UNIVERSITEIT GENT

FACULTEIT ECONOMIE
EN BEDRIJFSKUNDE

HOVENIERSBERG 24
B-9000 GENT
Tel. : 32 - (0)9 – 264.34.61
Fax. : 32 - (0)9 – 264.35.92

WORKING PAPER

The Impact of the Number of Performance Measures and Incentive Framing on Performance in a Multidimensional Task Environment

Peter Van de Weghe [1]

Werner Bruggeman [2]

August 2006

2006/404

[1] Department of Accounting and Corporate Finance; Ghent University
[2] Department of Accounting and Corporate Finance; Ghent University
We gratefully acknowledge the helpful comments of Frank Selto, Patricia Everaert, Patrick Van Kenhove and Marc Buelens on earlier drafts of this paper
Corresponding author: Peter Van de Weghe, Kuiperskaai 55E, 9000 Gent, Belgium; Peter.VandeWeghe@Ugent.be

D/2006/7012/49

ABSTRACT

This paper reports on the results of a 2x2 experiment, examining the impact of multiple performance measures (PM) and incentive framing on performance in a multidimensional environment. A comparison is made between the performance of people working under one PM and those working under three PM. Additionally, the performance of participants working under a bonus scheme is compared with a penalty-framed incentive scheme. The results indicate that it is better to use multiple PM, because it inspires higher effort intensity and better effort allocation. Furthermore, it is better to frame incentives as a penalty scheme, instead of as a bonus scheme, because loss aversion has a higher performance impact than perceived fairness. However, the difference between bonus and penalty schemes is not larger with multiple versus a single PM, because loss aversion under multiple PM is higher for bonuses than for penalties. Consequently, no additive effects were identified.

INTRODUCTION

Management control tries to motivate people to act in accordance with an organization's goals and strategies. Central to a management control system is the evaluation of employee performance against a standard and understanding how incentives affect employee behavior (Young and Lewis, 1995). First, performance must be measured; but, since employees usually operate in a multidimensional environment and perform several tasks within a job, or a single task with several performance dimensions (Feltham and Xie, 1994; Hemmer, 1996; Holmstrom and Milgrom, 1991; Sprinkle, 2003), the question arises whether to use one or multiple performance measures (PM). In many instances, research has focused on using one PM (Chow, 1983; Dillard and Fisher, 1990; Bailey et al., 1998) and remained silent on the possible effects of multiple PM, even though it might be useful to investigate the incremental value of additional PM (Sprinkle, 2003). The current paper tries to address Locke and Latham's call (1990) for more research on multiple goal settings. Second, when employing an incentive system, not only must performance be measured, but also incentives should be provided (Prendergast, 1999). Now, an employer either can use a bonus or a penalty frame as a financial incentive mechanism. Economic theory predicts that employees are invariant to incentive framing when the economic payoff is equal. However, Baker et al (1988) found that there is an apparent asymmetry between rewards and punishments, and that penalty schemes are not very often observed in practice. Based upon prospect theory (Tversky and Kahneman, 1986), Luft (1994) empirically demonstrated that employees prefer bonuses over penalties. Despite this bonus preference, management control literature has not looked at the performance effects of framing, even though evidence suggests that punishment can incite people to work harder (e.g. Hannan et al, 2005; Starks, 1987; Waller and Bishop, 1990).

Hence, this paper investigates in a multidimensional environment how performance is influenced by the number of PM and incentive framing. This research question is relevant because previous research focused on performance measurement systems with a single PM and on bonus incentive schemes. The general results of the laboratory experiment indicate that penalty systems induce higher performance than bonus systems and that performance will be higher when people are evaluated on multiple PM as opposed to a single PM. No interaction effects were found.

The remainder of this paper is organized as follows. The next section discusses the literature and develops some hypotheses for study. Second, the methodology of the experiment is described, and third, the results of the data analysis are provided. The last section presents a discussion and conclusion.

RELEVANT LITERATURE AND HYPOTHESIS DEVELOPMENT

Incentives and performance

Every organization must face control difficulties, because owners want to maximize firm value, while employees seek to maximize their own utility. Therefore, owners need to motivate their employees to pursue the same goals they do themselves, an objective which can be achieved by providing variable monetary incentives (Ashton and Ashton, 1995). The design of incentive contracts seems to be particularly valuable in addressing the monitoring problem for non-managerial personnel (Aron, 1990). Moreover, monetary incentives are considered to solve control issues, as they have a decision-influencing role (Sprinkle, 2003). Hence, the general expectation is that monetary incentives have a positive impact on effort and, consequently, on performance (Bonner and Sprinkle, 2002). However, the link between incentives and performance is not frequently observed in practice (Baker et al, 1988) and prior research also has revealed somewhat conflicting evidence on the effectiveness of monetary incentives. Some researchers argue that pay is not an effective motivator, because people's intrinsic motivation for a job is negatively affected by giving extrinsic monetary rewards (Deci, 1972; Lazear, 1991 ; Prendergast, 1999), while others argue that a fixed wage is preferable to variable pay, because workers might misallocate their effort in a multitask environment, as a result of inaccurately measured performance measures (Holmstrom and Milgrom, 1991).

Bonner et al (2000) reviewed 131 experiments in diverse settings and found that monetary incentives positively affected performance in hardly more than half of the experiments. They also noted that the likelihood of observing positive effects, as a result of monetary incentives on performance, depends upon the type of task and the type of incentive scheme. First, as tasks become cognitively more complex, the likelihood of a positive relationship diminishes. Therefore, giving financial incentives to employees working on an operational task will have a larger effect on performance than linking managers' pay to their judgment and decision performance. Indeed, providing incentives in judgment and cognitive tasks conceivably will not have a significant positive effect, because performance in these tasks often is not enhanced just by 'thinking harder' (Awasthi and Pratt, 1990). Second, quota schemes result in higher performance than piece rate, tournament or fixed pay schemes, because of the use of explicit goals in the quota schemes, as opposed to 'do-your-best' goals in the other schemes. Indeed, Locke and Latham (1990) propose, in their goal-setting theory, that people are best motivated by specific and challenging goals, as long as they are attainable. Hence, monetary incentives can be assumed to be performance enhancing when the task is not too complex, and wherever quota schemes with difficult, but attainable goals are employed.

Performance Evaluation: Multidimensionality

Before an incentive scheme can be implemented, people's performance must be evaluated (Prendergast, 1999). One option is to use performance measures (PM) and to set explicit goals to be achieved, since explicit goals elicit the highest level of performance. Most academic research has addressed performance measurement systems with one PM and on tasks that are one-dimensional; however, employees usually operate in a multidimensional environment and perform several tasks within a job, or a single task with several performance dimensions (Feltham and Xie, 1994; Hemmer, 1996; Holmstrom and Milgrom, 1991; Sprinkle, 2003). For example, a worker may be required to work on quantity and quality of output. In this case, the worker must make sure that enough products of good quality are produced. Now, if the employer pays the worker on the basis of one single PM (e.g. quantity), then performance only will be satisfactory for that one PM, and not for the other PM (e.g. quality), because pay for performance is too effective and incites people to do exactly what they are rewarded on (Baker et al, 1988). As a consequence, this leads to a decline in overall performance. To counter this problem, Holmstrom (1979) argues that all possible PM that reveal information about people's performance should be incorporated into the compensation contract. Similar to this informativeness principle, Bonner and Sprinkle (2002) state that it is preferable to, at least, contract on a limited dimension of performance than no dimension at all, because their review showed no evidence of incentives having a negative effect on performance dimensions that are not rewarded, but did show a positive effect of incentives on the rewarded performance dimension. Additionally, an increasing number of PM may enlarge the set of implementable actions which can lead to a more preferred action (Feltham and Xie, 1994); and using multiple PM also gives employees an indication of what the organization's priorities are (Ashford and Northcraft, 2003; Kernan and Lord, 1990). This allows them to better direct their effort towards relevant organizational goals (Bonner and Sprinkle, 2002; Sprinkle, 2003). Hence, effort intensity will increase and effort allocation will be optimized, by using multiple PM.

Now, if additional PM only can be used when they are informative, it is important to realize that PM have the highest incremental value when they are congruent, measurable and controllable. A performance measure is congruent if the impact that an employee's action has on the performance measure also has the required impact on the employer's payoff (Feltham and Xie, 1994). Hence, a congruent measure prompts employees to direct their effort in a way that benefits the employer. Additionally, if one of the dimensions of worker performance is difficult to measure properly, then a fixed wage is preferable over an incentive scheme (Holmstrom and Milgrom, 1991). Otherwise, workers will manipulate the system by reallocating their attention towards the easily-measured dimensions to the exclusion of any difficult one (Baker et al, 1988; Lazear, 1991). Lastly, when a measure is uncontrollable, either because employees have no impact upon the measure or because the measure itself is not accurate, employees shift their attention away from that measure and focus on controllable measures (Feltham and Xie, 1994; Prendergast, 1999).

Despite the possible positive effect of using more performance measures, it is valuable to exercise caution with respect to including too many measures, because of the danger of information overload (Emsley, 2003; Ittner and Larcker, 1998). Using multiple performance measures can reduce performance, due to limited personal capabilities and it also tends to increase the complexity of a task (Gilliland and Landis, 1992), which causes incentives to be less effective (Bonner et al, 2000).

In summary, theoretical and analytical research suggest that multiple PM should be used, because of their additional value to the organization. These findings can be formalized with the following main hypothesis:

H1: Using multiple PM leads to higher overall performance than using only a single PM in a multidimensional environment.

On the basis of the discussed literature, the following explanatory hypotheses for H1 can be formulated:

H1a: Using multiple PM leads to higher employee effort than using only a single PM in a multidimensional environment.

H1b: When operating in a multidimensional environment, using multiple PM induces employees to divide their effort more evenly over the different PM, while using only a single PM leads employees to focus more on that single aspect of the task.

Providing Incentives: Bonus versus Penalty

When an incentive scheme is designed, the organization must not only measure performance, it also must decide how to provide incentives (Prendergast, 1999). Now, previous research almost exclusively has focused on bonus incentive systems. However, a central question when studying incentives and motivation is the choice of carrot or stick (Aron and Olivella, 1994; Lazear, 1991; Young and Lewis, 1995). Employers not only have the option to reward people with bonuses when performance is good, but also to punish them with monetary penalties when performance is unsatisfactory. First, suppose that a worker earns a fixed wage of $50 and an additional $20 bonus if he reaches a predefined goal. Second, suppose that another worker earns a fixed wage of $70 and is subtracted a $20 penalty if he does not reach a specific goal. In both cases, the worker earns $50 when the goal is not achieved and $70 when it is achieved. According to the invariance principle that underlies rational choice theory, individuals should be indifferent to this alternative incentive framing, because the economic payoff is equal in both cases (Aron and Olivella, 1994; Baker et al, 1988; Lazear, 1991; Tversky and Kahneman, 1986). However, contrary to economic theory, prospect theory states that framing matters (Kahneman and Tversky, 1979; Tversky and Kahneman, 1986).

The Impact of Incentive Framing on Preference

The prospect theory utility function is concave for gains and convex for losses, and it is steeper for losses than for gains (Kahneman and Tversky, 1979). Prospect theory includes three main concepts. First, there are reflection effects; this means that, under conditions of uncertainty, people will shift from being risk averse to being risk seeking in the domain of a perceived loss. Second, there is non-linear weighting of probabilities, meaning that people overrate low probabilities and underestimate high probabilities, when making choices between alternatives. The third concept is loss aversion, which means that people's response to losses is more extreme than their response to gains. Loss aversion is an important concept, and has its applications in a wide variety of disciplines, whenever choice behavior is studied. Loewenstein (1988), for example, has shown that the delay premium which people want to receive for delaying consumption of an acquired good is higher than the speed-up cost they want to pay for accelerating consumption of an acquired good which initially was meant to be consumed in the future. In short, people expect to receive more to forsake a good, than they are willing to give for receiving a similar advantage. Translated to management control, this means that when people have to choose whether to work under a bonus or a penalty scheme, the majority prefers the bonus scheme, because people prefer positive connotations over negative ones, due to loss aversion (Luft, 1994). This could be the reason that penalties are infrequently observed in practice (Lazear, 1991; Baker et al, 1988). Nevertheless, Lazear (1991) suggests that bonus (penalty) schemes should be used when output below (above) some critical level has no value, while Aron and Olivella (1994) demonstrated analytically that bonus schemes are appropriate for non-production jobs and penalty schemes for unskilled jobs or aspects of highly skilled jobs that require effort, but no skill.

The Impact of Incentive Framing on Performance

This paper focuses on the performance effect of the bonus-penalty dichotomy, because of the scant attention to this issue in previous literature (Chow et al, 1988). It is not because penalties are not preferred or because they are not frequently used in practice, that they cannot have satisfactory performance effects. On the contrary, some research provides evidence that penalty schemes yield better performance results than bonus schemes. Waller and Bishop (1990) empirically determined that, in an intra-firm resource allocation setting, unit managers tend to misrepresent private information far less and consume fewer resources under a penalty scheme. Starks (1987) discovered, in an agency model, that fund managers who have to make investment decisions take the investors' interests more into account under a combined bonus and penalty system than under a pure bonus scheme. Analogously, Lewis (1980) showed that the threat of a penalty for poor performance is sufficient to prevent an agent from shirking, and thus can have a positive effect on performance. Further, Nalebuff and Stiglitz (1983) demonstrated that, when, as part of a contest compensation scheme, a penalty is given to the lowest ranked individual, performance will be higher than when a prize is given to the highest ranked individual. Finally, Hannan et al (2005) set up an experiment in which participants had to assign levels of effort in a one-dimensional task, working under either a bonus or penalty incentive scheme. In these conditions, the different incentive schemes were imposed on the participants, instead of given to them as a choice, as in Luft (1994). Hannan et al (2005) found that penalty schemes generate a greater level of positive effort and achieve a higher degree of performance than what is observed using a bonus scheme, even though participants still prefer a bonus over a penalty scheme. The reason for this finding is the net effect of two conflicting forces, loss aversion and perceived fairness, wherein the former effect dominates the latter.

Because of loss aversion, people will respond more extremely to losses than to gains and, thus, prefer minimizing losses over maximizing gains (Zelditch and Ford, 1994). After all, these losses are taken away from something a person already possesses, and according to the endowment effect, people value goods more when they own them than when they do not; consequently, out-of-pocket costs will be avoided more tenaciously than opportunity costs (Kahneman et al, 1986; Sunstein, 2001).

As a consequence, performance will be higher to keep the fixed wage under a penalty scheme, than to earn something additional under a bonus scheme.

Besides being loss averse, workers also tend to spend more effort (shirk more) when they feel (un)fairly treated. This corresponds to the concept of 'reciprocity', which states that people who feel they are being treated fairly by another party, will treat that party fairly in return (Fehr et al. 1997; Perugini et al. 2003; Rabin, 1993; Sprinkle, 2003). Similarly, people incurring losses from a penalty scheme often feel unfairly treated (Kahneman et al, 1986). They act accordingly, and perform less well than when they are evaluated under a bonus scheme, because people are willing to forego some money in order to punish unfair behavior (Rabin, 1993). However, it is plausible that the impact of perceived fairness on performance is mitigated, because reciprocity works both ways. Not only will workers perform worse when they feel unfairly treated, but employers also will not compensate their workers well when they observe shirking. Now, if workers anticipate this dual reciprocity, they will indeed perceive the bonus system to be fairer than the penalty system, but this will not necessarily result in higher performance under the bonus system. This means that the impact of fairness on performance will be very low or even inexistent and, thus, also lower than the performance impact of loss aversion. This leads to the following main hypothesis:

H2: Performance is higher under a penalty scheme than under a bonus scheme.

On the basis of the discussed literature, the following explanatory hypotheses for H2 can be formulated:

H2a: Loss aversion is higher under a penalty scheme than under a bonus scheme.

H2b: Perceived fairness is higher under a bonus scheme than under a penalty scheme.

The net effects of H2a and H2b will lead to H2, if and only if the following hypothesis holds:

H2c: Loss aversion has a larger positive impact on performance than perceived fairness.

Interaction Effects of Multidimensionality and Incentive Framing

It can be learned from previous literature, that multidimensionality primarily has focused on bonus incentive schemes, and that the little research conducted on the bonus-penalty dichotomy exclusively has focused on one PM conditions. Even though the main effects remain as hypothesized, it is also predicted that, when multidimensionality and incentive framing are combined, the performance difference between bonus and penalty schemes will be larger using multiple versus one PM. Again loss aversion and perceived fairness must be investigated to demonstrate this. First, according to hedonic framing (Thaler, 1999), utility is maximized when gains are segregated and losses integrated, because the prospect theory utility function is concave for gains and convex for losses. This finding would imply that penalties should be employed when one PM is used, and bonuses when multiple PM are used. However, Thaler (1999) found that subjects do not require losses to be integrated, because segregated losses makes subjects more sensitive to future losses. This means that loss aversion and consequent performance will be higher when multiple losses can be incurred.

Second, fairness does not only play a role when one PM is used, but also when multiple PM are used. Increasing the number of performance measures to evaluate worker performance may reduce the risk workers bear when doing their job (Feltham and Xie, 1994). Under multiple performance measures, people might disperse their risk of being penalized over different measures. Now, because reducing risk can affect people's perception of fairness, the feeling of unfairness will decrease when multiple PM are used, as opposed to one PM.

In sum, because perceived fairness will increase equally in both the bonus and penalty conditions when comparing one versus multiple PM, and loss aversion will increase more under a penalty scheme than under a bonus scheme, when comparing one versus multiple PM, it is predicted that the performance difference between people under a bonus versus a penalty scheme grows larger when multiple performance measures are used. These predictions can be formalized by the following interaction hypothesis:

H3: The difference in performance observed between a penalty scheme and a bonus scheme will be larger using multiple versus one PM.

On the basis of the discussed literature, the following explanatory hypotheses for H3 can be formulated:

H3a: Loss aversion will be higher using multiple versus one PM.

H3b: Perceived fairness will be greater using multiple versus one PM.

However, these two hypotheses are not sufficient to lead to H3. The next hypothesis, which is an extension of hypothesis H2c, also must be true:

H3c: The difference between the performance effect of loss aversion and perceived fairness is larger using multiple versus one PM.

This can be visualized as in Figure 1.

[Insert Figure 1 around here]

RESEARCH METHOD

Design and Participants

To test the afore-mentioned hypotheses, a 2 x 2, between-subjects laboratory experiment has been set up. The manipulated variables were 'number of performance measures' (one versus multiple PM) and 'type of incentive framing' (bonus versus penalty). An experiment was preferred over field data for a number of reasons. First, if concepts infrequently exist in the real world (e.g. penalty, multidimensionality), experiments are the most appropriate methodology to use (Kachelmeier and

King, 2002). Second, experiments provide relatively pure and uncontaminated results. Third, the effects of independent variables on a dependent variable can be reasonably isolated through experimental manipulation. Finally, causal relationships are more easily investigated in experiments (Sprinkle, 2003).

One hundred and eleven undergraduate students, all enrolled in a management control course, participated in this study. Fifteen were dropped from analysis due to their failure to correctly answer the manipulation questions (n = 4), due to misunderstanding the task (n = 4) or due to missing values (n = 7). Of the 96 remaining participants, the mean age was 20.98 years with standard deviation 1.45 and 32 of them were male, while 64 were female.[3]

Task

Low task complexity was desirable, since Bonner et al (2000) argue that the likelihood of detecting positive performance effects is larger for simple jobs than for complex ones, because a lack of effort in easy tasks can be countered by monetary incentives, while a lack of skill in difficult tasks cannot. Therefore, a production task was deemed to be most appropriate. Much management accounting research has used this kind of task (Chow, 1983; Dillard and Fisher, 1990; Bailey et al. 1998) because it is believed such tasks simulate an assembly-line setting, which is relevant to management accounting (Bonner et al, 2000). An adaptation of Chow's coding task (1983)[4] was employed, yet, instead of merely coding and decoding, participants in the current experiment actually constructed fictitious products physically, because dimensions were added to allow for testing multidimensionality concerns (Ashford and Northcraft, 2003).

In the experimental task, products were made by assembling colored cards (the product parts) and stapling them together. There were six different colors and each color could be used more than once or not at all in a product. The colors to be used and the order in which they had to be assembled were indicated by means of a product code, which consisted of a product number and ten letters. Six different letters were used in the codes, each representing one particular color. Thus, each product was

[3] Analysis revealed that men and women did not perform differently (t = 1.59; p = .12)
[4] A pretest was employed to discriminate between three different adaptations of Chow's task (1983), and the most suitable task was selected.

constructed of ten colored cards, in accordance with the product code. Once the product was assembled, the product number had to be written on the front of the product. After assembly, the cards were stapled together at the left side of the product. Finally, each colored card also had a particular cost per unit, and participants were asked to calculate each product's cost on the basis of the assembled product (for an example, see Appendix A).

Despite the fact that multiple subtasks had to be performed, which increases the task's component complexity (Wood, 1986), this task was considered to be a simple task, because the number of subtasks (three) is albeit not very high. In addition, concerning the other two elements of task complexity offered by Wood (1986), it is argued that the coordinative complexity of the task is low because of limited information-processing requirements placed on the participants, and that the dynamic task complexity is not even relevant because the state of the environment does not change during the experiment (Bonner et al, 2000).

Procedure

The experiment was run over three different sessions. When entering the room, students received a personal number, which was used for randomization, payment purposes and matching task performance with exit questions. Participants were assigned randomly to one of the four experimental conditions, which were randomized within each session. There was no significant impact of session on the results (F = .432; p = .65). The course of the experiment was divided into three parts.

First, students were provided with three envelopes and a bundle of ten pages, with instructions and some explanation of the task. One envelope contained the six stacks of colored cards, and was meant to contain the completed products afterwards. A second envelope was empty, into which the students could dispose of their trial products and their remaining unused colored cards at the end of the experiment. The third and last envelope contained the exit questionnaire. On the instruction sheets, participants were asked to act as if they were assembly line workers. The mission of the company was given,[5] so as to delineate organizational priorities and to make sure that participants knew what was important to their company (Ashford and Northcraft, 2003). Additionally, participants were informed

[5] The company's mission was to deliver as many good quality products as possible to its customers.

of the desire of the company to use correctly calculated product costs for an as-efficient-as-possible management accounting application. No further contextual details were provided, because this could be a potential source of contamination, since situational factors are very powerful in influencing people's behavior (Ross and Nisbett, 1991). Subsequent to the task content, a five-minute trial session provided enough time for participants to master the task and to correct wrong interpretations of the task. After that, the employed incentive system was explained.

Second, students worked for 15 minutes on the experimental task. When finished, they were asked to make sure the completed products were in the right envelope and the remaining unused colored cards were in the empty one, after which both envelopes were collected.

Third, students filled out the exit questionnaire, on which they had to respond to manipulation checks, questions concerning their understanding of the task, their understanding of the incentive scheme, and other exit questions (see below).

Manipulations (Independent Variables)

The participants' objective, for all conditions, was to achieve predefined goals with respect to certain performance measures. Specific and challenging (i.e. difficult, but attainable) goals were set, because such goals are supposed to have a larger effect on performance than 'do-your-best' goals (Locke and Latham, 1990; Bonner et al., 2000). Pay then was linked to goal attainment. The incentive scheme featured two factors, to say, the number of performance measures and the incentive frame.

Number of Performance Measures (NRPM)

The multidimensionality of PM compares the effects of using one PM versus multiple PM in a multidimensional environment. While one PM (1PM) speaks for itself, multiple PM was operationalized by using three PM (3PM). Indeed, it is useful to incorporate as many performance measures as possible, but the danger of information overload must be taken into account (Emsley, 2003; Ittner and Larcker, 1998). In addition, it turns out that, for non-management employees, performance peaks when the compensation plan uses three to five performance measures (McAdams and Hawk, 1994). For pragmatic reasons, the use of three performance measures was chosen.

In the 1PM conditions, students were evaluated only on the number of products they made (quantity). Although these participants only were evaluated relative to product quantity, they were, as stated in the company's mission, still expected to make good quality products and to calculate the right product cost. For these conditions, the goal to be reached was making 13 products.[6]

In the 3PM conditions, students were evaluated on three performance measures: the number of products made (quantity); the percentage of correctly assembled products (quality); and the percentage of correctly calculated costs (calculation).[7] The respective goals for these three measures were 11 products, 98% and 80%.[8] Following from these operationalizations, goal difficulty should be equal for both the 1PM and 3PM conditions.

Incentive Framing

Incentive framing was manipulated by employing either a bonus scheme or a penalty scheme. Students working under a bonus scheme could earn a certain amount above their base pay, if they attained their goal(s), while those working under a penalty scheme could lose a certain amount from their base pay, if they failed to meet their goal(s). Base pay was different in the bonus and penalty treatments, but the absolute amount of the variable pay was equal. Minimum pay was not equal to zero, but a positive enough amount to entice people to accept the contract. A variable component was provided to induce positive effort and performance effects.

---

[6] This goal was determined by analyzing a 'do-your-best' pretest, and corresponds with the 75th percentile on the number of products made. The 75th percentile is used to define a 'difficult but attainable' goal (Locke and Latham, 1990; Erez et al, 1990).

[7] As mentioned before, the cost of each product had to be calculated on the basis of the assembled product, instead of on the basis of the product code. This meant that, even if the order or the colors of the assembly was not correct, participants still could calculate a correct product cost. This fits reality, in which costs should be calculated on used resources, and not on planned resources.

[8] These goals were determined by analyzing a 'do-your-best' pretest and the overall combined goal difficulty corresponds with the 75th percentile. The 75th percentile is used to define a 'difficult but attainable' goal (Locke and Latham, 1990; Erez et al, 1990). The data on the three measures in the pretest initially were standardized and averaged over the three measures. Then, the 75th percentile of this combined score was found to be .58. This cut-off point was de-standardized again to find particular goal levels for each performance measure. This corresponded to 11 products, 98% correct products and 80% correct cost calculations.

Specifically, in the 1PM-Bonus condition, base pay was set at € 6[9] and another € 6 could be earned when 13 or more products were made. In the 3PM-Bonus condition, base pay was € 3 and participants could earn another € 3 per attained goal (irrespective of which goal). In the 1PM-Penalty condition, base pay was € 12 from which € 6 could be subtracted, if the student did not generate 13 products. Finally, in the 3PM-Penalty condition, base pay was € 12 and € 3 was subtracted per goal that was not reached (irrespective of which goal). Consequently, in every condition, the expected pay was equal and amounted to € 7.5 (see Appendix B).

Dependent Variables

Performance

Performance was measured relative to three components. First, there was the number of products made by the participants (quantity). Second, the percentage of correct products was calculated (quality) and third, there was the percentage of correct cost calculations (calculation). These components were measured for all four conditions, including the 1PM conditions, even though the pay of workers in these groups only depended upon their attainment of the quantity goal. This is logical, because companies are interested in maximizing overall performance and not performance relative to the component(s) used for calculating variable pay. It is important that workers perform well on all relevant components, i.e. quantity, quality and calculation.

To create an overall performance measure, people's performance on quantity, quality and calculation was combined. However, because these measures utilize different measurement units, their scores could not merely be averaged. A measure with a neutral measurement unit was needed. Therefore, the values on the different performance measures first were standardized by setting the mean value equal to zero and standard deviation to one. Then, the average over all three measures was calculated and a final performance measure generated. Now, the different conditions could be compared easily, utilizing the same measurement unit.

---

[9] € 1 ≈ $ 1.15

Effort

To avoid a black box approach that only looks at the effects of the treatments on performance, workers' level of effort also was registered. Therefore, it was decided to use a self-reported three-item measure of effort on a five-point Likert-type scale, based upon work published by Earley et al (1987). Effort was measured for each of the three requirements: quantity, quality, and cost calculation. Per requirement, the Cronbach's alphas for effort were .82 for quantity, .88 for quality and .89 for cost calculation.

Fairness

Fairness of the incentive system was measured by means of ten items on a seven-point scale. The items in this study were adapted from Smither et al (1993) and Paterson et al (2002). The answers on these items then were combined into one fairness construct, with a Cronbach's alpha of .92.

Loss Aversion

The measurement of loss aversion was somewhat more complex because – to the best of my knowledge – no widely-accepted measure for this construct exists. In addition, loss aversion is context-specific and, thus, should be assessed using a contextual measure. In fact, existent (psychological) research frequently merely assumes loss aversion to be present as a consequence of a negative frame, without measuring it (e.g. Benartzi and Thaler, 1995). Other research thinks about loss aversion more theoretically or qualitatively, also without measuring it quantitatively (e.g. Camerer, 2000; Kahneman et al, 1990). In addition, most existing research generally looked at the effects of loss aversion on preferences, while the present paper wanted to investigate its effects on performance.

Therefore, an adaptation of the measure used by Hannan et al (2005) was chosen. As such, loss aversion was measured using a single-item measure calibrated with a nine-point Likert-type scale. The question asked how disappointed people would be if they would not earn a bonus (receive a penalty).

Exit Questions

In addition, some demographic questions were asked, as well as questions to check basic assumptions and randomization. The last two categories all were measured using five-point Likert-type scales. The questions addressed variables like goal difficulty, goal commitment, task complexity, task interest and preference for bonus or penalty (see Appendix C).

RESULTS

Descriptive statistics

The descriptive statistics of dependent and independent variables, as well as of the randomization and assumption check variables are presented in Table 1. The values of the (in)dependent variables will be discussed in the hypothesis testing section.

[Insert Table 1 around here]

One-sample t-tests on the other variables reveal agreement with the assumptions and successful randomization. Indeed, the goals are perceived as difficult but attainable, because, on a five-point scale, goal difficulty is significantly higher than three ($t = 9.42$; $p < .001$) and significantly lower than five ($t = -13.82$; $p < .001$). Further, goal commitment and task interest are higher than average (three) ($t = 13.36$; $p < .001$ and $t = 7.75$; $p < .001$ respectively). The task is perceived as slightly more complex than average (three) ($t = 2.66$; $p = .009$), but still not too complex (four) ($t = -9.23$; $p < .001$), so that task complexity is not reducing the effectiveness of the incentive scheme[10]. Finally, it can be shown that actual pay (mean = 7.41) does not differ significantly from expected pay (= 7.5) ($t = -.40$; $p = .70$) and that actual pay does not differ between all four experimental conditions ($F = .87$; $p = .35$).

[10] Additional analysis reveals that the results of the present paper do not change when task complexity is used as a covariate.

Manipulation checks

A manipulation check was performed for incentive framing by asking what people's base pay was, whether they could earn extra money or money could be subtracted, and what the minimum and maximum pay was. Three people did not answer correctly and were excluded from analysis.

Another manipulation check concerned the number of performance measures and asked on how many performance measures they were evaluated. One participant in the 1PM treatment answered incorrectly '3' and was excluded from analysis.

Hypothesis Testing

The appropriate and relevant means for performance, loss aversion and fairness used in the following hypothesis testing are summarized in Table 2.

[Insert Table 2 around here]

Hypothesis 1 and its Explanations

Performance. H1 predicts that, in a multidimensional environment, using multiple PM leads to higher overall performance than using only a single PM. The results in Table 3 show that there is, indeed, a significant main effect of the number of performance measures on performance (F = 7.906; p = .006), in that it appears to be better to evaluate people on 3 PM (mean = .175) than on 1 PM (mean = -.175). Thus, H1 is supported by the experimental data.

[Insert Table 3 around here]

Effort. H1a predicts that, in a multidimensional environment, using multiple PM leads to higher employee effort than using only a single PM. To test this, a MANOVA was performed with the three types of effort as dependent variables and the treatments as independent variables. The results of this analysis are shown in Table 4 of which Panel B reveals that, in general, there is an effort difference between 1PM and 3PM (Wilks' Lambda = .917; p = .049). From Panel A, it can be seen that this difference implies that using 3PM elicits more effort than 1PM, which supports H1a. Further analysis of univariate tests per effort dimension, which are shown in Table 4 Panel C, can explain where the effort difference comes from. Therefore, a critical Hotelling's T² statistic ($\sqrt{T^2_{crit}}$ = 2.50) was compared to the individual t-values and showed that the difference between 1PM and 3PM primarily was caused by differences in quality effort (means: 3.44 vs 3.88; t = -2.64) and calculation effort (means: 3.60 vs 4.01; t = -2.56), while quantity effort (means: 3.79 vs 3.89; t = -.76) did not differ significantly between the 1PM and 3PM groups.

[Insert Table 4 around here]

H1b predicts that, when operating in a multidimensional environment, using multiple PM induces employees to divide their effort more evenly over the different PM, while using only a single PM leads employees to focus more on that single aspect of the task. Pairwise comparisons of effort means demonstrated that, under 1PM conditions, more effort was spent on quantity (mean = 3.79) than on the average of the other two aspects (mean = 3.52) (t = 2.04; p = .047). Conversely, in the 3PM conditions, no one pairwise comparison was significant and effort was more evenly distributed across all three components. Thus, H1b was supported by the data.

Hypothesis 2 and its Explanations

Performance. H2 predicts that performance is higher under a penalty scheme than under a bonus scheme. The results are depicted in Table 3 and reveal that there is a main effect of incentive framing (F = 4.721; p = .032). The group means demonstrate that performance is better if people are working under a penalty scheme (mean = .133) than under a bonus scheme (mean = -.133). The data support H2. One potential explanation for this result is the net effect of both loss aversion and feelings of fairness.

Loss aversion and fairness. Hypothesis H2a postulates that loss aversion is greater under a penalty scheme than under a bonus scheme. The results are shown in Table 5 and display no main effect of incentive framing on loss aversion (F = .546; p = .462), but an interaction effect between incentive framing and the number of performance measures (F = 5.035; p = .027). Hence, the data do not support H2a, but show that, for 1PM conditions, loss aversion is not significantly different between the two incentive schemes (t = -1.02; p = .312), whereas, under 3PM conditions, loss aversion is even greater for bonus than for penalty schemes (t = 2.20; p = .033).

[Insert Table 5 around here]

Hypothesis H2b predicts that bonuses are perceived to be a fairer incentive system than penalties. The results, which are exhibited in Table 6, show a main effect of incentive framing on fairness (F = 4.159; p = .044). Looking at the group means demonstrates that bonus schemes (mean = 3.47) are perceived to be significantly fairer than penalty schemes (mean = 2.94), thereby supporting H2b.

[Insert Table 6 around here]

Loss aversion, fairness and performance. To investigate H2c, which predicts that loss aversion has a larger positive impact on performance than fairness, performance was regressed on fairness and loss aversion. The results are presented in Table 7, Panel A. Loss aversion was found to have a significant positive effect on performance (coefficient = .079; t = 2.90; p = .005), while fairness has no significant effect on performance (coefficient = .074; t = 1.48; p = .142). As a consequence, the impact of loss aversion appears to be larger than that of fairness, so that H2c is supported by the experimental data.

[Insert Table 7 around here]

Hypothesis 3 and its explanations

Performance. Hypothesis H3 presumes that the performance difference between a penalty scheme and a bonus scheme will be larger when multiple PM are used versus one PM alone. However, the results in Table 3 show that there is no significant interaction effect, and that H3 is not supported (F = .42; p = .518), which also is depicted in Figure 2.

[Insert Figure 2 around here]

Loss aversion and fairness. H3a predicts that loss aversion will be greater with multiple PM versus one PM. It can be seen from Table 5 that H3a is supported (F = 8.485; p = .004) and, thus, that loss aversion is greater across 3PM conditions (mean = 6.06) versus 1PM conditions (mean = 4.71).

Hypothesis H3b predicts that perceived fairness will be greater using multiple versus one PM. The results, given in Table 6, show that fairness depends upon the number of performance measures (F = 18.09: p < .001). Fairness, indeed, is perceived to be greater under 3PM conditions (mean = 3.73) versus 1PM conditions (mean = 2.68). H3b is thus supported by the data.

Loss aversion, fairness and performance. H3c predicts that the difference between the performance effect of loss aversion and fairness is larger using multiple versus one PM. This was tested by regressing performance on fairness and loss aversion across both 1PM and 3PM conditions. The results are shown in Table 7, Panels B and C, respectively. For 1PM conditions, loss aversion had a significant impact on performance (coefficient = .116; t = 2.564; p = .014), while fairness had no impact (coefficient = .0199; t = .218; p = .829). For 3PM conditions, both perceived fairness and loss aversion exhibited no significant impact upon performance. As a consequence, H3c cannot be accepted.

DISCUSSION AND CONCLUSIONS

This study provides evidence regarding the performance effect of using multiple PM and incentive framing in a multidimensional environment. Prior research had paid scant attention to both incentive scheme parameters. An experiment was designed to compare the performance effects of using one versus three PM, and of assigning a bonus versus a penalty incentive scheme to subjects.

First, the results demonstrate that, in multidimensional environments, workers perform better when multiple PM are used instead of one PM. This higher performance appears to be caused by workers putting more effort into satisfying the additional measures used in the multiple performance measurement system, while people do not differ in their degree of effort relative to the common measure; this, in turn, is because this measure is linked to their incentive scheme in both instances. In addition, people who are evaluated on one PM mainly focus on that particular measure – without totally neglecting the other dimensions – while people working under multiple PM divide their effort more evenly over all dimensions. Thus, using multiple PM increases effort intensity and improves workers' effort allocation, because they know better what is important in the organization.

Second, it was demonstrated that penalty-framed incentives induce higher performance than bonus schemes, although the majority of participants would prefer to work under a bonus scheme,[11] something which agrees with Luft's (1994) findings. Despite the larger performance effect of penalties, the performance difference between bonuses and penalties is no larger under multiple PM versus one PM, which is contrary to prediction. There may be two reasons. First, loss aversion and fairness have no performance impact under multiple PM, while loss aversion has a larger impact than fairness under one PM. This implies that the net effect will not be larger under multiple PM. Second, under multiple PM, loss aversion is greater for bonus schemes than for penalty schemes, which also is contrary to prediction. A possible explanation may be that, under multiple PM, people are using mental accounting, and in particular hedonic framing, under which people cognitively segregate gains and integrate losses, because this maximizes the utility of joint outcomes (Thaler, 1999). Now when people do segregate their gains, loss aversion concerning not getting a bonus also will be higher, where if people integrate their losses, loss aversion concerning getting a penalty will be lower. In sum, these results explain why the performance difference between bonuses and penalties is not larger

[11] The mean preference for all conditions equals 1.93 on a 5-point scale, going from Bonus over Indifferent to Penalty. All conditions prefer bonus schemes over penalty systems, but this preference is significantly lower ($t = -2.56$; $p < .05$) for people who are evaluated on multiple PM (mean = 2.17) than for people evaluated on one PM (mean = 1.69)

under multiple versus a single PM. Hence, multiple PM and incentive framing do not have additive effects.

The results of the present paper have several important implications. First, the results can arouse more interest of accounting researchers in investigating multiple PM and penalty incentive schemes, because both incentive scheme features have some value in eliciting higher performance. Second, incentive schemes must be designed more carefully. Whenever a worker performs a multidimensional task and his superior opts to link pay to performance, the task should be analyzed very thoroughly, in order to know which dimensions should be incorporated into the incentive scheme. Indeed, a one-measure incentive scheme should not be used for a multidimensional task, since that does not provide satisfactory performance. Third, incentive scheme designers also should consider the possibility of using a different incentive frame. Although penalty frames are not preferred by workers, they do induce higher performance. However, it may be difficult to implement such an incentive scheme, because of union resistance and loss of work satisfaction. One option to mitigate this problem is to communicate the penalty system as a fair system in which workers cannot be punished if they exert a certain effort. As a consequence, resistance might diminish and penalty systems might be more easily implemented.

Certain features of the study reflect important restrictions that may limit the generalizability of the obtained findings. These limitations suggest a number of avenues for further inquiry. First, the multiple PM condition was operationalized by using three PM, a decision derived from review of prior research (McAdams and Hawk, 1994). While this applies to operational tasks, Tuttle and Burton (1999) found that people's minds can process between 7 and 9 measures efficiently during decision making. As a consequence, future research should investigate whether the obtained results still hold in a management setting or when more than three PM are used.

Second, the experiment did not take into account the transaction costs that accompany the different incentive frames. The possible effects of psychological or administration costs, which typically are higher for penalty versus bonus schemes (Frederickson and Waller, 2005), were not incorporated. Future research could examine how these higher costs could be countered. One option would be to employ a mixed incentive scheme, in which bonuses and penalties are combined – this only is possible when multiple PM are utilized. This kind of incentive scheme might benefit from the performance effects caused by the penalty frame, as well as from the reduced level of worker resistance caused by incorporating a bonus frame, although this should be further explored in future research.

Third, the participants in the experiment worked for only one round, which excluded labor market reputation concerns and possible learning effects (Brüggen, 2006) as well as long-term behavior inducing people to act myopically. Since optimal multi-period incentive schemes can be quite different from single-period contracts (Jacobides and Croson, 2001), future research should investigate the role of incentive framing and multiple PM in long-term, principal-agent relationships.

Finally, this study was set in a certain environment in which the employed PM were congruent, controllable and measurable. However, sometimes people face uncertainty in the environment or in the measurement system (Feltham and Xie, 1994) which could lead to different impacts of incentive framing (e.g. Frederickson and Waller, 2005) or multiple PM (e.g. Coronado and Krishnan, 2005). This uncertainty issue certainly warrants further investigations.

REFERENCES

Aron, D.J. (1990). Firm organization and the economic approach to personnel management. *The American Economic Review*, 80 (2), 23-27

Aron, D.J. and Olivella, P. (1994). Bonus and penalty schemes as equilibrium incentive devices, with application to manufacturing systems. *Journal of Law, Economics and Organization*, 10 (1), 1-34

Ashford, S.J. and Northcraft, G. (2003). Robbing Peter to pay Paul: Feedback environments and enacted priorities in response to competing task demands. *Human Resource Management Review*, 13, 537-559

Ashton, R.H. and Ashton, A.H. (1995), *Judgment and decision-making research in accounting and auditing*. Cambridge University Press

Awasthi, V. and Pratt, J. (1990). The effects of monetary incentives on effort and decision performance: The role of cognitive characteristics. *The Accounting Review*, 65 (4), 797-811

Bailey, C.B., Brown, L.D. and Cocco, A.F. (1998). The effects of monetary incentives on worker learning and performance in an assembly task. *Journal of Management Accounting Research*, 10, 119-131

Baker, G.P., Jensen, M.C. and Murphy, K.J. (1988). Compensation and incentives: Practice vs Theory. *Journal of Finance*, 43 (3), 593-616

Benartzi, S. and Thaler, R.H. (1995). Myopic loss aversion and the equity premium puzzle. *The Quarterly Journal of Economics*, 110 (1), 73-92

Bonner, S.E., Hastie, R., Sprinkle, G.B. and Young, S.M. (2000). A review of the effects of financial incentives on performance in laboratory tasks: implications for management accounting. *Journal of Management Accounting Research*, 13, 19-64

Bonner, S.E., and Sprinkle, G.B. (2002). The effects of monetary incentives on effort and task performance: theories, evidence, and a framework for research. *Accounting, Organizations and Society,* 27, 303-345

Brüggen, A. 2006. The role of labor market reputation concerns in multi-task settings. Working paper, Maastricht University.

Camerer, C.F. (2000). Prospect Theory in the wild. In Kahneman, D. and Tversky, A. (Eds.), *Choices Values and Frames* (pp. 288-299). New York, NY: Cambridge University Press

Chow, C.W. (1983). The effects of job standard tightness and compensation scheme on performance: an exploration of linkages. *The Accounting Review*, 58 (4), 667-685

Chow, C.W., Cooper, J.C. and Waller, W.S. (1988). Participative budgeting: Effects of a truth-inducing pay scheme and information asymmetry on slack and performance. *The Accounting Review*, 63 (1), 111-122

Coronado, F., and R. Krishnan. 2005. Task complexity and incentive contracting – An empirical investigation. Working paper, Michigan State University.

Deci, I. (1972). The effects of contingent and non-contingent rewards and controls on intrinsic motivation. *Organizational Behavior and Human Performance*, 8, 217-229

Dillard, J.F. and Fisher, J.G. (1990). Compensation schemes, skill level and task performance: An experimental examination. *Decision Sciences*, 21, 121-137

Earley, E.C., Prest, W. and Wojnaroski, P. (1987). Task planning and energy expended: Exploration of how goals influence performance. *Journal of Applied Psychology*, 72 (1), 107-114

Emsley, D. (2003). Multiple goals and managers' job-related tension and performance. *Journal of Managerial Psychology*, 18 (4), 345-356

Fehr, E., Gächter, S. and Kirchsteiger, G. (1997). Reciprocity as a contract enforcement device: experimental evidence. *Econometrica*, 65 (4), 833-860

Feltham, G.A. and Xie, J. (1994). Performance measure congruity and diversity in multi-task principal/agent relations. *The Accounting Review,* 69, 429-453

Frederickson, J.R. and Waller, W. (2005). Carrot or stick? Contract frame and use of decision-influencing information in a principal-agent setting. *Journal of Accounting Research*, 43 (5), 1-25

Gilliland, S.W. and Landis, R.S. (1992). Quality and quantity goals in a complex decision task: Strategies and outcomes. *Journal of Applied Psychology*, 77 (5), 672-681

Hannan, R.L., Hoffman, V.B., Moser, D.V. (2005). Bonus vs. Penalty: Does Contract Frame Affect Employee Effort? In Rapoport, A. and Zwick, R. (Eds.), *Experimental Business Research II*. Norwell, MA: Kluwer Academic Press, pp 151-169

Hemmer, T. (1996). On the design and choice of "modern" management accounting measures. *Journal of Management Accounting Research,* 8, 87-116

Hollenbeck, J.R. and Klein, H.J. (1987). Goal commitment and the goal-setting process: Problems, prospects and proposals for future research. *Journal of Applied Psychology*, 72 (2), 212-220

Holmstrom, B. (1979). Moral hazard and observability. *The Bell Journal of Economics*, 10, 74-91

Holmstrom, B. and Milgrom, P. (1991). Multitask principal-agent analyses: incentive contracts, asset ownership, and job design. *Journal of Law, Economics, and Organization,* 7, 24-52

Ittner, C.D. and Larcker, D.F. (1998). Innovations in performance measurement: Trends and research implications. *Journal of Management Accounting Research,* 10, 205-239

Jacobides, M.G. and Croson, D.C. (2001). Information policy: shaping the value of agency relationships. *Academy of Management Review*, 26 (2), 202-223

Kachelmeier, S.J. and King, R.R. (2002). Using laboratory experiments to evaluate accounting policy issues. *Accounting Horizons*, 16 (3), 219-232

Kahneman, D., Knetsch, J.L. and Thaler, R. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *The American Economic Review*, 76 (4), 728-741

Kahneman, D., Knetsch, J.L. and Thaler, R. (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of Political Economy*, 98, 1325-1348

Kahneman, D. and Tversky, A. (1979). Prospect Theory: An analysis of decision under risk. *Econometrica*, 47, 263-291

Kernan, M.C. and Lord, R.G. (1990). Effects of valence, expectancies, and goal-performance discrepancies in single and multiple goal environments. *Journal of Applied Psychology*, 75 (2), 194-203

Lazear, E.P. (1991). Labor economics and the psychology of organizations. *Journal of Economic Perspectives*, 5 (2), 89-110

Lewis, T.R. (1980). Bonuses and penalties in incentive contracting. *The Bell Journal of Economics*, 11 (1), 292-301

Locke, E.A. and Latham, G. (1990). *A theory of goal setting and task performance*. Englewood Cliffs, New Jersey: Prentice Hall

Loewenstein, G.F. (1988). Frames of mind in intertemporal choice. *Management Science*, 34 (2), 200-214

Luft J., (1994). Bonus and penalty incentives contract choices by employees. *Journal of Accounting and Economics,* 18 (2), 181-206

McAdams, J.L. and Hawk, E.J. (1994). *Organizational Performance & Rewards: 663 experiences in making the link*. New York, NY: American Compensation Association

Nalebuff, B.J. and Stiglitz, J.E. (1983). Prizes and incentives: Towards a general theory of compensation and competition. *The Bell Journal of Economics*, 14 (1), 21-43

Paterson, J.M., Green, A. and Cary, J. (2002). The measurement of organizational justice in organizational change programmes: A reliability, validity and context-sensitivity assessment. *Journal of Occupational and Organizational Psychology*, 75, 393-408

Perugini, M., Gallucci, M., Presaghi, F. and Ercolani, A.P. (2003). The personal norm of reciprocity. *European Journal of Personality*, 17, 251-283

Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economic Literature*, 37, 7-63

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review*, 83 (5), 1281-1302

Ross, L. and Nisbett, R.E. (1991). *The person and the situation: Perspectives of social psychology*. New York, NY: McGraw-Hill

Smither, J.W., Reilly, R.R., Millsap, R.E., Pearlman, K. and Stoffey, R.W. (1993). Applicant reactions to selection procedures. *Personnel Psychology*, 46, 49-76

Sprinkle, G.B. (2003). Perspectives on experimental research in managerial accounting. *Accounting, Organizations and Society*, 28, 287-318

Starks, L.T. (1987). Performance incentive fees: An agency theoretic approach. *Journal of Financial and Quantitative Analysis*, 22 (1) 17-32

Sunstein, C.R. (2001). Toward behavioral law and economics. In Gowda, R. and Fox, J.C (Eds.), *Judgments, Decisions, and Public Policy* (pp. 218-240). New York, NY: Cambridge University Press

Thaler, R.H. (1999). Mental Accounting matters. *Journal of Behavioral Decision Making*, 12, 183-206

Tuttle, B. and Burton, F.G. (1999). The effects of a modest incentive on information overload in an investment analysis task. *Accounting, Organizations and Society*, 24, 673-687

Tversky, A. and Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business*, 59 (4), 251-278

Waller, W.S and Bishop, R.A. (1990). An experimental study of incentive pay schemes, communication, and intrafirm resource allocation. *The Accounting Review*, 65 (4), 812-836

Winters, D. and Latham, G. (1996). The effects of learning versus outcome goals on a simple versus a complex task. *Group and Organization Management*, 21 (2), 236-250

Wood, R.E. (1986). Task complexity: Definition of the construct. *Organizational Behavior and Human Decision Processes*, 37, 60-82

Young, S.M. and Lewis, B. (1995). Experimental incentive-contracting research in management accounting. In Ashton, R.H. and Ashton, A.H. (Eds.), *Judgment and decision-making research in accounting and auditing*. Cambridge University Press, pp 55-75

Zelditch, M. Jr. and Ford, J.B. (1994). Uncertainty, potential power, and nondecisions. *Social Psychology Quarterly*, 57 (1), 64-73

Appendix A

An example of a product code is given below. The task is explained in more detail by means of this product code:

4. U X V V W Z U Y W X

- The product number is 4 and should be written on the first card of the product
- According to the assembly code (see below), the cards can be assembled in the following order
    - Yellow, Violet, Blue, Blue, Green, Red, Yellow, White, Green, Violet
- Then the cards must be stapled together
- Further, the product cost is calculated according to the cost code (see below):
    - In this case, this yields (if correctly assembled) a product cost of 145

Assembly code:	**Cost code:**
U = Yellow	Yellow = 27 €
V = Blue	Blue = 13 €
W = Green	Green = 9 €
X = Violet	Violet = 14 €
Y = White	White = 7 €
Z = Red	Red = 12 €

Appendix B

Expected pay in the 1PM conditions must be equal to expected pay in the 3PM conditions.

Now it is explained how the fixed and variable amounts of the incentive schemes in the experiment were established. This was done on the basis of the goal setting pre-test with 'do-your-best' goals.

Under 1PM conditions, the quantity goal of 13 products was attained in 25% of the cases, while 75% of the people did not achieve that number. This leads to the following statement:

$$\text{Expected Pay} = \frac{1}{4} * A + \frac{3}{4} * B \quad (1)$$

Where A (B) is the amount earned if the goal (not) is reached.

Under 3PM conditions, people earn different amounts according to the number of goals attained. Therefore, it is first examined what percentage of participants attains a certain number of goals. In the experiment, 33,33% of the workers failed to satisfy any goal, while 8,33% satisfied one goal (irrespective of which goal that was). Two goals were achieved by 33,33% and another 25% attained all three goals (which corresponds to the 75th percentile of the 1PM conditions). This leads to the following statement:

$$\text{Expected Pay} = \frac{1}{3} * W + \frac{1}{12} * X + \frac{1}{3} * Y + \frac{1}{4} * Z \quad (2)$$

Where W, X, Y and Z are the amounts that are earned when 0, 1, 2 or 3 goals are attained, respectively. If both the quantity goal attainability in the 1PM conditions and the combined three goal attainability in the 3PM conditions must be paid equally, it follows that:

$$A = Z \quad (3)$$

Further, the marginal monetary value of attaining a goal should be the same whether it is the first, the second or the third goal. This is because the experiment employs a bonus and a penalty scheme. In the case of unequal marginal monetary values, it could happen that the marginal value of attaining a third goal in the bonus condition would be higher than the marginal value of attaining the first two goals. When compared to the penalty condition, this would not be fair, because these subjects would be penalized larger for not attaining one of the three goals, while the marginal penalty would diminish when more goals are not met. Therefore, a constant marginal monetary value is needed. Therefore, let this value be V:

$$X = W + V \quad (4)$$

$$Y = X + V = W + 2V \quad (5)$$

$$Z = Y + V = W + 3V \quad (6)$$

If (3) and (6) are substituted into (1), this leads to:

$$\text{Expected Pay} = \frac{1}{4} * W + \frac{3}{4} * V + \frac{3}{4} * B \quad (7)$$

and (4), (5) and (6) substituted into (2) leads to:

$$\text{Expected Pay} = W + \frac{3}{2} * V \quad (8)$$

Consequently, the result of (1) = (2) (i.e. equality of pay and thus (7) = (8)) should give:

$$B = W + V \quad (9)$$

If V and W are chosen arbitrarily (dependent upon limited monetary resources, as an experimenter) both to be 3 €, then the earned monetary values per condition are as written below (in euros) with an **expected pay of 7.5 €**:

	Number of goals attained			
Condition	0	1	2	3
Bonus – 1PM	6	12		
Bonus – 3PM	3	6	9	12
Penalty – 1PM	6	12		
Penalty – 3PM	3	6	9	12

Appendix C

Exit Questions: Cronbach Alphas, factor loadings and References

Variable		**Cronbach Alpha**	**Factor Loadings***	**References**
Goal difficulty		.70		Winters and Latham (1996)
Quantity (factor 1)	The quantity goal was easy to attain (REVERSED)		.69	
	It was difficult to reach the number of predefined products		.92	
Quality (factor 2)	The quality goal was easy to attain (REVERSED)		.80	
	It was difficult to reach the percentage with respect to the quality of the assembly		.71	
Calculation (factor3)	The calculation goal was easy to attain (REVERSED)		.81	
	It was difficult to reach the percentage with respect to the accuracy of cost calculation		.78	
Goal commitment		.80		Hollenbeck and Klein (1987)
Quantity (factor 4)	I was strongly committed to attaining the quantity goal		.68	
	I didn't care if I would achieve the quantity goal or not		.80	
Quality (factor 5)	I was strongly committed to attaining the quality goal		.86	
	I didn't care if I would achieve the quality goal or not		.87	
Calculation (factor 6)	I was strongly committed to attaining the cost calculation goal		.86	
	I didn't care if I would achieve the cost calculation goal or not		.85	
Task complexity		.70		NA
(factor 7)	How complex was the exercise		.90	
	The task is difficult to do		.87	
Task interest		.85		Chow (1983); Tuttle and Burton (1999)
(factor 8)	The exercise was interesting		.47	
	The exercise was fun		.92	
	The exercise was boring (REVERSED)		.90	
Bonus-penalty preference		NA	NA	Luft (1994)

* The factor loadings were obtained by performing a Principal Component Analysis extracting eight factors and using a Varimax Rotation

Table 1 – Descriptive statistics

Variable	N	Min	Max	Mean	Std Dev
Quantity performance	96	7	14	10.17	1.50
Quality performance	96	.75	1.00	.95	.065
Calculation performance	96	.077	1.00	.74	.21
Overall performance	96	-2.24	1.27	.00	.66
Effort on quantity	96	1.67	5.00	3.84	.71
Effort on quality	96	1.00	5.00	3.66	.86
Effort on calculation	96	1.00	5.00	3.81	.80
Loss aversion	96	1.00	9.00	5.39	2.38
Fairness	96	1.00	6.00	3.20	1.30
Goal difficulty	96	2.00	5.00	3.81	.84
Goal commitment	96	1.50	5.00	3.98	.72
Task complexity	96	1.50	5.00	3.22	.82
Task interest	96	1.00	5.00	3.64	.81
Actual pay	96	3	12	7.41	2.30

Table 2 – Relevant means for Performance, Loss aversion and Fairness per treatment and per condition (standard deviations in parentheses)

		1PM		3PM		Total	
Bonus		N = 23		N = 25		N = 48	
	Performance	-.365	(.83)	.081	(.54)	-.133	(.72)
	Loss aversion	4.35	(2.41)	6.72	(2.13)	5.58	(2.54)
	Fairness	2.87	(1.30)	4.02	(1.24)	3.47	(1.39)
Penalty		N = 25		N = 23		N = 48	
	Performance	-.001	(.62)	.278	(.49)	.133	(.57)
	Loss aversion	5.04	(2.28)	5.35	(2.19)	5.19	(2.22)
	Fairness	2.50	(.99)	3.40	(1.18)	2.94	(1.17)
Total		N = 48		N = 48		N = 96	
	Performance	-.175	(.74)	.175	(.52)	.000	(.66)
	Loss aversion	4.71	(2.34)	6.06	(2.25)	5.39	(2.38)
	Fairness	2.68	(1.16)	3.73	(1.24)	3.20	(1.30)

Table 3 – ANOVA for Performance

Source of Variation	Type III Sum of Squares	df	Mean Square	F	p-value
Intercept	2.906E-04	1	2.906E-04	.001	.979
Framing	1.881	1	1.881	4.721	.032
NRPM	3.149	1	3.149	7.906	.006
Framing * NRPM	1.67	1	1.67	.420	.518

Table 4 – MANOVA for Effort

Panel A – Mean scores for Effort spent on the different components for the 1PM and 3PM conditions (standard deviations in parentheses)

NRPM	Effort on quantity	Effort on quality	Effort on cost calculation
1 (N = 48)	3.79 (.66)	3.44 (1.02)	3.60 (.92)
3 (N = 48)	3.89 (.75)	3.88 (.60)	4.01 (.62)

Panel B – Multivariate tests

Effect	Wilk's Lambda	F	Hypothesis df	Error df	p-value
Intercept	.022	1355.553	3	90	.000
Framing	.940	1.914	3	90	.133
NRPM	.917	2.720	3	90	.049
Framing * NRPM	.997	.099	3	90	.960

Panel C – Univariate tests of between-subjects effects

Source of Variation	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	p-value
Intercept	Quantity	1413.973	1	1413.973	2881.226	.000
	Quality	1285.067	1	1285.067	1845.000	.000
	Calculation	1387.847	1	1387.847	2257.599	.000
Framing	Quantity	1.910	1	1.910	3.891	.052
	Quality	2.009	1	2.009	2.884	.093
	Calculation	.809	1	.809	1.316	.254
NRPM	Quantity	.285	1	.285	.580	.448
	Quality	4.842	1	4.842	6.952	.010
	Calculation	4.036	1	4.036	6.565	.012
Framing * NRPM	Quantity	.043	1	.043	.087	.768
	Quality	.095	1	.095	.136	.713
	Calculation	.000	1	.000	.000	.988

Table 5 – ANOVA for Loss Aversion

Source of Variation	Type III Sum of Squares	df	Mean Square	F	p-value
Intercept	2757.275	1	2757.275	543.847	.000
Framing	2.770	1	2.770	.546	.462
NRPM	43.020	1	43.020	8.485	.004
Framing * NRPM	25.525	1	25.525	5.035	.027

Table 6 – ANOVA for Fairness

Source of Variation	Type III Sum of Squares	df	Mean Square	F	p-value
Intercept	981.627	1	981.627	702.179	.000
Framing	5.814	1	5.814	4.159	.044
NRPM	25.289	1	25.289	18.09	.000
Framing * NRPM	.387	1	.387	.277	.600

Table 7 – Regression of Performance on Fairness and Loss Aversion

Panel A –All subjects

	Coefficient	Std. Error	t-statistic	p-value
Intercept	-.663	.224	-2.960	.004
Fairness	.074	.050	1.481	.142
Loss Aversion	.079	.027	2.899	.005

Adjusted R^2 = .085

Panel B –Subjects with 1PM

	Coefficient	Std. Error	t-statistic	p-value
Intercept	-.773	.368	-2.098	.042
Fairness	.0199	.091	.218	.829
Loss Aversion	.116	.045	2.564	.014

Adjusted R^2 = .090

Panel C –Subjects with 3PM

	Coefficient	Std. Error	t-statistic	p-value
Intercept	-1.66	.319	-.521	.605
Fairness	.074	.061	1.202	.236
Loss Aversion	.011	.034	.323	.748

Adjusted R^2 = .021

Figure 1 – Predicted relationship

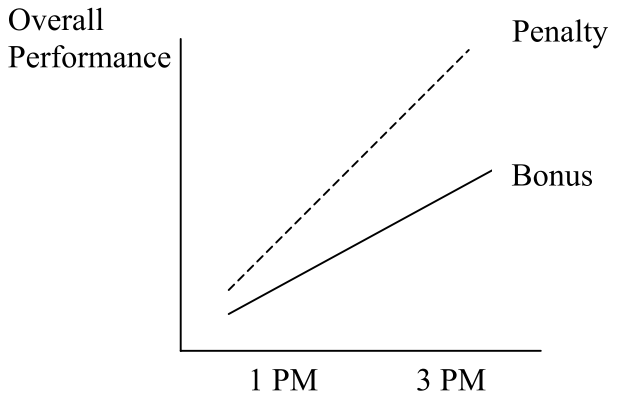

Figure 2 – Relationship based upon experimental results