

Overjustification in a Token Economy

David Greene
Carnegie-Mellon University

Betty Sternberg
School of Education, Stanford University

Mark R. Lepper
Stanford University

A token economy was designed to discover whether demonstrably effective reinforcement procedures would also produce an overjustification effect, indicated by a significant decrement in posttreatment engagement with previously reinforced activities, in the absence of perceived tangible or social rewards. Three different experimental token economy groups were compared with a single control group. Following baseline observations, a treatment phase was initiated, during which differential reinforcement was made contingent upon time spent with designated "target" activities. During this phase, subjects in all three experimental groups spent significantly more time with these activities than did the nondifferentially reinforced control subjects. Subsequently, after differential reinforcement was withdrawn, subjects in two of the three experimental groups spent significantly less time with their target activities than control subjects did, demonstrating that multiple-trial contingent reinforcement procedures are capable of producing overjustification effects. The relationship between these findings and the problem of achieving generalization of treatment effects from token economies is discussed.

Principles from attribution theory (Kelley, 1973) and self-perception theory (Bem, 1972), taken together, suggest that a person's intrinsic interest in an activity may be decreased by inducing him to engage in that activity as an explicit means to some extrinsic goal—a proposition that has been called the "overjustification" hypothesis (Lepper, Greene,

& Nisbett, 1973). If the justification provided to induce a person to engage in an activity were perceived to be unnecessarily high or otherwise psychologically "oversufficient," the person might come to infer that his actions were motivated by the contingencies of the situation, rather than by an intrinsic interest in the activity itself. Thus, a person induced to undertake an inherently desirable activity as a means to some ulterior end would no longer regard the activity as an end in itself.

To test this hypothesis, Lepper et al. (1973) introduced an attractive drawing activity into children's nursery school classrooms during "free play" periods and unobtrusively recorded measures of the children's interest in the activity. Youngsters showing initial interest were randomly assigned to one of three conditions. In the expected-award condition, children were asked to engage in the activity in order to obtain an extrinsic reward, a "Good Player" certificate; in the unexpected-award condition, they engaged in the same activity and received the same reward, but had no knowledge of the reward until after they had finished the activity; and in the no-award condition, children neither expected nor received a reward, but otherwise

This report is based on the first author's doctoral dissertation at Stanford. The second author generated the experimental materials and supervised the day-to-day operation of the experiment. The third author was chairman of the dissertation committee and participated actively in all phases of the study. Support came from National Institute of Mental Health Grants MH 24134 to Mark R. Lepper and MH 12283 to the Social Psychology Training Program at Stanford.

Grateful acknowledgement is made to Daryl J. Bem and J. Merrill Carlsmith for their contributions as members of the dissertation committee; to Stephen M. Johnson, Susan Roth, Richard Schulz, and Robert S. Siegler for their comments on an earlier draft; to John Delgado, principal, and M. Sue Bailey and Ann Jefferson, teachers, at the McKinley School for their gracious cooperation and assistance; and to Sandra Dubriel and Katie Godfrey for their help with data collection.

Requests for reprints should be sent to David Greene, Department of Psychology, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213.

duplicated the experience of subjects in the other conditions. These experimental treatments were administered in a room apart from the classrooms. Two weeks later, the drawing materials were again placed in the children's classrooms, and unobtrusive measures of post-experimental interest were recorded. As predicted, subjects who had agreed to engage in the activity in order to obtain the award subsequently spent significantly less time playing with the materials than did subjects in either of the other two conditions. Relative to their uniform baseline levels of interest, subjects in the expected-award condition showed a significant decrease from baseline to postexperimental observations, while subjects in the other two conditions showed no significant change.

This finding has been subject to considerable further scrutiny. On the one hand, its generality has been confirmed across a substantial number of procedural variations and subject populations. Deci and others (cf. Deci, 1971, 1975), for example, have obtained comparable results by paying college subjects to solve puzzles. Similarly, Lepper and Greene (1975) found that the presence of adult surveillance as well as the expectation of reward could decrease children's subsequent interest in playing with a target activity. On the other hand, it is clear that anticipated rewards will not undermine intrinsic motivation if they are not salient to the subject (Ross, 1975). Thus, in its current formulation (Lepper & Greene, 1976), the overjustification hypothesis applies to (a) activities of at least some initial interest to a subject, (b) conditions which make salient to a subject the instrumentality of engaging in a particular activity as a means to some extrinsic end, and (c) measures of subsequent engagement in situations where subjects do not expect extrinsic rewards.¹

Overjustification Studies and Token Economies

One important question raised but not answered in the present overjustification literature is its relevance to the systematic use of tangible reinforcement procedures in applied settings (Kazdin & Bootzin, 1972; O'Leary & Drabman, 1971). At first glance, the finding that salient expected rewards may

undermine intrinsic interest seems like a direct attack on the token economy establishment (cf. Levine & Fasnacht, 1974). However, a typical token reinforcement program and a typical overjustification study differ from each other in a number of potentially important ways. A consideration of these differences will help to elucidate the rationale for the present study.

The most obvious difference is between subjects selected for low rates of appropriate behavior versus subjects selected for high rates of appropriate behavior. Since the overjustification hypothesis presumes at least some intrinsic interest in a target activity, its domain would seem to exclude token economies whose subjects will not engage in appropriate behavior without extrinsic reinforcement (e.g., Ayllon & Azrin, 1965). To the extent that subjects in token economy programs are commonly selected for relatively (vs. absolutely) low rates of appropriate behavior, however, the applicability of the overjustification hypothesis remains an empirical issue.

A second major difference between token economies and overjustification studies is of potentially greater theoretical interest: the use of multiple-trial reinforcement procedures and the demonstration of a reinforcement effect via such procedures. These essential features of token economy programs have been notably absent from overjustification experiments. Indeed, two recent studies have sought to demonstrate that an overjustification effect would not occur when multiple-trial reinforcement procedures were employed. In one of these studies, Feingold and Mahoney (1975) found that five children increased (rather than decreased) their performance of a play activity after having been reinforced for playing with it. In the other study, Reiss and Sushinsky (1975, Experiment 2) found that "preferences" among songs established by discrimination training procedures transferred to a recognizably similar posttest situation. Both studies appear to demonstrate that overjustification effects should not be expected

¹ An extensive discussion of some relevant conceptual and definitional issues in the overjustification literature is available elsewhere (Lepper & Greene, 1976).

when token economy procedures are employed. However, as Lepper and Greene (1976) have noted, these studies differ from previous demonstrations of overjustification effects in several other ways, making interpretation of their findings equivocal.

One such difference concerns the use of intrasubject (vs. between-groups) designs in these studies. Although control groups are typically not necessary to the explicit objectives of token economies (Kazdin, 1973; Sidman, 1960), they are absolutely essential to studies designed to distinguish among the various multiple effects of particular experimental procedures. In the Lepper et al. (1973) study, for example, the effect of the expected-award manipulation is distinguishable from other effects of the experimental procedures (e.g., social feedback, increased task engagement, familiarity with the activity) because the design included one group of subjects who received the same reward unexpectedly and another group of subjects who received no reward at all. Unfortunately, neither the Feingold and Mahoney (1975) nor the Reiss and Sushinsky (1975, Experiment 2) study included subjects who did not receive contingent or differential reinforcement. Therefore, the data from these studies cannot answer the question of whether their reward procedures per se would have increased, decreased, or had no effect on subjects' subsequent interest, relative to appropriate control conditions in which subjects engaged in the target activity without expecting contingent reward.

Furthermore, these studies were testing a "straw-man" version of the hypothesis. The overjustification hypothesis does not predict that all rewards, or even all expected rewards, will undermine subsequent intrinsic interest (Greene, 1975; Lepper & Greene, 1976; Lepper et al., 1973; Ross, 1975); nor does it delineate procedures. On the contrary, the overjustification hypothesis presupposes that the effects of rewards on subsequent behavior are mediated by the information they convey concerning (a) a person's perceived motivation for engaging in the activity, (b) the person's competence at the activity, and/or (c) the subsequent probability of further extrinsic reinforcement for engagement in the activity—all of which will neces-

sarily depend on the context and manner in which the rewards are presented (cf. Bem, 1972; Feingold & Mahoney, 1975; Lepper & Greene, 1976). As one example, rewards contingent upon superior task performance, which provide an individual with significant information about his competence or ability at an activity, should be less likely to produce a decrement—or more likely to produce an increment—in subsequent intrinsic interest than rewards contingent upon task engagement per se (Deci, 1975; Lepper & Greene, 1976; Ross, 1976). Thus, an appropriate test of the overjustification hypothesis requires a comparison across experimental conditions in which other relevant factors have been held constant.

In addition, to provide data relevant to any hypothesis concerning intrinsic motivation, it is of central theoretical importance to distinguish between two classes of experimental settings: those in which extrinsic rewards are potentially available for the target activity versus those in which extrinsic rewards are not perceived to be available (Lepper & Greene, 1976; Lepper et al., 1973). There are, of course, many reasons why a person might engage in a target activity; however, the focus of the overjustification hypothesis is on intrinsic as opposed to extrinsic reasons. This distinction has been operationalized by defining intrinsically motivated behavior as that which occurs in the perceived absence of extrinsic rewards (Deci, 1971; Lepper et al., 1973; Ross, 1975). In the face of sufficiently powerful extrinsic rewards, individuals will often engage in activities in which they have no intrinsic interest. Therefore, assessments of subsequent intrinsic interest must be obtained in situations where subjects do not expect either tangible or social rewards to be contingent upon engagement in the activity.

Furthermore, to preclude such expectations, some provision must be made to minimize the confounding influence of potential artifacts, such as demand characteristics, subjects' reactivity to experimental procedures, and experimenters' expectancies (Johnson & Bolstad, 1973; Levine & Fasnacht, 1976). In social psychological experiments, failure to keep personnel "blind" to subjects' differential treatments may be sufficient grounds for out-

right rejection of a manuscript submitted for publication. In token economies, by contrast, personnel often have a justifiable "vested interest" in maintaining appropriate behaviors after tangible reinforcement has been withdrawn (O'Leary & Drabman, 1971). In perhaps the typical case, a teacher is trained in systematic observation techniques and contingent social approval behaviors as part of the token economy (Kazdin, 1973). In such a situation, it seems unreasonable to assume that the effects of this training will disappear when an attempt is made to reinstate pre-treatment baseline conditions, or that this situation provides an appropriate test of intrinsic interest (Lepper & Greene, 1976).

In short, a proper test of the relevance of the overjustification hypothesis to token economies requires a between-groups design and an experimental setting in which time spent with a target activity can reasonably be attributed to intrinsic motivation.

The Present Study

This kind of setting was created in two elementary classrooms by providing a time of day during which the entire class would partake of a set of four activities with a common focus. Within the set, individual children were free to choose which activities they would engage in. As long as no differential contingencies were imposed among the activities, we presumed that an individual's relative intrinsic interest in a particular activity was reflected in the time he or she spent playing with it. The dependent measure, then, was the amount of time spent playing with particular activities within the prescribed set. Given this situation, it was possible to introduce and later withdraw a system of reinforcement contingencies, and to observe its effects on children's immediate and subsequent interest in the various activities.

After baseline observations to determine initial relative preferences, children were randomly assigned to one of four groups, three experimental and one control. For subjects in each of the three experimental groups, two of the four available activities were designated on an individual basis as "targets." In the high-interest group, the target activities were the two which each child had played with

most often during the baseline phase. In the low-interest group, they were the two which each child had played with least often during baseline. In the choice group, they were the two chosen by each child individually after the baseline phase. During the treatment phase of the study, then, each experimental subject was differentially rewarded for spending time with his or her two target activities. For subjects in the control group, all four activities were designated as targets. During the treatment phase, these children were contingently rewarded for time spent with any of the four available activities. Thus, all four groups of children received contingent reinforcement during the treatment phase of the study. The difference between experimental and control subjects was that experimental subjects were *differentially* rewarded for playing with their target activities, whereas control subjects were *nondifferentially* rewarded for playing with any of the four available activities.

Two features of this design are critical. First, since target activities were different for different subjects within the same experimental group, it was possible to keep all personnel in the study from knowing any child's experimental condition. Second, the nondifferentially rewarded group provided a basis of comparison for each of the three experimental groups, controlling for effects of time and repeated measurement such as practice effects, satiation, and regression to the mean. Post-treatment data from each of the three experimental groups could therefore be compared with data from a group of subjects whose experiences had been otherwise identical.

We expected to demonstrate a reinforcement effect during the treatment phase in all three experimental groups. This effect would be indicated by a significant increase in time spent with their target activities by each group of experimental subjects, relative to the time spent with appropriately matched or yoked activities by the subjects in the control group. We could then discover whether the same procedures that had produced a reinforcement effect would also produce an overjustification effect after the contingencies had been withdrawn. This effect would be indicated by a significant decrease in time spent

with target activities by experimental subjects, relative to the time spent with appropriately matched or yoked activities by control group subjects.

METHOD

Experimental Setting and Subject Population

The study was conducted in an elementary school with an ongoing individualized mathematics program. During "math time," each child worked on one of 100 levels into which the elementary mathematics curriculum had been divided. A distinctive feature of the program was its explicit reliance on an elaborate system of extrinsic rewards, including a bi-weekly "Awards Assembly," when certificates and trophies were presented to all who had earned them since the last assembly. This system of extrinsic rewards provided a simple, natural way to deliver contingent extrinsic reinforcement during the treatment phase of the study.

Two classrooms served as the immediate experimental setting. Subjects were 44 fourth and fifth graders, selected after the baseline phase according to a procedure described below. The school population was predominantly of lower socioeconomic levels, with 40% of the families receiving welfare. Of the 44 subjects, 19 had Spanish surnames, 17 were Anglos, 7 were Black, and one was Asian. The mean total arithmetic grade equivalents on the California Test of Basic Skills for fourth graders and fifth graders, respectively, were 3.72 and 4.11 in the October preceding the study, and 4.51 and 5.54 in the May following the study.

Experimental Materials

The design required the constant availability of four different activities, comparable in initial interest. This requirement was met within the context of a "math lab." The four activities shared a common structure and format, each consisting of a set of manipulative materials, a sequence of "task cards" (instructing students how to use the materials) selected after extensive pretesting, and a folder containing "activity sheets." The specific materials were (a) geoboards (peg boards with rubber bands, to explore plane geometry); (b) Dienes blocks (cubes, bars, and larger cubes in different numerical bases); (c) attribute materials (items varying in color, shape, and size, to promote logical thinking and set theory ideas); and (d) tangrams (puzzle pieces in geometrical shapes, to be matched to various templates). Each child was provided with his own folder for each activity. Each folder was covered with a "log" sheet with columns headed "date," "time started," and "time finished." Each activity sheet had a place where the date on which it was completed was to be recorded. There were sufficient materials, task cards, and activity sheets to allow any child to choose to engage in any activity at virtually any time.

Procedure

Five school days before the first day of the baseline phase, the experimenter (the second author) introduced "some new math games" to each of two classrooms, in her capacity as administrator of the school's ongoing math program. For a regular, $\frac{1}{2}$ -hour period each day, she explained, four activities would be available for children to play with on their own, to help her "find out which games to use in our math program." One activity was introduced in some detail on each of 4 consecutive school days. After this training period, a 5th day was devoted to administering a simple questionnaire which asked about children's previous exposure to and anticipated liking of the four different activities.

Baseline. During the next 13 school days, children played with the four activities with no differential reinforcement contingencies, except with the constraint that each child was to try each of the activities at least once. At the end of this baseline phase, the amount of time children had spent with each activity was calculated. Children were then blocked into groups of four on the basis of the extent to which they had concentrated their time on their two most preferred activities and, within blocks, randomly assigned to conditions. A total of 51 children had been present in the two classes during the baseline period. Four children were dropped from the study, at the teachers' request, because of various learning problems; an additional three children, showing the most extreme concentrations of time on their two preferred activities, were eliminated to produce a balanced sample of 44 subjects. The sample included 11 subjects in each condition, 5 fourth graders and 6 fifth graders, or 5 boys and 6 girls.

Treatment. On the first day of the treatment phase, the experimenter handed out a sheet of paper to each child and asked the class to read it silently. She then went over this sheet, point by point, with the class, as follows:

[The experimenter] is very proud of the way you have all been helping her. Now, she is going to help you!

1. If you work with certain games for 3 hours, [the experimenter] will give you credit on your math award sheet for having completed 1 math level. [The award sheet was the basis on which children were given certificates and trophies at their Awards Assemblies.] (If you work 6 hours, you will get credit for 2 levels. If you work for 9 hours, you will get credit for 3 levels.) But, you'll only get credit if you write down correctly the time you start and stop, just as you have been doing. [Children had been recording data on their log sheets.]

2. [The observer] will be in this room every day during math lab to check on how well you keep track of the time you play with the games. If she finds that you do not record the time you played with the game correctly, you will not get any credit for playing with that game.

3. [An assistant] will keep track of how long you play with the games on a chart. [The experimenter held up before the class a 28×22 inch ($.71 \times .56$ m) bright orange posterboard chart with the title, "New Levels Completed for Math Awards Assembly." Each child's name was entered down one side, and headings across the top indicated how many hours and levels were completed.] When you come to class on Wednesday and Friday, you will see how long you have played and how much longer you have to play to get credit for a level.

4. As soon as you reach the 3-hour point on the chart, your teacher will give you a paper to bring to your math teacher. This paper will tell your math teacher to enter Math Lab on your award sheet in your math profile.

5. Next [the experimenter] is going to give you a sheet with your name on it and the list of games for which she will give you credit toward your math award(s).

At this point, the experimenter handed out a sheet in the format of a personalized letter to each child in the class. After the date and greeting, it said

You may play with any game you want to. Starting today, you will get 1 math level's credit for every 3 hours you play with: [the subject's target activities]. To help you remember, these folders have a big green X in the corner. [Target folders had been marked prior to class except as explained below.]

Subjects in the high-interest group were given the two activities with which they had spent the most time during baseline as their target activities. Low-interest group subjects were given the two activities with which they had spent the least time during baseline. For control subjects, all four activities were listed; thus, for these subjects, there remained no differential contingencies among the four activities. Subjects who had been assigned to the choice group were given a sheet with two blank lines. When all the sheets had been handed out, subjects in the choice group were asked to step outside the classroom with [the first author]. Meanwhile, the rest of the class proceeded to go to the math lab area as usual, where they found that the upper right corner of the log sheet on the cover of the folders for their target activities had been marked with a large green X.

Subjects in the choice group were told that they were being allowed to choose for themselves which two activities they would like to be the ones for which they could earn credit for levels. They were told to make the decision individually, and that different children might have different reasons for choosing their activities; in fact, each child might have a different reason for choosing each of his two activities. Many reasons were suggested, to emphasize the range of alternatives. Then these subjects were led back into their classroom and told to take the sheet with two blank lines to their seats, to think

about their choices by themselves, and then to write their choices on the sheet and bring it to [the first author] at the back of the classroom. As each child did so, the now-designated target folders were marked with a large green X and all four folders were given to the child to bring over to the math lab area. From this moment on, children in all four experimental groups were treated alike.

The treatment phase continued for 12 more days. For the first 3 of these days, the classroom teachers went over the experimental instructions at the beginning of the math lab period. The chart providing feedback to the children was updated every Tuesday and Thursday evening, so that upon first arriving in class Wednesday and Friday mornings, children would typically take note of their progress.

Withdrawal. On the first day of the withdrawal phase, the experimenter said that others in the school were resenting the "unfair advantage" that this and the other class were enjoying, and that she was inclined to agree that it would be unfair to continue such an advantage any longer. Therefore, she was taking down the chart as a sign that working with math lab activities would no longer earn credit for math levels. In addition, the upper right corner of the log sheet on the cover of all the folders had been clipped off, removing the discriminative cue to what had been target activities. On the other hand, she continued, there was no reason to remove the math lab materials from the classrooms; in fact, the teachers had been really happy with the math lab period. Children could continue to help her by working as before, or by making up their own activity sheets, which she would use in other classrooms.

After the 13th day of the withdrawal phase, all the folders were removed from the classrooms. The next day, the experimenter administered a questionnaire, telling children that their preferences would determine which activities stayed in their rooms, with the other activities destined for other classrooms. Before children left school on the day of this last questionnaire, the folders for the two activities which each class had "voted for" were returned to the math lab areas.

Data Collection

The dependent variable of primary interest was the time spent by each subject with each of the activities. During the baseline phase, children were instructed to record the date and time whenever they took one of their folders from or back to the math lab area, on the log sheet on the cover of the folder. Their recording behavior was monitored carefully and social approval was delivered contingently when appropriate recording behavior was observed. Although this procedure produced acceptably reliable data (83% to 91% agreement with data recorded by a classroom observer), the time and effort necessary to achieve the result were inordinate. Accordingly, from the first day of the treatment phase to the end of the study, time data were recorded by a classroom observer as well as by the students. In addition to

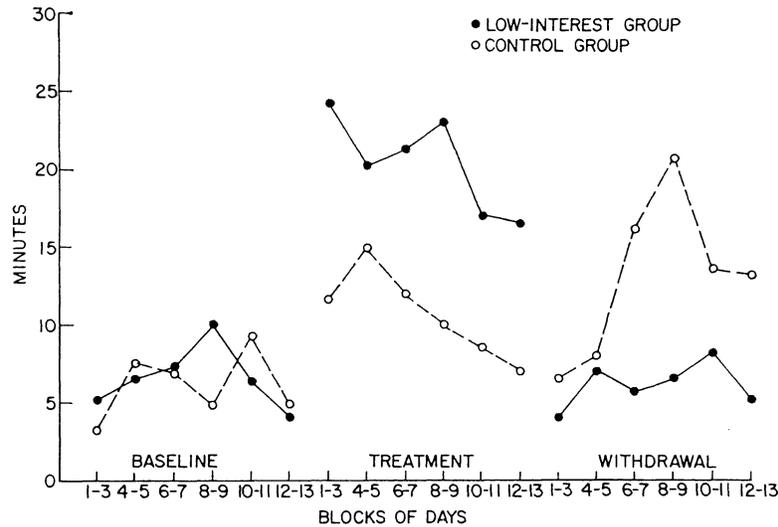


FIGURE 1. Mean time per day spent with target activities by low-interest group subjects and comparable data for control group subjects.

these data indicating the time each child spent with each of the activities, it was also possible to determine the rate at which the children completed the various activity sheets and to estimate the accuracy of the work they had done, allowing an examination of possible effects of the experimental treatments on performance.

RESULTS

Both within and across phases, the four activities varied as to how difficult they were to complete accurately (percent correct) or quickly (rate completed). To compensate for these differences among activities, several empirically derived weighting systems were applied to the performance data. Nevertheless, statistical analyses did not reveal any systematic differences in accuracy or rate of performance on target activities, either between groups or between phases. Nor did subjects' responses to the final questionnaire, scores on standardized tests, or progress in the school's regular math program show evidence of differences between groups. Although differences between groups on any of these global measures would have been of some interest, it is not surprising that none were found, since the differential contingencies in this study were designed to affect the amount of time spent with particular activities, rather than what was done with them. Significantly, there were

no reliable differences between groups or between phases in the mean time spent per day on all four activities combined; instead, the effects of our experimental manipulations were apparent only on the measure of greatest interest, the amount of time children spent with their target activities.

Within-Group Comparisons

A primary concern of the study was the demonstration of a reinforcement effect for each of the three experimental groups during the treatment phase. The relevant data are presented in Figures 1-3. These figures show the mean time per day spent with target activities during the three phases of the study by subjects in each of the three experimental groups, as well as comparable data for control group subjects. For each of the three experimental groups it can be observed that: (a) the mean time per day spent with target activities during the baseline phase, though variable from day to day, was essentially stable across days; (b) time on target during the treatment phase was greater than it had been during the baseline phase; and (c) removal of the contingency resulted in a decrease in time on target from treatment to withdrawal. Both these shifts occurred for all but 2 of the 33 experimental subjects. In

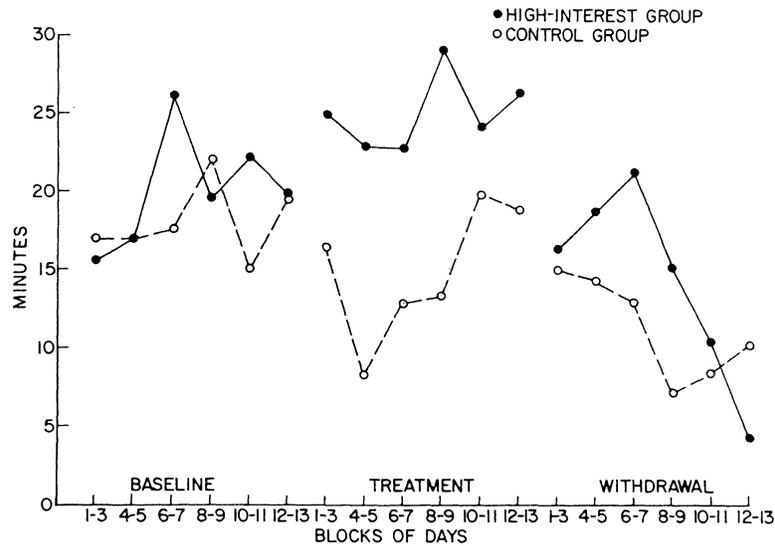


FIGURE 2. Mean time per day spent with target activities by high-interest group subjects and comparable data for control group subjects.

addition, *t* tests for correlated means established that all these within-group differences between phases were statistically significant at or beyond the $p < .01$ level.² Thus, the differential contingencies produced a clear reinforcement effect in each of the three experimental groups.³

It should be noted that the data for control group subjects in Figures 1-3 represent three different sets of data from the same subjects. Each set of data was generated by

² All significance levels are based on two-tailed tests.

³ The pattern of data for the low-interest group (Figure 1) is precisely that typically found in operant studies in applied settings. In fact, 1 of the 11 children in this group failed to respond to the treatment manipulation, a state of affairs quite familiar to researchers in this field (Kazdin & Bootzin, 1972; O'Leary, Becker, Evans & Saudargas, 1969). For the other 10 children in the group, the mean time per day with target activities during treatment was 22.4 minutes with a standard deviation of 8.2 minutes. For the "deviant" child, the mean time per day with target activities was only 1.2 minutes. Consequently, this subject's data were excluded from the low-interest group for purposes of subsequent between-groups comparisons. In addition, 3 subjects left school before the end of the study. Since all of them were in the control group, every effort was made to use all of their data whenever possible. Thus, for high-interest and choice group comparisons with control subjects during withdrawal, analyses are based on

following the rule that had determined which activities would be target activities for one of the experimental groups. For example, the control group data in Figure 1 show the mean time per day that these subjects spent with the two activities with which they spent the least time during the baseline phase, the "low-interest" activities of the control subjects. Similarly, data for control subjects in Figure 2 show the time they spent with their initially "high-interest" activities. The data for control subjects in Figure 3 show the time they spent with two activities yoked to the choices of target activities made by choice group subjects, in terms of the ranks of the chosen activities within each choice group subjects' baseline preferences. Since subjects were blocked before assignment to conditions, each experimental subject was matched to the control subject in the same block.

It should be clear that the data in Figures 1 and 2 for control group subjects are not independent, since for each subject the activities plotted in one figure are necessarily the two of the set of four not plotted in the other figure. What is represented twice in these two figures, then, is basically a single effect: Differences in the likelihood of playing with different activities during baseline tended to diminish by the end of the study for subjects

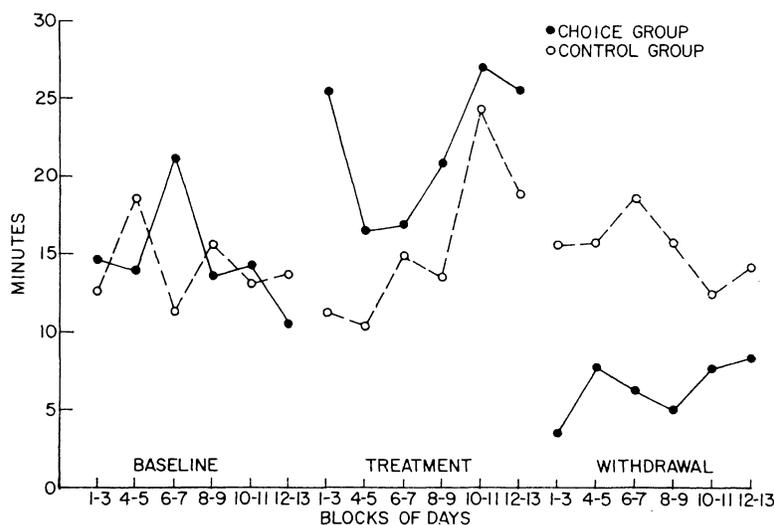


FIGURE 3. Mean time per day spent with target activities by choice group subjects and comparable data for control group subjects.

not exposed to differential contingencies during the treatment phase. Interestingly, there is no evidence of a shift in preferences for the activities yoked to the choices of choice group subjects (Figure 3). This difference between activities selected for initial extremity versus activities selected on a different basis is provocative, suggesting some kind of regression effect within the control group, although none of these within-group differences between phases (including baseline vs. withdrawal) attained statistical significance.

On the other hand, subjects in each of the three experimental groups spent less time with their target activities during the withdrawal phase than they had during the baseline phase. For the low-interest group this difference was not statistically significant, although inspection of Figure 1 suggests that interpretation of this lack of significance should make allowance for the restricted range between the initial baseline and zero. For the other two experimental groups, this posttreatment drop below baseline was statistically significant, $t(10) = 4.14$, $p < .01$, for choice

group subjects, and $t(10) = 2.38$, $p < .05$, for high-interest group subjects.

Within the logic of a within-group design, of course, these data may be viewed as evidence consistent with the overjustification hypothesis. However, the nondifferentially reinforced control subjects' data suggest the possibility that posttreatment decrements in interest were, at least in part, the result of boredom, satiation, or some other process ensuing directly from the high amount of time spent with target activities during the treatment phase. These explanations suggest a negative relationship between previous time spent with particular activities and subsequent interest in those activities. Specifically, correlations of time spent with target activities should be negative between treatment and withdrawal phases and/or between baseline and treatment phases combined and the withdrawal phase. Both these correlations were computed for each experimental group, the three experimental groups combined, and for each of the three sets of target activities for control group subjects. All of these 14 correlations were positive or no more negative than $r = -.05$; thus, they do not support this class of explanations.⁴

9 subjects per cell, including 1 who left school after 5 of the 13 days of that phase; during the first two phases of the study, these comparisons are based on all 11 subjects in each group; and low-interest group comparisons with control subjects are based on 1 fewer subject per cell in each phase.

⁴ In principle, of course, these correlations are not inconsistent with the possibility that all children proportionately lost interest in their initially pre-

TABLE 1
MEAN MINUTES PER DAY SPENT WITH TARGET ACTIVITIES BY EXPERIMENTAL SUBJECTS DURING
EACH PHASE AND COMPARISONS WITH CONTROL GROUP SUBJECTS

Target Activities	Phase		
	Baseline	Treatment	Withdrawal
High-probability during baseline			
Experimental (HI)	19.3	25.6	14.7
Control (C)	17.8	15.0	13.0
Chosen after baseline			
Experimental (CH)	14.5	21.9	6.8
Control (C) ^a	13.8	14.8	16.1
Low-probability during baseline			
Experimental (LO)	6.7	22.4	3.5
Control (C)	6.1	10.6	11.7
<i>t</i> tests			
HI vs. C (<i>n</i> = 11) ^b	1.57	2.83*	<1
CH vs. C (<i>n</i> = 11) ^b	<1	2.46*	3.55**
LO vs. C (<i>n</i> = 10) ^b	1.15	3.54**	2.54*

^a Target activities were yoked to ranks of preferences of activities chosen by experimental subjects.

^b Comparisons during the withdrawal phase are based on two fewer subjects per cell.

* $p < .05$.

** $p < .01$.

Between-Groups Comparisons

It should be emphasized, however, that within-group comparisons and between-groups comparisons are asking different questions. In the present study, our primary concern was whether any of the experimental treatments that produced a reinforcement effect would also produce an overjustification effect, as indicated by a significant decrease in time spent by experimental group subjects with their target activities, after the withdrawal of differential contingencies, relative to the time spent by control group subjects with appropriately matched or yoked activities. Since each experimental subject was matched with a control subject by the blocking procedure described earlier, *t* tests for correlated means were used to compare each experimental group with the control group during each phase. These comparisons, as well as the

cell means on which they are based, are presented in Table 1.⁵

During the baseline phase, of course, experimental and control subjects did not differ in the time they spent with the activities which would subsequently be designated as targets for experimental subjects. During the treatment phase, experimental group subjects spent significantly more time with their target activities than did the subjects in the control group ($p < .05$ or less in each case). During the withdrawal phase, subjects in two of the three experimental groups spent significantly less time with their target activities than did control group subjects matched or yoked to them. Although high-interest group subjects' time on target was significantly below their baseline level, $t(10) = 2.38$, $p < .05$, they did not differ from control subjects ($t < 1$). Subjects in the low-interest group and in the choice group, however, spent significantly less time with their target activities than did control subjects matched or yoked to them, $t(7) = 2.54$, $p < .05$, and $t(8) =$

ferred activities over time (Johnson, Note 1); this sort of explanation, however, does not account for the relative effects of the three experimental treatments. Moreover, studies designed explicitly to control for satiation have consistently found overjustification effects which could not be attributed to effects of increased task engagement (e.g., Calder & Staw, 1975; Lepper & Greene, 1975; Ross, 1975; Ross, Karniol, & Rothstein, 1976).

⁵ Each of the nine comparisons in Table 1 was also made separately for boys and girls and for fourth- and fifth-graders. In each of the four subgroups of subjects, the pattern of means is strikingly similar to that of Table 1 (see Greene, 1974).

3.55, $p < .01$, respectively.⁶ Thus, the same procedures which produced a reinforcement effect during the treatment phase also produced a posttreatment decrement in time spent with target activities (an overjustification effect) for subjects in two of the three experimental groups.

DISCUSSION

The present findings indicate that, under some conditions, multiple-trial contingent reinforcement procedures are capable of producing posttreatment decrements in engagement with previously reinforced activities. These results, obtained over the course of a 13-day period following the removal of demonstrably effective token programs, are clear evidence that overjustification effects are not limited to single-trial, "noncontingent" reward procedures, as Reiss and Sushinsky (1975) have contended. Instead, together with data from other studies of the subsequent effects of multiple-trial reinforcement procedures (Brownell, Colletti, Ersner-Hershfield, Hershfield, & Wilson, in press; Colvin, 1973; Johnson, Bolstad, & Lobitz, 1976; Meichenbaum, Bowers, & Ross, 1968), the present results demonstrate that the use of powerful systematic reward procedures to promote increased engagement in target activities may also produce concomitant decreases in task engagement, in situations where neither tangible nor social extrinsic rewards are perceived to be available.

These data are sufficient to indicate that the simple distinction between single-trial versus multiple-trial reinforcement procedures does not provide an adequate account of the conditions under which overjustification effects will be obtained. Unfortunately, however, the differences in results across the three experimental conditions do not allow an unambiguous conclusion concerning the effects of differences in subjects' levels of initial interest in the present situation. On the basis of previous laboratory data on overjustification (e.g., Calder & Staw, 1975), one would expect overjustification effects to be more likely among subjects rewarded for engagement in activities of relatively higher (vs. relatively lower) initial interest. Instead, during the withdrawal phase, high-interest sub-

jects showed no less interest than control subjects, while low-interest subjects showed significantly less subsequent interest than control subjects. In part, this reversal stems from the unanticipated general drift exhibited by control subjects over the course of the experiment, in which initial preferences tended to diminish over time. Therefore, interpretation of these particular differences is more problematic than would have been the case had control subjects' preferences remained constant throughout the study. On the one hand, this difference between the high- or low-interest groups may reflect a genuine effect of the interest manipulation; on the other hand, it may equally well reflect the relative insensitivity of the present design for assessing changes which parallel the temporal trends in the control condition.

In particular, the results of a recent doctoral dissertation (Colvin, 1973) indicate that some caution is appropriate in interpreting the present failure of the high-interest condition to produce a decrement in subsequent interest relative to control subjects, since Colvin's study provides a conceptual replication of the high-interest and control conditions in the present study and obtains clear evidence of significant overjustification effects on both within-group and between-groups comparisons. Colvin's subjects were elementary school children who had been selected for their demonstrated interest in art. They were asked to participate in the study on a voluntary basis and then randomly assigned to one of two experimental conditions. Token reinforcement systems were introduced after baseline observations had determined the children's relative preferences between two sets of different art materials. In one group, children were contingently reinforced for time spent with their preferred materials; in the other group, chil-

⁶ These differences between group means were paralleled by differences in the number of individual subjects in each group whose time with target activities decreased from baseline to withdrawal. This decrease occurred for 8 of 11 high-interest subjects versus 6 of 9 control subjects' "high-interest" activities, 7 of 10 low-interest subjects versus 3 of 8 control subjects' "low-interest" activities, and 9 of 11 choice group subjects versus 3 of 9 control subjects' activities yoked to their choices.

dren were noncontingently reinforced after each session. During this treatment phase, contingent token rewards proved highly effective in altering subjects' choices. Then, after 4 weeks, the token systems were abruptly terminated for both groups. Although the two groups did not differ in the percentage of time they spent with their preferred materials during the baseline phase, the group which had been contingently reinforced while the token systems were in operation spent significantly less time with these materials during an explicit "extinction" phase than did the group which had been noncontingently reinforced. In the latter group, moreover, initial activity preferences were maintained throughout the experiment. These findings provide evidence that token reinforcement procedures can produce decreases in posttreatment engagement with previously reinforced activities among subjects selected for an initial high interest in the activities.

At the same time, it should be noted that Colvin's study and the present one differed in several respects. For example, his procedure included two sets of art activities while ours included four sets of math activities, and his reinforcers were art and play materials while ours were rewards associated with the school's ongoing math program. In addition, Colvin's subjects were volunteers from a University School while ours were a "captive audience" from a public school population of relatively lower socioeconomic levels. It remains for further investigation to determine the significance, if any, of these differences in populations or procedures between the two studies.

Interestingly, compared to the data from the high-interest and low-interest conditions, subjects in the choice condition of the present study showed dramatic decreases in subsequent interest, which were apparent in both within-group and between-groups comparisons. Indeed, during the 13-day withdrawal phase, choice subjects' posttreatment engagement in the previously rewarded activities showed no overlap with baseline data from this same group or posttreatment engagement levels displayed by control subjects. In this light, it is important to note that the experimental procedure which distinguished this condition

involved more than the provision of an opportunity for subjects to select the target activities for which rewards were to be provided. Specifically, in this condition subjects were asked individually to think about the activities and the reward program, and were asked to make a public statement of their desire to receive rewards contingent upon engagement in particular activities. Such a procedure, which elicits an overt commitment from the subject and explicitly directs the subject's attention to the consequences of his behavior (cf. Collins & Hoyt, 1972; Kiesler, Nisbett, & Zanna, 1969; Zanna, Lepper, & Abelson, 1973), is reminiscent of procedures employed in previous laboratory demonstrations of overjustification (e.g., Lepper et al., 1973), in which subjects are asked to make an explicit acknowledgment of the instrumentality of their engagement in a particular activity prior to undertaking the activity.⁷

Certainly, the possibility that variations in the manner in which token reward programs are presented may affect subsequent interest following the withdrawal of a relatively long-term reward program deserves further empirical investigation (cf. Feingold & Mahoney, 1975). It would seem particularly important to note that there may be two operationally distinguishable components or stages to a subject's response to reward manipulations. Most research has focused on parameters of reward (or tasks, or subjects), in the hope that a set of boundary conditions might be

⁷ Note that the salience of instrumentality can and should be distinguished from the salience of particular rewards or contingencies. The distinction is analogous to Kruglanski's (1975) distinction between explanations of voluntary actions versus explanations of events or occurrences. In the former case, naive attribution is purposive or teleological, partitioning reasons into ends and means; specifically, distinguishing between endogenously (intrinsically) and exogenously (extrinsically) motivated actions. In the latter case, naive attribution is causal, distinguishing between personal and environmental explanations (cf. Heider, 1958; Jones & Davis, 1965). The overjustification hypothesis, of course, addresses the means-end distinction rather than the person-environment distinction (cf. Lepper, Greene, & Nisbett, 1973, p. 130). Therefore, the theoretically crucial issue is what makes instrumentality salient, rather than what makes particular stimuli salient (cf. Kruglanski, 1975, pp. 402-405).

established for specific effects. A perhaps more basic question is, when do people engage at all in the kind of cognitive "work" postulated by attribution and self-perception theories? If variations in the manner in which token reward programs are presented in fact determine whether subjects ever think about their reasons for engaging in activities, a closer look at such variables may resolve apparent discrepancies in the literature.

Overjustification Effects and Contrast Effects

In language more common to the applied reinforcement literature, the present study and related investigations (e.g., Colvin, 1973; Johnson et al., 1976) provide direct evidence of "contrast" effects from token economy procedures. A contrast effect is said to occur whenever components of a multiple schedule of reinforcement interact, such that a change in behavior frequencies, produced by a change in one component of the multiple schedule, is accompanied by a change in behavior frequencies in the opposite direction under the other, unchanged component (Dunham, 1968; Freeman, 1971; Reynolds, 1961). Operationally, then, any response suppression relative to baseline, following multiple-trial reinforcement procedures, is appropriately labeled an instance of "behavior contrast," as an empirical description of the directionality of the posttreatment effect. When, as in the present study, a between-groups design is used, the effect is typically described as "incentive contrast" (Black, 1968; Cox, 1975; Dunham, 1968). It should be emphasized that contrast effects are defined empirically, in terms of reinforcement procedures and behavior frequencies, rather than theoretically, in terms of particular hypothetical constructs or processes. Thus, there is no particular reason to suppose that all contrast effects with human subjects should necessarily be amenable to the same theoretical explanation (cf. Johnson et al., 1976).

Overjustification effects, on the other hand, are defined in terms of a specific set of theoretical constructs, rather than any particular observable manifestation of them. Overjustification effects have been demonstrated, for example, as a consequence of the imposition

of a variety of extrinsic constraints, including "enforced rehearsal" (Rosenhan, 1969), surveillance (Lepper & Greene, 1975), externally imposed deadlines (Amabile, DeJong, & Lepper, 1976), and salient, expected rewards (e.g., Deci, 1971; Lepper et al., 1973; Ross, 1975), as well as multiple-trial contingent reinforcement procedures (Colvin, 1973; the present study). Similarly, the predicted consequences of these manipulations have been assessed by diverse dependent measures, including qualitative indices of task performance (e.g., Garbarino, 1975; Kruglanski, Friedman, & Zeevi, 1971; Lepper et al., 1973) and questionnaire instruments (e.g., Calder & Staw, 1975; Kruglanski, Alon, & Lewis, 1972), as well as behavioral measures of subsequent intrinsic interest (e.g., Deci, 1971; Lepper et al., 1973; Ross, 1975). It would appear, therefore, that overjustification effects and contrast effects should be characterized as constituting two conceptually distinct domains which partially overlap. Within the overjustification domain, behavioral effects which eventuate from multiple-trial contingent reinforcement procedures may be accurately described as instances of behavioral or incentive contrast. Conversely, within the contrast domain, the overjustification hypothesis affords one plausible theoretical account for at least some of the effects with human subjects (cf. Johnson et al., 1976).

Further study of the possible role of cognitive mediation in contrast effects would seem to be warranted by the current literature on attempts to achieve generalization from token economy programs. A recent scholarly review of this literature (Kazdin, 1975) offered the following appraisal:

Amid the enthusiasm over the progress already made in token economy research and the exciting trends, there remains a major void. There have been relatively few advances in developing a behavioral technology which can be used effectively to maintain behavior and to ensure transfer of training to settings where contingencies are not rigidly programmed. (p. 263)

In this context, the clear implication of an attributional perspective is to favor strategies for achieving generalization of treatment effects that focus on subjects' cognitions about

their reasons for engaging in target behaviors, rather than strategies that focus on programming the posttreatment environment (Greene, 1974; Kopel & Arkowitz, 1975; cf. Kazdin, 1975; Kazdin & Bootzin, 1972). From this perspective, generalization to nonprogrammed settings is more likely to occur when training procedures induce subjects to make endogenous rather than exogenous attributions (Kruglanski, 1975) about their reasons for engaging in target behaviors. Exemplary of such procedures are: (a) the use of minimal and naturally available rather than overly powerful and arbitrary reinforcers (e.g., O'Leary, Drabman, & Kass, 1973), (b) the use of various "fading" techniques in which extrinsic rewards are gradually phased out (e.g., Drabman, Spitalnik, & O'Leary, 1973), and (c) the use of self-control and self-reinforcement techniques to replace or supplant externally imposed reinforcement programs (e.g., Brownell et al., in press; Drabman et al., 1973).

In the present study, of course, theoretical objectives dictated some departures from informed token economy practice (cf. Kazdin, 1975; O'Leary, in press), which should be acknowledged in any attempt to evaluate the applied significance of the present findings. For example, to provide a situation in which either increases or decreases in relative interest would be apparent, the present study employed "normal" subjects and a limited set of four mathematics activities of comparable initial interest. Similarly, to assess the effects of the reward system per se—with other potentially confounding factors (e.g., feedback about one's competence) held constant—the present reinforcement contingencies were based on time spent with target activities rather than a performance-based criterion. Nor was any attempt made to withdraw the reward system gradually or otherwise induce subjects to attribute their behavior during the treatment phase to intrinsic factors.

Thus, the procedures employed in the present study do not constitute an optimal strategy for achieving generalization of treatment gains to unprogrammed settings. Had the program been designed to promote subjects' feelings of competence via performance-contingent reinforcement, or to enhance subjects'

feelings of personal responsibility for their behavior via extended "fading" procedures, for example, the results might well have been different. Considerable caution should therefore be exercised in extrapolating from the present study, and the present results should not be taken as evidence that sensitively designed token reinforcement systems will always, or even typically, produce decrements in subsequent intrinsic interest.

At the same time, however, the present study also differed from most existing token programs in its aim to distinguish between intrinsically and extrinsically motivated behavior. Consequently, there was no attempt made to maintain treatment gains by substituting either contingent social approval or other naturally available reinforcers for the withdrawn token system; on the contrary, every effort was made in the present study to eliminate such extrinsic incentives during the posttreatment phase. Indeed, to accomplish this objective, classroom personnel were kept "blind" to subjects' treatment conditions throughout the study, and were explicitly instructed not to provide any differential social reinforcement contingent on a child's choice of activities. From a theoretical point of view, these were the departures from typical token economy practice that merit the greatest attention. Given the applied objectives of token reinforcement programs, their implementation has rarely provided the theoretically appropriate conditions for testing the overjustification hypothesis (Greene, 1974; Lepper & Greene, 1976). But if the same reinforcement procedures may have quite different effects on intrinsic as opposed to extrinsic motivation, research presumed to evaluate token reinforcement programs should include measures capable of assessing intrinsic as well as extrinsic motivation.

Conclusions

A proper test of the relevance of the overjustification hypothesis to applied token reinforcement programs requires a between-groups design and an experimental setting in which time spent with an activity can reasonably be attributed to intrinsic (vs. extrinsic) motivation. The present study included these features in an applied reinforce-

ment program extending over 9 weeks in two public elementary school classrooms. In two of the three experimental conditions, demonstrably effective multiple-trial reinforcement procedures produced posttreatment decrements in intrinsic interest in previously reinforced activities, relative to nondifferentially reinforced control subjects. These findings demonstrate that typical token economy procedures are capable of producing overjustification effects under some conditions. Precise specification of these conditions awaits further research.

REFERENCE NOTE

1. Johnson, S. M. Personal communication, November 25, 1974.

REFERENCES

- Amabile, T. M., DeJong, W., & Lepper, M. R. Effects of externally imposed deadlines on subsequent intrinsic motivation. *Journal of Personality and Social Psychology*, 1976, 34, 92-98.
- Ayllon, T., & Azrin, N. H. The measurement and reinforcement of behavior of psychotics. *Journal of the Experimental Analysis of Behavior*, 1965, 8, 357-383.
- Bem, D. J. Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6). New York: Academic Press, 1972.
- Black, R. W. Shifts in magnitude of reward and contrast effects in instrumental and selective learning: A reinterpretation. *Psychological Review*, 1968, 75, 114-126.
- Brownell, K., Colletti, G., Ersner-Hershfield, R., Hershfield, S. M., & Wilson, G. T. Self-control in school children: Stringency and leniency in self-determined and externally-imposed performance standards. *Behavior Therapy*, in press.
- Calder, B. J., & Staw, B. M. Self-perception of intrinsic and extrinsic motivation. *Journal of Personality and Social Psychology*, 1975, 31, 599-605.
- Collins, B. E., & Hoyt, M. F. Personal responsibility-for-consequences: An integration and extension of the "forced compliance" literature. *Journal of Experimental Social Psychology*, 1972, 8, 558-593.
- Colvin, R. H. *Imposed extrinsic reward in an elementary school setting: Effects on free-operant rates and choices*. Unpublished doctoral dissertation, Southern Illinois University, 1973.
- Cox, W. M. A review of recent incentive contrast studies involving discrete-trial procedures. *The Psychological Record*, 1975, 25, 373-393.
- Deci, E. L. Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, 1971, 18, 105-115.
- Deci, E. L. *Intrinsic motivation*. New York: Plenum, 1975.
- Drabman, R. S., Spitalnik, R., & O'Leary, K. D. Teaching self-control to disruptive children. *Journal of Abnormal Psychology*, 1973, 82, 10-16.
- Dunham, P. J. Contrasted conditions of reinforcement: A selective critique. *Psychological Bulletin*, 1968, 69, 295-315.
- Feingold, B. D., & Mahoney, M. J. Reinforcement effects on intrinsic interest: Undermining the overjustification hypothesis. *Behavior Therapy*, 1975, 6, 367-377.
- Freeman, B. J. Behavioral contrast: Reinforcement frequency or response suppression? *Psychological Bulletin*, 1971, 75, 347-356.
- Garbarino, J. The impact of anticipated reward upon cross-age tutoring. *Journal of Personality and Social Psychology*, 1975, 32, 421-428.
- Greene, D. Immediate and subsequent effects of differential reward systems on intrinsic motivation in public school classrooms (Doctoral dissertation, Stanford University, 1974). *Dissertation Abstracts International*, 1974, 35, 4626B. (University Microfilms No. 75-6854)
- Greene, D. Comment upon Feingold and Mahoney's "Reinforcement effects on intrinsic interest: Undermining the overjustification hypothesis." *Behavior Therapy*, 1975, 6, 712-714.
- Heider, F. *The psychology of interpersonal relations*. New York: Wiley, 1958.
- Johnson, S. M., & Bolstad, O. D. Methodological issues in naturalistic observation: Some problems and solutions for field research. In L. A. Hamerlynck, L. C. Handy, & E. J. Mash (Eds.), *Behavior change: Methodology, concepts, and practice*. Champaign, Ill.: Research Press, 1973.
- Johnson, S. M., Bolstad, O. D., & Lobitz, G. K. Generalization and contrast phenomena in behavior modification with children. In E. J. Mash, L. A. Hamerlynck, & L. C. Handy (Eds.), *Behavior modification and families*. New York: Brunner/Mazel, 1976.
- Jones, E. E., & Davis, K. E. From acts to dispositions. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2). New York: Academic Press, 1965.
- Kazdin, A. E. Methodological and assessment considerations in evaluating reinforcement programs in applied settings. *Journal of Applied Behavior Analysis*, 1973, 6, 517-531.
- Kazdin, A. E. Recent advances in token economy research. In M. Hersen, R. M. Eisler, & P. M. Miller (Eds.), *Progress in behavior modification* (Vol. 1). New York: Academic Press, 1975.
- Kazdin, A. E., & Bootzin, R. R. The token economy: An evaluative review. *Journal of Applied Behavior Analysis*, 1972, 5, 343-372.
- Kelley, H. H. The processes of causal attribution. *American Psychologist*, 1973, 28, 107-128.
- Kiesler, C. A., Nisbett, R. E., & Zanna, M. P. On inferring one's beliefs from one's behavior. *Journal of Personality and Social Psychology*, 1969, 11, 321-327.

- Kopel, S. A., & Arkowitz, H. The role of attribution and self-perception in behavior change: Implications for behavior therapy. *Genetic Psychology Monographs*, 1975, 92, 175-212.
- Kruglanski, A. W. The endogenous-exogenous partition in attribution theory. *Psychological Review*, 1975, 82, 387-406.
- Kruglanski, A. W., Alon, S., & Lewis, T. Retrospective misattribution and task enjoyment. *Journal of Experimental Social Psychology*, 1972, 8, 493-501.
- Kruglanski, A. W., Friedman, I., & Zeevi, G. The effects of extrinsic incentives on some qualitative aspects of task performance. *Journal of Personality*, 1971, 39, 606-617.
- Lepper, M. R., & Greene, D. Turning play into work: Effects of adult surveillance and extrinsic rewards on children's intrinsic motivation. *Journal of Personality and Social Psychology*, 1975, 31, 479-486.
- Lepper, M. R., & Greene, D. On understanding "overjustification": A reply to Reiss and Sushinsky. *Journal of Personality and Social Psychology*, 1976, 33, 25-35.
- Lepper, M. R., Greene, D., & Nisbett, R. E. Undermining children's intrinsic interest with extrinsic rewards: A test of the overjustification hypothesis. *Journal of Personality and Social Psychology*, 1973, 28, 129-137.
- Levine, F. M., & Fasnacht, G. Token rewards may lead to token learning. *American Psychologist*, 1974, 29, 816-820.
- Levine, F. M., & Fasnacht, G. Reply. *American Psychologist*, 1976, 31, 90-92.
- Meichenbaum, D. H., Bowers, K. S., & Ross, R. R. Modification of classroom behavior of institutionalized female adolescent offenders. *Behavior Research and Therapy*, 1968, 6, 343-353.
- O'Leary, K. D. Token reinforcement programs in the classroom. In T. Brigham & C. Catania (Eds.), *The analysis of behavior: Social and educational processes*. New York: Irvington-Naiburg/Wiley, in press.
- O'Leary, K. D., Becker, W. C., Evans, M. B., & Saudargas, R. A. A token reinforcement program in a public school: A replication and systematic analysis. *Journal of Applied Behavior Analysis*, 1969, 2, 3-13.
- O'Leary, K. D., & Drabman, R. Token reinforcement programs in the classroom: A review. *Psychological Bulletin*, 1971, 75, 379-398.
- O'Leary, K. D., Drabman, R. S., & Kass, R. E. Maintenance of appropriate behavior in a token program. *Journal of Abnormal Child Psychology*, 1973, 1, 127-138.
- Reiss, S., & Sushinsky, L. W. Overjustification, competing responses, and the acquisition of intrinsic interest. *Journal of Personality and Social Psychology*, 1975, 31, 1116-1125.
- Reynolds, G. S. Behavioral contrast. *Journal of the Experimental Analysis of Behavior*, 1961, 4, 57-71.
- Rosenhan, D. Some origins of concern for others. In P. A. Mussen, J. Langer, & M. Covington (Eds.), *Trend and issues in developmental psychology*. New York: Holt, Rinehart & Winston, 1969.
- Ross, M. Salience of reward and intrinsic motivation. *Journal of Personality and Social Psychology*, 1975, 32, 245-254.
- Ross, M. The self-perception of intrinsic motivation. In J. H. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research*. Hillsdale, N.J.: Erlbaum, 1976.
- Ross, M., Karniol, R., & Rothstein, M. Reward contingency and intrinsic motivation in children: A test of the delay of gratification hypothesis. *Journal of Personality and Social Psychology*, 1976, 33, 442-447.
- Sidman, M. *Tactics of scientific research*. New York: Basic Books, 1960.
- Zanna, M. P., Lepper, M. R., & Abelson, R. P. Attentional mechanisms in children's devaluation of a forbidden activity in a forced-compliance situation. *Journal of Personality and Social Psychology*, 1973, 28, 355-359.

(Received March 29, 1976)