MARGARET G. MELOY, J. EDWARD RUSSO, and ELIZABETH GELFAND MILLER*

Researchers have long debated the methodological necessity of monetary incentives in experimental research. The current work shows that financial incentives not only can fail to improve task performance but also can worsen it. Three studies verify that incentives can elevate mood and that this mood enhancement contributes to worsened task performance. The authors discuss implications for the use of incentives in experimental research.

Monetary Incentives and Mood

INCENTIVES AND DECISION MAKING

A diverse body of literature suggests that monetary incentives are beneficial in studies of decision making to stimulate the increased effort that people expend in the "real world." This view has been embraced by experimental economists, who view financial incentives as essential to the validity of the research. Empirically, performance-based incentives have successfully increased the search for more external information (Hulland and Kleinmuntz 1994) and the accuracy of prediction and probability matching tasks (Hogarth et al. 1991; Wright and Aboul-Ezz 1988). Similarly, they have reduced the preference reversal phenomenon (Ordóñez et al. 1995), the framing effect in judgment tasks (Levin, Chapman, and Johnson 1988), and the time to reach an economic equilibrium in double oral auctions (Jamal and Sunder 1991).

Further bolstering the value of incentives is the belief that even if incentives fail to enhance performance (e.g., because of a performance ceiling), they can do no harm. Thus, even if the benefits of incentives are not universally present (and such a universal claim is impossible to validate), the absence of harm completes the argument for their necessity. Any effective counterargument must demonstrate not only that incentives are sometimes ineffective but also that they actually degrade performance in at least some common situations.

In contrast to economists, psychologists have resisted using financial inducements in experimentation on the grounds that incentives can negatively affect task performance. First, a monetary payment may undermine intrinsic motivation to engage in the task, which in turn leads to poorer performance (Bahrick 1954; Deci 1971; Eisenberger and Cameron 1996; Gneezy and Rustichini 2000; Heyman and Ariely 2004; Lepper, Greene, and Nisbett 1973; McGraw 1978; Ryan and Deci 2000). Second, incentives may hurt performance if individuals exert "too much effort" when a simpler heuristic would suffice (Arkes, Dawes, and Christiansen 1999; Camerer and Hogarth 1999). Psychologists who have tried to provide a framework for understanding when incentives help versus hurt have concluded that aspects of the decision environment, such as complexity, presence of feedback, time pressure, and risk, affect the specific impact of the incentive on performance (e.g., Ashton 1990; Hogarth et al. 1991). Shah, Higgins, and Friedman (1998, p. 285) conclude that the "perceived value of an incentive lies in the extent to which it supports an individual's goals."

The current work contributes to the understanding of incentives by identifying a third way that they can lead to degraded performance. In particular, we suggest that monetary incentives can elevate mood and that this elevated mood can lead to performance decrements in some tasks. By examining this third path, we hope to bring psychologists and economists closer together in their understanding of the impact of incentives on decision making and to suggest that using incentives in experimental work can have unintended effects.

Mood as Mediator

To establish mood as mediating the negative impact of incentives on performance, the current work focuses on tasks for which increased effort enhances task performance

*Margaret G. Meloy is Assistant Professor of Marketing, Smeal College of Business, Pennsylvania State University (e-mail: mmeloy@psu.edu). J. Edward Russo is Professor of Marketing, Professor of Management and Organizations, and S.C. Johnson Family Chair of Management, Johnson Graduate School of Management, Cornell University (e-mail: jer9@cornell.edu). Elizabeth Gelfand Miller is Assistant Professor of Marketing, Carroll School of Management, Boston College (e-mail: milleliz@bc.edu). The authors thank Colin F. Camerer, Kurt A. Carlson, Ronald L. Hess, George Loewenstein, William T. Ross, and William D. Schulze for helpful comments on prior drafts and Hans Baumgartner, Katrina J. Bens, Girish Mallapragada, and Sriram Venkataraman for help with the structural equation models. They also thank Christine Anderson-Cook and Jennifer Chang Coupland for use of their students for in-class data collection. Portions of this work were presented at the annual meeting of the Society for Consumer Psychology in San Antonio, Tex., in February 2000 and at the biennial meeting of Behavioral Decision Research in Management in Tucson, Ariz., in May 2000. This work was supported by three grants: NYC-121424 from the United States Department of Agriculture to the first author, SBR-9730237 from the National Science Foundation to the second author, and SES-00112039 from the National Science Foundation to the first and second authors.

but elevated mood degrades it. Focusing on these tasks enables us to examine the opposing forces of effort and mood. Because the mood literature is not unified with respect to the mechanisms through which a good mood affects decision making (for comprehensive reviews of mood and decision making, see Clore and Martin 2001; Forgas 2001; Lewis and Haviland-Jones 2000), we do not attempt to explain the specific process by which mood leads to this degraded performance. Rather, we attempt to demonstrate empirically that a financial incentive can enhance mood, which sometimes can worsen task performance.

The literature contains no demonstrations of the complete causal chain from performance-based incentives to elevated mood to degraded performance. However, the separate links have been observed in at least one robust decision bias, the predecisional distortion of information (Russo, Medvec, and Meloy 1996; Russo, Meloy, and Medvec 1998). In a choice between two alternatives, predecisional distortion is the systematic evaluation of new information to favor the currently preferred alternative. Unlike phenomena that rely on a prior preference for one option, individuals tend to engage in predecisional distortion as soon as they develop a preference for one alternative, no matter how tentative (Russo, Medvec, and Meloy 1996; Russo, Meloy, and Medvec 1998).

Carlson and Russo (2001) show that prospective jurors in a legal case distorted evidence in support of a tentative leaning toward guilt or innocence despite repeated instructions to avoid coming to a premature conclusion. Using the same framework, Hope, Memon, and McGeorge (2004) find that jurors exposed to negative pretrial publicity distorted testimony to support the prosecution, resulting in elevated rates of a guilty verdict. Finally, in work involving product choices, Carlson, Meloy, and Russo (2006) find that when feature information for two options was ordered to create an early advantage for an inferior alternative, a suboptimal choice was made.

We chose predecisional distortion as the outcome measure for Studies 1 and 2 because it is a measurable bias with material consequences and because increased effort (through an accountability manipulation) has been shown to reduce it (Russo, Meloy, and Wilks 2000) and an experimentally induced elevated mood has been shown to aggravate it (Meloy 2000). The fundamental issue of whether an incentive improves or degrades performance depends on whether the positive impact of greater effort in reducing the bias dominates or is dominated by a negative impact of an elevated mood in exacerbating it.

Stepwise Evolution of Preference Method

To trace the binary decision process in detail and measure the distortion of new information as it is processed, the current work employs a method known as the stepwise evolution of preference (SEP; Meloy and Russo 2004). This method has two essential characteristics: First, the two alternatives being considered are jointly presented, one attribute at a time. Second, after each attribute is reviewed, three responses are recorded. An example may best convey the method.

Experimental participants are instructed to make a choice between one of two fine-dining restaurants. Information is provided about both restaurants, and each attribute or unit of information contains all the relevant facts for both options. Six attributes are typically provided; in the case of restaurants, for example, these could be ambiance, dining area, menu, parking, reservations, and special requests. The menu, for example, may be as follows:

> Restaurant G's daily menu contains a wide variety of entrées, including poultry, pork, beef, vegetarian items, and pasta dishes. In addition, it always has two fish specials, one vegetarian special, and one or two pasta specials. It also has two signature dishes, one a beef delicacy and one vegetarian, which are unique to the restaurant. Overall, Restaurant G's food is described as outstanding.
>
> Restaurant Z's main entrée menu is fairly limited, with a few poultry, pork, and beef dishes; several pasta dishes; and two different dinner salads. One of its specialties is a braised chicken creation that cannot be found anywhere else. It has many dinner specials each day, including vegetarian items; one pasta special; several fish specials; and numerous game dishes, such as venison, rabbit, duck, and quail. In general, Restaurant Z serves distinctive dishes that are memorable.

After each attribute is considered, participants provide three responses. First, they evaluate the attribute's diagnosticity on a scale from 1 ("strongly favors Restaurant G") to 9 ("strongly favors Restaurant Z"). Second, they identify the restaurant that is "leading" on the basis of all the attributes they have seen up to that point. This response uses a horse-race metaphor to convey the idea that the "lead" might change as more information becomes available. Finally, participants express their confidence in the leader by indicating how much of $10 they would bet on the leader to "win the horse race." A bet of $5 on each restaurant signals that the restaurants are neck and neck. After reviewing the six attributes, participants choose one restaurant.

STUDY 1

In Study 1, we examine the impact of a standard performance-based incentive, a reward for the accuracy of a task's output, on the predecisional distortion of information. We use a straightforward two-condition design (i.e., incentive versus no incentive). If a monetary incentive is effective in increasing effort in the absence of a positive impact on mood, distortion should be reduced, as it was in Russo, Meloy, and Wilks's (2000) accountability condition. However, if the incentive also elevates mood, distortion might be higher in the incentive condition (Meloy 2000). The resulting effect on distortion should depend on the resolution of the forces of increased effort and elevated mood.

Participants

A total of 108 people participated in Study 1. We assigned 33 participants to a control condition; they made no choices (as we explain subsequently). We randomly assigned all others to either an incentive (n = 43) or a no-incentive (n = 42) condition. Each experimental participant made two choices, one between a pair of fine-dining restaurants and one between two resort destinations for a spring-break trip. Each session took approximately 25 minutes.

Incentive and No-Incentive Tasks

Participants in the two experimental conditions (i.e., with or without an incentive) read a role-playing scenario that explained the task and the criterion for an accurate choice.

The role-playing scenario involved auditioning to become a reviewer for a "Places to Stay and Places to Eat" guide. For the audition, participants evaluated the profiles of various hotels and restaurants from a different community and picked the better hotel or restaurant in each pair. We told them, "Since the situations are real cases from a different community that have already been rated by a separate panel of experts, there is a 'right' decision in each case. Your two choices will be compared to the correct answers in order to test how successful you have been for the audition. This test will determine whether you will be selected to serve on the panel of reviewers."

For the incentive condition, we told participants, "We are prepared to pay you depending on how many choices you answer correctly. Your selections will be compared to the ratings of the panel of experts. If your choices match the evaluations of the expert panel, we will pay you as follows. If you get 1 out of the 2 choices correct, you will be paid $2. If you get both of the decisions correct, you will be paid $12." We selected this payout structure to stress the importance of getting both decisions "right."

Participants reviewed six attributes of information before making each decision. For the resort hotel choice, the six attributes were ambiance, beach, distance to amenities, swimming pools, rooms, and services. We gathered the three SEP measures after participants reviewed each attribute, that is, the evaluation of the information on the nine-point scale (i.e., diagnosticity), designation of the leader, and confidence in the leader as reflected in the "bet." We randomly assigned the initial order in which participants reviewed the attributes, and we reversed this order for half of the participants to counterbalance the attributes' serial position. We wrote each attribute of information to be roughly equivocal, favoring neither alternative by much, if at all. This nondiagnosticity of the information facilitated its distortion (Russo, Meloy, and Medvec 1998) and enabled us to detect the relative impacts of effort and mood more easily.

Following the last attribute, we asked participants to identify their final choice. They also reported the time (in minutes) devoted to the choice task. After participants made both choices, they completed Peterson and Sauber's (1983) Mood Short Form to assess current mood. We also gathered additional demographic information at this time.

No-Choice Control

We neither instructed nor allowed the no-choice group to make a choice. It provided both an improvised standard of decision accuracy and the unbiased estimates of information diagnosticity from which we could calculate distortion. We needed to devise a criterion of accuracy to compute the performance-based reward in the incentive condition. We derived this standard from the measure of each attribute's diagnosticity on the nine-point scale. To preclude the possibility of any preference for one option forming spontaneously (see Russo, Meloy, and Medvec 1998), we showed the control participants a new pair of alternatives for every attribute (e.g., Restaurants T and H for Attribute 1, Restaurants B and W for Attribute 2, and so on).

The control participants' mean evaluations of the information provided an unbiased estimate of the true diagnosticity of each attribute. We then defined accuracy as the choice of the alternative favored by the sum of the six individual diagnosticities from the no-choice control evaluations. Because the determination of accuracy was based on the mean unbiased estimates of the no-choice control group, it was consensual only. No objective measure of accuracy was available. However, none of the results that follow depend on absolute level of performance, only on the difference between the standard choice and the incentive group. As we stated previously, participants in the incentive group received $2 for one accurate choice and $12 if both decisions were correct. The specific payout scheme for the incentive condition provided an expected payout value of approximately $5 per person (i.e., [.50 × $2]) + [.25 × $12]).

Calculation of Distortion

The attribute evaluations of the no-choice control participants also provided the standard against which we measured predecisional distortion. We adjusted the evaluation of choice subject$_i$ on attribute$_j$ ($Eval_{ij}$) by subtracting the average evaluation of those in the no-choice control condition for attribute$_j$ ($Eval_j$). The absolute difference $|Eval_{ij} - Eval_j|$ was then signed to reflect support for or opposition to the leading alternative. If the difference pointed toward the currently preferred option (as indicated by the leader in the horse race after the prior attribute), the absolute level of distortion was positively signed. However, if $Eval_{ij}$ was less favorable to the preferred option than was $Eval_j$, the absolute level of distortion was negatively signed. For example, suppose that after the second attribute, a participant favored Hotel J. Suppose further that the control group yielded 5.55 as the unbiased diagnosticity of the third attribute, which slightly favored Hotel C on the nine-point scale. If the decision maker rated the third attribute as a 3 (favoring Hotel J) on this same scale, the magnitude of the distortion for this piece of information would be +2.55 units.

Note that we could not calculate distortion in two cases. Because the signing of distortion requires a leader, we could not calculate distortion for the first attribute. The same rationale held when the two alternatives were tied in preference, as indicated by a bet of $5 on each alternative. With the exception of these ties, we calculated distortion across Attributes 2–6. We averaged attribute-level distortions across each choice to create a participant-level distortion.

Results

Increased effort. To test for a positive impact of the incentive on effort, we examined the self-reported time on the task for each choice. The monetary incentive increased the mean time that participants spent on each choice from 5.0 minutes (in the no-incentive condition) to 7.1 minutes. This difference was a substantial jump of 42% ($t(164) = 6.06$, $p < .001$).[1] Thus, the incentive had the intended impact on effort.

Mood. Mood was also elevated in the incentive condition. On a five-point scale, with higher scores reflecting a more positive mood, participants averaged 3.40 in the no-incentive condition but 3.87 in the incentive condition ($t(168) = 4.57$, $p < .01$). Thus, participants who received a

[1]For completeness, note that several participants failed to record their time spent on the task. The averages we report include only the participants who provided time measures.

monetary incentive for being accurate in their choice were happier than those who did not receive an incentive.

Distortion. What was the resolution of the competing forces of increased effort and enhanced mood on the predecisional distortion of information? The baseline condition (i.e., participants without any incentive) distorted each piece of information nearly half a unit (.40) on the nine-point diagnosticity scale. However, when participants were provided with a monetary incentive for accuracy, distortion was substantially greater (.76; $t(167) = 2.121$, $p < .05$).[2] Thus, although participants in the incentive condition put greater effort into the choice task (which should have reduced the bias), the net result was an increase in predecisional distortion, which was a decrement in performance.

Mediation analysis. We conducted a mediation analysis using the procedures that Baron and Kenny (1986) outline to test the proposed link from incentive to mood to worsened performance. When we regressed mood on incentive, standardized $\beta = .39$ ($t(168) = 5.54$, $p < .001$); when we regressed distortion on incentive, standardized $\beta = .16$ ($t(168) = 2.12$, $p < .05$); and when we regressed distortion on both mood and incentive, standardized $\beta = .31$ ($t(167) = 2.04$, $p < .05$), and standardized $\beta = .10$ ($t(167) = 1.17$, $p > .20$), respectively. Baron and Kenny's (1986) three requirements for mediation were met. For completeness, we also tested time, but its effects were nonsignificant (standardized $\beta = -.06$, $p > .50$). The Sobel (1982) test revealed a significant, indirect effect of mood ($z = 2.66$, $p < .01$). We then conducted a LISREL analysis to gather more precise measures of fit.[3] The best-fitting model had mood as the sole mediator (for fit, comparative fit index [CFI] = .99, and nonnormed fit index [NNFI] = .99; $\chi^2 = 1.13$, $p > .20$; Browne and Cudek 1993; for error, root mean square error of approximation [RMSEA] = .01; Bentler and Bonett 1980).

Discussion

The mediation analysis and path model made clear that the promise of a monetary incentive enhanced participants' moods and that this contributed to a decline in task performance. Because the direction of causation from mood to elevated distortion was previously established, Study 1 provides clear evidence that the use of an incentive for accuracy can lead to degraded performance, in part, because it elevates mood.

Study 1 used a standard incentive based on the output of the task as the proper measure of performance. However, the source of the problem occurred when participants evaluated individual attributes in a biased fashion. Would the incentive have succeeded in reducing distortion if it had been targeted at those attribute evaluations? To answer this question, we ran a second study in a more objective evaluative domain, scholarship applicants, which included numeric information. Participants were "auditioning for a position on the University Scholarship Committee"; we told them that their responses would be compared with the outcomes assigned by the previous year's committee. Each pair of scholarship applicants was described by six attributes: extracurricular activities, letters of recommendation, performance on exams, personal interview, statement of purpose, and work experience. Each participant made two choices between pairs of applicants. We counterbalanced the order of the attributes and the order of the choices.

We paid those who received the targeted incentive instructions on the basis of their attribute evaluations on the nine-point diagnosticity scale. We deemed them to be accurate when their attribute evaluations were in the appropriate third of the scale. The payout formula was $\Sigma(n - 4)$, where n was the number of attributes correctly evaluated (out of 12) and 4 was the chance base rate; the maximum possible payout was $36. Again, the expected payout value was approximately $5. The decision-based incentive was $2 for one accurate decision and $12 for two accurate decisions. Before participants reviewed any of the scholarship applicants, we provided them with a practice task to expose them to typical scholarship information. Otherwise, the procedures were the same as those in Study 1.

A total of 193 students participated in this follow-up study for extra course credit. We replicated the main findings of Study 1; there was greater effort (56 versus 44 seconds per attribute, $p < .01$), greater mood (4.0 versus 3.6, $p < .01$), and higher mean distortions (.44 versus .20, $p < .05$) in the incentive conditions than in the no-incentive condition. Again, using mediation analysis, we found a mediating role of mood (all $p < .05$) (CFI = .96, and NNFI = .89; $\chi^2 = 3.37$, $p < .10$; RMSEA = .04), establishing that the link from incentive to elevated mood to degraded performance is robust. More important, there were no differences between the decision incentive and the attribute incentive on any of the dependent measures (all *p*s > .30). The targeted incentive was no more successful at reducing the biased evaluation of information than was the conventional outcome incentive.

The results of Study 1 and its replication made it clear that we needed to seek an incentive method that could achieve the goal of increasing effort without the parallel mood enhancement. Two changes to the incentive instruction might reduce the mood effect. First, we could frame the incentive negatively. That is, instead of allowing participants to earn or gain money for correct answers, we could endow them with a fixed sum at the beginning of the study and then penalize them for each performance error. Although this switch from a gain frame to a loss frame should be effective in increasing effort, its potential downside is that the increased stress from the financial losses may induce a negative mood, at least for some participants. However, this is an empirical matter.

The second potential remedy is to inform participants well before the actual study that they will be paid for their performance. This should remove the pleasant surprise of the potential to earn money as one cause of mood elevation. The downside of such forewarning, and the reason we did not use it in Study 1 and the follow-up, is that the mood of the no-incentive participants (i.e., the controls) may be depressed when their expectation of earning money is deflated. However, it may still be possible to reduce the depression of mood by forewarning participants only of the

[2]To probe this result in greater detail, we examined distortion and the bets (i.e., confidence) at the attribute level. Did the incentive increase confidence, which in turn increased distortion? We regressed distortion on both confidence and incentive. There was no impact of the incentive on either the slope or the intercept of the regression equation (both $p > .50$). We could not attribute the increase in distortion in the incentive condition to increasing confidence.

[3]We included time spent on the task in an initial path model, but it did not reflect the best fit or minimum error.

possibility of a performance-based payment, with a candid disclosure that not all participants can be paid. Again, the issue should be settled empirically. These two methodological adjustments motivated Study 2.

STUDY 2

In Study 2, we tested the ameliorative impact of two alternative incentive manipulations in an effort to disrupt the causal chain from incentive to good mood to reduced task performance. We reframed the incentive from a gain to a loss, and to reduce the mood-elevating effect of a pleasant surprise, we forewarned participants of the incentive.

Experimental Tasks and Procedures

All experimental participants completed one of the scholarship choices we discussed previously from the follow-up to Study 1 (and two other tasks that were irrelevant to the present work). We made one change in the "leader" and "bet" questions to better capture the relative confidence in the tentative leader.[4] We asked the participants in Study 2 to imagine the scholarship candidates as two runners in a 60-yard dash. To convey the progress of the race over time, participants placed the leading candidate's identifying letter at the appropriate yard mark in the race. That is, participants placed the leader at the 10-yard mark after the first attribute, at the 20-yard mark after the second, and on through to the 60-yard mark after they evaluated the last attribute. Each participant then indicated how far behind the other candidate was by placing its letter somewhere on the track behind the leader's position. If the two candidates were tied, participants indicated this by putting one candidate exactly above (i.e., equal to) the other.

With the exception of the change in the leader questions, we used the same SEP method as in Study 1. At the conclusion of the task, participants completed a mood scale (Allen and Janiszewski 1989), answered demographic questions, and responded to suspicion and forewarning checks.

Framing the Incentive

Study 2 contrasted the standard gain framing of the incentive with a loss version. We gave participants a fixed sum and penalized them for each incorrect response. In the "earn" or gain-framed condition, we told participants that they would have an opportunity to earn $7 for being accurate in the choice task. In the penalty or loss-framed condition, we told them, "In this task, you will be endowed with cash. If you are inaccurate in your choice, you will be penalized. The money remaining at the end of the session is yours to keep. It is in your best interest to be as accurate as possible in order to avoid the penalty." The payoff structure for the loss-framed condition mirrored that of the gain-framed condition. We randomly assigned half of the participants to each of the two incentive frames.

The Element of Surprise

In general, mood researchers agree that happy surprises (e.g., finding a dime in a coin return) have positive impacts on mood. In the absence of the element of surprise, standard mood inductions can be less effective (e.g., expecting to find a dime in the coin return). To examine the element of surprise as a contributor to the incentive-induced mood effects of Study 1, we included a forewarned condition in Study 2. We told half of the experimental participants in advance of the opportunity to "receive cash" for their participation in the research. We did not forewarn the other half, who learned of the incentive only when they read the instructions in class.

The result of the framing and forewarning variables is a 2×2 experimental design. We required two control conditions to complete the design. Note that because we found no difference between a decision incentive and the targeted attribute-based incentive in the follow-up to Study 1, we adopted the decision incentive for the scholarship choice in Study 2.

No-Incentive and No-Choice Conditions

The calculation of distortion and the determination of accuracy for the scholarship decision required that we gather information evaluations from a no-choice control group specific to this pool of participants. As in Study 1, participants in this condition evaluated the six attributes of information in the absence of choice. The baseline comparison group for all the analyses in Study 2 was a no-incentive choice condition. We gave these participants the same set of tasks, and they answered the same questions as those in the incentive group. The only difference was the absence of an incentive for being accurate. It is their responses, particularly their baseline level of distortion, against which we made all comparisons.

Participants

We recruited 221participants from an Introductory Statistics class at a large research university. They participated in Study 2 as part of an in-class exercise on decision making and probabilistic judgments. Students had been told in advance that the researcher would come to class and administer a survey. Each session lasted approximately 40 minutes. At the conclusion of the data collection, we extensively debriefed participants and gave them a short lecture on decision-making biases.

Participants were spread across six sections of the same course. We assigned the first of the six sections that met during the period of the study to the no-choice control condition. These participants (n = 20) provided the baseline measures of information diagnosticity and the standard of accuracy for the scholarship choice in the remaining sections. We assigned the second section (n = 42) to the no-incentive condition. We hoped to gather the no-incentive condition's responses before word leaked out about the incentive. We also did not want the incentive and no-incentive conditions running in the same section in case someone raised a question about the incentive. (We had visions of a mutiny among the no-incentive participants if that had occurred.) We randomly assigned participants in Sections 3 and 4 (n = 97 total) to either the earn or the lose "incentive-surprise" conditions. Finally, we randomly assigned students in Sections 5 and 6 (n = 62) to either the earn or the lose "incentive-forewarned" conditions. In Sections 5 and 6, we made an announcement that "by coming to class and completing the survey, you will have an opportunity to receive cash for being accurate in a research study

[4]Recall that the normative bet would always be $10 on the leader. Although we had not observed participants responding in this way, the scale was called into question by an experimental economist. To satisfy any paradigmatic issues that might arise, we changed the scale.

on decision making." We provided no additional information. We asked students in these sections to keep that information to themselves because the remaining sections might not have the opportunity to receive money. (With particularly tight budgets at the university, the students found this explanation for the request for discretion entirely plausible.)

Results

Manipulation checks. In the debriefing questionnaire, we asked participants in the incentive conditions whether they had heard anything about the study before their arrival at class that day. We further asked if they had heard anything about receiving money for being accurate. One participant in the forewarned condition claimed that the incentive was a surprise, and two participants in the surprise conditions claimed that they had heard about the study from friends in another section. We eliminated these three participants from all analyses.[5] Again, we tested for an effect of the order of attributes, but we found none (all *p*s > .40). Thus, all analyses focused exclusively on the framing of the incentive (earn versus penalty) and the element of surprise (surprise versus forewarned).

Effort. As in Study 1, we asked participants to record their time on the task for each portion of the study. Unfortunately, many participants did not have a timepiece, and clocks were missing in three of the six classrooms. Despite this handicap, we could still use a 2 × 2 analysis of variance (ANOVA) to test the available data for differences in time across the four incentive groups. The ANOVA revealed that neither the element of surprise nor the framing of the incentive had any impact on the time spent on the task ($F(1, 142) = 1.32$, $p > .20$; $F(1, 142) = .02$, $p > .50$, respectively). The interaction was also nonsignificant ($p > .80$). The mean total time on all tasks, averaged across all four incentive conditions, was approximately 8.7 minutes. In contrast, those in the no-incentive condition spent, on average, 5.3 minutes on the same set of tasks. The incentive was successful in motivating participants to work harder across all the tasks ($t(186) = 1.66$, one-sided $p < .05$).

Distortion. The main question of Study 2 was whether either the loss frame or forewarning would reduce the deleterious effect of an incentive on the distortion of information. A 2 × 2 ANOVA revealed neither a main effect of incentive framing ($F(1, 152) = .02$, $p > .50$) nor a main effect of surprise ($F(1, 152) = .12$, $p > .50$) on predecisional distortion. The interaction was also nonsignificant ($F(1, 152) = .43$, $p > .50$). Thus, neither a loss version of the incentive nor forewarning was effective in reducing predecisional distortion. Over all four incentive conditions, predecisional distortion was almost twice as great (.50) when we provided participants with *any* incentive as when they completed the same tasks in the no-incentive condition (.27). The difference between the combined incentive and the no-incentive conditions was significant ($t(195) = 2.00$, $p < .05$).

Mood. We turned to mood for the explanation of the observed differences. The 2 × 2 ANOVA again revealed no main effects of either the surprise ($F(1, 150) = 2.02$, $p > .15$) or the incentive ($F(1, 150) = 1.35$, $p > .20$) frame on mood. Participants in the incentive conditions were happier (5.49 on a seven-point scale) than their no-incentive counterparts (5.15; $t(194) = 2.34$, $p < .05$). A Dunnett's test revealed that this result held for all but the penalty-forewarned condition (critical value = 2.20, $p > .15$), resulting in a significant interaction ($F(1, 150) = 5.78$, $p < .05$). Participants in the penalty-forewarned condition were less happy (5.07) than any of their counterparts (Dunnett's test critical value = 2.06, $p < .05$). However, this condition incongruously paired the second greatest distortion of the 2 × 2 design with the lowest mood. We have no explanation for this result.

Again, we used Baron and Kenny's (1986) method to verify a mediating relationship of mood. When we regressed mood on incentive, standardized $\beta = .25$ ($t(194) = 3.52$, $p < .001$); when we regressed distortion on incentive, standardized $\beta = .14$ ($t(194) = 1.90$, $p < .06$); and when we regressed distortion on mood and incentive, standardized $\beta = .23$ ($t(193) = 3.20$, $p < .01$), and standardized $\beta = .09$ ($t(193) = 1.24$, $p > .20$), respectively. We again tested time, and its effects were nonsignificant (standardized $\beta = -.07$, $p > .30$). The Sobel (1982) test again revealed an indirect effect of mood ($z = 3.03$, $p < .01$). Providing participants with an incentive led to an increase in mood, resulting in increased predecisional distortion. We again conducted LISREL analyses with mood and time as potential mediators of the relationship between incentive and distortion.[6] As with Study 1, the best-fitting path (CFI = .99, and NNFI = .99; $\chi^2 = 1.91$, $p > .15$; RMSEA = .03) included mood as the sole mediator of the relationship between the incentive and distortion. Again, the power of the mood effect dominated the power of the increased effort to reduce distortion.

Discussion

Study 2 provides clear evidence that any incentive, regardless of its framing or of whether participants are forewarned, can have negative impacts on information processing in decision making. The presence of any form of the incentive elevated mood and degraded performance.

Although the results of Studies 1 and 2 clearly demonstrate the links from incentive to elevated mood to increased distortion of information, we believed that this result would be considerably stronger if it were also shown to occur in another task. This desire for generality motivated the next study.

STUDY 3

Study 3 examines the links from incentive to mood and from mood to overconfidence. The phenomenon of overconfidence is one of the most robust findings in the decision and judgment literature (Alba and Hutchinson 2000; Klayman et al. 1999). In their work on reference group neglect, Camerer and Lovallo (1999) find that incentives do not reduce overconfidence, but other work suggests that among experts, a financial incentive for performance reduces this bias (Murphy and Winkler 1984). We hoped that in the pres-

[5]We had run the two surprise conditions in the largest sections of the class because we expected a higher level of cross-talk among the students. Although we collected all the data for this study in a 24-hour period, we expected word to leak out, necessitating the elimination of a larger proportion of the "surprise" participants than actually resulted. This is the reason for the unbalanced sample sizes between the forewarned and the surprise incentive conditions.

[6]Key measures of error (e.g., RMSEA = .08) and fit (e.g., NNFI = .93) for the LISREL model with both time and mood suggested that we seek a better-fitting model.

ence of an objectively correct numeric answer, the incentive would strengthen effortful processing effects and provide a rigorous test of the net effect of an incentive on both mood and effort.[7]

Overconfidence

The current study examines overconfidence, which we assess by setting confidence intervals. We provided participants with a series of factual items, such as the year in which Mozart was born, and asked them to specify a lower limit and an upper limit such that they were 90% confident that the correct answer fell somewhere in their interval. In a ten-item task, perfect accuracy consists of exactly nine of the ten ranges containing the true value. That is, the correct answer should fall either above the upper limit or below the lower limit for one and only one interval. However, studies employing this method typically report that four to six of the intervals fail to contain the true value rather than the ideal of one (Klayman et al. 1999). Because the proffered intervals are too narrow, they reflect a bias toward overconfidence or the belief that the uncertain quantities can be more tightly bound than is justified.

Incentive and the payout structure. Recall that in Study 2, we found no effect of forewarning or framing of the incentive (gain or loss) on mood or performance. As such, we did not forewarn participants and included only the more familiar gain frame in Study 3. The measure of task accuracy was the number of intervals (out of ten) that correctly captured the true value of the item. The specifics of the payoff structure were as follows: If either one or two of the ten intervals contained the true value, participants earned $1, three to four correct resulted in a gain of $2, five to six correct resulted in earning $3, exactly seven correct resulted in a $5 gain, and eight resulted in a $7 gain. Exactly nine intervals containing the true value was the precise goal and resulted in earning all $9 (the maximum). However, if participants' intervals were too broad and all ten intervals contained the true value, their earnings dropped to $4. Again, we attempted to motivate participants with a gradated payoff structure.

Participants

A total of 72 students participated in the study for extra course credit at a large university. The course from which we drew participants required introductory statistics as a prerequisite. This enabled us to preclude the possibility of differences in exposure to the notion of a confidence interval. Each session lasted approximately 20 minutes.

Procedures

Participants came to the lab in groups of approximately six. After they set their upper and lower limits in the confidence quiz, they filled out Peterson and Sauber's (1983) Mood Short Form, completed demographic information, and were debriefed. We paid all participants in the incentive condition for their accuracy at the end of the session. We also paid participants in the no-incentive condition $5 for their participation at the end of the session and swore them to secrecy.

Results

Effort. Participants in the incentive and no-incentive conditions self-reported time spent on the task. Again, the incentive had the desired impact on effort. The students who received an incentive spent more time setting upper and lower limits (9.1 minutes) than their no-incentive counterparts (6.8 minutes) ($t(71) = 3.31$, $p < .01$).

Overconfidence. The mean number of intervals containing the true value was 4.74 in the no-incentive condition, a typical result. In contrast, the incentive group created intervals that contained the true value only 1.71 times, on average. Performance was worsened when participants were given a financial reward for accuracy ($t(71) = 9.04$, $p < .01$). Thus, overconfidence is another task in which an incentive degrades performance.

Mood. The demonstration of an increase in overconfidence in the presence of an incentive left open the issue of mood. When we replicated the effects in Studies 1 and 2, mood was again elevated in the incentive condition (3.48) relative to the no-incentive condition (3.03; $t(71) = 2.35$, $p < .05$). Participants who received a financial incentive were relatively happier.

Mediation. To link clearly the incentive to mood to increased overconfidence (i.e., fewer intervals containing the true value), we again conducted mediation analysis (Baron and Kenny 1986). As expected, we found a significant impact of the incentive on mood (standardized $\beta = .27$, $t(71) = 2.35$, $p < .05$), a significant relationship between the incentive and degraded performance on the confidence quiz (standardized $\beta = -.75$, $t(71) = 9.53$, $p < .01$), and a significant relationship between mood and overconfidence when we included both mood and incentive in the regression (standardized $\beta = -.17$, $t(70) = 2.02$, $p < .05$; standardized $\beta = -.71$, $t(70) = 8.44$, $p < .01$, respectively). Based on the Sobel (1982) test, the effect of this mediating relationship meets the criterion for significance ($z = 1.83$, one-sided $p < .05$), albeit weakly.

We again conducted LISREL analyses to examine mood and time as mediators of the relationship. Unlike Studies 1 and 2, the best-fitting path included time as a mediator (CFI = .95; NNFI = .91; $\chi^2 = 16.23$, $p > .10$; RMSEA = .08). In addition to the usual impact of incentive on mood ($t(70) = 2.18$, $p < .05$), the incentive positively affected the time spent on the task ($t(70) = 9.04$, $p < .01$). In turn, increased time spent on the task led to a weak direct effect on accuracy ($p > .30$). However, increased time negatively affected mood ($t(70) = -2.04$, $p < .05$). Unlike Studies 1 and 2, there was greater downward pressure on mood when time increased. Although this dampened the impact of mood on overconfidence, the mediating role of mood between incentive and accuracy remained significant, though only marginally ($t(70) = -1.85$, one-sided $p < .10$).

Study 3 verified that the link from incentive to elevated mood to degraded performance generalizes to a second phenomenon, namely, overconfidence. Participants were more overconfident in the presence of the incentive, in part, because their moods were elevated. This effect was dampened by the increased effort exerted, but nonetheless, the links from incentive to elevated mood to degraded performance remained.

GENERAL DISCUSSION

Despite the wealth of research indicating that performance-based incentives encourage people to process

[7]Note that there is no published evidence to suggest that a good mood causes people to be overconfident, but it has been shown that a good mood causes the related phenomenon of optimism (Martin et al. 1993; Nygren et al. 1996).

information rigorously, a monetary incentive may lead to biased information processing and overconfidence because of an elevated mood. In the current work, mood effects were uniformly present whenever there was an incentive, regardless of type. Furthermore, we measured mood at the end of the study when its effect may have been weaker. As such, the true impact of the incentive on mood is likely to be understated in the current work (Baron and Kenny 1986).

Our results have several implications for the use of incentives in experimental research. First, providing participants with performance-based monetary incentives may cloud the understanding of the underlying relationships among variables. In the current study, the detrimental impacts of enhanced mood swamped the positive motivational impacts of the incentive on effort. Thus, at a minimum, if incentives are considered a "requirement," mood should be monitored. Then, any confounding effects should be analyzed and, if possible, covaried out. We also advocate the inclusion of a no-incentive control condition for comparison purposes. Although this adds another condition to the experimental design, it seems essential in identifying the true impacts of the incentive on performance.

However, the deleterious effect of mood in the present work does *not* suggest that research studies involving incentives are invalid. Although the incentives in our studies were real, the choices that participants made in Studies 1 and 2 were hypothetical, and interim feedback was nonexistent. In a task with real consequences, richer information sources, and more continuous feedback, the mood effects might disappear. However, we caution that most experimental studies face conditions that are more similar to ours than to the latter.

The results of Study 3 suggest that the deleterious effects of an incentive on mood are not necessarily unusual. Recall that in Study 3, we hoped that skilled participants (those with some training in statistics) and the presence of a more objective measure of accuracy would test the balance between effort and mood as rigorously as was possible. Again, the increase in mood dominated the power of the incentive to increase effortful processing.

An open issue is the characteristics of the tasks whose performance is worsened by a more positive mood. Our studies do not provide sufficient evidence to outline the characteristics of tasks that are either susceptible or insensitive to a deleterious effect of elevated mood, but they provide evidence that mood measurement is essential in any study that involves incentives. For example, Beeler and Hunton (1997) find that a financial incentive exacerbated the escalation of commitment phenomenon; they report that the incentive led people to continue to "throw good money after bad." Unfortunately, Beeler and Hunton do not assess mood. Paralleling this result, Isen and Stokker (1993) report that a good mood exacerbated the problem of escalation of commitment, leading participants to continue to fund projects that were not worthy of the additional investment. However, these researchers did not use monetary incentives. As further advances are made in understanding exactly how mood affects decision making, the ability to predict which tasks are susceptible to mood effects will become more feasible.

Negating Mood Effects

How might incentives be used to increase effort without a concomitant elevation of mood? Although two natural remedies, reframing and forewarning, failed in Study 2, other possible tactics remain. For example, when designing research studies, researchers might include a simple mood manipulation in the incentive-absent condition. This should reveal whether both mood and incentives have similar impacts on the phenomenon of interest. Again, the cost of this approach is adding a mood condition to the experimental design.

Another approach is to elicit a more analytic mode of processing. For example, participants might be instructed to outline their choice strategy before reviewing specific information. Alternatively, priming an analytic mode of processing might also hold promise. At a nonconscious level, priming an accuracy goal (Bargh and Chartrand 1999) might improve task performance. Finally, because accountability has been shown to increase effortful processing in a productive way, an accountability manipulation might be used in conjunction with the incentive. Participants who have to explain and justify their strategy to another person might adopt a more analytic process that is less susceptible to the mood effects that result from the incentive. Note, however, that these tactics are all designed to change the fundamental process to become more analytic, presumably something that an incentive alone accomplishes.

Conclusion

The current work establishes that incentives may alter mood, and depending on the behavior of interest, an altered mood may change performance. Attention to this possibility seems essential if the impact of incentives is to be fully understood. In many tasks, an elevated (or depressed) mood will have little or no effect on performance, but it would be wise to establish this in advance. For tasks in which mood is known to affect performance, a minimum recommendation is that mood should be monitored and analyzed for such an effect. If degraded performance occurs, there may be no fully satisfactory way to deal with the incentive-induced change in mood (and with its consequences). Finally, it may be useful to reexamine prior failures of incentives, informed by specific predictions from the mood literature. Such reinterpretations may resolve anomalous results and point the way toward a more effective use of incentives.

REFERENCES

Alba, Joseph W. and J. Wesley Hutchinson (2000), "Knowledge Calibration: What Consumers Know and What They Think They Know," *Journal of Consumer Research*, 27 (September), 123–56.

Allen, Chris T. and Chris A. Janiszewski (1989), "Assessing the Role of Contingency Awareness in Attitudinal Conditioning with Implications for Advertising Research," *Journal of Marketing Research*, 26 (February), 30–43.

Arkes, Hal R., Robyn M. Dawes, and Caryn Christiansen (1986), "Factors Influencing the Use of a Decision Rule in a Probabilistic Task," *Organizational Behavior and Human Decision Processes*, 37 (February), 93–110.

Ashton, Robert H. (1990), "Pressure and Performance in Accounting Decision Settings: Paradoxical Effects of Incentives, Feedback and Justification," *Journal of Accounting Research*, 28 (Supplement), 148–80.

Bahrick, Harry P. (1954), "Incidental Learning Under Two Incentive Conditions," *Journal of Experimental Psychology*, 47 (3), 170–72.

Bargh John A. and Tanya Chartrand (1999), "The Unbearable Automaticity of Being," *American Psychologist*, 54 (July), 462–79.

Baron, Reuben M. and David. A. Kenny (1986), "The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations," *Journal of Personality and Social Psychology*, 51 (December), 1173–82.

Beeler, Jesse D. and James E. Hunton (1997), "The Influence of Compensation Method and Disclosure Level on Information Search Strategy and Escalation of Commitment," *Journal of Behavioral Decision Making*, 10 (June), 77–91.

Bentler, P.M. and Douglas G. Bonett (1980), "Significance Tests and Goodness of Fit in the Analysis of Covariance Structures," *Psychological Bulletin*, 88 (November), 588–606.

Browne, Michael W. and Robert Cudeck (1992), "Alternative Ways of Assessing Model Fit," *Sociological Methods and Research*, 21 (2), 230–58.

Camerer, Colin F. and Robin M. Hogarth (1999), "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework," *Journal of Risk and Uncertainty*, 19 (December), 7–42.

——— and Dan Lovallo (1999), "Overconfidence and Excess Entry: An Experimental Approach," *American Economic Review*, 89 (March), 306–318.

Carlson, Kurt A., Margaret G. Meloy, and J. Edward Russo (2006), "Leader-Driven Primacy: Using Atttribute Order to Affect Consumer Choice," *Journal of Consumer Research*, forthcoming.

——— and J. Edward Russo (2001), "Biased Interpretation of Evidence by Mock Jurors," *Journal of Experimental Psychology: Applied*, 7 (June), 91–103.

Clore, Gerald L. and Leonard L. Martin (2001), *Theories of Mood and Cognition: A User's Guidebook*. Mahwah, NJ: Lawrence Erlbaum Associates.

Deci, Edward L. (1971), "Effects of Externally Mediated Rewards on Intrinsic Motivation," *Journal of Personality and Social Psychology*, 18 (April), 105–115.

Eisenberger, Robert and Judy Cameron (1996), "Detrimental Effects of Reward: Reality or Myth?" *American Psychologist*, 51 (November), 1153–66.

Forgas, Joseph P. (2001), *Handbook of Affect and Social Cognition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Gneezy, Uri and Aldo Rustichini (2000), "Pay Enough or Don't Pay at All," *Quarterly Journal of Economics*, (August), 791–810.

Heyman, James and Dan Ariely (2004), "Effort for Payment: A Tale of Two Markets," *Psychological Science*, 15 (11), 787–93.

Hogarth, Robin M., Brian H. Gibbs, Craig R. McKenzie, and Margaret A. Marquis (1991), "Learning from Feedback: Exactingness and Incentives," *Journal of Experimental Psychology: Learning, Memory and Cognition*, 17 (July), 734–52.

Hope, Lorraine, Amima Memon, and Peter McGeorge (2004), "Understanding Pretrial Publicity: Predecisional Distortion of Evidence by Mock Jurors," *Journal of Experimental Psychology: Applied*, 10 (2), 111–19.

Hulland, John S. and Donald N. Kleinmuntz (1994), "Factors Influencing the Use of Internal Summary Evaluations Versus External Information in Choice," *Journal of Behavioral Decision Making*, 7 (June), 79–102.

Isen, Alice M. and Laura G. Stokker (1993), "The Influence of Positive Affect and Meaningful Loss on Decisions to Cut One's Losses in Sunk-Cost Situations," paper presented at the Annual Meeting of the Society for Judgment and Decision Making, Washington, DC (November 7–8).

Jamal, Karim and Shyam Sunder (1991), "Money Vs. Gaming: Effects of Salient Monetary Payments in Double Oral Auctions," *Organizational Behavior and Human Decision Processes*, 49 (June), 151–66.

Klayman, Joshua, Jack B. Soll, Claudia Gonzalez-Vallejo, and Sema Barlas (1999), "Overconfidence: It Depends on How, What and Whom You Ask," *Organizational Behavior and Human Decision Processes*, 79 (September), 216–47.

Lepper, Mark R., David Greene, and Richard E. Nisbett (1973), "Undermining Children's Intrinsic Interest with Extrinsic Rewards: A Test of the 'Overjustification' Hypothesis," *Journal of Personality and Social Psychology*, 28 (October), 129–37.

Levin, Irwin P., Daniel P. Chapman, and Richard D. Johnson (1988), "Confidence in Judgments Based on Incomplete Information: An Investigation Using Both Hypothetical and Real Gambles," *Journal of Behavioral Decision Making*, 1 (March), 29–41.

Lewis, Michael and Jeannette M. Haviland-Jones (2000), *Handbook of Emotions*, 2d ed. New York: The Guilford Press.

Martin, Leonard L., David W. Ward, John W. Achee, and Robert S. Wyer (1993), "Mood as Input: People Have to Interpret the Motivational Implications of Their Moods," *Journal of Personality and Social Psychology*, 64 (March), 317–26.

McGraw, Kenneth O. (1978), "The Detrimental Effects of Reward on Performance: A Literature Review and a Prediction Model," in *The Hidden Costs of Reward: New Perspectives on the Psychology of Human Motivation*, Mark R. Lepper and David Greene, eds. Hillsdale, NJ: Lawrence Erlbaum Associates, 33–60.

Meloy, Margaret G. (2000), "Mood-Driven Distortion of Product Information," *Journal of Consumer Research*, 27 (December), 345–59.

——— and J. Edward Russo (2004), "Binary Choice Under Instructions to Select Versus Reject," *Organizational Behavior and Human Decision Processes*, 93 (March), 114–28.

Murphy, Allan and Robert Winkler (1984), "Probability Forecasting in Meteorology," *Journal of the American Statistical Association*, 79 (September), 489–500.

Nygren, Thomas E., Alice M. Isen, Pamela J. Taylor, and Jessica Dulin (1996), "The Influence of Positive Affect on the Decision Rule in Risk Situations: Focus on Outcome (and Especially Avoidance of Loss) Rather Than Probability," *Organizational Behavior and Human Decision Processes*, 66 (April), 59–72.

Ordóñez, Lisa D., Barbara A. Mellers, Shi-Jie Chang, and Jordan Roberts (1995), "Are Preference Reversals Reduced When Made Explicit?" *Journal of Behavioral Decision Making*, 8 (December), 265–77.

Peterson, Robert A. and Matthew H. Sauber (1983), "A Mood Scale for Survey Research" in *AMA Educators Proceedings*, Patrick Murphy, O.C. Ferrell, Gene R. Laczniak, Robert F. Lusch, Paul F. Anderson, Terence A. Shimp, and Russell W. Belk, and Charles B. Weinberg, eds. Chicago: American Marketing Association, 409–414.

Russo, J. Edward, Victoria H. Medvec, and Margaret G. Meloy (1996), "The Distortion of Information During Decisions," *Organizational Behavior and Human Decision Processes*, 66 (April), 102–110.

———, Margaret G. Meloy, and Victoria H. Medvec (1998), "Predecisional Distortion of Product Information," *Journal of Marketing Research*, 35 (November), 438–52.

———, ———, and T. Jeffrey Wilks (2000), "Predecisional Distortion of Information by Auditors and Salespersons," *Management Science*, 46 (February), 13–27.

Ryan, Richard M. and Edward L. Deci (2000), "When Rewards Compete with Nature: The Undermining of Intrinsic Motivation and Self-Regulation," in *Intrinsic and Extrinsic Motivation: The Search for Optimal Motivation and Performance*, Carol Sansone and Judith M. Harackiewicz, eds. San Diego: Academic Press, 13–54.

Shah, James, E. Tory Higgins, and Ronald S. Friedman (1998), "Performance Incentives and Means: How Regulatory Focus Influences Goal Attainment," *Journal of Personality and Social Psychology*, 74 (February), 285–93.

Sobel, M.E. (1982), "Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models" in *Sociological Methodology*, Samuel Leinhardt, ed. San Francisco: Jossey-Bass, 290–310.

Wright, William F. and Mohamed E. Aboul-Ezz (1988), "Effects of Extrinsic Incentives on the Quality of Frequency Assessments," *Organizational Behavior and Human Decision Processes*, 41 (April), 143–52.