Review of Educational Research
Spring 1996, Vol. 66, No. 1, pp. 39–51

The Debate About Rewards and Intrinsic Motivation: Protests and Accusations Do Not Alter the Results

Judy Cameron and **W. David Pierce**
University of Alberta

A prevailing view in education and social psychology is that rewards decrease a person's intrinsic motivation. However, our meta-analysis (Cameron & Pierce, 1994) of approximately 100 studies does not support this position. The only negative effect of reward occurs under a highly restricted set of conditions, circumstances that are easily avoided. These results have not been well received by those who argue that rewards produce negative effects under a wide range of conditions. Lepper, Keavney, and Drake (1996), Ryan and Deci (1996), and Kohn (1996) have suggested that the questions asked in our meta-analysis were inappropriate, that critical studies were excluded, that important negative effects were not detected, and that the techniques used in our meta-analysis were unsuitable. In this response, we show that the questions we asked are fundamental and that our meta-analytic techniques are appropriate, robust, and statistically correct. In sum, the results and conclusions of our meta-analysis are not altered by our critics' protests and accusations.

Our research (Cameron & Pierce, 1994) has clearly touched a nerve. The results of our meta-analysis indicate that rewards can be used effectively to enhance or maintain an individual's intrinsic interest in activities. These findings are challenging to those who espouse the view that rewards and reinforcement are generally detrimental to a person's intrinsic motivation. Our article has drawn criticism because the data from approximately 100 experiments show that there is only one small negative effect of reward, an effect that is highly circumscribed and easily avoided. This finding is disconcerting to those who contend that the negative effects of reward are substantial, generalized and occur across many conditions.

Our analysis of 20 years of research is the most extensive review of the literature on rewards and intrinsic motivation to date. Because of its thoroughness, the data, analysis, and conclusions must be taken seriously. Faced with the evidence, researchers who have argued that rewards produce harmful effects under a wide range of conditions are put in a difficult position. One option they can take is to reanalyze the data in an attempt to show that rewards have strong negative effects on intrinsic motivation. Our data are readily available for additional analyses, and our procedures are clearly outlined in the original article. Failing this option, a second strategy is to suggest that the findings are invalid due to intentional bias, deliberate misrepresentation, and inept analysis. Our critics have chosen the second strategy.

Lepper, Keavney, and Drake (1996); Ryan and Deci (1996); and Kohn (1996) have responded to the results of our meta-analysis by accusing us of asking inappropriate questions, omitting important moderator variables, excluding critical experiments, and contradicting other reviews on the topic. In addition, they criticize our meta-analytic procedures and decisions as flawed.

In response to these criticisms, we show that all relevant studies were included in our analyses and that the questions and reward conditions we assessed expand on previous reviews to provide a more comprehensive picture of the effects of rewards on intrinsic motivation. We answer the statistical concerns of our critics and show that our analysis is appropriate, accurate, and robust. Most importantly, we show that none of the objections raised by our critics negates our findings.

The results and conclusions of our meta-analysis remain important, especially for those involved in education and other applied settings. An issue of prime concern to educators is how to use rewards effectively to promote learning without disrupting students' intrinsic interest. Contrary to Ryan and Deci's (1996) claim that our "theoretical position acknowledged no conditions under which one should expect negative effects" (p. 33), our results provide important clarifications about the conditions under which rewards produce positive or negative effects on intrinsic motivation. Of primary importance in classroom situations is the finding that rewards can be used to maintain or enhance students' intrinsic interest in schoolwork. Verbal praise and performance feedback increase the value of an activity. When tangible rewards are offered contingent on level of performance or are given unexpectedly, students remain motivated in the subject area. A slight negative effect can be expected when a teacher offers a tangible reward without regard to the students' level of performance. Under this condition, when the rewards are withdrawn, students will continue to like their schoolwork as much as others, but they may spend slightly less time on it in a free period. This negative effect can be easily prevented by offering students rewards for successful solution of problems, completion of work, or for attaining specified levels of performance on particular tasks. The point is that teachers can reward the level and quality of students' work without disrupting motivation and interest in learning. These conclusions are not altered by the comments of Kohn, Ryan and Deci, and Lepper et al.

In the following commentary we address our critics' concerns. Our response is organized in two sections; the first deals with the general issues that have been raised by our critics, and in the second we focus on specific statistical criticisms.

General Issues

The Overall Question

One issue of contention involves our decision to begin our meta-analysis by investigating the overall effect of reward on intrinsic motivation (overall effect hypothesis). Lepper and his colleagues state that "to ask about the 'overall' or 'in general' effects of rewards or reinforcers is to pose a fundamentally meaningless question" (p. 7). They argue that the question is senseless and misleading, a view echoed by Kohn and by Ryan and Deci.

We maintain that the overall effect hypothesis is central to an understanding of this area of research. One reason is practical. Many educators, parents, and administrators have adopted Kohn's (1993) position that overall, rewards and incentive systems are harmful. In the present context, this stance means that rewards negatively affect students' intrinsic interest, a question of overall effect. Others involved in education are still open to the possibility that rewards may be beneficial. A classroom teacher who wishes to implement an incentive system is first of all interested in whether rewards disrupt intrinsic interest in the subject matter. Of course, it may be advantageous to target particular subgroups or implement additional measures, but the question of the overall effect of reward is crucial to one's teaching strategy.

Another reason to address the main effect hypothesis is that academic journals, introductory textbooks, newspapers, and some of our critics continue to point to the overall negative or harmful effects of reward and reinforcement. In a prominent scientific journal, *Nature*, we learn that "it has been repeatedly shown that if people are rewarded for performing a task they find intrinsically pleasurable, they do it less, not more" (Sutherland, 1993, p. 767). A major introductory psychology textbook informs us that

> when an extrinsic reward is given, the motivation becomes extrinsic and the task itself is enjoyed less. When the extrinsic rewards are withdrawn, the activity loses its material value. . . . The moral is: *A reward a day makes work out of play.* (Zimbardo, 1992, p. 454, italics in the original)

Even in this issue of *Review of Educational Research*, Kohn asserts that "there is more than adequate justification for avoiding the use of incentives to control people's behavior, particularly in a school setting" (p. 3).

These examples are but a small sample of the claims made about the overall effects of reward. Many university students, educators, and parents have been exposed to this negative main effect assumption and base their own understanding and use of rewards on it. Social policy in our schools and other institutions reflects these beliefs. Because of this, an analysis of the general effects of reward is warranted.

In their critiques of our meta-analysis, Lepper et al. and Ryan and Deci indicate that they and others have long recognized that the negative overall effect hypothesis is incorrect. Nonetheless, numerous writers interpret the research findings as indicative of an overall negative effect and decry the use of rewards in educational and work settings (e.g., see Kohn, 1993). As a result, many parents, teachers, and others are reluctant to use rewards—any rewards—under any circumstances! Lepper and his colleagues suggest that reversing this incorrect conclusion will be harmful. They imply that we are trying to propagate our own myth—that rewards have no negative effects. We do not want to add any more myths to this research area. So let us be clear in stating that our research demonstrates that rewards have either positive or negative effects depending on the way they are administered. Importantly, the only negative effect of reward on intrinsic motivation occurs under a circumscribed set of conditions, namely, when rewards are tangible and promised to individuals without regard to any level of performance.

The Role of Moderator Variables

A major focus of our meta-analysis was to assess the effects of various moderator variables. The moderators we included (type of reward, reward expectancy, and reward contingency) were chosen because of their theoretical and practical importance in the literature on intrinsic motivation as well as replication over a number of experiments. Our results indicate that the detrimental effects of reward are limited and depend on multiple moderators. All of our critics, Lepper et al., Ryan and Deci, and Kohn, are concerned that we failed to assess the impact of additional important moderators. The implication of their comments is that decremental effects of reward occur under numerous conditions and are far more widespread than our analysis suggests. Interestingly, however, as we describe below, an analysis of additional moderators would, in fact, show the opposite.

Lepper et al. point to studies that assessed the impact of initial task interest and reward salience on intrinsic motivation. Other moderator variables hypothesized to influence intrinsic motivation include reward attractiveness, presence or absence of the experimenter, task difficulty, reward magnitude, and so on. It is critical to point out that the few studies designed to investigate the impact of these moderators typically begin with the one condition that produces a negative effect. Furthermore, such moderators have been shown to enhance, mitigate, or reverse the negative effects of expected, tangible, noncontingent reward. For example, Ross (1975) found that salient rewards make the negative effect of tangible, expected, noncontingent reward greater. McLoyd (1979), on the other hand, demonstrated that individuals offered a noncontingent, tangible reward experienced an increase in intrinsic motivation when the task was less interesting, while Williams's (1980) research indicated that the negative effects of tangible, expected, noncontingent reward could be offset by offering attractive rewards. In other words, the variables we have not assessed are moderators that have typically been added to the conditions that produce the single negative effect of reward found in our meta-analysis. Thus, an analysis of studies that included moderators that increase the negative effects of expected, tangible, noncontingent reward would serve to place further restrictions on the circumstances under which rewards undermine intrinsic motivation. That is, the negative effect phenomenon may be even more circumscribed than our data indicate, a finding contrary to the implications hinted at by our critics.

Presently, however, there is no way to assess the theoretical or applied importance of these moderator variables. This is because only one or two studies have replicated the same moderator procedures on a common dependent measure of intrinsic motivation. If the effects of moderators such as reward salience, reward attractiveness, and so on were systematically replicated, a subsequent meta-analysis could be conducted to determine the conditions that moderate the negative effect on intrinsic motivation of tangible, expected, noncontingent rewards when they are removed. Of course, such an analysis would simply extend our findings and show that tangible, expected, noncontingent rewards produce negative effects on intrinsic motivation only when other conditions are present. For example, in terms of reward attractiveness, Williams's (1980) research shows that when tangible, expected, noncontingent, *unattractive* rewards are given, intrinsic motivation decreases; the same reward condition with *attractive* rewards does not produce a decrement. Although present theoretical accounts (e.g., cognitive evaluation theory, the overjustification hypothesis) may be able to organize such circumscribed effects, the theories would become less and less generalizable. In applied settings, negative effects of reward on intrinsic motivation would depend on so many conditions that there would be little need for concern.

Both Kohn (1996) and Ryan and Deci (1996) raise the question of moderators in the context of our finding that verbal praise produces positive effects both on the free time students spend on tasks and on attitude measures of intrinsic motivation. Specifically, they claim that verbal praise directed at controlling student behavior has negative effects on intrinsic motivation, whereas informational praise does not. We did not conduct an analysis on the control-informational dimension of verbal reward because these variables appear in only one or two studies. In addition, most research on this topic has been conducted without adequate no-feedback control groups (e.g., Ryan, 1982). Until a sufficient number of experiments with control groups are conducted, a meta-analysis of conditions that have few replications would not be reliable or beneficial to our understanding of reward and intrinsic motivation. We note, however, that although there are so few studies on this topic, the effects of controlling and informational verbal reward were analyzed in a recent meta-analysis by Tang & Hall (1995). They found no significant effects on either of these dimensions.

In sum, although our meta-analysis was designed to assess the effects of several moderators on reward and intrinsic motivation, Lepper et al., Ryan and Deci, and Kohn have suggested that many additional important moderators were omitted. As we have shown, an analysis of additional moderators would not alter our conclusions or change any of the results of our meta-analysis. That is, negative effects of reward on intrinsic motivation are highly conditional and occur solely in the presence of multiple moderators. In educational settings, negative effects can be avoided by praising students for their work and making tangible rewards contingent on performance.

Our Findings in Context

Both Ryan and Deci and Lepper et al. argue that our findings contradict previous narrative reviews and other meta-analyses of reward and intrinsic motivation. Lepper et al. are not consistent on this point, and in a later section of their critique they concede that "other recent meta-analyses, . . . as well as numerous previous narrative reviews, have reached exactly [our] conclusion" (p. 7). In this section, we show that our results are in accord with other summaries of reward and intrinsic motivation and that our review advances the knowledge in this area. We briefly comment on three other meta-analyses on this topic (Rummel & Feinberg, 1988; Tang & Hall, 1995; Wiersma, 1992).

The most recent meta-analysis on rewards and intrinsic motivation, conducted by Tang and Hall (1995), was designed to test several theoretical propositions about the overjustification effect. Fifty studies were included, largely a subset of the experiments examined in our review. One analysis concerned assessing the effects of expected, tangible, task-contingent (noncontingent) reward on the free time measure of intrinsic motivation. Tang and Hall found a negative effect, as did we. Also, in accord with our findings, they found no detrimental effect with unexpected, tangible reward. It is difficult to compare our findings on the effects of verbal reward on free time with their study, because their analysis included only two effect sizes (their result was not significant).

Tang and Hall (1995) reported a negative effect on the free time measure for performance-contingent reward, whereas we found no significant effect. This difference in findings is due to Tang and Hall's classification of performance-contingent reward as well as to their omission of several relevant studies. Of the seven studies that Tang and Hall analyzed as performance contingent, six are actually task-contingent reward procedures, as defined by Deci and Ryan (1985). We used Deci and Ryan's definitions and identified 10 studies of performance-contingent reward; overall, there was no evidence of a negative effect. Additional measures of intrinsic motivation (e.g., attitude toward task) that we examined were not reported by Tang and Hall.[1]

The meta-analyses by Wiersma (1992) and Rummel and Feinberg (1988) were discussed in our original article (Cameron & Pierce, 1994). Wiersma analyzed 20 studies, and Rummel and Feinberg analyzed 45 studies. We cannot compare our findings with those of Rummel and Feinberg, because they averaged over different dependent measures of intrinsic motivation. Our meta-analysis shows that this is inappropriate, because the free time and attitude measures do not necessarily covary with the same experimental treatment. In addition, in both Rummel and Feinberg's and Wiersma's analyses, many of the effect sizes reported came from studies where one reward condition was compared to another reward condition. The lack of a no-reward group makes a comparison of findings problematic. Wiersma does, however, report effect size estimates for six experiments on free time that compared a no-reward condition to an expected, tangible, noncontingent reward condition. Though we have not conducted a meta-analysis on his results, we computed the average of the six independent effects sizes and found a negative effect, a finding compatible with our original conclusions.

All in all, our findings for rewards that are tangible, expected, and noncontingent are consistent with other meta-analyses. Our research, however, went beyond an analysis of the one negative reward procedure and assessed the effects of reward under a variety of conditions. In terms of other reward procedures (e.g., verbal reward, performance-contingent reward) and other measures of intrinsic motivation (e.g., attitude toward a task), we failed to find any detrimental effects on intrinsic motivation. That is, our study showed that most reward procedures can be used to maintain or enhance intrinsic motivation; the negative effect other reviews have detected is only a small part of a larger picture. Thus, our meta-analysis provides a more complete account of the effects of rewards on intrinsic motivation.

The Completeness of Our Review

A criticism put forward by Kohn, as well as by Ryan and Deci, is that we failed to include several critical experiments in our meta-analysis. The implication is that had such studies been included, our results would have been different.

Kohn cites a number of studies that he believes we have overlooked. Most of these studies were located in our original search and were not included in our meta-analysis because of the lack of an adequate no-reward control condition. In addition, as we reported in our original article, our meta-analysis included studies published up to and including 1991. The studies from the period 1992–1994 cited by Kohn (Boggiano et al., 1992; Freedman, Cunningham, & Krismer, 1992; Gottfried, Fleming, & Gottfried, 1994) were, of course, not included. Of these, Freedman et al. varied the amount of reward but had no nonreward control group. The article by Boggiano et al. reported past research in order to develop a theory or model of students' achievement patterns. Gottfried et al. examined parental motivational practices; their study did not include any of the reward conditions or dependent measures that we analyzed in our meta-analysis. Earlier studies by Birch, Marlin, and Rotter (1984) and Fabes, Fultz, Eisenberg, May-Plumlee, and Christopher (1989) concerned food preferences and prosocial behavior, respectively. Clearly, all these studies are off topic. Other papers that Kohn cites as missing are, in fact, included in our analyses (a list of all studies is presented in Cameron & Pierce, 1994, pp. 399–403).

In contrast to Kohn, Lepper et al. charge us with including too many "bad" studies. An essential criterion of a reliable meta-analysis, however, is that all the studies done in a field are examined, independently of one's own theoretical position and the degree to which the results of any particular study may be promising. We have met this criterion. In fact, our meta-analysis on the effects of rewards on intrinsic motivation is the most comprehensive review of this literature to date. The results are based on a large number of studies, and, to our knowledge, no relevant published studies were omitted. Due to the large sample of studies included in our analyses, any single study that may have been overlooked would not alter the conclusions. Overall, our results were based on all the available evidence, and the findings are central to an understanding of the effects of rewards on intrinsic motivation.

Meta-Analytic Issues

In addition to the general criticisms discussed above, Lepper and his associates object to our use of meta-analysis for assessing the research on the effects of rewards on intrinsic motivation. In particular, they contend that the distributions of effect sizes in our article indicate that meta-analytic tests should not have been conducted. In accord with Ryan and Deci (1996) and Kohn (1996), they further suggest that the statistical procedures used in our meta-analyses must be flawed. Specifically, they criticize the technique of aggregating effect sizes within a single study when moderator variables are present. In this section, we respond to our critics' meta-analytic and statistical concerns. We show that our analyses are appropriate, that the data are approximately normal and homogeneous, that inclusion or exclusion of outliers does not alter the results, and that our procedures yield correct estimates for the effects of rewards on intrinsic motivation at each level of analysis.

The Appropriateness of Meta-Analysis

There are two main issues that concern Lepper et al. with regard to our use of meta-analytic techniques for assessing the effects of rewards on intrinsic motivation. First, they suggest that the apparent normality of our distributions for the critical measures of intrinsic motivation (free time, attitude) is deceptive. Their second concern is that the data are not homogeneous (equal spread of effect sizes) and that meta-analytic tests should therefore not have been performed.

As Lepper et al. acknowledge (p. 13–14), our distributions of effect sizes approximate a normal shape. However, they attribute the normality of these distributions to the inclusion of "pure zero cases" and random estimates. They argue that our inclusion of "pure zero cases" in our graphic portrayal of effect sizes (Cameron & Pierce, 1994, Figures 1 and 2) guarantees a normal distribution around the value of zero. Pure zero cases refer to studies that did not provide sufficient information to calculate effect sizes or random estimates (4 cases for free time and 17 cases for attitude). The truth is that we did not include pure zero cases in these figures. This is clearly stated on pages 379 and 384 of our original article. The normality of the distributions centering around zero is not due to pure zero cases. Thus, Lepper and his associates need not be concerned.

In terms of our use of random estimates of effect sizes, our procedure is innovative and may be more appropriate than merely assigning a zero effect to the experiment or omitting the study itself. The procedure depended on the information available in each study. When t or F values were nonsignificant and were reported as less than some value (e.g., < 1), a random number between 0.01 and that value was selected; and an effect size was then calculated. In other cases, t or F values were not available, but means or directions of means were reported. In these situations, a random number between 0.01 and the critical value of t or F at $p = .05$ was drawn, and an effect size was then calculated. (For more information, see Cameron & Pierce, 1994, p. 376).

With regard to the normality of our distributions, it is important to note that the direction of effect for random estimates was always known. If more studies had had negative effects, the distribution would have been pulled in that direction. The actual shape of the distribution shows that positive and negative effect sizes occurred with similar frequency. This is based not on our use of random estimates but on the actual direction of effects reported in such studies. In other words, the use of random estimates in no way biases the results toward an average zero effect size. The normality of the distributions centering around zero is not due to this, and, again, there is no need for concern. The point is that the effect size distributions approximated a normal shape, and meta-analytic tests could be used with confidence.

Although Lepper et al. agree that our distributions are normal, they argue that our data are heterogeneous (lacking equal spread) and therefore inappropriate for meta-analysis. Our decision to use meta-analytic procedures involved a consideration of several issues. Initially, we were concerned with the normality of the distribution of effect sizes. We showed that the distributions were approximately normal and reported the degree of kurtosis and skewness of the free time distribution in the original article (p. 381). Next we considered the results of the Q test for homogeneity. It is well known that this test is liberal in the sense that the null hypothesis (homogeneity) is too often rejected (Hunter, Schmidt, & Jackson, 1982). Because of this problem, we set the critical value of Q farther out on the chi-square distribution, just below the value at the .01 level (that is, $p > .01$).

Homogeneity was achieved by excluding extreme effect sizes. The exclusion of outliers is not unusual and is recommended by Hedges (1987) as a method for obtaining more equal spread of the effect sizes. To assess any biases due to the removal of outliers, we reported all analyses with extreme values included and excluded. In addition, we identified the studies with extreme values and discussed the conditions that may have led to these atypical results. Inspection of our original article shows that the results do not change to any extent by excluding outliers.

The validity of our meta-analysis is also increased by the use of the CL statistic (McGraw & Wong, 1992). CL is another way to express effect size. Importantly, McGraw and Wong conducted 118 tests (simulations) to show that the CL statistic is robust with respect to violations of normality and homogeneity. Because of this, we used CL in all our analyses and reported results identical to those of the other meta-analytic tests.

In sum, the distribution of effect sizes for the critical measures of intrinsic motivation approximated a normal shape. The normality was not due to the inclusion of "pure zero cases" or random estimates as Lepper and his associates have suggested. Homogeneity of effect sizes was achieved by excluding outliers. All results were reported with outliers included and excluded; our findings were not altered to any extent by the exclusion of outliers. In addition, given our use of the CL statistic, we are confident that our analyses are appropriate and that the results are accurate and valid.

Aggregation of Effect Sizes in Meta-Analysis

Lepper et al., Ryan and Deci, and Kohn are critical of the method of aggregating effect sizes within a study to yield a single estimate for each meta-analytic test. They contend that such procedures yield inaccurate estimates of the effects of reward on intrinsic motivation. Underlying this criticism is the supposition that the effects of important moderators and interactions were not detected in our analyses. Again, the implication of these comments is that negative effects of reward are more prevalent than our results communicate.

In response to this concern, we first note that aggregation of effect sizes within a study is a common procedure in meta-analysis that avoids violation of the assumption of independence (Cooper, 1989; Hedges & Olkin, 1985). The procedures for aggregation are clearly described in our original article (pp. 376–377). It is important to point out that a serious statistical violation occurs when more than one effect size from an individual experiment is entered into a single meta-analysis. Typically, in such cases, a control group is compared with more than one experimental treatment within a study, several effect sizes are calculated, and each is entered into a single meta-analytic test. The major problem is that the effect sizes are not independent (errors among observations are correlated). If the dependencies in such data were properly accounted for, the error term would become larger and mean effect sizes would become smaller. Another problem is that a particular study will contribute more weight to the overall meta-analytic outcome than a study yielding only one effect size. Other meta-analyses on reward and intrinsic motivation favored by Lepper et al. (p. 5) have violated the assumption of independence by entering several (sometimes over 10) effect sizes from one study into a single meta-analytic test (e.g., Rummel & Feinberg, 1988; Tang & Hall, 1995). The implication is that conclusions based on these meta-analyses could be incorrect.

The way to achieve independence and at the same time retain effect sizes for an analysis of the impact of various moderators is to (a) aggregate them into a single estimate for an overall analysis of the effects of rewards on intrinsic motivation and (b) conduct further analyses of the effects of various moderator variables. For factorial designs, the main effect of reward is entered into an analysis of the overall effects of reward; interaction effects that have been replicated in a sufficient number of experiments are then analyzed separately. These are the procedures we used in our meta-analyses. As we indicated previously, the moderators we analyzed (reward type, reward expectancy, and reward contingency) were chosen because of their theoretical and applied importance as well as replication.

Lepper et al. are concerned that aggregation of the moderators (rather than separate analyses) yields inaccurate estimates of the effects of reward on intrinsic motivation (p. 11–13). As mentioned earlier, the moderators not assessed in our analyses (e.g., presence of experimenter, reward attractiveness, salience, distraction, etc.) have appeared in only one or two studies, and in these studies they have been added to the tangible, expected, noncontingent reward condition to decrease, mitigate, or increase the negative effect. In terms of such studies, it is possible to obtain an unbiased estimate of the effect size of tangible, expected, noncontingent reward. When the results are pooled across all studies, the effects of any additional moderators are averaged out. That is, although any one of these manipulations may push intrinsic interest up (e.g., reward attractiveness) or down (e.g., surveillance, reward salience) in a given study, their effects are expected to cancel out across many studies. In other words, the best estimate of the effect size of tangible, expected, noncontingent reward when additional moderators are present is the average of all the comparisons of the rewarded conditions with nonrewarded control groups.

Of course, additional meta-analyses could be conducted on the effects of these moderators if they were sufficiently replicated. As we pointed out, however, because they are added to the one reward procedure that produces a reliable negative effect, the results would show that decremental effects of reward on intrinsic motivation depend on even stricter conditions than our analysis indicates. This is demonstrated in Lepper et al.'s analysis of three factorial experiments (Calder & Staw, 1975; Loveland & Olley, 1979; McLoyd, 1979) that crossed initial task interest (high, low) with reward (reward, no reward). Lepper et al. (p. 10) show that in these three studies, rewarding activities with high intrinsic interest yields a large negative effect size. In contrast, rewarding a task with low initial interest produces a positive effect size. In each of these studies, the reward procedure involved tangible, expected, noncontingent (or task-contingent) rewards—the one procedure that produces a negative effect on the free time measure of intrinsic motivation.

Thus, if Lepper et al.'s analysis is reliable, the results indicate that tangible, expected, noncontingent rewards are harmful only when delivered for more interesting tasks. It is worth mentioning here, however, that a study excluded in Lepper et al.'s analysis (Mynatt et al., 1978) also crossed task interest with tangible, expected, noncontingent reward but found positive effects of reward for both low- and high-interest tasks. Given that there are so few studies of the interest variable, the results from this one study could substantially alter Lepper et al.'s conclusions about the importance of level of task interest when rewards are tangible, expected, and noncontingent.

In summary, the procedures used in our meta-analysis yield correct estimates for the effects of reward on intrinsic motivation at each level of analysis. Our critics have implied that analyses of additional moderators and interactions would yield more general negative effects of reward on intrinsic motivation. However, as we have shown, further analyses would actually reveal that positive effects of reward are more general and that decremental effects of reward occur under even more restricted circumstances than our results indicate.

Conclusion

A prominent view in education and social psychology is that rewards decrease a person's intrinsic motivation. Our meta-analysis of 20 years of research suggests that this view is incorrect. The findings from approximately 100 studies indicate that rewards can be used effectively to enhance or maintain intrinsic interest in activities. The only negative effect of reward occurs under a highly specific set of conditions, circumstances that are easily avoided. Not surprisingly, these results have not been well received by those who argue that rewards produce negative effects on intrinsic motivation under a wide range of conditions.

In response to the findings, Lepper, Keavney, and Drake (1996), Ryan and Deci (1996), and Kohn (1996) have suggested that the questions asked in our meta-analysis were inappropriate, that critical studies were excluded, that important negative effects were not detected, and that the techniques used in our meta-analysis were unsuitable. In this response, we have shown that the questions asked are fundamental to an understanding of the relationship between rewards and intrinsic motivation and that our meta-analytic techniques are appropriate, robust, and statistically correct. Our meta-analysis includes all relevant studies on the topic, and the results clearly show that negative effects of rewards occur under limited conditions. All told, the results and conclusions of our meta-analysis are not altered by our critics' protests and accusations.

Our findings have important practical implications. In applied settings, the results indicate that verbal rewards (praise and positive feedback) can be used to enhance intrinsic motivation. When tangible rewards (e.g., gold stars, money) are offered contingent on performance on a task or are delivered unexpectedly, intrinsic motivation is maintained. A slight negative effect of reward can be expected when tangible rewards are offered without regard to level of performance. Under this condition, when the rewards are withdrawn, individuals report as much interest in the activity as those in a nonrewarded group, but they may spend slightly less time on it in a free period.[2] This negative effect can be prevented by rewarding people for completing work, solving problems successfully, or attaining a specified level of performance. In other words, rewards can be used effectively in educational and other applied settings without undermining intrinsic motivation.

Notes

[1]Tang and Hall (1995) reported effect sizes for questionnaire measures of intrinsic motivation. The studies they analyzed used questionnaire items to index attributions of causality; moral obligation; attitude toward the task; perceptions of luck, ability, effort, and difficulty; feelings of competence; negative affect; self-esteem; and so on. Tang and Hall combined the effect sizes of all these measures and reported meta-analyses based on this composite index. They did not examine attitude toward the task separately, as we did. Thus, we cannot compare our findings on the attitude measure of intrinsic motivation.

[2]It may be informative to consider how serious the negative effect of expected, tangible, noncontingent reward on free time really is. How much less time would students spend on academic subjects if a teacher implemented this reward procedure and then removed it? Results from our meta-analysis indicate that the average effect size for a comparison between people who receive an expected, tangible, noncontingent

reward and nonrewarded individuals on time on task following withdrawal of reward is –0.26.

In the original experiments, time on task was typically measured over an 8-minute period. In order to convert the effect size of –0.26 to real time, one needs to know the pooled standard deviation of rewarded and nonrewarded groups. Because many researchers reported only *t* or *F* statistics, we will use a well-designed study by Pretty and Seligman (1984) to estimate a pooled standard deviation. Their study reported two experiments with large sample sizes and readily available statistical information. Both experiments compared a condition of expected, tangible, noncontingent reward (N = 30) with a nonrewarded control group (N = 30) on 8 minutes of free time. The pooled standard deviation was 2.6 minutes.

Using this estimate of error, we are able to convert the negative effect size from the meta-analysis into real time. An effect size of –0.26 would mean that in an 8-minute period, the average individual who is promised a noncontingent, tangible reward will spend about 41 seconds less time on the task when the reward procedure is withdrawn than the average nonrewarded individual. Given this result, what would happen if a teacher implemented this incentive procedure in a reading program and then removed it? According to the estimate, students who are offered gold stars for reading would spend about 3 minutes, 25 seconds less time reading in a 40-minute free-choice period than students not given the incentive. Of course, this is a hypothetical example, but it does illustrate the magnitude of this negative effect size in terms of real time.

References

Birch, L. L., Marlin, D. W., & Rotter, J. (1984). Eating as the "means" activity in a contingency: Effects on young children's food preference. *Child Development, 55,* 431–439.

Boggiano, A. K., Shields, A., Barrett, M., Kellam, T., Thompson, E., Simons, J., & Katz, P. (1992). Helplessness deficits in students: The role of motivational orientation. *Motivation and Emotion, 16,* 271–296.

Calder, B. J., & Staw, B. M. (1975). Self-perception of intrinsic and extrinsic motivation. *Journal of Personality and Social Psychology, 31,* 599–605.

Cameron, J., & Pierce, W. D. (1994). Reinforcement, reward, and intrinsic motivation: A meta-analysis. *Review of Educational Research, 64,* 363–423.

Cooper, H. M. (1989). *Integrating research: A guide for literature reviews* (2nd ed.). Beverly Hills, CA: Sage.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior.* New York: Plenum.

Fabes, R. A., Fultz, J., Eisenberg, N., May-Plumlee, T., & Christopher, F. S. (1989). Effects of rewards on children's prosocial motivation: A socialization study. *Developmental Psychology, 25,* 509–515.

Freedman, J. L., Cunningham, J. A., & Krismer, K. (1992). Inferred values and the reverse-incentive effect in induced compliance. *Journal of Personality and Social Psychology, 62,* 357–368.

Gottfried, A. E., Fleming, J. S., & Gottfried, A. W. (1994). Role of parental motivation practices in children's academic intrinsic motivation and achievement. *Journal of Educational Psychology, 86,* 104–113.

Hedges, L. (1987). How hard is hard science, how soft is soft science? The empirical cumulativeness of research. *American Psychologist, 42,* 443–455.

Hedges, L., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies.* Beverly Hills, CA: Sage.

Kohn, A. (1993). *Punished by rewards.* Boston: Houghton Mifflin.
Kohn, A. (1996). By all available means: Cameron and Pierce's defense of extrinsic motivators. *Review of Educational Research, 66,* 1–4.
Lepper, M. R., Keavney, M., & Drake, M. (1996). Intrinsic motivation and extrinsic rewards: A commentary on Cameron and Pierce's meta-analysis. *Review of Educational Research, 66,* 5–32.
Loveland, K. K., & Olley, J. G. (1979). The effect of external reward on interest and quality of task performance in children of high and low intrinsic motivation. *Child Development, 50,* 1207–1210.
McGraw, K. O., & Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin, 111,* 361–365.
McLoyd, V. C. (1979). The effects of extrinsic rewards of differential value on high and low intrinsic interest. *Child Development, 50,* 1010–1019.
Mynatt, C., Oakley, D., Arkkelin, D., Piccione, A., Margolis, R., & Arkkelin, J. (1978). An examination of overjustification under conditions of extended observation and multiple reinforcement: Overjustification or boredom? *Cognitive Therapy and Research, 2,* 171–177.
Pretty, G. H., & Seligman, C. (1984). Affect and the overjustification effect. *Journal of Personality and Social Psychology, 46,* 1241–1253.
Ross, M. (1975). Salience of reward and intrinsic motivation. *Journal of Personality and Social Psychology, 32,* 245–254.
Rummel, A., & Feinberg, R. (1988). Cognitive evaluation theory: A meta-analytic review of the literature. *Social Behavior and Personality, 16,* 147–164.
Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology, 43,* 450–461.
Ryan, R. M., & Deci, E. L. (1996). When paradigms clash: Comments on Cameron and Pierce's claim that rewards do not undermine intrinsic motivation. *Review of Educational Research, 66,* 33–38.
Sutherland, S. (1993). Impoverished minds. *Nature, 364,* 767.
Tang, S., & Hall, V. (1995). The overjustification effect: A meta-analysis. *Applied Cognitive Psychology, 9,* 365–404.
Wiersma, U. J. (1992). The effects of extrinsic rewards in intrinsic motivation: A meta-analysis. *Journal of Occupational and Organizational Psychology, 65,* 101–114.
Williams, B. W. (1980). Reinforcement, behavior constraint, and the overjustification effect. *Journal of Personality and Social Psychology, 39,* 599–614.
Zimbardo, P. G. (1992). *Psychology and life* (13th ed.). New York: Harper Collins.

Authors

JUDY CAMERON is Assistant Professor, Department of Educational Psychology, University of Alberta, 6-102 Education North, Edmonton, Alberta, Canada T6G 2G5; judy.cameron@ualberta.ca. She specializes in educational psychology.

W. DAVID PIERCE is Professor, Centre for Experimental Sociology, 1-48 Tory, University of Alberta, Edmonton, Alberta, Canada T6G 2H4; dpierce@gpu.srv.ualberta.ca. He specializes in social psychology.

Received November 6, 1995
Accepted November 20, 1995