Journal of Personality and Social Psychology
1971, Vol. 18, No. 1, 105–115

EFFECTS OF EXTERNALLY MEDIATED REWARDS ON INTRINSIC MOTIVATION[1]

EDWARD L. DECI[2]

University of Rochester

Two laboratory experiments and one field experiment were conducted to investigate the effects of external rewards on intrinsic motivation to perform an activity. In each experiment, subjects were performing an activity during three different periods, and observations relevant to their motivation were made. External rewards were given to the experimental subjects during the second period only, while the control subjects received no rewards. Of interest was the difference in the experimental group's motivation between Period 1 and Period 3, relative to the difference in the control's. The results indicate that (*a*) when money was used as an external reward, intrinsic motivation tended to decrease, whereas (*b*) when verbal reinforcement and positive feedback were used, intrinsic motivation tended to increase. Discrepant findings in the literature were reconciled using a new theoretical framework which employs a cognitive approach and concentrates on the nature of the external reward.

If a boy who enjoys mowing lawns begins to receive payment for the task, what will happen to his intrinsic motivation for performing this activity? Or, if he enjoys gardening and his parents seek to encourage this by providing verbal reinforcement and affection when he gardens, what will happen to his intrinsic motivation for gardening? These are examples of the classical problem concerning the effects of external rewards on intrinsic motivation.

One is said to be intrinsically motivated to perform an activity when he receives no apparent rewards except the activity itself. This intrinsic motivation might be either innate or learned (White, 1959). It is not the purpose of this study to deal with the specific nature of, or development of, intrinsic motivation, but rather, it assumes that at a given time a person can be intrinsically motivated to do an activity, and it then asks: What are the effects of external rewards on this motivation?

In the two examples of the boy, he is performing the activity for no apparent rewards except the enjoyment he experiences in the activity or the feelings of satisfaction he gets from the activity. External rewards are then introduced. Will his intrinsic motivation for the activity decrease, increase, or remain the same? Even though this problem is discussed in many psychology books—the conclusion being that external rewards decrease intrinsic motivation (e.g., Atkinson, 1964; deCharms, 1968; Murray, 1964)—a review of the literature reveals that it has yet to receive a definitive empirical answer.

Consider first the possibility that external rewards do decrease intrinsic motivation. deCharms (1968) proposed that when external rewards are given for an intrinsically motivated activity, the person perceives that the locus of control or the knowledge or feeling of personal causation shifts to an external source, leading him to become "a pawn" to the source of external rewards. His behavior is then motivated by the external reward, rather than by his intrinsic interest. Similarly, Festinger (1967) stated that external rewards affect the person's concept of why he is working and his attitude toward the work. Reasoning from his theory of cognitive dissonance (Festinger, 1957), he also predicted that external rewards should decrease intrinsic motivation. Both deCharms and Festinger cited the work of Harlow and his colleagues in support of their conclusions. Since the Harlow studies are frequently cited in discussions of

[1] These studies were conducted at Carnegie-Mellon University. The author wishes to thank Victor H. Vroom and Daryl J. Bem for helpful suggestions about the research and about earlier drafts of the manuscript.

[2] Requests for reprints should be sent to the author, Department of Psychology, University of Rochester, River Campus Station, Rochester, New York 14627.

this topic, it is appropriate to review them in some detail.

In the first experiment, Harlow, Harlow, and Meyer (1950) found that monkeys will work on a "puzzle apparatus" over an extended period for no apparent reward except the activity itself. They accounted for this by positing a manipulation drive which can be categorized as an intrinsic motive. The four experimental and four control monkeys were observed manipulating the apparatus for portions of 13 days. On the fourteenth day, each experimental monkey was placed under 22 hours of food deprivation and observed for two 5-minute periods, at the beginning of which the apparatus was baited with a raisin. The control animals performed for no food rewards.

The experimenters reported a disruption effect for the food group; that is, there were more errors and fewer solutions in the food than in the no-food group. The errors were, however, of a systematic nature; the monkeys tried first to manipulate the piece closest to the food. Immediately after the second 5-minute period with food, the apparatus was reassembled without food in the presence of each monkey, and he was observed for another 5-minute period. There were fewer successful manipulations during the last 5-minute period than during a similar 5-minute period for a nonfood group. However, the experimental monkeys did perform better than they had in the two 5-minute food sessions.

In another study from the Harlow laboratory, Gately (1950) observed four monkeys for periods of 5 minutes each and noted that when their puzzles were baited with one raisin at the beginning of each period, they responded more in the first minute of the 5-minute period than in the last minute.

Although these two studies do appear to support the hypothesis that extrinsic rewards decrease intrinsic motivation, there are several factors which weaken that conclusion. There were very few subjects used in each study. Also, periods of observation were only 5 minutes each, and the disruption could have been a short-term effect caused by one of the following processes. First, for example, the monkeys had been food deprived for 22 hours, so adding the motivation of food to the intrinsic motivation may have put them past the optimal point of the Yerkes-Dodson (1908) inverted-U relationship between motivation and performance. A second possibility is that the sight of food for the deprived animal could have increased emotionality, such that the decreased activity was a manifestation of a "recovery" period. Still another possibility is that the monkeys were merely relaxing to digest their food.

Furthermore, a third study in the Harlow series (Davis, Settlage, & Harlow, 1950), which is discussed below, actually found an increase in performance during the third period. So, the direct evidence supporting the prediction that external rewards decrease intrinsic motivation is rather weak.

There is some indirect evidence occasionally cited which supports a related hypothesis, namely that *insufficient* external rewards will lead to *enhanced* intrinsic motivation (for example, Aronson & Mills, 1959; Weick, 1964, 1967a; Zimbardo, Weisenberg, Firestone, & Levy, 1965). However, this evidence is still somewhat equivocal (see Weick, 1967a, for a review), and there is no persuasive theoretical argument which indicates that the processes causing this phenomenon would also cause a decrease in intrinsic motivation when sufficient external rewards are present.

In summation, then, a review of the available evidence, both direct and indirect, reveals that the support for the prediction that external rewards decrease intrinsic motivation is not substantial.

The second possible effect of external rewards on intrinsic motivation is that of enhancement. Woodworth (1918) has suggested that many activities, regardless of the initiating motive, become intrinsically interesting. Allport (1937) gave the name "functional autonomy" to the Woodworth notion.

Learning theorists in the behavioral tradition would also make this prediction and claim that continued pairing of an activity with a reward would cause the activity to become a secondary reinforcer and strengthen the rewarding property of the activity (see

Keller, 1969). Ferster and Skinner (1957), Morse (1966), and Uhl and Young (1967) have demonstrated that animal behavior strengthened by intermittent food rewards persists after the food is removed.

Further, in the third experiment from the Harlow laboratories mentioned above (Davis et al., 1950), the researchers obtained results which seem to support the enhancement of intrinsic motivation by external rewards. For this study they used two normal and six brain-lesioned monkeys in the experimental group and the same number of each in the control group. During the first and third periods, the apparatus contained no food, whereas during the second it was baited with one raisin every 10 minutes for the experimental but not for the control group. The authors found that after a temporary disruption following the introduction of food, performance increased for the experimental group. During the third period when once again there were no external rewards, the performance continued to increase.

The work of developmental psychologists interested in the process of internalization would also suggest that external rewards will enhance or create intrinsic motivation. Aronfreed (1968) summarized a large number of studies (primarily with animals) which reported the persistence of behavior after the rewards had been removed. The studies were presented as support for the development of internal control of behavior through the continued reinforcement (generally intermittent) of the desired behavior. This leads developmental psychologists in the social learning tradition to state that "Young children would quickly acquire dispositions to maintain certain forms of behavior for considerable periods of time without reinforcing consequences [Aronfreed, 1968, p. 23]."

The final alternative, that extrinsic rewards do not affect intrinsic motivation, is not a theoretically derived propostion, but simply serves as the null hypothesis in the other experiments already discussed.

It seems clear, after reviewing the literature, that there is no definitive answer to the question of the effects of externally mediated rewards on intrinsic motivation. Nevertheless, the pattern of findings suggests two propositions which may serve to reconcile the conflicting results.

First, the finding of continued activity after removal of rewards in studies with subhuman species may be accounted for more parsimoniously by the process of resistance to extinction than by an increase in intrinsic motivation. It is suggested that human data might be more fruitfully viewed using a cognitive approach, since humans have greater cognitive control over their own behavior and motivation. A cognitive approach to this problem would focus on the changes in the phenomenological interpretation of the task following the introduction of external rewards.

Second, it is suggested that distinctions should be made among the different kinds of external rewards. If a person's cognitive evaluation of different external rewards is different, it is possible that different rewards would have different effects on the person's intrinsic motivation. It may be that money and closely related tangible rewards have some peculiar property associated with them which affects intrinsic motivation differently from nontangible rewards.

Specifically, it is suggested that money is frequently used as a means of "buying" services which would probably not otherwise be rendered. Perhaps, then, the presence of money as an external reward suggests to the subjects that they "should probably not render this activity without pay," that is, they should not be so intrinsically motivated to do the activity.

This could lead the subjects to a process of cognitive reevaluation of the activity from one which is intrinsically motivated to one which is motivated by the anticipation of money.

If it is true that money acts differently from other external rewards, then the findings related to internalization need not be inconsistent with the predicted decrease in intrinsic motivation caused by money.

In experiments with children, the external rewards tend to be verbal reinforcement, social approval, and so on. These external rewards are less likely to be perceived by the person as controlling his behavior, so there would be no stimulus to initiate the process of cognitive reevaluation.

The difference, then, between money and social approval as external rewards lies in the person's perception of the locus of control of his behavior. While with money he could readily come to accept the reward as the reason for his behavior, he is less apt to do this when the rewards are verbal reinforcements. In the latter case, the approval rewards may not be phenomenologically distinct from the feelings of satisfaction which the person gets from performing the activity.

The research to be reported in this study investigated the effects of external rewards on intrinsic motivation from the point of view of the new framework just presented. It was first concerned with the effects of money on a person's intrinsic motivation. Hypothesis I is based on the frequently made assertion that external rewards decrease intrinsic motivation to perform a task, and on the conception of money which many people are thought to hold.

Hypothesis I

If a person is engaged in some activity for reasons of intrinsic motivation, and if he begins to receive the external reward, money, for performing the activity, the degree to which he is intrinsically motivated to perform the activity decreases.

The second hypothesis is based on the work of the developmental learning theorists. It follows from the framework presented above and deals with social approval as an external reward.

Hypothesis II

If a person is engaged in some activity for reasons of intrinsic motivation, and if he begins to receive external rewards in the form of verbal reinforcement and positive feedback for performing the activity, the degree to which he is intrinsically motivated to perform the activity is enhanced.

These two hypotheses were tested by three experiments, two in the laboratory and one in the field. The general paradigm was the same for all three. The behavior of subjects was observed during three different periods. First, they were performing at an operant level for no apparent external reward. Then in the second period, the experimental subjects were rewarded for the activity, while the controls received no reward; and finally rewards were halted, and the level of activity continued to be observed.

EXPERIMENT I

Method

The subjects in this laboratory experiment were introductory psychology students, who were fulfilling a course requirement by participating.

Each of the 24 subjects was told at the time he was recruited that he would have to participate in three separate sessions of not more than 1 hour each on 3 different days. Twelve of the subjects were in the control group and 12 in the experimental group. All of the subjects drawn from any one section of introductory psychology were in the same condition, to minimize the possibility of their finding out that there were two conditions to the experiment. However, subjects were taken from six different sections, to minimize the possibility that there would be confounding due to section.

During the three sessions of the experiment, the subjects were engaged primarily in working on a puzzle called Soma, produced by Parker Brothers. This task was selected because it seemed that most college students would be intrinsically motivated to do it. The puzzle is made up of seven pieces of plastic. Each piece is made to look like it is composed of three or four 1-inch cubes, and there are a total of twenty-seven 1-inch cubes in the seven pieces. These seven pieces can be fitted together to form millions of different configurations—including, for example, a cube 3 inches on each side.

During the experimental sessions, the subjects were asked to reproduce, using the puzzle pieces, various configurations which had been drawn on paper. During each session, a subject worked on four such configurations. The time to complete each configuration was measured with a stopwatch, and if a subject was unable to do a configuration within 13 minutes, he was stopped and shown how to do it. This allowed him to see that all of the configurations were possible. At the end of each configuration, subjects were told how long it had taken them to do it.

During the first of the three sessions for each subject, the experimental and control people did the same thing. After they entered the experimental room, they were seated at a table with the puzzle pieces in front of them, three drawings of configurations to the right of them, the latest issues of *New Yorker, Time,* and *Playboy* to their left, and the experimenter on the opposite side of the table. They were told that they would spend all three sessions using the pieces of plastic to form various configurations, such as the sample they were shown.

During the second session, the experimental people were paid $1 for each configuration which they were able to reproduce within the 13-minute time

limit, whereas the control people were given the same configurations without pay. The experimental people were merely told at the beginning of the second session that the format of the session would be the same as it had been in the first, but that they would receive $1 for each puzzle they solved within the 13-minute time limit.

Finally, in the third session both groups were given more configurations, but neither group received pay. The experimental subjects were told that they would not be paid for the third session because there was only enough money available to pay them for one of the sessions.

Each subject was given the same 12 configurations, 4 during each session. The three sets of 4 configurations each remained constant; however, an ordering of the three sets over the three sessions was made so as to have an equal number of each of the six permutations in the control and experimental groups.

To summarize, each subject was in three sessions working on puzzle configurations. The only difference between the control and the experimental groups was that the experimental subjects were paid for their performance during Session 2. Therefore, during the second session the experimental subjects were performing both for intrinsic motives and for external rewards.

To obtain the measure of motivation used in this study, the experimenter left the room for 8 minutes in the middle of each session. The pretext was as follows. The subjects were told that the experiment was designed to investigate several concepts related to problem solving. Different concepts were associated with different configurations. The experimenter would give the subject two configurations and then, based on the data from these, determine what were the appropriate configurations for the rest of the hour, in order to look at the concepts of interest. The way of determining the appropriate configurations was to feed the data into a computer through a teletype and have a program select the configurations. To do this it was, of course, necessary for the experimenter to leave the room.

As he left the room he said, "I shall be gone only a few minutes, you may do whatever you like while I am gone." They could read magazines, work on the puzzle, stare around the room, and so on.

The primary measure of motivation was the amount of time during the 8-minute free choice situation which they spent working on the puzzle. This was determined by the experimenter, who observed and timed them through a one-way window while the subjects thought he was at the teletype.

Since any subject who was unable to do a configuration within the 13 minutes allowed was shown the solution, the possibility that the Zeigarnik (1927) effect would influence whether or not he worked on the puzzle in the 8-minute free choice period was minimized.

The configurations whose drawings were on the table during the entire experiment were impossible ones to do. This precluded the possibility that a subject would finish a configuration and have that as a causal factor in determining whether or not he continued working on the puzzle.

The criterion selected for whether or not a subject was working on the task during the free choice period was that he be manipulating and looking at one or more of the pieces. If a subject was reading or daydreaming and had a hand on a puzzle piece, he was judged not to be working on the puzzle.

At the end of each session, the subjects were asked to rate on a 9-point scale the degree to which they found the task interesting and enjoyable. This information was collected to verify empirically the assumption that the subjects were intrinsically interested in the task.

Results

The first hypothesis posits that when money is used as an external reward for an activity, one's intrinsic motivation for that activity decreases. Experiment I has investigated this proposition.

The subject's motivation for the activity was measured by the number of seconds he spent on the activity during the free choice period. Table 1 presents the average number of seconds spent on the puzzle by the experimental and control groups during the three time periods.

As one would expect, when the external rewards were introduced to the experimental people during Time 2, their motivation increased. They spent an average of 67.7 seconds more working on the puzzle during the free choice period in Session 2 than in Session 1, while the control people spent about the same amount of time in Session 2 as in Session 1.

Then as Hypothesis I predicts, when the rewards were removed for Time 3, motivation

TABLE 1

MEAN NUMBER OF SECONDS SPENT WORKING ON THE PUZZLE DURING THE EIGHT-MINUTE FREE CHOICE PERIODS

Group	Time 1	Time 2	Time 3	Time 3 − Time 1
Experimental (n = 12)	248.2	313.9	198.5	−49.7
Control (n = 12)	213.9	205.7	241.8	27.9
		$E(T_3 - T_1) - C(T_3 - T_1)$		−77.6 sec.* (SE = 58.5)

Note.—The higher the score, the higher the motivation.
* $p < .10$, $df = 22$, one-tailed t test.

dropped to a level considerably lower than it had been during Time 1. Since each subject was his own control, in that his behavior in Period 3 was compared to his behavior in Period 1, and since there was a control group doing the same activity for no pay to control for extratreatment effects such as boredom and learning, the appropriate statistic for testing Hypothesis I is: experimental (Time 3 motivation − Time 1 motivation) minus control (Time 3 motivation − Time 1 motivation). Hypothesis I predicts that this will be significantly negative.

The test of the hypothesis is also presented in Table 1. The value of the statistic $E(T_3 - T_1) - C(T_3 - T_1)$ is 77.6 seconds, which, when tested by the Student t distribution, yields significance at the .10 level. While this does not reach the customary .05 level, it does nevertheless give some support to the hypothesis. It is also possible that a ceiling effect was obscuring the strength of the experimental effect. Two of the experimental subjects spent all 8 minutes (480 seconds) of free choice time in each of the three sessions working on the puzzle, and a third experimental subject spent more than 7 minutes, 45 seconds of each session, working on the puzzle.

The assumption that the task used in the experiment would be intrinsically interesting for the subjects was correct. At the end of each session, the subjects rated the degree to which they found the task interesting and enjoyable on a 9-point scale, where 9 represents "extremely interesting and enjoyable," 5 represents "neutral," and 1 represents "extremely uninteresting and unenjoyable."

At all three sessions, both the experimental and control subjects found the task interesting and enjoyable. The session averages ranged from 7.25 to 8.00 and were not significantly different between the groups or among the sessions.

The data from Experiment I do offer some support for Hypothesis I. There did seem to be a decrease in intrinsic motivation for the activity following the experience with monetary rewards. However, there were only 24 subjects in the experiment, and the significance was only at the .10 level, so additional evidence on the validity of the finding is needed.

EXPERIMENT II: A FIELD REPLICATION

Method

The setting for this experiment was a biweekly college newspaper. The subjects were eight students who worked on two four-man staffs of headline writers. One staff worked for the Tuesday edition and the other for the Friday. At no time were the two groups brought together. Throughout the 5-month period when the experiment was in progress, the students remained unaware that they were being subjects.

During the fall semester of 1969, the last 19 issues of the paper (10 Tuesdays and 9 Fridays) were used as part of the experiment. The tenth Friday edition was canceled unexpectedly for reasons unrelated to the experiment. The first 6 issues (3 of each) were not used, so that the subjects could have an opportunity to become accustomed to the task. The task simply involved writing headlines according to prescribed rules. The students, all freshmen and sophomores, reported at a stated time to a room which had been designated for the headline staff, and they worked until all the headlines had been written.

The Tuesday staff served as the experimental group, and the Friday staff, as the control group. The 10 weeks were broken into three periods, with 4 weeks in the first period and 3 weeks in each of the other two.

During all three periods, the subjects wrote headlines, and although they did not know it, the time it took them to write each headline was being recorded. The experimenter, who was thought by everyone except the editor-in-chief to be a staff member, was in the room as the "supervisor" and surreptitiously recorded the time it took for each headline to be written.

The performance on the headlines was used as the measure of motivation. It was assumed that the more quickly someone performed, the more highly motivated he was to do this task.

Absences were also recorded and used as a measure of attitudes. This measure is commonly used in research on organizations (Ross & Zander, 1957).

The only difference between the experimental group and control group was that the experimental group was paid 50¢ per headline written during the second period. They were told that the paper's budget was divided into funds, such that certain amounts had to be used for certain things, and one of the funds had money in it which had to be used up by the end of the semester, so the editor-in-chief had decided to use it to pay the production staff. They were cautioned not to tell others, since it might cause animosity from people on the other staffs who were not being paid. After 3 weeks had passed, they were told that all the money had been exhausted and that they would no longer be paid.

The experimental design was identical to that of the preceding laboratory experiment, and the data were analyzed in the same way to test Hypothesis I. The major difference in the two experiments was that motivation was measured differently. Performance scores were used in the second experiment because it was done in the field and the experimenter did not have enough control to set up an exactly analogous measure.

The field experiment was carried one step further than the laboratory experiment in order to investigate the stability of whatever effect might appear. At the beginning of the second semester, which began 5 weeks after the end of Time Period 3, the staffs began working again. For the first 2 weeks of this semester, times for writing headlines were again recorded and used as the Time 4 measure of motivation.

Results

The advantages and disadvantages of field versus laboratory experiments are well-known and widely discussed. Field experiments are conducted in real, ongoing situations which are often free of demand characteristics (Orne, 1962) and the first day at work syndrome (Weick, 1967b); however, all field experiments share the problem of loss of some control. This was no exception.

Since the fourth period was conducted during a different semester than the first three, and since the subjects had new class schedules which conflicted with headline-writing sessions, it was not possible to record absences meaningfully. Therefore, the attitude score was available for the first three periods only, although the motivation scores were available for all four periods.

About halfway through the first semester, in Period 2, one of the control subjects was given a new and more important assignment on the newspaper, so she had to give up writing headlines. Since the editor who gave the assignment was unaware that the experiment was being conducted, there was no way to prevent this move.

Further, another of the control subjects was absent during both sessions of Time 3, so there was no motivation data for him from that period. He was, of course, included in the absences analysis, but could not be included in the motivation analysis. Also, during the second semester he had a schedule conflict which prevented him from working on the staff, so there were no Time 4 measures for him.

Since the experiment was begun with four control subjects, there were three included in the absences analysis and two in the motivation analysis.

One of the experimental subjects was absent during all three sessions of Period 3; however, he did attend the sessions in Time 4. It therefore seemed useful to use his motivation data, since it was available for Time 4. Consequently, to get a Time 3 motivation score for him, a weighted average of his first two periods was used. The relation of this score to the Time 1 score is opposite to what Hypothesis I predicts, so the scores of the other subjects would have to be even greater in the predicted direction to substantiate the hypothesis.

Table 2 presents the data for the motivation scores in the field experiment and the test of Hypothesis I. The general trend in the control group during the study was an increase in performance—with a slight reversal at Time 4. While the experimental group showed the same increase from Time 1 to Time 2, this trend did not continue after the

TABLE 2

MEAN NUMBER OF MINUTES SPENT TO WRITE EACH HEADLINE

Group	Time 1	Time 2	Time 3	Time 4	Time 3 − Time 1	Time 4 − Time 1
Experimental ($n = 4$)	22.39	20.34	21.35	20.05	−1.04	−2.34
Control ($n = 2$)	22.19	20.97	12.60	13.79	−9.59	−8.40
				$E - C$	8.55 min.** ($SE = 1.99$)	6.06 min.* ($SE = 3.08$)

Note.—The lower the score, the higher the motivation.
* $p < .10$.
** $p < .01$, $df = 4$, one-tailed t test.

reward of money had been removed. For testing Hypothesis I, the statistic $E(T_3 - T_1) - C(T_3 - T_1)$ is 8.55 minutes, which is significant at the .01 level.

One might wonder, since two of the subjects in the control group did not work during the third period, whether this opens the possibility that the difference which appeared (as predicted by Hypothesis I) may be partially due to the fact that the control group had to work faster (since there were fewer of them) to get all the headlines written in a reasonable amount of time. That is, part of the difference may have been due to the fact that the number of people present was a covariate. However, this was not the case. The average time to write each headline for the experimental and control groups was computed as a function of the number of people present during the session. The data show quite clearly that the number of people present does not act as a covariate with the performance scores.

The appropriate statistic for determining whether the effect continues over time is $E(T_4 - T_1) - C(T_4 - T_1)$. The value of this (as shown in Table 2) is 6.06 minutes, which is significant at the .10 level.

These data give support to the first hypothesis and indicate that the decreased intrinsic motivation in the experimental group relative to the control group is more than just a temporary phenomenon.

Table 3 presents the percentages of absences from sessions during each of the first three periods in the experimental and control groups. The denominator of each fraction is equal to the number of subjects times the number of sessions in that period. Absences increased for both groups during Time 2. One might expect that the introduction of pay would have caused the experimental group's absences to decrease (i.e., their attitudes or satisfaction to increase). However, this was not the case. Nevertheless, at Time 3 after the pay had been removed, the experimental group's absences increased considerably, while the control group's decreased slightly. For testing the attitudinal counterpart (or satisfaction counterpart) of Hypothesis I, the same statistic (now in percentages) is appropriate. This yields a value of 47%, which is significant at the .10 level ($df = 5$).

TABLE 3

PERCENTAGE OF SUBJECTS ABSENT FROM HEADLINE-WRITING SESSIONS

Group	Time 1	Time 2	Time 3	Time 3 − Time 1
Experimental ($n = 4$)	15	30	75	60
Control ($n = 3$)	20	44	33	13
		$E(T_3 - T_1) - C(T_3 - T_1)$		47* ($SE = 31$)

Note.—The denominators of the fractions yielding these percentages are equal to the number of subjects times the number of sessions in each of the time periods.
* $p < .10, df = 5$.

The results of the first two experiments give support to Hypothesis I. Money does seem to negatively affect one's intrinsic motivation for an activity as suggested by Harlow (1953) and predicted by Festinger (1967) and deCharms (1968). It is now appropriate to turn to a test of Hypothesis II, which states that intrinsic motivation increases if the interpolated external reward takes the form of social or verbal reinforcement rather than money.

EXPERIMENT III

Method

The design for this experiment was virtually identical to the first laboratory experiment. Again, there were 12 subjects in the experimental group and 12 in the control group, all of whom were from an introductory psychology class. This experiment was done during a different semester, however, when the class had a higher percentage of engineering students and a lower percentage of fine arts students.

Finally, the rewards used in this experiment were verbal rather than monetary. Thus, at the beginning of the second session, the experimental subjects were told that they had done very well on the puzzles during the first session, and at the end of each puzzle which they solved, they received verbal reinforcement (e.g., "That's very good") and positive (usually false) feedback (e.g., "That's much better than average for this configuration"). In the event that a subject did not solve a configuration, he was told that it was one of the most difficult of all the configurations used and that most people were unable to solve it, so he had not done badly.

During the third session the experimental group, like the control group, received no reinforcement. As in the first experiment, the measure of motivation was the number of seconds spent on the puzzle during the 8-minute free choice period.

In summary, both the experimental and control people solved puzzles for three sessions as in Ex-

TABLE 4

MEAN NUMBER OF SECONDS SPENT WORKING ON THE PUZZLE DURING THE EIGHT-MINUTE FREE CHOICE PERIODS

Group	Time 1	Time 2	Time 3	Time 3 − Time 1
Experimental ($n = 12$)	134.0	146.7	129.3	−4.7
Control ($n = 12$)	246.8	146.1	64.7	−182.1
	$E(T_3 - T_1) - C(T_3 - T_1)$			177.4 sec.* (SE=88.1)

Note.—The higher the score, the higher the motivation.
* $p < .05$, $df = 22$, one-tailed t test.

periment I. The main difference was that the experimental group received verbal reinforcement and positive feedback during Session 2. The expectation from Hypothesis II was that the experimental group's intrinsic motivation in Session 3 minus its intrinsic motivation in Session 1, as compared to the control's, would be significantly positive.

Results

Table 4 reports the average number of seconds of free choice time spent on the puzzle by the experimental and control groups during the three periods. For the control subjects there was a steady decrease in motivation scores over the three periods. The experimental subjects, however, showed no decrease in motivation for this task over the three sessions.

The statistic $E(T_3 - T_1) - C(T_3 - T_1)$ is again appropriate for testing Hypothesis II, which predicts that it will be significantly greater than zero. The test of this hypothesis also appears in Table 4. The value of the statistic is 177.4 seconds, which is significant at the .05 level and gives support to the hypothesis.

During Time 1, the scores for the experimental group were lower than those for the control group, even though the subjects were randomly assigned to conditions and the first session was run blind. Since there may have been regression toward the mean or a boredom effect (either of which could cause a difference in the predicted direction, given that the initial scores were not the same), it would be useful to remove the variance attributable to the initial score and then test Hypothesis II again. This can be done by multiple regression for difference scores using the following model:

$$\text{Diff.} = b_0 + b_1 \text{ (Treatment)} + b_2 \text{ (Initial)} + E.$$

Treatment assumes the value 0 for control subjects and 1 for experimental subjects, and initial is the Time 1 score.

After b_2 (the variance accounted for by initial scores) is removed, b_1 is 76.5, which is significant at the .10 level. So, removing the variance associated with the initial score lowers the significance, but still there is some support for the hypothesis.

Another interesting aspect of the data is found in the results for the control groups. The control groups did the same thing in both studies. In the first, the scores remained nearly the same over the three periods, but in the present study there was a marked decrease over the three periods. It is difficult to account for this, although the difference in the subject population which was mentioned earlier may be responsible. Thus, in the control group for the first experiment, there were 7 people in the arts and 5 in the technical fields. For the arts subjects, the Time 1 and Time 3 scores were 134.3 and 208.7, respectively, and for the technical subjects, the scores were 323.2 and 288.0, respectively. Judging from this small amount of data, the arts students seem to spend more free choice time working on the puzzle in Session 3 than in Session 1, relative to the technical students. In the control group for the third experiment, there were only 2 arts students and 10 technical ones. Hence, one finds the general pattern for the whole group (that is, Time 3 lower than Time 1) which appeared in the technical subgroup of Experiment I. When all of the data for the two lab experiments are broken down into arts versus technical students, the same general difference between the two categories is apparent. This appears in Table 5.

The statistic [Technical (Difference 3 − 1) − Arts (Difference 3 − 1)] is 79.9. This gives a t value of 1.33, which is significant at the .10 level, one-tailed. So this indicates that perhaps there is a difference between technical and arts students on this task. The dif-

TABLE 5

MEAN NUMBER OF SECONDS OF FREE CHOICE TIME SPENT WORKING ON THE PUZZLES BY TECHNICAL AND ARTS STUDENTS AT TIME 1 AND TIME 3 IN BOTH LABORATORY EXPERIMENTS

Experiment	Arts			Technical		
	Time 1	Time 3	*n*	Time 1	Time 3	*n*
Monetary (Exp. I)	277.0	230.7	7	207.8	153.4	5
Verbal (Exp. III)	79.5	304.5	2	144.9	94.3	10
Control (Exp. I & III)	195.2	175.5	9	251.4	139.8	15
Total	214.2	211.3	18	208.6	126.9	30

ference, if it is general, may be explained by the fact that technical students are more at ease with the puzzles because of their backgrounds, so they use it a lot at first, but become less motivated as they feel they are mastering it. On the other hand, the arts students would not be so at ease with it, and they would feel less motivated at first because they are not as good at it; however, as they get better they become more motivated. If this pattern holds across large samples, it would explain a large part of the difference between the two control groups. It would not, however, vitiate the general findings of these experiments, since the mix of arts and technical students was the same in the experimental and control groups *within* each of the experiments.

The other measure in the third experiment (taken at the end of each session) was the subjects' ratings on a 9-point scale of the degree to which they found the task interesting and enjoyable. The scores were virtually identical to those for the first laboratory experiment. Again, there were no significant differences among the three periods or between the two groups. The average scores for the periods ranged from 7.42 to 8.25.

DISCUSSION

This series of studies has been conducted to investigate the effects of externally mediated rewards on intrinsic motivation, and to test a framework reconciling what appeared to be contradictory findings from past experiments. The picture may be coming clearer, since the experiments gave support to the two hypotheses in this paper. The first two experiments suggest that when money is used as an external reward for some activity, the subjects lose intrinsic motivation for the activity. On the other hand, the third experiment suggests that when verbal reinforcement and positive feedback are used as the external rewards, the subjects' intrinsic motivation seems to increase relative to the nonrewarded subjects'.

It appears that money—perhaps because of its connotation and use in our culture—may act as a stimulus which leads the subjects to a cognitive reevaluation of the activity from one which is intrinsically motivated to one which is motivated primarily by the expectation of financial rewards. In short, money may work to "buy off" one's intrinsic motivation for an activity. And this decreased motivation appears (from the results of the field experiment) to be more than just a temporary phenomenon.

On the other hand, rewards such as social approval do not seem to affect a person's phenomenology in the same way. He will continue to be intrinsically motivated, since he is less likely to think of affection or verbal approval as a control mechanism.

It would now be useful to reconsider the experiments reviewed in the Introduction to see whether their findings are consistent with this new framework.

The Davis et al. (1950) results seem best explained by resistance to extinction. Then, in the Harlow et al. (1950) and the Gately (1950) studies, the disruption effect following the introduction of food rewards can be interpreted as a temporary phenomenon. Recall that in the Davis et al. study, there was a temporary disruption at the beginning of the food-rewarded period which was soon overcome by a high level of responding. It is likely that the Harlow et al. and the Gately phenomena were the same as this initial disruption.

The large body of developmental data on internalization reported in Aronfreed (1968) confirms that rewarded behavior (especially if rewarded intermittently) persists even after the rewards have been removed. Similar results were found in Experiment III of the present series. The extrinsic rewards often used by parents (or psychologists) for children are verbal reinforcement, love, and so on.

Money is seldom used, so we would expect persisting behavior following rewards as predicted by Hypothesis II. Even if money were found to create or enhance internal control in children, this could be attributed to the fact that money has not yet taken on the same connotation for children as for adults.

The theoretical framework which received some support from the three experiments does seem to provide a means for reconciling the results of previous experiments.

REFERENCES

ALLPORT, G. W. *Personality: A psychological interpretation.* New York: Holt, 1937.

ARONFREED, J. *Conduct and conscience.* New York: Academic Press, 1968.

ARONSON, E., & MILLS, J. The effect of severity of initiation on liking for a group. *Journal of Abnormal and Social Psychology,* 1959, **59**, 177–181.

ATKINSON, J. W. *An introduction to motivation.* Princeton: Van Nostrand, 1964.

DAVIS, R. T., SETTLAGE, P. H., & HARLOW, H. F. Performance of normal and brain-operated monkeys on mechanical puzzles with and without food incentive. *Journal of Genetic Psychology,* 1950, **77**, 305–311.

DECHARMS, R. *Personal causation: The internal affective determinants of behavior.* New York: Academic Press, 1968.

FERSTER, C. B., & SKINNER, B. F. *Schedules of reinforcement.* New York: Appleton-Century-Crofts, 1957.

FESTINGER, L. *A theory of cognitive dissonance.* Evanston, Ill.: Row, Peterson, 1957.

FESTINGER, L. The effect of compensation on cognitive processes. Paper presented at the McKinsey Foundation Conference on Managerial Compensation, Tarrytown, New York, March 1967.

GATELY, M. J. Manipulation drive in experimentally naive rhesus monkeys. Unpublished master's thesis, University of Wisconsin, 1950.

HARLOW, H. F. Motivation as a factor in the acquisition of new responses. *Nebraska Symposium on Motivation,* 1953, **1**, 24–49.

HARLOW, H. F., HARLOW, M. K., & MEYER, D. R. Learning motivated by a manipulation drive. *Journal of Experimental Psychology,* 1950, **40**, 228–234.

KELLER, F. S. *Learning: Reinforcement theory.* (2nd ed.) New York: Random House, 1969.

MORSE, W. H. Intermittent reinforcement. In W. K. Honig (Ed.), *Operant behavior: Areas of research and application.* New York: Appleton-Century-Crofts, 1966.

MURRAY, E. J. *Motivation and emotion.* Englewood Cliffs, N. J.: Prentice-Hall, 1964.

ORNE, M. T. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist,* 1962, **17**, 776–783.

ROSS, I. C., & ZANDER, A. Need satisfaction and employee turnover. *Personnel Psychology,* 1957, **10**, 327–338.

UHL, C. N., & YOUNG, A. G. Resistance to extinction as a function of incentive, percentage of reinforcement, and number of nonreinforced trials. *Journal of Experimental Psychology,* 1967, **73**, 556–564.

WEICK, K. E. Reduction of cognitive dissonance through task enhancement and effort expenditure. *Journal of Abnormal and Social Psychology,* 1964, **68**, 533–539.

WEICK, K. E. Dissonance and task enhancement: A problem for compensation theory? *Organizational Behavior and Human Performance,* 1967, **2**, 189–208. (a)

WEICK, K. E. Organizations in the laboratory. In V. H. Vroom (Ed.), *Methods of organizational research.* Pittsburgh: University of Pittsburgh Press, 1967. (b)

WHITE, R. W. Motivation reconsidered: The concept of competence. *Psychological Review,* 1959, **66**, 297–333.

WOODWORTH, R. S. *Dynamic psychology.* New York: Columbia University Press, 1918.

YERKES, R. M., & DODSON, J. D. The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparative and Neurological Psychology,* 1908, **18**, 459–482.

ZEIGARNIK, B. Das behalten erledigter und unerledigter handlungen. *Psychologische Forschung,* 1927, **9**, 1–85.

ZIMBARDO, P. G., WEISENBERG, M., FIRESTONE, I., & LEVY, B. Communicator effectiveness in producing public conformity and private attitude change. *Journal of Personality,* 1965, **33**, 233–255.

(Received June 30, 1970)