JOURNAL OF MANAGEMENT ACCOUNTING RESEARCH
Volume Twenty
2008
pp. 153–168

Contracting Frame and Individual Behavior: Experimental Evidence

Bryan K. Church
Georgia Institute of Technology

Theresa Libby
Wilfrid Laurier University

Ping Zhang
University of Toronto

ABSTRACT: This paper reports the results of an experiment examining the effect of the framing of incentive contracts on individual behavior. We examine two budget-based incentive contracts that, though economically equivalent, are framed differently. Previous research documents that individuals prefer bonus-framed to penalty-framed contracts (Luft 1994; Hannan et al. 2005). We explore whether these preferences affect effort expended on a task where increased effort results in increased performance. In addition, we test whether these preferences motivate effort differentially in the presence and absence of an effective financial incentive for performance. Consistent with prospect theory predictions, results indicate the penalty-framed contract motivates higher task performance than the bonus-framed contract for individuals whose performance falls within the bonus or penalty range (i.e., where financial incentives are effective in motivating performance). Performance did not differ according to contract frame for individuals whose performance enabled them to receive the maximum payment or for individuals whose performance resulted in them receiving the minimum payment (i.e., where financial incentives were not effective in motivating performance). Although prior research indicates contract framing affects contract preferences, our results indicate these preferences may not result in differences in individual performance unless effective financial incentives are also utilized.

Keywords: framing; bonus; penalty; incentive contract.

Data Availability: Contact the authors.

INTRODUCTION

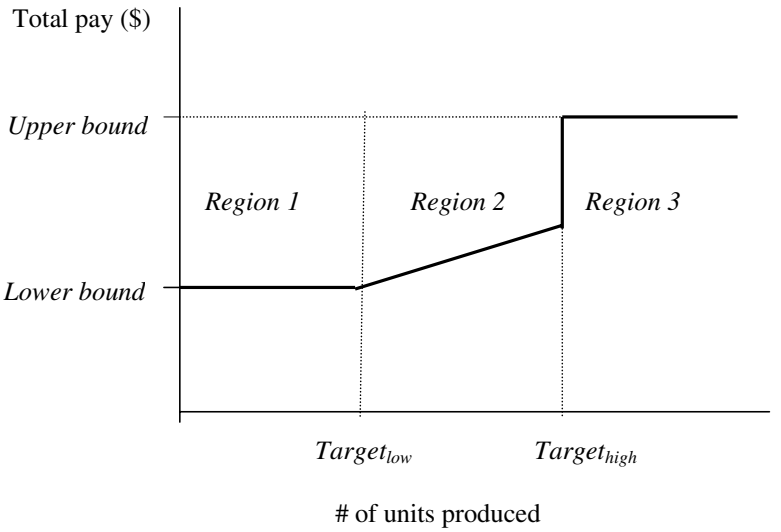
In the typical pay-for-performance compensation scheme, a manager's total cash compensation includes a salary plus a bonus based on meeting or exceeding a minimum performance hurdle. Once this hurdle is reached, the bonus increases as performance increases until the bonus is capped at some maximum level (Jensen 2001). In this paper, we examine (1) the effect of bonus-versus-penalty framing of such a contract on the effort individuals expend and (2) how the amount of effort individuals expend to increase their performance might change depending on the level of the performance hurdles that are set (relative to the individual's actual ability) and the contract frame (bonus or penalty).

According to prospect theory, individuals have a greater subjective disutility for changes in wealth perceived as losses than for changes perceived as gains, even if the losses or gains leave them in the same final wealth position (Kahneman and Tversky 1979); therefore, individuals should work harder to avoid a penalty than to achieve a bonus of an equivalent dollar amount. Luft (1994) finds that individuals prefer bonus-framed to economically equivalent penalty-framed contracts, and Hannan et al. (2005) find that individuals select a higher level of costly effort to avoid a penalty than to receive a bonus of an equivalent dollar amount.

In the current study, we examine the effect of bonus and penalty contract frame on the amount of effort individuals are willing to exert on an experimental production task. This study extends prior work in this area in three main ways. First, we introduce a contract form that archival research indicates is commonly used in practice (e.g., Healy 1985; Holthausen et al. 1995; Murphy 2000; Jensen 2001). This contract form differs from those previously used in the experimental incentive contracting literature in that it caps the variable component (bonus or penalty) of compensation at some maximum level. With such a contract, the variable portion of compensation depends on performance relative to a low and high pre-specified budget target (see Figure 1). When performance falls below the low target or above the high target, there are no effective incentives for additional effort. When performance falls between the two targets, additional effort translates into higher performance and consequently higher pay (i.e., incentives are effective).

We examine whether framing effects observed in prior literature occur only in regions where incentives are effective in motivating effort (i.e., Region 2 of Figure 1) or whether contract framing affects employees' attitudes toward the whole contract, not just in the effective incentive region. Given the suggestions in Luft (1994) concerning language effects of contract frame on individual preferences, it may be that the contract frame colors the

FIGURE 1
Incentive Contract Form



employee's reaction to the contract over the whole incentive range, not just between the upper and lower bounds on incentive pay. Luft (1994, 198) suggests that bonus/penalty labeling effects ''are a potential confounding factor in studies in which labeling language covaries with institutional structure, information levels, payoff functions, etc.'' For example, bonus and penalty contracting language may affect both individual preferences as well as individual effort independent of the effect of financial incentives embedded in the contract. By utilizing a contract with three different incentive regions, we can investigate employees' level of effort depending upon where the target is set, providing additional insight into the impact of contract frame on effort. Specifically, we examine incentive effects of bonus/penalty language both within and outside of the regions of the incentive contract where changes in performance lead to changes in pay.

Second, we extend Hannan et al. (2005), concerning the effects of penalty-framed contracts on effort choice, by examining behavior in a ''real effort'' setting. While Hannan et al. (2005) use a traditional ''effort choice'' setting to test their hypotheses, we note a call in the literature for incentive contracting research that examines the robustness of results based on effort choice alone as an indicator of performance effects of various incentive contract forms (Van Dijk et al. 2001). In an ''effort choice'' setting, participants choose an effort level from a menu of effort choices. Each level of effort has a related cost which captures the negative utility derived from exerting effort. Although higher levels of effort are more costly, they also increase the probability of receiving higher pay. In a ''real effort'' setting, participants actually perform some sort of work moving the experiment from a hypothetical choice setting to a setting that better resembles an actual workplace.

Third, to our knowledge, few if any prior studies in the experimental incentive contracting literature in accounting have examined the effects of a kinked incentive contract with both a low target and a cap on variable pay. While we believe the examination of incentive effects of this contract form is important given it is often used in practice, this methodological choice introduces new challenges. Specifically, behavior is likely to be different in the flat regions of the compensation function (i.e., expected performance below the low target or above the high target) and in the sloping region (i.e., expected performance between the two targets). Clean tests of predictions about behavior therefore need to partition the experimental participants accurately into regions (e.g., separate those with insufficient ability to meet the low target, even under incentive conditions, from those who can move their performance up to the sloping part of the function with additional effort). Moreover, targets need to be set relative to participants' abilities so that an adequate number of participants will fall into each region. As our experiment illustrates, neither of these tasks is simple.

In our experimental setting, M.B.A. student participants practice an effort-sensitive clerical task over three learning periods. In the fourth period, participants are randomly assigned either a bonus-framed or penalty-framed version of a budget-based incentive contract with regions of effective and ineffective incentives. Participants perform the task for one period under the assigned incentive contract to examine whether the number of symbols translated will differ between the groups depending on contract frame and effectiveness/ineffectiveness of incentives embedded in the incentive contract.

Results indicate that the amount of effort individuals expend on the experimental task is affected by the framing of the incentive contract when task performance falls between the low and the high targets (i.e., within Region 2 of Figure 1). Within this range where monetary incentives are effective, individuals assigned the penalty-framed contract exert significantly more effort than individuals assigned the bonus-framed contract, as predicted.

On the other hand, when monetary incentives are no longer effective, the contract frame does not appear to affect the effort exerted by individuals (i.e., in Regions 1 and 3 of Figure 1).

The remainder of the paper is organized as follows: In the next section, we provide a literature review that serves as a basis to develop the research hypothesis. We then include a section that describes the research method followed by a presentation of the experimental results. In the final section, we discuss these results and provide suggestions for future research.

LITERATURE REVIEW AND HYPOTHESIS DEVELOPMENT

The question of how to best structure an incentive contract to align individual and organizational goals has a long history in the incentive contracting literature in accounting. For example, Demski and Feltham (1978) demonstrate the effects of budgets and standards as motivational devices in situations where ability and effort are difficult to observe. Specifically, they examine the effect of including an accounting-based target in incentive contracts to motivate effort and align individual and organizational goals. The typical contract of this type then will include a fixed wage plus a bonus scaled to the difference between actual performance and the budget target (for a review of the literature in this area, see Young and Lewis 1995).

Few contracts between employees and employers actually include penalties for poor performance (Baker et al. 1988), although such contracts are often observed between the organization and its suppliers and customers. This is a particularly interesting observation given (1) rational decision theory predicts that individuals will be indifferent to incentives framed as bonuses or penalties as long as they have equivalent wealth effects (Baker et al. 1988) and (2) economically equivalent bonus and penalty forms of these contracts can be easily constructed (Luft 1994).

To illustrate, consider the following simple example: In this example, we characterize a bonus contract as a base salary (W_b) plus a bonus (B), where B is contingent on an observable performance measure. A penalty contract may be characterized as a base salary (W_p) less a penalty (P), where P is contingent on an observable performance measure. The two contracts are economically equivalent as long as:

$$W_p = W_b + B \text{ and } W_b = W_p - P.$$

Thus, $B = P$ and $W_p \geq W_b$.[1] According to rational decision theory, the magnitude of B or P affects individuals' contract choices, but the labeling (bonus versus penalty) should be irrelevant.

Research in behavioral decision theory, however, suggests that choices can be affected by the characterization of available options (e.g., Tversky and Kahneman 1981, 1986): that is, individuals' choices are not independent of the way that options are represented or described. For example, Thaler (1980) argues that consumers prefer the term ''cash discount'' to ''credit-card surcharge'' because the former connotes an opportunity cost of using a credit card whereas the latter connotes an out-of-pocket cost. Analogous to Thaler's (1980) discussion of cash versus credit-card prices, individuals facing an incentive contract choice

[1] In our setting, the bonus (penalty) is not incurred until a pre-specified threshold level of performance is achieved (not achieved). Therefore, an additional condition for this relation to hold is that the threshold at which a bonus is earned or a penalty is incurred is the same for both contract frames.

may prefer bonuses because such contracts provide for the opportunity to earn additional pay, whereas penalty contracts impose a cost for failing to achieve a targeted level of performance. This line of reasoning is also consistent with prospect theory arguments that the subjective disutility of a perceived loss (i.e., a decrement in wealth) exceeds the subjective utility of a perceived gain (i.e., an increment in wealth), even when the economic incentives are equal for a given level of effort (Kahneman and Tversky 1979). Empirical evidence supporting this view is provided by Luft (1994) who finds that individuals prefer a bonus-framed contract to a penalty-framed contract, even if they receive the same pay under the two contracts for a given level of effort.

Relying on prospect theory, we examine whether framing a budget-based incentive contract in bonus versus penalty terms might also affect the amount of effort individuals are willing to exert in meeting or exceeding budget targets. As suggested by Hastie and Dawes (2001), we operationalize the framing effect through the creation of alternate statements of the same problem which causes a shift in the individual's reference point. Specifically, the reference point manipulation shifts the individual's perception of an identical outcome so that it seems like a gain when compared with the low reference point, but a loss when compared to a high reference point. To operationalize this idea, we design a budget-based incentive contract with an upper and lower bound on pay (see Figure 1). These boundaries are linked to actual performance relative to specified targets.

In a related study, Hannan et al. (2005) examine the effect of individual preferences for bonus- over penalty-framed contracts on effort choice and whether predictions from prospect theory hold in an incentive contracting setting. In their study, experimental subjects are paid under either a bonus-framed or economically equivalent penalty-framed ''all-or-nothing''-style contract. Participants are presented with a menu of effort choices as well as the cost of each effort choice. Each individual's payoff is a function of his or her effort choice and the outcome of a random draw used to determine the final payoff. When making their effort choice, participants are aware that the probability of a positive outcome (i.e., higher pay) increases as the level of effort selected (and its related cost) increases. Results indicate that, consistent with prospect theory, individuals choose a higher level of effort and, thereby, incur a higher cost to reduce the probability of a penalty than to increase the probability of a bonus of an equivalent dollar amount.

We expect to replicate the results of Hannan et al. (2005) for participants whose performance falls in the region of our incentive contract where incentives are effective (i.e., where performance falls between the low and high targets). That is, we expect individuals paid under a penalty-framed contract to exert higher effort and consequently attain higher levels of performance and related incentive pay, than individuals paid under an equivalent bonus-framed contract when performance falls within Region 2 of Figure 1. Our research hypothesis regarding Incentive Contract Region 2 is stated formally as follows:

H1a: When individuals' abilities allow their marginal level of effort to increase their monetary payoff, individuals compensated under a penalty-framed incentive contract will exert more effort than individuals compensated under an economically equivalent bonus-framed incentive contract.

While the structure of the incentive contract utilized by Hannan et al. (2005) allows them to examine the effect of contract framing when financial incentives are effective at motivating effort, our setting also allows us to examine the effects of the bonus/penalty frame in contract regions where financial incentives are ineffective at motivating effort. We

believe this is important given the arguments in Luft (1994) that labeling language (bonus/penalty) triggers different networks of associations for the individual based on prior experience. These associations may color the individual's reaction to the bonus or penalty frame even in regions of the contract where financial incentives are ineffective in motivating effort (i.e., Regions 1 and 3). According to Frederickson and Waller (2005), an individuals' reaction to contract framing can have both psychological and economic effects. Given the structure of our incentive contract, we are able to examine whether the psychological effects created by contract labels (bonus and penalty) affect the level of effort an individual is willing to exert independent of the effect of the financial (i.e., economic) incentive to exert such effort. If so, we would expect to find results in Regions 1 and 3 of Figure 1 similar to those predicted for Region 2: that is, individuals compensated under a penalty-framed contract will exert a higher level of effort and, therefore, display a higher level of performance than those compensated under a bonus-framed contract even when additional effort will not increase their monetary payoff. This hypothesis is posed to complement H1a and is stated formally as follows:

H1b: When individuals' abilities do not allow their marginal level of effort to increase their monetary payoff, individuals compensated under a penalty-framed incentive contract will exert more effort than individuals compensated under an economically equivalent bonus-framed incentive contract.

METHOD

Participants

We recruited 68 M.B.A. students at a medium-sized public university to participate in the study. The experiment was administered to participants in groups of 15 to 20 and students were paid for participating. All participants received a base payment of $10 for the first three experimental periods plus a variable payment in the fourth period depending on performance. The maximum amount that could be earned in the fourth period was $20. Participants earned an average of $25.86, with a range of $20 to $30, for participating approximately 80 min. In addition to the experimental earnings, participants received a raffle ticket providing them with a chance to win a $150 cash prize. The raffle was included to encourage participation in the study.

Procedures

Participants were randomly assigned to an experimental session. Contracting frame was manipulated between sessions. Thirty-six participants completed the experiment under a bonus-framed contract and 32 under a penalty-framed contract. At the beginning of each period, instructions were distributed and read aloud by the researcher. The experiment included a symbol translation task similar to that developed by Chow (1983). Participants were given a collection of worksheets containing lines of symbols and a translation key. They were instructed to translate the symbols into letters of the alphabet using the key. The translation key and several sample lines of symbols from the worksheets used are provided in the Appendix.[2]

The experiment consisted of four periods. In each period, participants were provided with worksheets that included eight rows of symbols, with ten symbols per row. In each period they were given five minutes to translate as many symbols as possible. Participants

[2] Although the order of symbols that appeared on the worksheets differed between periods, the same translation key was used in all four periods.

were instructed to work sequentially, translating all symbols on the first line of the first page before moving to the second line and so forth. After five minutes were up, participants were given an answer key to check their work. They recorded the number of symbols correctly translated on their experimental materials. The researchers also subsequently re-checked participants' work.

The first three periods were identical and provided the participants with experience performing the symbol translation task. Participants received a flat fee of $10 for completing the first three periods. They were instructed to carefully practice the task in these periods as their ability to perform the task would affect their experimental earnings in the fourth period.[3]

After the completion of the third period, the experimental materials differed between participants in the bonus and penalty conditions. In addition, in the fourth period participants were compensated based on their performance. Those in the bonus condition were compensated based on the following contract (in its general form):

$$\text{Payment} = W_b \text{ if } M \leq T_l,$$

$$W_b + \lambda(M - T_l) \text{ if } T_l < M < T_h,$$

$$W_b + \lambda(T_h - T_l) + B \text{ if } M \geq T_h,$$

where T_l is a low (easily achievable) target, T_h is a high (difficult-to-achieve) target, λ is a positive constant, W_b is the base wage under the bonus contract, B is the size of the bonus, and M is the actual number of symbols translated correctly.

Participants in the penalty condition were compensated based on a different contract, as expressed below (in its general form):

$$\text{Payment} = W_p - P - \lambda(T_h - T_l) \text{ if } M \leq T_l,$$

$$W_p - P - \lambda(T_h - M) \text{ if } T_l < M < T_h,$$

$$W_p \text{ if } M \geq T_h,$$

where W_p is the base wage under the penalty contract, P is the size of the penalty, and M and T are as defined previously. To ensure that the two contracts provide identical incentives, the following relations hold:

$$W_b = W_p - P - \lambda(T_h - T_l) \textit{ and } W_b + B + \lambda(T_h - T_l) = W_p$$

and, thus, $B = P$ and $W_p \geq W_b$.

Based on pilot tests of this experimental task, the easily achievable target, T_l, was set at 140 symbols translated and the difficult-to-achieve target, T_h, was set at 220 symbols translated. Under the bonus contract, participants received a fixed payment of $10 ($W_b$) plus a bonus of 2.5 cents (λ) per symbol translated between 140 and 219 symbols (to a

[3] A pilot study, which included 12 participants drawn from the same population as the actual participants in this study, suggested that task performance levels off after three periods.

maximum of $2). They received an additional fixed bonus of $8 for translating 220 symbols or more (i.e., exceeding T_h), resulting in a total possible bonus of $10.

Under the penalty contract, participants received a fixed payment of $20 ($W_p$) less a penalty, where the penalty was determined as follows. If participants' correctly translated at least 220 symbols (T_h), the penalty was zero. If they correctly translated between 140 and 219 symbols, the penalty was $8 plus a piece-rate penalty of 2.5 cents (λ) multiplied by the difference between 220 symbols and the number of symbols correctly translated (to a maximum of $2). If participants correctly translated less than 140 symbols (T_l), no additional penalty was imposed (i.e., the maximum penalty was $10). Under both contracts, if participants failed to translate at least 140 symbols, the low target, they received $10 and if they translated 220 symbols or more, they received $20. The dollar amounts were selected to ensure participants were interested in exerting effort on the task.

Subsequently, participants performed the symbol translation task for the fourth time. At the conclusion of the fourth period, participants completed a post-experiment questionnaire, which included a manipulation check, demographic questions, and various questions to elicit participants' perceptions of the experiment. After completing the questionnaire, participants were paid and given one raffle ticket, which provided them the opportunity to win a $150 cash prize.[4]

RESULTS

Manipulation Check

All participants were required to calculate expected pay under three different sets of data concerning hypothetical levels of actual performance using the contract frame (bonus or penalty) under which they were paid during the treatment period. These computations were made just before the beginning of the fourth (treatment) period. All participants correctly calculated their pay using these examples. Accordingly, participants understood the nature of the payment under their assigned contract frame.[5]

Descriptive Statistics

Descriptive statistics by period are included in Table 1 (Panel A). In the first three periods, mean performance was 106.97 symbols translated (std. dev. = 20.21), 138.16 symbols translated (std. dev. = 25.53), and 167.60 symbols translated (std. dev. = 31.76)

[4] Because we infer differences in the degree of effort motivated by the differentially framed incentive contracts, we considered allowing participants to make a real choice between leisure and effort. In the fourth period, we allowed 30 participants to choose whether to work on the translation task (effort) or to spend time reading from a selection of magazines provided (leisure). Only three of the 30 participants took the leisure opportunity for at least some portion of the fourth period. Notably, one participant (in the penalty condition) selected the leisure activity for the entirety of the period decoding zero symbols and accepting the base wage. Given that participant reacted in a more extreme manner to the incentive contract than the other participants, we excluded that participant from the data analyses that follows. We repeated the data analysis including this participant and inferences are unaffected. We also ran all our tests reported below including the leisure-choice variable. The variable never approached statistical significance ($p > 0.50$) and inferences for the contracting-frame variable were unaffected.

[5] As part of the post-experiment questionnaire, we asked participants to indicate whether the incentive contract used in the fourth (treatment) period specified a base amount (1) plus amounts for meeting or exceeding assigned targets or (2) less amounts for failing to meet or exceed assigned targets. Thirty-three of 36 participants in the bonus condition responded correctly to this question. By comparison, only 16 of 31 participants responded correctly in the penalty condition. The observed difference in response to the post-experiment question may arise because participants are not accustomed to penalty-framed contracts; such contracts seem to be used infrequently. We repeated all analyses excluding those who answered the post-experiment question incorrectly and found that inferences were unaffected.

TABLE 1
Descriptive Statistics

	Mean			Std. Dev.			Minimum		Maximum	
	Total	**Bonus**	**Penalty**	**Total**	**Bonus**	**Penalty**	**Bonus**	**Penalty**	**Bonus**	**Penalty**
Period 1 Performance	106.97	111.17	102.09	20.21	20.21	19.40	76	59	151	142
Period 2 Performance	138.16	141.00	134.87	25.53	26.06	24.91	78	50	199	190
Period 3 Performance (Performance capability) (P3)	167.60	169.42	165.48	31.76	29.95	34.13	80	80	226	272
Period 4 (Treatment) Performance (P4)	185.01	184.50	185.61	40.59	44.14	36.76	80	124	259	284
Normalized Performance (P4/P3)	1.12	1.11	1.14	0.27	0.34	0.17	0.59	0.86	2.88	1.75
Total Earnings	$25.94	$26.21	$25.62	$4.04	$4.20	$3.88	$20.00	$20.00	$30.00	$30.00
n	36	31								

Performance represents the number of symbols correctly decoded by participants in each experimental period. Bonus and Penalty are defined by an economically equivalent contract framed in either bonus or penalty terms.

respectively. The first three experimental periods were included as learning periods. It is interesting to note that by the end of the third learning period, the average performance fell between the target levels (T_l and T_h) used in the incentive contract. Further inspection of the data indicates that 57 of 67 participants achieved a performance level in the third period between T_l and T_h. Hence, the targets used in the incentive contracts appear to be reasonable.[6] The fourth (treatment) period is of primary interest because the bonus or penalty contract was introduced in this period. Mean performance in the fourth period was 185.01 symbols translated (std. dev. = 40.59).

In this experiment, we test for differences in *normalized performance*, measured as the ratio of performance in Period 4 (the treatment period, denoted P4) to performance in Period 3 (before the experimental treatment is introduced, denoted P3), depending on contract frame.[7] We note that P3 performance represents ability plus the level of effort individuals are willing to exert for fixed pay. Period 4 (P4) performance represents ability at the task plus the additional effort motivated by the incentive compensation, plus potentially some additional, small amount of learning. For the most part, the increase from P3 to P4 represents the effort increase induced by the incentive. Based on our pilot testing, we have reason to believe that the learning effect during P4 is small and, insofar as the learning effect is not a function of effort, it is expected to be similar across bonus and penalty conditions. Therefore, it should not be of major concern when comparing bonus and penalty conditions.

Overall, mean normalized performance (P4/P3) was 1.12 (std. dev. = 0.27). The descriptive data suggest that, on average, participants exerted more effort in P4 (i.e., under the budget-based incentive contract) than in the P3.

Hypothesis Tests

First, we performed an analysis of variance (ANOVA) in which the dependent variable was normalized performance and the independent variable was contracting frame (bonus or penalty). Results are presented in Table 2. We find that contract frame does not significantly affect normalized performance on the experimental task in the overall sample, $F(1, 65) = 0.193$, $p = 0.662$.[8]

Recall that our hypotheses require that we test for differences in performance depending on contract frame for those individuals falling within and outside of Region 2 of the incentive contract. To test these hypotheses, we would ideally like to partition participants into incentive contract regions based on their actual ability independent of the effects of contract frame. One measure of this ability might be participants' performance in P3 (i.e., our performance capability measure), although this measure is potentially quite noisy since participants received no direct incentives for producing output in P3 other than a fixed payment covering P1–P3 in total. An alternative would be to partition the sample based on P4 performance. This measure can also be problematic because it may commingle incentive effects and measurement issues (i.e., additional effort in P4 may be reflective of skill *and*

[6] Of the ten participants whose performance fell outside of the target levels, the performance of seven of them was below T_l (140 symbols) and the performance of three of them was equal to or above T_h (220 symbols). Of these participants, six were in the bonus condition and four were in the penalty condition.

[7] Defining normalized performance as the difference between performance in P4 and in P3 gives qualitatively similar results.

[8] We repeat all analyses reported in this section of the paper using nonparametric Mann-Whitney tests and the results (not tabulated) are unaffected.

TABLE 2
Analysis of Variance for Normalized Performance by Contract Frame

Source	df	Mean Square	F-Statistic	p-value
Panel A: Complete Sample				
Contract frame	1	0.014	0.193	0.662
Error	65	0.075		
Panel B: Region 2 (using P3 performance to partition sample)				
Contract frame	1	0.018	1.168	0.285
Error	54	0.015		
Panel C: Region 2 (using P4 performance to partition sample)				
Contract frame	1	0.055	6.568	0.014
Error	40	0.008		

Normalized performance is measured as P4 performance/P3 performance
Contract frame is defined as an economically equivalent contract framed in either bonus or penalty terms.
Region 2 is as defined in Figure 1.
No significant differences were found depending on contract frame in Regions 1 or 3 (i.e., the flat regions of the incentive contract as defined in Figure 1)

incentives effects, not just the latter). As a consequence, in the tests below, we report results using both methods to partition the data, after which we apply additional sensitivity tests.[9]

Using P3 to partition the data (i.e., treating participants with P3 performance between 140 and 220 as falling in Region 2), we find that 56 participants fall in Contract Region 2 (refer to Figure 1), where incentives are sufficient to motivate effort. Eight participants are categorized in Region 1 (i.e., performance falls below the low target), and three participants are categorized in Region 3 (i.e., performance falls above the high target or above the cap on incentive pay). Next, we perform two separate ANOVAs: one for the 56 participants identified as having performance in Contract Region 2 in P3 and one for the 11 participants whose performance fell within Contract Region 1 or Contract Region 3 in P3. For Contract Region 2, the effect of contract frame on performance is not significant at conventional levels: $F(1, 54) = 1.168$, $p = 0.285$ (see Table 2, Panel B) nor is it significant for Contract Regions 1 and 3 combined, $F(1, 9) = 0.001$, $p = 0.996$. Further inspection of the data indicates that, from P3 to P4, 17 participants perform at a level that moves them to a higher contract region (elaborated below). Thus, up to 17 participants could be misclassified; i.e., identified as being in a lower region than is actually the case. The effect is potentially significant in that 17 participants represent 25 percent of our sample.[10] The potential for misclassification suggests that the findings need to be interpreted carefully.

[9] While we might have asked participants to provide performance expectations for P4 performance after completing P3 and used this measure to partition the sample, results reported by Luft (1994) indicate there may be a frame-dependent differential bias in stated expectations about future performance making participants' expectations a potentially dubious measure of future performance.

[10] Of the 17 participants, nine are in the penalty-framed contract condition and eight in the bonus-framed contract condition.

We repeat the analyses using the alternative P4 performance measure to partition the data (i.e., treating participants whose P4 performance was between 140 and 220 as falling into Region 2). In this case, 42 participants fall into Contract Region 2. Eight participants' performance falls below the low target and 17 participants performed above the high target. For participants in Contract Region 2, contract frame is statistically significant, $F(1, 40) = 6.568$, $p = 0.014$ (see Table 2, Panel C). For participants in the other regions, contract frame is insignificant, $F(1, 23) = 0.013$, $p = 0.911$. The findings are consistent with H1a, but not H1b; however, caution is advised in interpreting the results.[11] Due to the mixed results and concerns over the means of partitioning the data, we conduct additional analyses. Specifically, we investigate the sensitivity of our findings using various other sample partitions.

Sensitivity Tests

A fundamental issue is whether performance in P4 includes incentive effects *and* effort adjustments, which are unrelated to incentive effects. To test the sensitivity of our results, we performed the hypothesis tests again after partitioning the sample into contract regions based on performance in P3 with adjustment, as described below.

Some participants who performed below a threshold in P3 are likely able to perform above it in P4 because of the additional effort induced by the introduction of incentives and perhaps because of continued learning. As mentioned earlier, 17 of 67 participants in fact move into a higher contract region between P3 and P4: 13 go from Region 2 to 3, three go from Region 1 to 2, and one goes from Region 1 to 3.[12] Accordingly, we partition the data using cutoff points to define contract regions that are below the contract threshold points (i.e., the low and high targets of 140 and 220 symbols).[13]

Because most participants who change regions move to region 3, the greatest potential for misclassification occurs at the upper cutoff point (i.e., between Regions 2 and 3 in Figure 1). As such, we focus our attention on testing the sensitivity of our results to various upper cutoff points. In addition, we examine the effect of moving the lower cutoff point down to 120 symbols from 140 symbols. By altering the lower cutoff point in this way, we find that four participants are re-classified from Contract Region 1 to Contract Region 2.[14] The choice of other plausible lower cutoff points had little effect on our results.

Next, we examine the effect on our results of using several plausible alternative upper cutoff points from 170 to 210 symbols in increments of ten symbols. As the upper cutoff is lowered, the number of participants in Region 2 is reduced. The findings, using each set of cutoff points (lower and upper) are presented in Table 3.

In all cases, contract frame is statistically significant at $p < 0.10$; moreover, performance for participants in Region 2 is higher under the penalty contract as predicted in H1a. For participants in Contract Regions 1 and 3, the ANOVA results (not tabulated) indicate

[11] These analyses are repeated using a difference measure (P4–P3) as the dependent variable, as opposed to normalized performance (P4/P3). The results are unaffected.

[12] Breaking the data down by experimental condition, eight participants assigned to the penalty-framed contract move from Region 2 to Region 3 and one moves from Region 1 to Region 3. Three participants assigned to the bonus-framed contract move from Region 1 to Region 2 and five participants move from Region 2 to Region 3.

[13] We acknowledge that the use of arbitrary cutoff points may still result in classification errors because effort levels may not be uniform across participants (i.e., performance in P3 may be reflective of ability for some participants, but not others).

[14] For these four participants, two have performance in P4 in Contract Region 2 (149 and 162 symbols, respectively) and two have performance in Contract Region 1 (80 and 113 symbols, respectively).

TABLE 3
Results of Sensitivity Tests of the Effect of Contract Frame on Normalized Performance using Alternative Cutoff Points to Partition the Sample for Contract Region 2

Source	df	Mean Square	F-Statistic	p-value
Contract Region 2 Defined by Cutoff Points of 120 and 210 Symbols				
Contract frame	1	0.058	2.848	0.097
Error	57	0.020		
Contract Region 2 Defined by Cutoff Points of 120 and 200 Symbols				
Contract frame	1	0.062	2.927	0.093
Error	51	0.021		
Contract Region 2 Defined by Cutoff Points of 120 and 190 Symbols				
Contract frame	1	0.081	3.647	0.062
Error	47	0.022		
Contract Region 2 defIned by Cutoff Points of 120 and 180 Symbols				
Contract frame	1	0.081	3.418	0.072
Error	42	0.024		
Contract Region 2 Defined by Cutoff Points of 120 and 170 Symbols				
Contract frame	1	0.121	5.196	0.029
Error	32	0.743		

Normalized performance is measured as P4 performance/P3 performance
Contract Region 2 is as defined in Figure 1.
Contract frame is defined as an economically equivalent contract framed in either bonus or penalty terms.
No significant differences were found depending on contract frame in Regions 1 or 3 (i.e., the flat regions of the incentive contract as defined in Figure 1) using these alternative cutoff points to partition the sample.

that contract frame never achieves statistical significance ($p > 0.60$). These findings fail to support our hypothesis H1b that framing language would affect individual effort even when monetary incentives alone were ineffective at motivating effort.

While the various cutoff points used in our sensitivity analysis could still result in misclassification, we take comfort in the fact that the findings are robust: i.e., inferences are similar across the different cutoff point specifications. As suggested earlier, the potential for misclassifying participants is greatest when distinguishing those in Regions 2 and 3.[15] By using a cutoff that is lower than 220 symbols as the high target, we may incorrectly categorize participants in Region 3 instead of Region 2: the possibility increases as the cutoff point for the high target is lowered.[16] Such misclassification works against H1a because sample size is reduced in Region 2, which lessens statistical power. By comparison, the potential misclassification works in favor of H1b. If contract frame impacts behavior, the effect is most likely to manifest itself for participants in Region 2 (who may be incorrectly included in Region 3). Yet, the results of the sensitivity tests are to the contrary. The findings

[15] Altering the cutoff for the low target has a minimal effect because few participants move between contract regions.
[16] The potential for misclassification appears to be similar across the experimental conditions.

suggest that contract frame affects participants' behavior, but only if marginal effort allows them to increase their monetary payoff.

DISCUSSION

In this study, we examine the effect of contract framing in bonus or penalty terms on individual effort. In particular, we examine this issue within the context of a budget-based incentive contract containing performance hurdles (including a cap on pay) that define the amount of performance-related pay individuals receive. The experiment described here was inspired, to a great degree, by the work of Luft (1994) exploring the effects of language (i.e., the nature of bonus and penalty labels) on individuals' preferences for bonus-framed over economically equivalent penalty-framed contracts. Although Hannan et al. (2005) find that individuals are more likely to select a costly level of effort to avoid a penalty than to receive a bonus, the conditions under which subordinate effort will be affected by framing and when it will not remain an open question.

Consistent with prospect theory (Kahneman and Tversky 1979), our findings indicate that individuals will exert more effort to avoid a penalty than to receive a bonus when financial incentives provided by the contract are effective, but not when they are ineffective (i.e., outside of the pay-for-performance range). This finding is contrary to the notion that bonus and penalty framing will color an individual's reaction to the overall contract regardless of the incentive region in which the individual's performance falls (i.e., above the high target and at maximum pay).

Our study has limitations: First, while our choice to increase the external validity of our experiment through the use of real effort and a kinked incentive contract with a cap on pay allowed us to examine the issues in the paper on a more real-world basis than in previous studies, it also presented several challenges. Examples of these challenges include difficulties in disentangling effort and skill, making the partitioning of the sample into incentive regions difficult. In addition, the cap on pay may have created the potential for ceiling effects on performance in Region 3 (i.e., above the high target) that may not have allowed us to find support for our hypothesis concerning performance in that region independent of the effect of contract frame.[17] Even so, we believe that our study provides additional evidence that contract framing affects the amount of effort individuals will exert when incentives are effective at motivating effort.

Our finding that contract frame is irrelevant once the high target is achieved also has implications for practice. Specifically, if a large proportion of managers can achieve the maximum payout (i.e., fall within Region 3 of Figure 1), then contract framing will not matter to the additional effort they are willing to exert. Any additional effort they do exert after pay is capped may be related to intrinsic as opposed to extrinsic factors which are manager-specific (e.g., feelings of satisfaction from doing a good job). In these situations, the organization is better off offering bonus contracts since employees generally prefer them and there should be no difference in effort due to contract frame above the cap. Even so, it is likely that many managers in real organizations will fall within the incentive range (i.e., Region 2) and are likely to be subject to the framing effects documented in our study.

[17] Alternatively, it could be that the small number of participants whose performance fell within Region 3 of the incentive contract limited our power to find statistically significant differences between bonus and penalty contract frames in that region. Future research should take this possibility into account.