The overlearning reversal effect in monkeys provided a salient irrelevant dimension[1]

M. R. D'AMATO
RUTGERS UNIVERSITY

Abstract

Capuchin monkeys were first trained on a color discrimination and then reversed. Later these same colors formed irrelevant background cues for a form discrimination. Despite the ready availability of the irrelevant (color) dimension, the controls reversed the form discrimination just as fast as overtrained Ss. Yet the controls reverted to color earlier than the overtrained Ss and responded to it more frequently during the first 80 reversal trials.

Problem

The role of overlearning in discriminated reversal learning is still an unsolved issue in spite of the relatively large amount of research that has been devoted to the problem. As things now stand an overlearning reversal effect (ORE—facilitation of reversal learning by overtraining) is much more likely to occur in visual discrimination problems than in spatial discrimination tasks (cf. D'Amato & Schiff, 1964). But even within a visual discrimination setting failures to obtain an ORE have not been uncommon (e.g., Cross et al, 1964; D'Amato & Schiff, 1965; Erlebacher, 1963).

At present there are two major potential explanations of the ORE which could conceivably delineate the conditions under which this effect may be expected. Theios & Blosser (1965) place the major burden on a single variable, magnitude of reward, which, according to these investigators, must be relatively large in order for an ORE to occur. It is unlikely, however, that the entire pattern of results in the area of the ORE can be explained by this single variable.

A somewhat more general and perhaps more promising hypothesis has been advanced by Sutherland and Mackintosh (e.g., Mackintosh, 1965) and, in a more mathematical fashion, by Lovejoy (in press). Essentially their notion is that the ORE can be expected only where at least one irrelevant stimulus dimension exists, a dimension to which Ss are likely to attend during reversal learning. The general point is that nonovertrained Ss are likely to revert to such dimensions when, during reversal training, the responses learned during acquisition of the discrimination are no longer effective. Overlearning, on the other hand, serves to keep Ss attending (and therefore responding) to the relevant stimulus dimension, thus preventing their switching to other, irrelevant, dimensions which presumably only forestalls solution of the reversal problem.

The purpose of the present study was to evaluate this hypothesis by building in, so to speak, an irrelevant dimension with which Ss previously had had substantial discriminative experience. Without some assurance that a readily available irrelevant dimension exists, assessment of the theory is difficult.

Method

Eight experimentally naive Capuchin monkeys (according to the supplier, *Cebus apella*, captured in Peru), about 3-4 lbs. in weight, were trained on a red (S+) vs. green (S-) discrimination in a specially devised testing chamber. The discriminanda were displayed on 2 of 5 projectors, the pair of projectors used changing from trial to trial. S's response was pressing a plastic key mounted on the face of the projectors. After reaching criterion, Ss were reversed on the color discrimination, green now being positive. There then intervened a form discrimination which is irrelevant to present purposes.

Next, Ss were trained on the test discrimination, between a circle and a triangle, each stimulus serving as S+ for half of the Ss. The essential point is that these forms were embedded on red and green backgrounds which together constituted an irrelevant dimension, one to which Ss, by virtue of their past experience with these cues, presumably had a strong disposition to attend and to respond.

After reaching criterion on the compound stimuli, 4 of the 8 Ss (the control group) were reversed, color remaining irrelevant. Before being placed on reversal the other 4 Ss received a number of overtraining trials (from 160 to 480), amounting for each S to approximately twice the number of trials required to reach acquisition criterion.

The procedural details were as follows. Forty trials were given daily, and to reach criterion S had to make 10 correct responses in succession during a single day's trials. To initiate a trial, S activated a centrally located microswitch, which illuminated 2 of the 5 projectors with S+ and S-. A correct response was rewarded with approximately .5 cc water; an incorrect response resulted in a 1-min. time out. The stimuli were automatically programmed and presented by means of a tape block reader (D'Amato, in press).

Results

Considering first reversal performance on the initial color discrimination, the mean number of trials to reversal criterion was 198.5 for the Ss that later served as controls on the test discrimination and 193.9 for the Ss later to comprise the overtrained group.

On the test discrimination the control group required 96.5 trials on the average to reach acquisition criterion compared to 124.0 for the overtrained Ss, a nonsignificant difference. On the other hand the overtrained Ss required substantially more trials to reverse than the controls, in terms of group means 325.0 vs. 218.3. In view of the differences in acquisition scores, however, a measure of reversal learning was employed which took acquisition differences into account; namely, the reversal ratio, a ratio formed by the number of trials required to reverse over the number needed for acquisition. The mean reversal ratios for the two groups were quite close, 2.57 for the controls and 2.72 for the overtrained Ss. Thus in terms of absolute or relative measures of speed of reversal learning, there is no sign of an ORE in the data.

Analysis of the data of individual Ss showed considerable differences in their tendencies to respond to the color cues during acquisition and reversal of the test discrimination. We took as indication that S was responding to color any sequence of 5 or more trials on which the chosen discriminanda had the same color background (hereafter called a "color" response). It might be noted that training trials were assembled in such a way that at most 4 trials in succession could occur with the same color cue being associated with the positive (or negative) form stimulus. In order to get some indication as to whether the two groups differed with respect to their responses to the irrelevant color dimension during reversal we calculated for each S the number of reversal trials to the first color response. The highest score in the control group was 54 and the lowest in the overtrained group, 72. The respective medians were 34.5 and 76.5, a difference significant at about the .02 level by the U test. As a measure of the persistence of color responding in the two groups we determined the total number of color responses made during the first 80 reversal trials (approximately the maximum number of reversal trials available for all Ss). For this calculation a string of, say, 10, responses to the same colored compound stimulus was counted as 2 color responses. The average number of color responses was 3.25 in the control group and only 0.50 in the overtrained Ss, a difference significant at the .02 level. (Interestingly, of the 71 color responses made during acquisition and the first 80 trials of reversal 68 were to green, the last reinforced color.)

Discussion

In terms of our definition of a color response the data suggest that, as required by the hypothesis under consideration, during reversal learning the controls reverted more quickly than the overtrained Ss to the irrelevant dimension of color and, at least for the first 80 reversal trials, more frequently. Nevertheless, there was no hint of an ORE in the trials-to-criterion data. The warranted conclusion seems to be that the presence of an irrelevant dimension to which the controls, during reversal learning, have a demonstrably greater tendency to respond than overtrained Ss is still no guarantee that overtraining will facilitate reversal learning.

It is conceivable, of course, that the tendency to respond to color which was built up by the earlier training afforded Ss was insufficient to maintain in the control group responding to this dimension to the point where an ORE would occur. However, the mean reversal ratio attained by the controls, 2.57, shows that they had considerable difficulty in reversing and suggests that there was ample opportunity for responding to irrelevant cues, if not color, then position, sequence cues, etc.

An important shortcoming of the present method of attempting to assess the role played by irrelevant cues is that S's response to such dimensions must be inferred from a relatively arbitrary response criterion. Five responses in a row to the same color component of a compound stimulus may be reasonably convincing evidence that S is responding to this aspect of the discriminandum. But there is no reason why S might not shift his attention much more rapidly, attending to color for only a trial or two, then turning to position, and so on. With irrelevant dimensions inextricably associated with the relevant dimension, such rapid fluctuations in the basis of S's response become impossible to track and disentangle. Consequently, for a more precise evaluation of the theory proposed by Sutherland, Mackintosh and Lovejoy it will be necessary somehow to disassociate the irrelevant dimensions from the relevant cues so that S's response, and by inference his attention, to the former can be positively identified.

References

Cross, H. A., Flicking, R. M., Carpenter, J. B., & Brown, L. T. Discrimination reversal performance in squirrel monkeys as a function of prereversal experience and overlearning. *Psychon. Sci.*, 1964, 1, 353-354.

D'Amato, M. R. Direct programming of multiple stimuli—the tape block reader. *J. exp. Anal. Behav.*, in press.

D'Amato, M. R., & Schiff, D. Further studies of overlearning and position reversal learning. *Psychol. Rep.*, 1964, 14, 380-382.

D'Amato, M. R., & Schiff, D. Overlearning and brightness discrimination reversal. *J. exp. Psychol.*, 1965, 69, 375-381.

Erlebacher, A. Reversal learning in rats as a function of percentage of reinforcement and degree of learning. *J. exp. Psychol.*, 1963, 66, 84-90.

Lovejoy, E. An analysis of the overlearning reversal effect. *Psychol. Rev.*, in press.

MacKintosh, N. J. Overtraining, reversal, and extinction in rats and chicks. *J. comp. physiol. Psychol.*, 1965, 59, 31-36.

Theios, J., & Blosser, D. An incentive model for the overlearning reversal effect. *Psychon. Sci.*, 1965, 2, 37-38.

Note

1. Supported by National Science Foundation Grant GB-2784. The data were collected by Nancy Killian and J. Fazzaro.