Response sequence learning as a function of primary versus conditioned reinforcement

STEVE LUCK, MARIANNE COLGROVE, and ALLEN NEURINGER
Reed College, Portland, Oregon

Two groups of pigeons were required to generate a fixed sequence of responses on three keys, for example, middle-left-right. One group received a small food reward (S_food) following each correct response except the terminal one, which was followed by a large food reward. The second group received conditioned reinforcement from an overhead light (S_light) for each correct response, with the terminal correct response followed by both S_light and the large food reward. We manipulated length of sequence (3 or 7 responses) and duration of required interresponse interval (IRI; 1 to 9 sec). S_light contingencies generated more accurate performances than did S_food when sequence length was 3 responses but not when it was 7 responses. IRI duration influenced accuracy under the S_light contingencies but not under S_food. These results show that conditioned reinforcers sometimes generate more accurate sequence learning than do primary reinforcers, and that schedule contingencies influence which type of feedback will optimize performance. The results parallel those from the matching-to-sample and conditional discrimination literature.

Pigeons, rats, and monkeys can learn complex sequences of responses (e.g., Capaldi,erry, & Davidson, 1980; Hulse, 1978; Richardson & Bitter, 1982; Richardson & Kresch, 1983; Richardson & Wazak, 1981; Straub, Seidenberg, Bever, & Terrace, 1979; Straub & Terrace, 1981; Terrace, 1984). The present study compares two methods of teaching such complex sequences. Under one method, each correct response in an experimenter-defined sequence is followed by food, a primary reinforcer. Under the other, each correct response is followed by an overhead light, a conditioned reinforcer, with the last correct response in the sequence producing primary reinforcement as well. The central question is whether primary reinforcing feedback is more or less effective than conditioned reinforcing (or informational) feedback for the acquisition of complex response chains. Some previous studies provided no differential feedback prior to the final response in a sequence (e.g., Straub et al., 1979), whereas others provided differential feedback for every correct or incorrect response (e.g., Richardson & Bitter, 1982). To date, however, no study has compared the relative effectiveness of primary versus conditioned reinforcement as feedback under a complex sequence learning task.

Since the strength of a conditioned reinforcer derives from association with the primary reinforcer, it might be expected that primary reinforcement would generally be more effective than the derivative, and presumably weaker, conditioned reinforcing feedback. However, primary reinforcers may also have disadvantageous side effects (e.g., Balsam & Bondy, 1983; Catania, 1973; Schwartz, 1982a, 1982b). Furthermore, in some complex learning situations, such as matching-to-sample and conditional discrimination, conditioned reinforcers have been shown to generate higher response accuracies than primary reinforcers. For example, Ferster (1960) used a matching-to-sample task in which correct responses produced a light as conditioned reinforcer, with a fixed ratio of these correct matches leading to primary reinforcement. Intermediate ratio values produced higher response accuracy than longer and shorter ratios, including continuous reinforcement (CRF). Thus, in some instances, conditioned reinforcement led to higher accuracy than primary reinforcement. Stubbs and Galloway (1970) showed that Ferster's method of following correct responses with a conditioned reinforcer yielded higher accuracy in a fixed ratio 50 (FR 50) conditional discrimination task relative to the same task without conditioned reinforcers. However, variables such as fixed-ratio length, intertrial interval (ITI), and level of training appear to interact with the conditioned versus primary reinforcement dimension in determining response accuracy on complex discrimination tasks (e.g., Fujita, 1985; Nevin, Cumming, & Berryman, 1963; Thomas, 1979).

The present research varied both number of responses per sequence and minimum required interresponse interval in asking whether primary reinforcers are generally more effective in training sequences of responses or whether, as in matching-to-sample and conditional discrimination learning, there are parameters under which conditioned reinforcing feedback yields higher levels of accuracy.

**METHOD**

**Subjects**

Twelve pigeons bred in the Reed College pigeon colony were housed in individual cages and maintained at approximately 80%
of their ad-lib weights. Prior to this study, they had had extensive experience in other experiments on operant behavior.

Apparatus

The apparatus was a standard aluminum and Plexiglas pigeon operant chamber measuring 30 × 40.5 × 30 cm. Three Gerbrands keys were located on the front wall, 12 cm from the floor and spaced 7 cm apart in a horizontal row. The translucent keys were lighted from behind with 120-V ac lights colored green, red, and blue from left to right. A Gerbrands food hopper was located below the row of keys, 7 cm from the floor. A white overhead light was located above the front ceiling of the chamber. The apparatus was housed in a sound-attenuating chamber and controlled by a Commodore microcomputer and relay equipment in an adjoining room. Masking noise was produced by a Gerbrands white-noise generator.

Procedure

Subjects were divided arbitrarily into two groups of 6 (one pigeon in the S\textsuperscript{Light} group died in the middle of the experiment and therefore only 5 S\textsuperscript{Light} subjects were used in some conditions). One group (S\textsuperscript{Light}) received overhead light feedback for all correct responses in a sequence and overhead light plus 6-sec access to food for the terminal correct response. The other group (S\textsuperscript{Food}) received 2-sec of food for all correct responses except for the terminal one, which was followed by 6-sec of identical food. Subjects in both groups had to learn to peck the three keys in predefined order, with sequence lengths being either 3 responses or 7 responses. For example, one 3-response sequence was middle-left-right (M-L-R). Under this schedule, if the first response was to the middle key, it was counted as correct. Responses to left or right keys prior to the correct response were counted as incorrect. Similarly, once a middle response had been emitted, responses on the middle and right keys were counted as incorrect and the first left response to occur was correct, and so forth. Thus, although incorrect responses were counted and led to consequences different from those of correct responses (see below), they did not reset the sequence. The absence of a reset-following-error, or correction, contingency was necessary to prohibit the S\textsuperscript{Food} animals from perseverating on the initial response key, which would result in reinforcement after every second response. Each completed sequence contained exactly 3 correct responses (or 7, when sequence length 7 was employed) but an indeterminate number of incorrect responses.

For the animals in the S\textsuperscript{Food} group, every correct response in a sequence was followed by a 2-sec presentation of grain reinforcement, except for the final response of each sequence, which was followed by 6-sec of grain. The longer duration of this final reinforcer was intended to make the consequences of the final response of the sequence discriminably different from the rest of the sequence (see Catania, 1963; Neuringer, 1967). A minimum required interresponse interval (IRI), during which keylights were darkened and keypecks ignored, preceded every response. Therefore, the consequence of a correct response was food (2- or 6-sec presentations) followed by the IRI, whereas incorrect responses produced only the IRI.

For the birds in the S\textsuperscript{Light} group, each correct response was followed by a 2-sec illumination of the overhead light, except for the final correct response of each sequence, which was followed by a 6-sec presentation of both the overhead light and the food hopper. As under S\textsuperscript{Food}, all responses were preceded by an IRI and incorrect responses were followed by only the IRI. The main difference, therefore, between S\textsuperscript{Light} and S\textsuperscript{Food} contingencies was that under the S\textsuperscript{Light} condition, correct responses that did not terminate the sequence were followed by overhead light whereas under the S\textsuperscript{Food}, all correct responses were followed by food.

The main questions of interest were whether accuracy of performance differed under S\textsuperscript{Food} versus S\textsuperscript{Light} contingencies and whether accuracy depended upon sequence length and a minimum required interresponse interval. The pigeons experienced five experimental conditions in the following order: (1) 3-response sequence (L-R-M) with a 3-sec IRI; (2) 7-response sequence (M-R-L-R-M-L) with a 3-sec IRI; (3) 7-response sequence (M-R-L-R-M-L) with a 1-sec IRI; (4) 3-response sequence (M-L-R) with a 9-sec IRI; (5) 3-response sequence (M-R-L) with a 1.5-sec IRI. Each subject served in the experiments 3 days per week, with at least one "off" day preceding each "experimental" day, so that all birds were at their experimental weights (80% ad-lib body weights) prior to each session (the weight gains within a session being greater, of course, for the S\textsuperscript{Food} birds), and each session lasted for 150 correct responses or for a specified length of time, whichever came first. The time limit varied between 45 and 70 min, depending on performance and weight gain, but was always equivalent between the two groups. Each condition was maintained until accuracy reached asymptote, usually between 15 and 25 sessions.

RESULTS

Accuracy was defined as percent correct responses, calculated by dividing number of correct responses in a session by total number of correct plus incorrect responses. Each subject’s accuracy was calculated for each session, and the means of the 6 subjects’ accuracies (or 5 S\textsuperscript{Light} animals after the death of one) were then calculated for each group separately. All behavioral measures reported below were averaged over the final four sessions for each subject to provide an estimate of asymptotic performance.

The first main result was that when sequence length was 3, the S\textsuperscript{Light} conditioned reinforcement generated higher accuracies than did the S\textsuperscript{Food} primary reinforcement, across all IRIs, as shown in Figure 1. The second main finding was that the magnitude of this difference decreased as the IRI duration increased. These results were statistically supported by a mixed-design two-factor analysis of variance (ANOVA), which showed significant main effects of reinforcement—S\textsuperscript{Food} versus S\textsuperscript{Light}; [F(1,9) = 32.43, p < .0005] and IRI duration [F(2,18) = 18.62, p < .0001] and a significant reinforcement × IRI duration interaction [F(2,18) = 7.18, p < .01]. The third main finding was that the S\textsuperscript{Light} group performed less accurately under the 7-response sequence than under the 3-response sequence, but that there was no change in the S\textsuperscript{Food} group’s performance as a function of response sequence length. An S\textsuperscript{Food} versus S\textsuperscript{Light} reinforcement × sequence length ANOVA performed on the 3- and 7-response sequences with 3-sec IRI produced significant reinforcement [F(1,10) = 7.40, p < .05] and sequence length [F(1,10) = 107.39, p < .0001] main effects and a significant interaction [F(1,10) = 82.69, p < .0001]. Further analytical comparisons showed that accuracy in the S\textsuperscript{Food} group was statistically identical under 3- and 7-response sequence lengths (F < 1), while accuracy of the S\textsuperscript{Light} group decreased as sequence length increased [F(1,5) = 195.93, p < .0001]. All 10 conditions of the experiment generated mean sequence performances that were significantly more accurate than expected by chance at p < .05, one-tailed, indicating that learning occurred reliably under all conditions.

Figure 2 shows response rate, defined as mean number of responses per minute. The S\textsuperscript{Light} group’s response rate was lower than that of the S\textsuperscript{Food} group in four of the
five conditions, but an ANOVA on the two reinforcement contingencies (S\textsuperscript{Food} vs. S\textsuperscript{Light}) and five conditions (sequence and IRI combinations) showed only marginally significant reinforcement \([F(1,9) = 3.56, p < .10]\) and reinforcement \(\times\) condition interaction \([F(4,36) = 2.07, \ p < .11]\) effects.

Several authors have noted that when subjects are required to make a fixed number of correct discriminations for reward—an FR discrimination task—accuracy tends to increase as the potentially reinforced trial approaches (Mintz, Mourer, & Weinberg, 1966; Nevin et al., 1963; Thomas, 1979). Figure 3 shows percentage of total errors made as a function of position in the sequence under four of the five conditions (error data were not available from the 3-response, 3-sec IRI condition). These results are, at one level of analysis, consistent with those of the
previous studies. For both $S^{\text{Food}}$ and $S^{\text{Light}}$ groups, percentage of errors was always greater during the first response following reinforcement than during the last response prior to reinforcement. Thus, at this level of analysis, errors decreased as subjects neared the 6-sec reinforcer. However, some of the functions were not monotonic. In particular, for the $S^{\text{Food}}$ subjects under three of the four conditions shown, the maximum number of errors occurred during the penultimate response position, at Position 2 under the 3-response sequence (in Figure 3b but not in Figure 3a), and Position 6 under the 7-response sequences (in Figures 3c and 3d). Indeed, in the two 7-response conditions for $S^{\text{Food}}$ birds, probability of errors during the penultimate position was almost twice that during any other position. For subjects in the $S^{\text{Light}}$ condition, percentages of errors were monotonically decreasing when there were 3 responses per sequence, but bitonically related to position under the 7-response conditions, with most errors at Position 4 (Figure 3c) or Position 5 (Figure 3d). Significant interactions were found between $S^{\text{Food}}$ and $S^{\text{Light}}$ reinforcement contingencies and serial position for both sequence lengths [3-response, $F(2,18) = 4.07, p < .05$; and 7-response, $F(6,60) = 10.71, p < .0001$].

In addition to this difference in the serial dependency of errors, the $S^{\text{Food}}$ and $S^{\text{Light}}$ groups manifested a second difference in types of errors. We calculated the number of responses that were incorrect repetitions of the

![Bar chart](image)

Figure 3. Percentage of total errors made at each position in the sequence in (a) 3-response sequence with 1.5-sec IRI; (b) 3-response sequence with 9-sec IRI; (c) 7-response sequence with 1-sec IRI; and (d) 7-response sequence with 3-sec IRI (mean ± standard error).
previous correct response, calling such errors "response-repetition errors." Schwartz (1982a, 1982b) showed that primary reinforcers sometimes lead to repetitions of the previously reinforced response even when such repetitions are not reinforced, and we hypothesized that the SFood contingency would produce more response repetition errors than SLight. Figure 4 displays the mean of the proportion of total errors attributable to response repetition in each condition (except for the 3-response 3-sec IRI condition, for which no error data were collected). As predicted, SFood produced a higher proportion of response repetition errors than did SLight for both 3-response sequences \( F(1,9) = 32.29, p < .0005 \) and 7-response sequences \( F(1,10) = 8.50, p < .02 \). Both groups exhibited a higher proportion of these errors in the 7-response sequence than in the 3-response sequence \( F(1,9) = 39.82, p < .0001 \).

**DISCUSSION**

The present sequence-learning task can most easily be compared to other complex tasks, such as matching-to-sample and conditional discrimination, if we consider the contingencies of primary reinforcement. Under the 3-response requirement, SLight animals received food reward for every third correct response, an FR 3 schedule; under the 7-response requirement, food reward was given for every seventh correct response, an FR 7 schedule. The SFood schedule may be viewed as an FR 1, or continuous reinforcement (CRF), contingency, since food reward was provided following every correct response. Viewed in this manner, the present results appear to parallel those from the matching-to-sample literature in four basic ways. First, if the SFood condition is considered as FR 1 reinforcement for correct responses, then intermediate reinforcement ratios generated higher response accuracies than did either higher or lower ratios. When every third correct response led to food reinforcement (SLight, 3-response sequence), accuracy was higher than when every seventh correct response (SLight, 7-response sequence) or every response (SFood condition) led to primary reinforcement (Figure 1). This result is similar to matching-to-sample tasks where intermediate reinforcement ratios also produced the highest response accuracy (e.g., Ferster, 1960; Fujita, 1985; Thomas, 1979).

The finding that SLight generated fewer response repetition errors than SFood (Figure 4) may partly explain the higher accuracy of SLight subjects. However, since the SLight subjects made fewer of these errors even when their accuracy was lower than that of the SFood birds (in the 7-response sequences), differences in response repetition cannot be entirely responsible for the overall differences in accuracy. Data on response repetition errors are not available from the matching-to-sample literature, but, given the present results, we predict higher frequencies of such errors under CRF than under ratio reinforcement of matching-to-sample and conditional discrimination learning.

Second, IRI duration had no effect on response accuracy in the SFood condition, but did significantly affect accuracy under the SLight condition (Figure 1). Under SLight, long IRIs combined with relatively high reinforcement ratios (7 response) to cause the least accurate performance. In a matching-to-sample case, Thomas (1979) also found that 5-sec versus 25-sec ITI durations did not affect accuracy in the CRF contingency and that long ITIs and high ratios of responses to reinforcers caused the worst performance. Indeed, in Thomas's study, long ITI durations
with very high reinforcement ratios (FR 60) could not maintain responding. (Note that Thomas varied minimum intertrial time, an interval following each trial, whereas we varied minimum interresponse time. Both manipulations affected response and reinforcement rates.)

One possible explanation for the differential effect of IRI on accuracy under SLight versus SFood conditions is that long IRI durations under SLight led to relatively low reinforcement densities, whereas reinforcement densities were relatively high in SFood throughout the experiment. For example, with a 9-sec IRI, SLight subjects could be rewarded with food only after a minimum of 31 sec, whereas SFood subjects could be rewarded after a minimum of 9 sec. This might also explain the lower response rate seen in the SLight group under the longer IRI durations and longer sequence lengths, both of which decrease SLight reinforcement density.

Third, under SLight conditions, responses were more accurate at the end of the sequence than at the beginning (Figure 3). In the matching-to-sample case as well, Nevin et al. (1963), Mintz et al. (1966), and Thomas (1979) report that subjects became more accurate as they neared the reinforced trial. However, in addition to this "recency effect" (few errors at the end of the sequence), a "primacy effect" (relatively few errors at the beginning of the sequence) was observed, especially in the 7-response, SLight conditions (hatched bars in Figures 3c and 3d). This function is similar to those observed in studies of animal (e.g., Santiago & Wright, 1984) and human (e.g., Baddeley, 1986) memory. The extremely high percentage of errors made by SFood animals in the penultimate positions of both 7-response sequences may be an exaggerated form of this same primacy-recency function (as in Wright, Santiago, Sands, Kendrick, & Cook, 1985).

Fourth, as Ferster (1960) reported for matching-to-sample, responding under SLight conditions was faster at short sequence lengths than under long sequence lengths (although this difference only approached statistical significance in the present experiment).

Three aspects of these results deserve special attention. First, to maximize accuracy of sequence performance, it is not necessary to maximize primary reinforcement and, under some conditions, quite the opposite is true. The response repetition errors caused by presentation of food contingent on all correct responses instead of every nth correct response can be considered another "negative side effect" of primary rewards (see Balsam & Bondy, 1983; Catania, 1973; Lepper & Green, 1978; Schwartz, 1982a, 1982b). In both sequence learning and conditional discrimination tasks, it may be more efficacious to present conditioned reinforcement or positive informational feedback (Luck, Colgrove, & Neuringer, 1983) than to give primary reinforcement contingent upon all correct responses. The present experiment replaced food with an overhead light, the latter ostensibly providing subjects with the same "correct or incorrect" information as the food, but the light did not produce the negative side effects of the primary reward (e.g., response repetition errors).

Second, the type of reinforcement—primary versus conditioned—significantly influenced the distribution of errors. The greatest number of errors occurred earlier in a sequence of responses when conditioned reinforcers (SLight) were employed than when every correct response resulted in primary reinforcement (SFood).

Third, the present results showed that when every correct response in a complex sequence was followed by primary reinforcement, accuracy of responding was less sensitive to reinforcement and task parameters than when only some correct responses produced primary reinforcement. Under SFood reinforcement, response accuracy was relatively low throughout the experiment and accuracy was not influenced by changes in either number of responses in the sequence or minimum interresponse interval. On the other hand, SLight contingencies resulted in marked sensitivity to both of these parameters.

These results may be relevant to the education of people, where response accuracy also sometimes appears to be better controlled by conditioned reinforcers or informational feedback than by the primary reinforcer itself (see Davidson & Osborne, 1974; Lepper & Green, 1978; and Schwartz, 1982b; but see also Getsie, Langer, & Glass, 1985). The locus of maximal learning difficulty may also be influenced by the type of feedback for correct responses—primary versus conditioned reinforcement. Finally, when primary reinforcers are presented frequently as feedback for correct responses, motivation and response rates may be high but response accuracy may be unaffected by the reinforcing feedback. Consequently, complex response sequences may best be taught by providing the lowest frequency of primary reinforcement necessary to maintain responding together with a high frequency of informational feedback for correct responses.

REFERENCES


(Manuscript received April 27, 1987; revision accepted for publication September 22, 1987.)