probability there in the slightest. The behavior of repeating previously reinforced choices was not one the experimenter set out to reinforce, yet it occurred with persistently high probability. It is tempting to speculate that we have here an effect, repetition of the just-reinforced response, that is not explained by any higher-order reinforcement contingency, but is the basis of reinforcement itself; a simple "strengthening" effect that does not itself require differential reinforcing contingencies. After all, we cannot without danger of infinite regress in reasoning, explain *all* potentially nameable behaviors by their effects on reinforcement probability; how then should we explain the "behavior" of repeating those choices that increase reinforcement probability? One has to stop somewhere with a pre-existing mechanism, not itself requiring reinforcement, and it seems useful to label this "simple strengthening".

Such a "simple strengthening" effect of reinforcement has often been proposed (e.g., Hull, 1943), and with respect to certain dimensions of responding, such as the speed of running or the rate of lever pressing, the evidence for it is already strong. For example, an increase in reinforcer size causes an increase in running speed (Logan, 1960), and in the rate of key pecking (Shettleworth and Nevin, 1965), even when there is no differential reinforcement for these rate increases. Despite the formulation of a "micromolar" theory (Logan, 1960) in an effort to explain these and related instances, it has become clear that not all variations in rate and "vigor" of performance can be explained by differential reinforcement (Logan, 1960, p. 179; Williams, 1966). It seems that certain "strengthening" effects of reinforcement occur even when such "strengthening" is not demanded by the schedule.

With respect to choice behavior, the subject of the present investigation, it is more difficult to reach a firm conclusion from existing data. Of course, animals are known to repeat reinforced responses in discrimination tasks of many different kinds, but this fact is ambiguous, for such behavior receives differential reinforcement. The way around this problem is to look for "strengthening" effects in a schedule that does not differentially reinforce repetition, and this is what was attempted in the present experiment. The conclusion is that simple strengthening does indeed exist in the two-lever situation used here and has quite a pronounced effect, the repetition probability in the first ordinal position exceeding 0.7. However, there is evidence that the strengthening effect of a single reinforcement declines spontaneously with schedule exposure (the downward drift in Sessions 1 to 18). There is similar evidence for the strengthening effect of runs of reinforcement.

Repetition of the just-reinforced response is a behavior the rat brings to the two-lever choice situation. If the schedule encountered is random, the repetition behavior declines somewhat in probability. This finding, as already mentioned, agrees with work on rate measures. In a wider sense, it forms part of the growing literature concerned with the adaptive mechanisms that animals bring with them to the learning situation (Hinde and Stevenson-Hinde, 1973).

REFERENCES

Hinde, R. A. and Stevenson-Hinde, J. Constraints on learning. London: Academic Press (1973).

Hull, C. L. Principles of behavior. New York: Appleton-Century-Crofts, 1943.

Killeen, P. Preference for fixed-interval schedules of reinforcement. Journal of the Experimental Analysis of Behavior, 1970, 14, 127-131.

Lea, S. E. G. and Morgan, M. J. The measurement of rate-dependent changes in responding. In R. M. Gilbert and J. R. Millenson (Eds.), Reinforcement: behavioral analyses. New York: Academic Press (1972). Pp. 129-145.

Logan, F. A. Incentive. New Haven: Yale University Press (1960).

Menlove, R. L., Moffitt, M., and Shimp, C. P. Choice between concurrent schedules. Journal of the Experimental Analysis of Behavior, 1973 19, 331-344.

Nevin, J. A. Interval reinforcement of choice behavior in discrete trials. Journal of the Experimental Analysis of Behavior, 1969, 12, 875-885.

Shettleworth, S. J. and Nevin, J. A. Relative rate of response and relative magnitude of reinforcement in multiple schedules. Journal of the Experimental Analysis of Behavior, 1965, 8, 199-202.

Shimp, C. P. Probabilistically reinforced choice behavior in pigeons. Journal of the Experimental Analysis of Behavior, 1966, 9, 443-455.

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

Williams, D. R. Relation between response amplitude and reinforcement. Journal of Experimental Psychology, 1966, 71, 634-641.

Williams, B. A. Probability learning as a function of momentary reinforcement probability. Journal of the Experimental Analysis of Behavior, 1972, 17, 363-368.