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 Uncertainty on Incentive Framing Effectiveness in a Multidimensional Task Environment

Peter Van de Weghe [1]

Werner Bruggeman [2]

July 2006

2006/399

---

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/2006/7012/44

ABSTRACT

In this paper, an experiment was conducted in a multidimensional environment to examine the incentive framing effectiveness under conditions of uncertainty, as opposed to certainty. Whereas previous research generally has treated uncertainty as an indivisible concept, this paper proposes a framework from which it is clear that several sources of uncertainty exist, each influencing a different part of the proposed effort-outcome relationship. A comparison was made between certainty and uncertainty, stemming from uncontrollable factors respectively imperfect monitoring. The results indicate that it is valuable to use penalty-framed incentives under certainty conditions, and that performance is higher under certainty than under conditions with either source of uncertainty. The reason lies in a higher level of effort intensity and more efficient effort allocation. Furthermore, it seems that penalty schemes induce higher performance than bonuses under imperfect monitoring, while incentive framing has no effect under uncertainty stemming from uncontrollable factors. The latter results, under uncertainty conditions, can be explained by differences in risk attitude and perceived risk.

INTRODUCTION

A fundamental issue in management accounting is how to design incentive schemes that motivate workers to act in accordance with organizational goals. Because employees currently 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), it is most valuable to investigate incentive schemes in the light of multidimensional environments.

Incentives generally are assumed to have positive effects on performance, but there are certain factors that influence their effectiveness (Bonner et al, 2000; Bonner and Sprinkle, 2002). One of these factors is incentive framing, i.e. whether the incentive scheme is presented as a bonus or a penalty system (Aron and Olivella, 1994; Lazear, 1991; Young and Lewis, 1995). Although previous research has found that bonus schemes are preferred over penalty schemes (Luft, 1994), there also is evidence that, when a scheme is imposed, a penalty system incites greater levels of effort and performance than a bonus system (Hannan et al, 2005). These results generally were observed to occur under certain and stable conditions, whereas people's behavior also can be affected by uncertainty about future outcomes of current actions (Selto and Cooper, 1990). In addition, research on uncertainty primarily has focused on bonus schemes, finding a negative relationship between uncertainty and the use and effectiveness of incentive compensation (Banker and Datar, 1989; Feltham and Xie, 1994).

To overcome the limitations of previous research, this paper investigates how uncertainty can have an impact on the incentive framing effectiveness in multidimensional environments. The existing literature treats uncertainty as one indivisible concept, but it must be recognized that there are different sources of uncertainty (Prendergast, 2002). The current paper presents a framework of uncertainty which shows several possible sources of uncertainty. The basis of the framework is that people expend effort on a task relative to their expectations of the outcome (usually pay), an outcome which, in turn, is influenced by effort-performance, performance-evaluation and evaluation-outcome relationships. However, these variables and relationships are subject to a number of uncertainties. Indeed, uncertainty can exist with respect to person variables (e.g. risk attitude), task variables (e.g. set of implementable actions), environmental variables (e.g. assigned goals) and incentive scheme variables (e.g. rewarded dimension of performance), all of which have an uncertain impact upon effort, and upon the effort-performance and outcome-effort relationships. Furthermore, uncontrollable factors can have a direct influence on performance, while they also can affect the performance-evaluation and outcome-effort relationships, just like imperfect monitoring of performance. Finally, uncertainty also can arise with respect to the evaluation-outcome relationship, by discerning performance measures from reward measures. It is argued that, when investigating uncertainty, one should recognize these different sources of uncertainty and make a distinction between them, because they influence human behavior in different ways. Because this paper investigates incentive framing effectiveness under uncertainty, it is most interesting to look at the impact of the sources of uncertainty that influence the performance evaluation; that is to say, uncertainty driven by uncontrollable factors and imperfect monitoring.

Hence, a controlled laboratory experiment was employed to examine the effects of the sources of uncertainty on incentive framing effectiveness. In general, the results indicate that performance is higher when there is no uncertainty than when uncertainty exists. In addition, under a certainty condition, people perform better when their incentive scheme is framed as a penalty system than as a bonus system. The same pattern occurs under uncertainty stemming from imperfect monitoring, where people perform better under penalties than under bonuses. In contrast, under uncertainty due to uncontrollable factors, incentive framing will not have different effects on performance.

The remainder of this paper is organized as follows. The first section discusses the literature and develops an uncertainty framework and some hypotheses. In the second section, a description of the methodology used is presented, and the third section provides the results of the data analysis. The last section presents a discussion and conclusions.

LITERATURE REVIEW AND HYPOTHESES

Incentive framing under certainty

One of the central issues in incentive scheme design is that of carrot and stick (Aron and Olivella, 1994; Lazear, 1991; Young and Lewis, 1995). When considering incentive framing, there basically are two possible frames, bonus and penalty, which means that employees can be rewarded for satisfactory performance or be punished when performance is poor. Indeed, it is possible to give a fixed wage of $50 to an employee and reward him with an additional $20 bonus when he achieves a predefined goal; while, on the other hand, it also is possible to give a fixed wage of $70, but punish an employee by imposing a $20 penalty when a goal is not met. Because the expected economic payoff is equal in both incentive frames, economic theory predicts that individuals should be indifferent between them (Aron and Olivella, 1994; Baker et al, 1988; Lazear, 1991; Tversky and Kahneman, 1986). However, according to Kahneman and Tversky's (1979) prospect theory, framing actually does matter. The prospect theory value function is concave for gains and convex for losses, and steeper for losses than for gains. The latter feature means that people's response to losses is more extreme than to equal-sized gains, a phenomenon called loss aversion. Loss aversion leads to people preferring bonuses over penalties (Luft, 1994). However, despite this bonus preference, Hannan et al (2005) empirically demonstrated that penalty schemes encourage workers to work harder than bonus schemes, because loss aversion, which is higher under penalties, has a larger impact upon performance than fairness, which is higher under bonuses. In addition, Zelditch and Ford (1994) have shown that people rather minimize losses than maximize gains. This existing research on incentive framing exclusively has been situated under certainty, i.e. no uncertainty, so that the first hypothesis can be stated as follows:

H1: Under certainty, performance is higher under a penalty scheme than under a bonus scheme.

Uncertainty

Uncertainty can affect people's behavior, because it makes them insecure about future outcomes of current actions (Selto and Cooper, 1990). Therefore, uncertainty also can attenuate incentives' effectiveness. The predominant claim is that uncertainty leads to less incentive compensation (Banker and Datar, 1989; Feltham and Xie, 1994; Coronado and Krishnan, 2005), whereby these analytical and empirical findings imply that incentives have less performance enhancing value under uncertainty than under certainty. It is argued that the relationship between uncertainty and incentives is qualified by the different circumstances under which uncertainty arises, i.e. the sources of uncertainty.

Sources of uncertainty

In the accounting literature, there is no generally agreed upon definition of uncertainty. Nonetheless, uncertainty frequently is characterized as a state of the environment that adds noise to any performance measures that might be employed, thereby making them less informative (e.g. Banker and Datar, 1989; Feltham and Xie, 1994). However, although I agree that uncertainty can add noise to performance measures, the issue of uncertainty is more complex than treating it as one indivisible concept, like previous research has done. Indeed, according to Princeton's lexical reference system, Wordnet (http://wordnet.princeton.edu), uncertainty can be defined as "being unsettled or in doubt or dependent on chance". Now, based on this definition, different sources of uncertainty that may arise in management accounting are identified. Therefore, a framework is developed to show where uncertainty can arise in management accounting. This framework is shown in Figure 1.

[Insert Figure 1 about here]

The basis of the framework is that people expend effort on a given task relative to their expectations of the outcome (usually pay) (Bonner et al, 2000; Bonner and Sprinkle, 2002), and that the outcome is dependent upon the effort expended. The effort-outcome relationship can be broken down further into an effort-performance relationship, a performance-evaluation (measurement) relationship, and an evaluation-outcome relationship (Naylor et al, 1980). These are the main relationships that exist in the use of incentives as a motivational tool. Each of these relationships can suffer from uncertainty, as described later in this paper.

Variables that can create uncertainty about effort and the effort-performance relationship are person variables, task variables, environmental variables and incentive scheme variables (Bonner and Sprinkle, 2002). Examples of person variables are people's skill level and their risk attitude. It is obvious that if a person lacks the necessary level of skill, performance will not be highly correlated with effort (Bonner et al, 2000). Similarly, the effort spent on a task can be influenced by a person's risk attitude (Frederickson and Waller, 2005). Task variables also have an influence on effort and its relationship with performance. If a task is too complex, then people will not spend much effort on it; or, even if they do, their effort may not be fully converted into performance. Connected with this task complexity is uncertainty about the set of actions to be performed by the employee (Jacobides and Croson, 2001; Prendergast, 2002). When there is uncertainty about what the agent should do or on what he should focus, it is doubtful that high effort will lead to desired performance, both because of lower effort intensity and sub-optimal effort allocation. Coronado and Krishnan (2005), for example, argued that, when jobs are complex, this can lead to more uncertainty, because it is unclear what should be done or what is important. In this case, complex jobs will be more outsourced and less delegated within the firm, which results in less incentive compensation. This opinion is similar to those expressed by Bonner et al (2000) and Prendergast (1999), who assert that the more complex a task is, the less effective incentives will be. Additionally, one environmental variable is assigned goals, which can lead to uncertainty by being too difficult to attain, resulting in workers just 'giving up' (Locke and Latham, 1990). Other sources of uncertainty can be incentive scheme variables; for example, the rewarded dimension of performance. When the wrong kind of performance is rewarded, effort and its relationship with performance also face some uncertainty (Feltham and Xie, 1994, Hemmer, 1996, Holmstrom and Milgrom, 1991). These variables usually are not viewed as sources of uncertainty; but, actually, they are, since they render people uncertain as to whether their spent effort will be converted into subsequent performance, or they make them uncertain about spending any effort at all.

Some kind of uncertainty also can occur in the performance-evaluation relationship, or in performance itself. Performance can be influenced by uncontrollable factors, like the weather (climate) or machines that are not working properly, inducing higher uncertainty for employees (Frederickson and Waller, 2005; Prendergast, 2002). Concerning the performance-evaluation relationship, uncertainty also exists that is caused by uncontrollable factors; however, now it concerns uncertainty in developing performance measures (Feltham and Xie, 1994; Jacobides and Croson, 2001;

Prendergast, 2002). When designing an incentive scheme, one should strive to incorporate controllable performance measures. However, this is not always possible, as in the case of an assembly line worker whose performance is measured imprecisely by a failing performance measurement system. This undeniably leads to uncertainty which, in a multidimensional environment, can lead to people focusing on the easily-measured and controllable measures (Feltham and Xie, 1994), thereby creating considerable distortion between principal and agent goals. As such the set of states of nature over which the employee has no control, can lead to substantial uncertainty concerning performance measures (Waller, 1995).

Another example of a source of uncertainty in the performance-evaluation relationship is imperfect monitoring, which concerns the use of already developed performance measures, but in which employees face uncertainty because their performance is only partially, and not fully measured (Aron and Olivella, 1994; Jacobides and Croson, 2001; Holmstrom, 1979). Companies can, for example, opt not to monitor every single product made by their workers, but to use some kind of statistical sampling due to, for example, cost concerns[3]. When dealing with imperfect monitoring, workers face uncertainty, in that they are not informed which products will be checked or evaluated. Hence, imperfectly-monitored performance measures will be noisy, because they are less informative and possibly present an untrue view of performance.

Additionally, there can be uncertainty in the evaluation-outcome relationship, because a distinction can be made between performance measures and reward measures (Baiman, 1990). It can happen that it is not well-specified in advance which of the performance measures will be used in the compensation scheme or what the outcome will be (pay, promotion,…), leading to considerable uncertainty.

The last relationship that can experience uncertainty is that between outcome and effort. The uncertainties influencing this relationship also have an influence on the previously-mentioned relationships. These uncertainties involve person variables, task variables, environmental variables and

[3] Indeed, the principal must decide both how much to invest in the firm's production process and how much to invest in the monitoring system (Baiman, 1990). Therefore, an organization can decide to lower its monitoring intensity and invest more in its production process.

incentive scheme variables, as well as whatever uncertainty caused by uncontrollable factors and imperfect monitoring.

It is clear that, when studying uncertainty, one should make a distinction between different sources of uncertainty, because they are different in their origins and effects. Now, because this paper investigates incentive framing effectiveness under uncertainty, it is most interesting to look at the impact of the sources of uncertainty that influence performance measures, because those performance measures have the greatest direct effect upon outcome. In addition, Bonner and Sprinkle (2002) document that the performance-evaluation and evaluation-outcome relationships generally are minimized in laboratory studies of incentive effects, which warrants taking a further look at exactly that problem of incorrect performance measurement. Now, the uncertainty driven by uncontrollable factors and imperfect monitoring will be discussed.

Incentive framing under uncertainty stemming from uncontrollable factors

When workers perform a multidimensional task, some dimensions may be fully controlled and others not. This creates goal incongruence between principal and agent, whereby workers spend more effort on the more controllable performance measures and less on the uncontrollable dimensions (Brüggen, 2006; Feltham and Xie, 1994; Holmstrom and Milgrom, 1991). The skewed effort allocation that exists with this type of uncertainty leads to worse overall performance, as opposed to what occurs with situations of certainty (Jacobides and Croson, 2001). Prendergast (1999) also found analytically that, if there is measurement error, effort always is below the first best level reached under certainty. This can be stated formally in the following hypothesis:

H2: Task performance under uncertainty stemming from uncontrollable factors will be lower than task performance under certainty.

When performing a task, one can choose to be either diligent or to shirk. Under certainty, people expect a positive correlation between effort and outcome (Bonner and Sprinkle, 2002). This means that people who are diligent expect to achieve a positive outcome, while people who shirk do not expect a positive outcome. However, under uncertainty, the situation is different. Consider the problem independent from possible incentive framing effects. If an agent opts to be diligent, then he expects, due to uncontrollable factors, that there is a large possibility that he will be evaluated as diligent, but also a small possibility that he will be evaluated as shirking. If the agent chooses to shirk, then he expects, due to uncontrollable factors, that there is a large possibility that he will be viewed as shirking, but also a small possibility that he may be perceived as being diligent. Consequently, opting to be diligent is considered the less risky choice, versus shirking which can be considered a risky option.

When the bonus-penalty dichotomy and prospect theory are applied to this problem, it is clear that bonus and penalty systems will have different implications when there is uncertainty induced by uncontrollable factors. Indeed, Frederickson and Waller (2005) employed an experimental principal-agent setting, in which a bonus or a penalty was linked to an ex post monitored signal of an uncontrollable state (e.g. climate). One experimental group was provided with a bonus when the state signal was unfavorable, which meant that when the uncontrollable state likely was unfavorable to the agent's performance, this was counterbalanced by a bonus. Similarly, for the other group, a penalty was applied when the state signal was favorable, which meant that when the uncontrollable state likely was favorable to the agent's performance, this was counterbalanced by a penalty. Their results showed that agents systematically underweight the state signal under penalties, which implies that they overweight the chance of experiencing a penalty. Meanwhile, the state signal generally is optimally weighted when there are bonuses; this implies a correct weighting of the chance of experiencing a bonus. Given the knowledge that subtle changes in task or context (e.g. incentive framing) can induce remarkable changes in perceived risk (Selto and Cooper, 1990), the findings by Frederickson and Waller (2005) imply that perceived risk[4] will be higher under penalties than under bonuses. In short, people working under a penalty system will perceive the chance of being falsely accused of shirking as greater than people working under a bonus system will.

---

[4] Perceived risk is the subjective estimation of probabilities under uncertainty (Das and Teng, 2004)

Additionally, Tversky and Kahneman (1986) employed a set of differently framed choice problems under uncertainty, and found that, when the problem is stated in positive terms, people react risk averse; whereas problems stated in negative terms lead to risk seeking behavior[5]. The intuition is that, when people are diligent, they perceive the chance of being falsely accused of shirking as low under a bonus and high under a penalty system. The end result is that people working under a bonus system will tend to be diligent, and people working under a penalty system will tend to shirk, which is similar to the giving-up-phenomenon induced by tournaments (Baker et al, 1988) or by goals that are too difficult to attain (Locke and Latham 1990). Combining the findings of Frederickson and Waller (2005) and Tversky and Kahneman (1986) leads to the following hypothesis:

H3: Under uncertainty stemming from uncontrollable factors, performance is higher under a bonus scheme than under a penalty scheme.

Incentive framing under uncertainty stemming from imperfect monitoring

Another source of uncertainty is imperfect monitoring, whereby only a fraction of a worker's performance is monitored. With imperfect monitoring, the possibility of undetected shirking exists and people will try to game the system (Aron and Olivella, 1994; Jacobides and Croson, 2001). Lowering monitoring intensity decreases incentives to exert effort (Shapiro and Stiglitz, 1984), while under perfect monitoring, people have no incentive to shirk (Ahn and Faith, 1996). Generally, when performing a multidimensional task, people will focus their attention on the measures that are fully monitored, while they will shirk on the imperfectly monitored measures, resulting in a misallocation of effort (Feltham and Xie, 1994; Holmstrom and Milgrom, 1991). As a result, overall performance will be lower under imperfect monitoring, which corresponds again with Prendergast's (1999) logic that effort will be below the first best level when measurement error occurs. This can be written formally as the hypothesis:

---

[5] Risk attitude (i.e. being risk seeking, risk neutral or risk averse) is the construct that determines or reflects an individual's willingness to bear the consequences of a risky choice (Selto and Cooper, 1990)

H4: Task performance under uncertainty stemming from imperfect monitoring will be lower than performance under certainty.

The literature on efficiency wage models (e.g Altenburg and Straub, 1998; Nantz and Sparks, 1990; Shapiro and Stiglitz, 1984) shows that, when workers' effort and performance only can be imperfectly monitored, the threat of dismissal can induce workers to supply effort and to be diligent on the job. The reason is that, when imperfect monitoring occurs, people will shirk in equilibrium because when they are fired, another employer will hire them again under the same wage conditions. Therefore, companies should associate some kind of penalty with unemployment. One option is to raise wages compared to those in equilibrium, which implies that when fired, they lose the higher wage. Now, if all companies raise their wages, demand for labor decreases and unemployment results. In that way, the threat of dismissal induces workers to be diligent. While dismissal is a nonmonetary penalty, economic analysis by Aron and Olivella (1994) reveals how monetary bonuses and penalties can be used in compensation. These authors determined that, under conditions of imperfect monitoring, bonuses are appropriate for non-production jobs, whereas penalties are best for unskilled jobs or aspects of highly skilled jobs that require diligence, but no skill. This means that, for workers, penalty schemes probably induce higher levels of performance than bonus schemes.

Again, as in the case of uncontrollable factors, one can choose to be either diligent or to shirk, when there is imperfect monitoring. When a worker is paid under a bonus scheme and chooses to be diligent, this means that he spends his effort on all performance measures, including the imperfectly monitored ones. Because his effort must be divided over all performance measures, and because set goals are difficult, though attainable, it is difficult for the average worker to reach all goals. In combination with imperfect monitoring, this leads to an average or above-average bonus. In contrast, when the average worker spends less effort and opts to shirk on imperfectly-monitored measures, there is a small chance that no shirking will be detected and that the maximum bonus will be earned. Tversky and Kahneman (1992) performed an experiment in which students had to indicate their preference between sure outcomes and risky prospects. Analysis showed that for small probabilities of gains, people are risk seeking. Translated to the context of imperfect monitoring, this means that the risky prospect of earning a high bonus while shirking is preferred over a sure average bonus while being diligent. The reason is that they overestimate the low probabilities of prospects (Tversky and Kahneman, 1992) and, thus, underestimate the probability of detection, which means that they perceive the risk to be low. Therefore, they will try to game the system and shirk.

Similarly, average workers paid under a penalty scheme and choosing to be diligent will face difficulties reaching all performance goals. In combination with imperfect monitoring, this leads to an average or below-average penalty. Now when the worker chooses to shirk on the imperfectly monitored measures, a small chance exists that all shirking will be detected and that the maximum penalty will be applied. The previously-mentioned experiment of Tversky and Kahneman (1992) also showed that, for small probabilities of loss, people are risk averse. Translated to the context of imperfect monitoring, this means that the risk averse option of no or an average penalty under diligence is preferred over the risky prospect of receiving a high penalty while shirking. The reason is that workers tend to overestimate low probabilities of prospects (Tversky and Kahneman, 1992) and, thus, overestimate the probability of detection. Therefore, they will not try to game the system and will be diligent.

Here again, the underestimation (overestimation) of the probability of detection under bonus (penalty), shows that subtle changes in task or context (e.g. incentive framing) can induce remarkable changes in perceived risk (Selto and Cooper, 1990). Therefore, the following hypothesis are stated:

H5: Under uncertainty stemming from imperfect monitoring, performance is higher under a penalty scheme than under a bonus scheme.

RESEARCH METHOD

Design and participants

To test the hypotheses, a between subjects laboratory experiment was set up. A 2 x 3 factorial design was used to cross three levels of ‘uncertainty’ (certainty, uncertainty coming from uncontrollable factors, and uncertainty coming from imperfect monitoring) with two levels of ‘incentive framing’ (bonus and penalty). Consequently, six conditions were employed, resulting in two 2 x 2 experiments, because each source of uncertainty was compared separately with the certainty control group as comparison between both sources of uncertainty has no value.

One hundred thirty-seven undergraduate students, enrolled in a cost calculation course, participated in this study. A total of 14 students were dropped from analysis, for a variety of reasons, which included: failure to answer manipulation questions correctly (n = 5); not understanding the task (n = 1); not following the product sheet order (n = 3); having participated in a previous version of the experiment (n = 3); or missing values (n = 2). Among the 123 remaining participants, the mean age was 20.11 years, with a standard deviation of 1.50. Fifty-five of the 123 (44.7%) were male, while 68 were female.

Task

Because the present paper investigates worker behavior in a multidimensional environment, the experimental task used was an adaptation of Chow's (1983) decoding task, because it simulates an assembly line setting and has been used frequently in accounting literature (e.g. Bailey et al., 1998; Dillard and Fisher, 1990). In the adapted task, participants had to make fictitious products, according to a particular code. Indeed, 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 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 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).

Procedure

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 six experimental conditions. The course of the experiment was divided into three parts. First, students received three envelopes and a bundle of 11 pages, which included 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, in which the students could dispose of their trial products and the remaining unused colored cards at the end of the experiment. The third and last envelope contained an exit questionnaire. In the instruction sheets, they were asked to act as if they were assembly line workers. The mission of the company was given[6] 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 of the company's desire to use correctly calculated product costs for an as-efficient–as-possible management accounting application. No further contextual details were provided, as this could be a possible source of contamination, since situational factors are very powerful in influencing people's behavior (Ross and Nisbett, 1991). Subsequent to receiving task content, a five-minute trial session provided enough time for all workers to master the task and to correct incorrect interpretations of the task. After that, the employed performance measurement and incentive system was explained.

In the second phase of the experiment, 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 that the remaining unused colored cards were in the empty one, after which both envelopes were collected.

As the third and final experimental phase, students filled out the exit questionnaire.

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

Manipulations (Independent Variables)

Because the present paper is related to a multidimensional environment, the task employed included multiple dimensions. This was operationalized using three dimensions, because for nonmanagement employees, performance peaks when the compensation plan uses three to five performance measures (McAdams and Hawk, 1994) and information overload was to be avoided (Emsley, 2003; Ittner and Larcker, 1998). Those three dimensions lead to the use of 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 participants' objective was to reach predefined goals on each performance measure. Because specific and challenging (i.e. difficult but attainable) goals are supposed to have a larger effect on performance than just 'do your best' goals (Locke and Latham, 1990; Bonner et al., 2000), the overall goal difficulty level was set at the 75th percentile. The individual goal levels were deducted from a 'do your best' pretest as follows. First, the data on the three measures 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 resulted in goals of 11 products for quantity, 98% for quality, and 80% for the cost calculation measure. Subsequent to performance measurement, compensation was linked to goal attainment.

The first manipulation concerns uncertainty and relates to performance evaluation. The second manipulation concerns incentive framing and relates to incentive provision.

Uncertainty (see Appendix B)

Certainty. Actually, certainty truly is not manipulated, but is a control condition. This means that performance in this condition is fully and precisely measured, without noise. Participants working in

[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 wasn't correct, participants still could calculate a correct product cost. This fits reality, where costs should be calculated on used resources and not on planned resources.

this condition were informed that each product would be monitored and that there would be no measurement error when their performance was measured.

Uncertainty stemming from Uncontrollable Factors (UF). It is assumed that quantity is easily measured, just by counting the number of products made, which implies that uncertainty has no effect on this performance measure. In contrast, the quality of products and the correctness of cost calculation are susceptible to a certain degree of uncertainty stemming from uncontrollable factors. As a result, the performance measurement for quantity is similar to that of under certainty. Uncertainty stemming from uncontrollable factors only is applied to the quality and the cost calculation measures, not to the quantity measure. Hence, to operationalize uncertainty from uncontrollable factors to the quality and cost calculation measures, measurement error was applied (Banker and Datar, 1989) by using a normally-distributed error term (Feltham and Xie, 1994; Prendergast, 1999; Prendergast, 2002). This can be written as follows:

$$Y_{oi} = Y_{ai} + \varepsilon_i$$

where i is the performance measure for the quality and cost calculations; Y_{oi} is the observed performance; Y_{ai} is the actual performance; and $\varepsilon_i \sim N(0, \sigma_i^2)$ represents the measurement error. To make σ_i^2 credible and acceptable, standard deviations of 5.97% for the quality measure and 24.40% for the cost calculation measure were used, which were the standard deviations found in the pretest. When Y_o should fall out of the range [0, 1], it would be truncated to the closest value lying in this interval.

Participants were told that their performance was measured with error and that this error was normally distributed with a standard deviation of 5.97% for the quality measure and 24.40% for the cost calculation measure. In addition, to make the effects of the normal distribution more conceivable, they also were notified that 68% of the errors would fall within the [-5,97%, +5,97%] and [-24,40%, +24,40%] intervals respectively, whereas 95% would fall within the [-11,94%, +11,94%] and [-48,80%, +48,80%] intervals.

Uncertainty stemming from Imperfect Monitoring (IM). As in the case of uncertainty derived from uncontrollable factors, quantity is assumed not to be susceptible to uncertainty, as it only contains the number of products made. In contrast, the quality and cost calculation measures can be costly to observe and, thus, are susceptible to imperfect monitoring. This imperfect monitoring is operationalized by only checking 30% of the products for quality and cost calculation, which was determined as follows. The pretest learned that the average worker makes ten products, of which two are not perfect (either bad quality or incorrect cost calculation). This leads to the question, how many products should be monitored to have a fifty-fifty probability of detecting at least one imperfect product. Now, because monitoring products for quality and cost calculations is characterized by trials in which there are just two mutually exclusive outcomes possible (correct or incorrect), the binomial distribution can be drawn upon and the following equation for binomial probabilities can be solved for 'n':

$$P(k,n) = \frac{n!}{k!(n-k)!} p^k q^{(n-k)}$$

where P(k,n) = the probability of detecting k incorrect products when n products are monitored = .50, k = 0, p = the average percentage of incorrect products = .20, and q = 1–p. It was determined that n = 3; consequently, three products out of ten should be monitored to have a fifty percent chance of detecting incorrect products. Therefore, 30% was the monitoring intensity used in the present experiment.

The participants in this condition were told that it was too costly for the company to monitor each product separately and that only 30% of the assembled products would be checked on quality and cost calculation accuracy. They also were informed that their performance percentage would be calculated by dividing the number of correct products monitored by the total number of products, and not the number of monitored products.

Incentive Framing

The second manipulation was incentive framing, which was manipulated by employing either a bonus or a penalty scheme based on goal attainment. This goal attainment depends on a comparison of performance as measured under the (un)certainty conditions with predefined goals. Students working under a bonus scheme could earn a certain amount above their base pay when they attain one or more goals, while those working under a penalty scheme could have a certain amount subtracted from their base pay when they fail to attain these goals. 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 amount to let people accept the contract. A variable component was provided to induce positive effort and performance effects.

Specifically, participants working under the bonus system had a base pay of € 3 and could earn an additional € 3 per attained goal (irrespective of which goal). On the other hand, participants working under the penalty system had a base pay of € 12 and € 3 were subtracted per goal that wasn’t reached (irrespective of which goal). Consequently, pay was equal in every bonus and penalty condition, qualified by the source of uncertainty.

Dependent variable

Performance

Performance was measured in terms of 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). To create an overall performance measure, each student’s performance on all three measures was combined. However, because these measures have different measurement units, their scores could not be averaged or summed. A measure with a neutral measurement unit was needed. Therefore, the values for each of the different performance measures initially were standardized so that they all had a mean value of zero and a standard deviation of one. Then, the average of the three measures was calculated to generate a final performance measure. Additionally, performance with respect to each of the three conditions could be compared easily, using the same measurement unit.

Exit questions

The exit questionnaire included some demographic questions and manipulation checks concerning the sources of uncertainty and incentive framing, as well as questions to probe each participant's understanding of the task and the incentive scheme.

In addition, to verify results on the performance measures, some questions were incorporated to measure a participant's level of effort, because effort and performance should be highly and positively correlated to minimize uncertainty relating to the effort-performance relationship. Because of the simplicity of the tasks (Brüggen, 2006), there being the same established goals for all conditions, and the randomization of person variables, it was expected to observe this positive effort-performance correlation. Effort was measured with respect to each performance dimension using a three-item five-point Likert-type scale, based on work by Earley et al (1987). The Cronbach's alphas were .81 for quantity, .78 for quality and .83 for cost calculation.

Further, situation-specific measures for risk attitude and perceived risk were developed, because both variables were assumed to be situation-dependent (Selto and Cooper, 1990; Hanoch et al, 2006). For this reason and because comparisons over the uncertainty conditions have no value with respect to the hypotheses, both risk attitude and perceived risk were measured differently for each uncertainty manipulation. The measures (see Appendix C) were established after qualitative pre-testing on a group of four students. Risk attitude, in each case, was measured by means of two items on a five-point Likert-type scale, directly asking about the extent to which participants took risks or not. This measure is in line with the previously-stated definition of risk attitude (see footnote 11). The reliability of the risk attitude measure exhibited Cronbach's alphas of .75 for the certainty condition, .75 for UF and .90 for IM. Perceived risk was measured in two ways, because existing research in diverse fields yields no agreement on whether to use Likert-type scales (Chaudhary et al, 2004; Gerber and Neeley, 2005; Siegrist et al, 2005) or a subjective estimation of probabilities (Slovic, 1987; Das and Teng, 2004).

Perceived risk under certainty was measured by asking to what extent participants expect measurement errors or the detection of bad products. Perceived risk for UF was measured by asking how strongly they expected the random measurement error to be in their favor or to their detriment. This expectation of beneficial versus detrimental measurement error is perfectly and negatively correlated, implying that all subjects in the UF conditions expected measurement error one way or another. Because perceived risk usually refers to downside variances (Das and Teng, 2004), only the measure for unfavorable measurement error was used for analysis. At last, perceived risk for IM was measured by asking to what extent participants expected bad products to be detected, if they made such products. Both the five-point Likert-type scales and the probability estimation showed these variables to be highly correlated (with correlations between .56 and .87, all of them being significant at the .01 level), so only Likert-type measures were used for further analysis.

Control of other sources of uncertainty

As previously explained and in accordance with Figure 1, there are other possible sources of uncertainty when investigating incentive effects. To make sure that the experimental results were due to the employed manipulations alone, the other sources had to be properly controlled by measurement, careful design or randomization.

Because all participants were randomly assigned to one of the experimental conditions, the person variables they bring to the experiment should have been controlled by randomization, except for risk attitude which is a variable that is controlled by measurement, because it is a situation-specific variable, influenced by the respective incentive frames. Additionally, uncertainty about task variables, environmental variables and incentive scheme variables were controlled by the design of the experiment, because for each condition the same task was to be performed, the same institutional detail was given, the same goals were established, and people were rewarded relative to the same dimensions of performance and received equal pay according to their performance.

Furthermore, any direct effect of uncontrollable factors on performance was ruled out, because the manipulation of uncontrollable factors only pertained to uncertainty about the performance measures, and, thus, the performance-evaluation relationship. Consequently, performance itself could not be affected by random measurement error.

Finally, there was no uncertainty possible with respect to the evaluation-outcome relationship, because all performance measures were incorporated into the incentive plan. As such, the outcome was totally and perfectly correlated with the performance evaluation.

As a result, after controlling for the uninvestigated sources of uncertainty, Figure 1 can be downsized to Figure 2, which only depicts the relevant sources of uncertainty examined in the present paper.

[Insert Figure 2 about here]

RESULTS

Descriptive statistics

Descriptive statistics on dependent variables and relevant exit questions are given in Table 1. Note that the three performance variables are described here in their original measurement units, number (quantity) and percentages (quality, calculation), while further analysis is done with the standardized overall performance variable.

[Insert Table 1 around here]

To verify that uncertainty between effort and performance was minimized, correlations between both variables and their respective dimensions were calculated. These are given in Table 2 and show that all dimensions of effort and performance are significantly correlated with their respective counterpart.

[Insert Table 2 around here]

Manipulation checks

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

Another manipulation check concerned uncertainty. First, a general question asked whether the performance measurement method led to uncertainty (on a five-point Likert-type scale). People in the certainty condition (mean = 2.35) felt significantly less uncertainty than people in the UF (mean = 3.17; t = -3.12; p = .003) or IM (mean = 3.21; t = -3.30; p = .001) conditions. Second, to ensure that people in the UF and IM conditions felt the kind of uncertainty that had been intended, they were asked whether 'monitored products were inspected without error' (UF: mean = 1.98; IM: mean = 3.93; t = -12.10; p < .001) and whether 'quality and cost calculations were checked for each assembled product' (UF: mean = 3.85: IM: mean = 2.05; t = 12.17; p < .001). As such, manipulations were perceived as intended.

Hypothesis testing

The appropriate and relevant means for performance, effort, risk attitude and perceived risk used in the following hypothesis testing are summarized in Table 3.

[Insert Table 3 around here]

Hypothesis 1

H1 predicted that under certainty, performance will be higher under a penalty scheme than under a bonus scheme. Considering performance under certainty shows that penalty schemes (mean = .553) indeed exerted a significantly greater effect on performance than bonus schemes (mean = .045) (t = 375; p = .001), such that that H1 is supported.

Hypothesis 2 and 3 (see Figure 3)

Hypotheses 2 and 3 concerned the comparison of the certainty conditions with the UF conditions. To test these hypotheses, an analysis of covariance (ANCOVA) was performed, for which the results are provided in Table 4. The dependent variable is overall performance, with incentive framing (bonus-penalty) and uncertainty (certainty-UF) as independent variables, while risk attitude served as the covariate, meant to remove unexplained variation from the error term.

[Insert Figure 3 around here]

[Insert Table 4 around here]

Hypothesis 2. H2 predicted that task performance under uncertainty stemming from uncontrollable factors will be lower than task performance under certainty. From Table 4, it can be seen that the uncertainty main effect is significant (F = 78.462; p < .001), confirming that performance is higher under certainty (mean = .299) as opposed to under UF (mean = -.300); thus, supporting H2.

To investigate where this result comes from, the difference in effort allocation was investigated between both certainty and UF. The multivariate analysis of covariance (MANCOVA), with all three effort dimensions as dependent variables, framing and uncertainty as independent variables, and risk attitude as the covariate, is shown in Table 5 Panel A. It can be seen that both certainty and UF differ with respect to effort spent (Wilk's Lambda = .63; p < .001). Univariate tests of between-subject effects per effort dimension are given in Panel B, but to judge their significance, a Hotelling's T^2 statistic ($\sqrt{T^2_{crit}}$ = 2.89) was generated. Comparing the individual t-values with this test statistic reveals that the people under certainty spend significantly more effort than under UF, only on quality (means: 3.75 vs 3.43; t = 2.98) and cost calculation (means: 3.65 vs 3.11; t = 5.19), but not on quantity (means: 3.63 vs 3.41; t = 2.07).

[Insert Table 5 around here]

Additionally, it is interesting to inspect how effort is allocated within each treatment of certainty and UF, because it is expected that subjects in the certainty conditions divide their effort equally over all dimensions, while those in the UF conditions are expected to spend more effort on quantity, as opposed to quality and cost calculation. This was investigated by employing pairwise t-tests per uncertainty condition (not tabulated). Under certainty, it seems that the effort spent on quantity (mean = 3.63), quality (mean = 3.75) and cost calculation (mean = 3.65) is not significantly different (p-values between .31 and .84), indicating that subjects expend equal effort towards all three dimensions. Subjects working under UF conditions seem to expend more effort on quantity (mean = 3.41; t = 2.81; p = .008) and quality (mean = 3.43; t = 2.94; p = .02) than on calculation (mean = 3.11), but there appears to be no difference between the effort spent on quantity and quality (t = -.12; p = .91).

Hypothesis 3. H3 predicted that, for people working under a situation of uncertainty stemming from uncontrollable factors, performance will be higher among those working under a bonus scheme than those working under a penalty scheme. Given H1, which is confirmed by the main effect of framing in Table 4, H3 thus investigates the interaction between framing and uncertainty. From Table 4, it can be gleaned that there, indeed, is a significant interaction effect (F = 4.118; p = .046); however, it is not completely in the predicted direction, as contrasts reveal that under conditions of UF, bonus systems (mean = -.34) do not elicit higher performance than penalty schemes (mean = -.26) (t = -.43; p = .67). Therefore, the reason for the significant interaction effect is not because under UF, bonuses elicit higher performance than penalties (as predicted), but because penalties do not cause higher performance than bonuses (as under certainty). As such, H3 is not supported by the experimental data.

According to theory development, risk considerations can explain for behavior under UF. Independent samples t-tests reveal that, within the UF treatment, risk attitude is not significantly higher for penalties (mean = 2.79) than for bonuses (mean = 2.68) (t = .45; p = .65); while it was expected that people under penalties would be risk seeking and those working under bonuses risk averse. In addition, perceived risk was expected to be higher for a penalty scheme than for a bonus scheme, which was supported by the data (means: 3.38 vs 2.55; t = 3.40; p = .002).

Hypothesis 4 and 5 (see Figure 4)

Comparable to the previous section, the results for hypotheses 4 and 5 also are presented together, because they concern the comparison between certainty and IM conditions. To test these hypotheses, an ANCOVA was performed, for which the results are given in Table 6. The dependent variable was overall performance, with incentive framing (bonus-penalty) and uncertainty (certainty-IM) as independent variables, while risk attitude served as the covariate, meant to remove unexplained variation from the error term.

[Insert Figure 4 around here]

[Insert Table 6 around here]

Hypothesis 4. H4 predicted that task performance under uncertainty stemming from imperfect monitoring will be lower than task performance under certainty. Table 6 shows that the uncertainty main effect is significant ($F = 23.118$; $p < .001$), confirming that performance is higher under certainty (mean = .299) as opposed to under IM (mean = .008), and thus supporting H4.

Deeper analysis shows how both certainty and IM differ with respect to effort allocation. The employed MANCOVA, with all three effort dimensions as dependent variables, framing and uncertainty as independent variables, and risk attitude as the covariate, is displayed in Table 7 Panel A. It can be seen that both certainty and IM differ with respect to effort spent (Wilk's Lambda = .84; $p = .004$). Univariate tests of between-subject effects per effort dimension are presented in Panel B; but, to judge their significance, a Hotelling's T^2 statistic ($\sqrt{T^2_{crit}} = 2.89$) was generated. Although univariate tests show significance for quality and cost calculation, comparing the individual t-values with the Hotelling's test statistic reveals that no single effort dimension causes the multivariate results that were found. It, thus, appears to be the combination of effort on quantity (means: 3.63 vs 3.50; $t = 1.24$), quality (means: 3.75 vs 3.50; $t = 2.24$) and cost calculation (means: 3.65 vs 3.37; $t = 2.70$) that is responsible for the result that effort spent is higher under conditions of certainty versus under IM.

[Insert Table 7 around here]

Additionally, it is interesting to see how effort is allocated within each treatment of respective certainty and IM, because it is expected that subjects in the certainty conditions divide their effort equally over all dimensions, while those in the IM conditions are expected to expend more effort on quantity, as opposed to quality and cost calculation. This was investigated employing pairwise t-tests per uncertainty condition (not tabulated). As reported earlier, under certainty conditions, effort spent on quantity, quality and cost calculation is not significantly different, indicating that subjects spend equal effort over all three dimensions. For subjects working under IM conditions, it also seems that they divide their effort equally over all three effort dimensions (p-values between .281 and 1.00). However, more thorough analysis shows that, for the IM-bonus condition, the differences between effort spent on quantity (mean = 3.51) and cost calculation (mean = 3.11) on the one hand, and between quality (mean = 3.51) and cost calculation on the other hand are statistically significant ($t = 2.11$; $p = .05$) and marginally significant ($t = 1.77$; $p = .09$), respectively; whereas there is no difference between the amount of effort expended towards quantity and quality. In contrast, for the IM-penalty condition, there were no statistically significant differences between the effort spent on all three dimensions (p-values between .384 and 1.00).

Hypothesis 5. H5 predicted that, for people working under a situation of uncertainty stemming from imperfect monitoring, performance will be higher among those working under a penalty scheme than those under a bonus scheme. Looking at Table 6, one can detect that the main effect of framing is significant ($F = 25.416$; $p < .001$), while the interaction effect of framing and uncertainty is insignificant ($F = .607$; $p = .438$). Given H1, this implies that, under IM, performance is higher for penalties (mean = .35) than for bonuses (mean = -.33). These results support H5.

According to theory development, risk considerations can explain behavior under IM. Independent samples t-tests reveal that, within the IM treatment, risk attitude is significantly higher for bonuses (mean = 3.19) than for penalties (mean = 2.45) ($t = 2.73$; $p = .009$), and that perceived risk is significantly higher for penalties (mean = 3.24) than for bonuses (mean = 2.48) ($t = 3.12$; $p = .003$).

DISCUSSION AND CONCLUSIONS

This paper provides experimental data to verify the incentive framing effectiveness under situations of certainty versus uncertainty. While common research treats uncertainty as a comprehensive indivisible concept, this paper discerns two important sources of performance evaluation uncertainty, each having different implications with respect to worker performance and incentive framing effectiveness. The two sources investigated were uncertainty derived from uncontrollable factors and uncertainty stemming from imperfect monitoring.

The results show that, when people are working under certain conditions, a penalty incentive frame generally will result in higher performance than when incentives are framed as a bonus.

When comparing performance under certainty with performance under uncertain performance evaluation due to uncontrollable factors, it is clear from the results that workers will perform better in the certain environment. The reason is twofold, and concerns different effort intensity and effort allocation. First, an uncertain environment makes people spend significantly less effort on the two dimensions that are measured with error (quality and cost calculation) as opposed to people working under certainty; while the effort spent on the number of products (quantity) is not dependent upon whether one is working under conditions of certainty or uncertainty. The second reason is that the certainty treatment urges people to divide their effort equally over all three performance dimensions, while people who are confronted with uncertainty coming from uncontrollable factors, spend much less effort on cost calculation than on quantity and quality. The latter result probably is because the possible measurement error is higher for cost calculation than for quality. Further, the results also show that, under this type of uncertainty, penalty schemes lose their advantage of higher effectiveness versus bonus schemes, as is the case in certain environments. However, although perceived risk was found to be higher for penalties than for bonuses, the loss in effectiveness was not as large as predicted, because penalties did not invoke a higher risk attitude than bonuses. The potential of uncontrollable measurement error made people lose their motivation and perform worse than under certainty, irrespective of incentive framing.

When comparing performance under certainty with performance under uncertain performance evaluation due to imperfect monitoring, results reveal that the certain environment incites the highest level of performance. As in the case of comparison with uncertainty coming from uncontrollable factors, the reason lies in greater effort intensity. However, concerning effort allocation, there is a difference between bonus and penalty schemes. It seems that people working under a bonus scheme, indeed, spend much less effort on cost calculation, while people working under a penalty scheme divide their effort almost equally over the different dimensions. The smaller effort under uncertainty due to imperfect monitoring can for penalty schemes only be attributed to less effort intensity and not worse effort allocation. In addition, it was shown that, with imperfect monitoring, the effectiveness of a penalty scheme is greater than with a bonus scheme, because perceived risk is higher and risk attitude lower for penalties. Thus, with imperfect monitoring, people who are paid according to a bonus scheme expect that bad products will not be detected, which makes them try to game the performance measurement; while those under a penalty scheme expect that bad products will be detected, which makes them to perform well and expend their effort proportionally, as under certainty.

The results described in the present paper have a number of important implications. First, although it never was intended to cover the full range of sources of uncertainty, this paper tries to open up a discussion on uncertainty, and shows that it is important for researchers, as well as managers, to distinguish its different sources, and to acknowledge that several management accounting and control principles have different consequences or should be implemented in different ways, depending upon the kind of uncertainty that is faced. This paper only scratched the surface of this problem, but further research could elaborate more on this interpretation of uncertainty, which could be applied to different problems. Second, companies should recognize that workers perform better when they are paid according to a penalty system than to a bonus system, even when uncertainty arises due to imperfect monitoring. This result counters the predominant claim that incentives are not useful under uncertainty (Coronado and Krishnan, 2005), and uncovers the necessity of thinking critically with respect to potential sources of uncertainty. Because perfect monitoring sometimes can be detrimental to the principal (Jacobides and Croson, 2001), or because monitoring costs have to be cut, it could be useful to employ a system of imperfect monitoring in combination with a penalty incentive scheme; the former to overcome the disadvantages of too much monitoring, and the latter to remain at a high level of performance. The third implication is that, when an organization faces uncertainty due to uncontrollable factors, it should not choose a penalty scheme, because it does not elicit higher performance and does raise more resistance than bonus schemes (Luft, 1994).

Finally, this study is subject to a few limitations. First, as with most experiments, the participants were asked to work in a laboratory setting on a highly simplified task, in order to increase internal validity. However, this leads to lower external validity, which reduces the generalizability of the results. Second, the paper did not take into account the disutility of effort (Hannan et al, 2005) or other psychological or administration costs that people bear in real life (Frederickson and Waller, 2005), because the intention was to investigate the pure effect of the manipulations on effort and performance, albeit creating a mere simplification of reality. Further research could address this problem in a similar setting, and look at whether the net effect of both higher costs and higher performance associated with penalties is positive or negative, and whether it changes with respect to different sources of uncertainty. Third, the experiment has been conducted only during one period, such that long term effects were not observable. It is plausible that, in the long term, incentive framing also would have an impact on competences, employee satisfaction, retention or competitiveness. These are possible areas for further research. Fourth, the experiment in this paper employed a pure bonus and penalty system. However, further research should investigate whether a mixed bonus-penalty system yields similar results or not, because a mixed system likely is more feasible and acceptable to implement than a pure penalty system; and because it is unknown whether framing effectiveness would change compared to pure bonus and penalty schemes. The fifth and last limitation concerns the fact that the paper only encompasses two sources of uncertainty, while many more exist. Further research should try to address other sources of uncertainty, while making recommendations on whether to use penalty frames or bonus ones.

REFERENCES

Ahn, S.C., and Faith, R.L. (1996). Optimal labour contracts and involuntary unemployment under costly and imperfect monitoring. *Economica,* 63, 569-588

Altenburg, L. and Straub, M. (1998). Efficiency wages, trade unions, and employment. *Oxford Economic Papers – New Series*, 50 (4), 726-746

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

Baiman, S. (1990). Agency research in managerial accounting: A second look. *Accounting, Organizations and Society*, 15 (4), 341-371

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

Banker, R.D. and Datar, S.M. (1989). Sensitivity, precision, and linear aggregation of signals for performance evaluation. *Journal of Accounting Research*, 27 (1), 21-39

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.

Chaudhary, N.K., Solomon, M.G., and Cosgrove, L.A. (2004). The relationship between perceived risk of being ticketed and self-reported seat belt use. *Journal of Safety Research*, 35, 383– 390

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

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

Das, T.K. and Teng, B. (2004). The risk-based view of trust: a conceptual framework. *Journal of Business and Psychology*, 19 (1), 85-116

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

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

Gerber, B.J. and Neeley G.W. (2005). Perceived Risk and Citizen Preferences for Governmental Management of Routine Hazards. *The Policy Studies Journal*, 33 (3), 395-418

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

Hanoch, Y., Johnson, J.G., and Wilke A. (2006). Domain specificity in experimental measures and participant recruitment. Psychological Science, 17 (4), 300-304

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

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

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

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

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

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

Nantz, K. and Sparks, R. (1990). The labor-managed firm under imperfect monitoring: Employment and work effort responses. *Journal of Comparative Economics*, 14 (1), 33-50

Naylor, J.C., Pritchard, R.D. and Ilgen, D.R. (1980). *A theory of behavior in organizations*. New York, NY: Academic Press

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

Prendergast, C. (2002). The tenuous tradeoff between risk and incentives. *Journal of Political Economy*, 110 (5), 1071-1102

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

Selto, F.H. and Cooper, J.C. (1990). Control of risk attitude in experimental accounting research. *Journal of Accounting Literature*, 9, 229-264

Shapiro, C. and Stiglitz, J.E. (1984). Equilibrium unemployment as a worker discipline device. *The American Economic Review*, 74 (3), 433-444.

Siegrist, M., Gutscher, H., and Earle, T.C. (2005). Perception of risk: the influence of general trust, and general confidence. *Journal of Risk Research*, 8 (2), 145-156

Slovic, P. (1987). Perception of risk. *Science*, 236, 280-285

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

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

Waller, W.S. (1995). Decision-making research in managerial accounting: Return to behavioral-economics foundations. In Ashton, R.H. and Ashton, A.H. (Eds.), *Judgment and decision-making research in accounting and auditing*. Cambridge University Press, pp 29-54

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

Summary table of the uncertainty operationalizations

	Certainty	Uncontrollable factors	Imperfect monitoring
Quantity		- No random error term is added - All products are monitored	
Quality	- No random error term is added - All products are monitored	- A random error term is added - All products are monitored	- No random error term is added - 30% of the products is monitored
Cost calculation	- No random error term is added - All products are monitored	- A random error term is added - All products are monitored	- No random error term is added - 30% of the products is monitored

Appendix C

Risk attitude – Certainty

1. Because performance was perfectly measured, you didn't dare to take the risk to shirk (REVERSED)
2. Because performance was perfectly measured, you played it safe by trying to reach your goals by being diligent (REVERSED)

Perceived risk – Certainty

1. You expect random measurement error in your performance measurement (scale).
2. You expect that, if you make bad products, they will be detected (scale).
3. What is the chance that random measurement error occurs in your performance measurement (probability)?
4. What is the chance that, if you make bad products, they will be detected (probability)?

Risk attitude – UF

1. Because the possibility of measurement error exists, you consciously took the risk of trying to reach your goals through shirking.
2. Because the possibility of measurement error exists, you played it safe by trying to reach your goals by being diligent (REVERSED).

Perceived risk – UF

1. You expect the measurement error to be in your favor (REVERSED) (scale).
2. You expect the measurement error to be to your detriment (scale).
3. What is the chance that the measurement error will be in your favor (REVERSED) (probability)?
4. What is the chance that the measurement error will be to your detriment (probability)?

Risk attitude – IM

1. Because of imperfect monitoring, you took the risk of trying to reach your goals through shirking.
2. Unless imperfect monitoring, you played it safe by trying to reach your goals by being diligent (REVERSED).

Perceived risk – IM

1. You expect that, if you make bad products, they will be detected, unless imperfect monitoring exists (scale)
2. What is the chance that, if you make bad products, they will be detected through imperfect monitoring (probability)?

Table 1 – Descriptive statistics

Variable	N	Min	Max	Mean	Std Dev
Quantity performance	123	7	13	9.61	1.32
Quality performance	123	.63	1	.94	.08
Calculation performance	123	.33	1	.73	.14
Quantity effort	123	2	5	3.51	.60
Quality effort	123	2	5	3.56	.57
Calculation effort	123	2	4.67	3.37	.63
Overall effort	123	2.56	4.67	3.48	.39
Risk attitude	123	1	4.50	2.82	.84
Perceived risk of uncontrollability*	81	1	5	2.54	1.00
Perceived risk of imperfect monitoring**	82	1	5	3.33	1.01
Age	123	18	24	20.11	1.50

* This is the perceived risk about the expectation of measurement errors as measured in the certainty and UF conditions

** This is the perceived risk about the expectation of detection of bad products as measured in the certainty and IM conditions

Table 2 – Correlations between Effort and Performance

		Effort				Performance			
		Overall	Quantity	Quality	Calculation	Overall	Quantity	Quality	Calculation
Effort	Overall	1	.64**	.57**	.71**	.75**	.48**	.33**	.59**
	Quantity		1	.03	.21*	.47**	.74**	.01	.14
	Quality			1	.13	.34**	-.09	.60**	.13
	Calculation				1	.61**	.25**	.06	.84**
Performance	Overall					1	.62**	.54**	71**
	Quantity						1	-.08	.24**
	Quality							1	.09
	Calculation								1

* correlation is significant at the .05 level

** correlation is significant at the .01 level

Table 3 – Relevant means for Performance, Effort, Risk attitude and Perceived risk per treatment and per condition (standard deviations in parentheses)

		Bonus		Penalty		Total	
Certainty		N = 20		N = 20		N = 40	
	Performance	.045	(.45)	.553	(.41)	.299	(.49)
	Effort						
	Overall	3.52	(.26)	3.82	(.38)	3.68	(.36)
	Quantity	3.55	(.68)	3.70	(.60)	3.63	(.64)
	Quality	3.58	(.49)	3.92	(.48)	3.75	(.51)
	Cost calculation	3.45	(.47)	3.85	(.49)	3.65	(.52)
UF		N = 20		N = 21		N = 41	
	Performance	-.337	(.59)	-.264	(.49)	-.300	(.53)
	Effort						
	Overall	3.27	(.39)	3.36	(.28)	3.32	(.34)
	Quantity	3.37	(.49)	3.46	(.63)	3.41	(.56)
	Quality	3.40	(.70)	3.46	(.43)	3.43	(.57)
	Cost calculation	3.05	(.60)	3.16	(.59)	3.11	(.59)
	Risk attitude	2.68	(.73)	2.79	(.83)	2.73	(.78)
	Perceived risk	2.55	(.83)	3.38	(.74)	2.98	(.88)
IM		N = 21		N = 21		N = 42	
	Performance	-.331	(.68)	.346	(.51)	.008	(.68)
	Effort						
	Overall	3.38	(.38)	3.53	(.38)	3.46	(.39)
	Quantity	3.51	(.59)	3.49	(.64)	3.50	(.61)
	Quality	3.51	(.65)	3.49	(.53)	3.50	(.59)
	Cost calculation	3.11	(.70)	3.61	(.50)	3.37	(.65)
	Risk attitude	3.19	(.75)	2.45	(.99)	2.82	(.94)
	Perceived risk	2.48	(.81)	3.24	(.77)	2.86	(.87)
Total		N = 61		N = 62		N = 123	
	Performance	-.210	(.60)	.206	(.58)	.00	(.62)
	Effort						
	Overall	3.39	(.36)	3.57	(.40)	3.48	(.39)
	Quantity	3.48	(.59)	3.55	(.62)	3.51	(.60)
	Quality	3.50	(.61)	3.62	(.52)	3.56	(.57)
	Cost calculation	3.20	(.62)	3.54	(.60)	3.37	(.63)

Table 4 – ANCOVA for Performance for subjects Certainty and UF

Source of Variation	Type III Sum of Squares	df	Mean Square	F	p-value
Intercept	8.653	1	8.653	73.678	.000
Risk attitude	9.315	1	9.315	79.318	.000
Framing	1.542	1	1.542	13.134	.001
Uncertainty	9.215	1	9.215	78.462	.000
Framing * Uncertainty	.484	1	.484	4.118	.046

Table 5 – MANCOVA for Effort for subjects Certainty and UF

Panel A – Multivariate tests

Effect	Wilk's Lambda	F	Hypothesis df	Error df	p-value
Intercept	.06	403.99	3	74	.000
Risk attitude	.70	10.55	3	74	.000
Framing	.89	3.06	3	74	.033
Uncertainty	.63	14.60	3	74	.000
Framing * Uncertainty	.97	.65	3	74	.587

Panel B – Univariate tests of between-subjects effects

Source of Variation	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	p-value
Intercept	Quantity	105.337	1	105.337	329.228	.000
	Quality	92.261	1	92.261	337.665	.000
	Calculation	94.134	1	94.134	364.467	.000
Risk attitude	Quantity	3806	1	3.806	11.896	.001
	Quality	1.196	1	1.196	4.376	.040
	Calculation	2.958	1	2.958	11.454	.001
Framing	Quantity	.258	1	.258	.805	.372
	Quality	.744	1	.744	2.723	.103
	Calculation	1.229	1	1.229	4.757	.032
Uncertainty	Quantity	1.376	1	1.376	4.300	.042
	Quality	2.424	1	2.424	8.871	.004
	Calculation	6.964	1	6.964	26.962	.000
Framing * Uncertainty	Quantity	.003	1	.003	.009	.926
	Quality	.261	1	.261	.956	.331
	Calculation	245	1	.245	.947	.334

Table 6 – ANCOVA for Performance for subjects Certainty and IM

Source of Variation	Type III Sum of Squares	df	Mean Square	F	p-value
Intercept	15.464	1	15.464	154.392	.000
Risk attitude	13.637	1	13.637	136.154	.000
Framing	2.546	1	2.546	25.416	.000
Uncertainty	2.316	1	2.316	23.118	.000
Framing * Uncertainty	.061	1	.061	.607	.438

Table 7 – MANCOVA for Effort for subjects Certainty and IM

Panel A – Multivariate tests

Effect	Wilk's Lambda	F	Hypothesis df	Error df	p-value
Intercept	.056	419.840	3	75	.000
Risk attitude	.663	12.696	3	75	.000
Framing	.892	3.031	3	75	.035
Uncertainty	.837	4.872	3	75	.004
Framing * Uncertainty	.932	1.821	3	75	.151

Panel B – Univariate tests of between-subjects effects

Source of Variation	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	p-value
Intercept	Quantity	126.439	1	126.439	393.613	.000
	Quality	102.368	1	102.368	360.684	.000
	Calculation	110.144	1	110.144	417.064	.000
Risk attitude	Quantity	6.024	1	6.024	18.753	.000
	Quality	1.255	1	1.255	4.420	.039
	Calculation	3.302	1	3.302	12.504	.001
Framing	Quantity	.133	1	.133	.414	.522
	Quality	.154	1	.154	.543	.464
	Calculation	2.230	1	2.230	8.445	.005
Uncertainty	Quantity	.491	1	.491	1.529	.220
	Quality	1.420	1	1.420	5.005	.028
	Calculation	1.927	1	1.927	7.297	.008
Framing * Uncertainty	Quantity	.615	1	.615	1.916	.170
	Quality	.937	1	.937	3.302	.073
	Calculation	.004	1	.004	.017	.898

Figure 1: Uncertainty framework

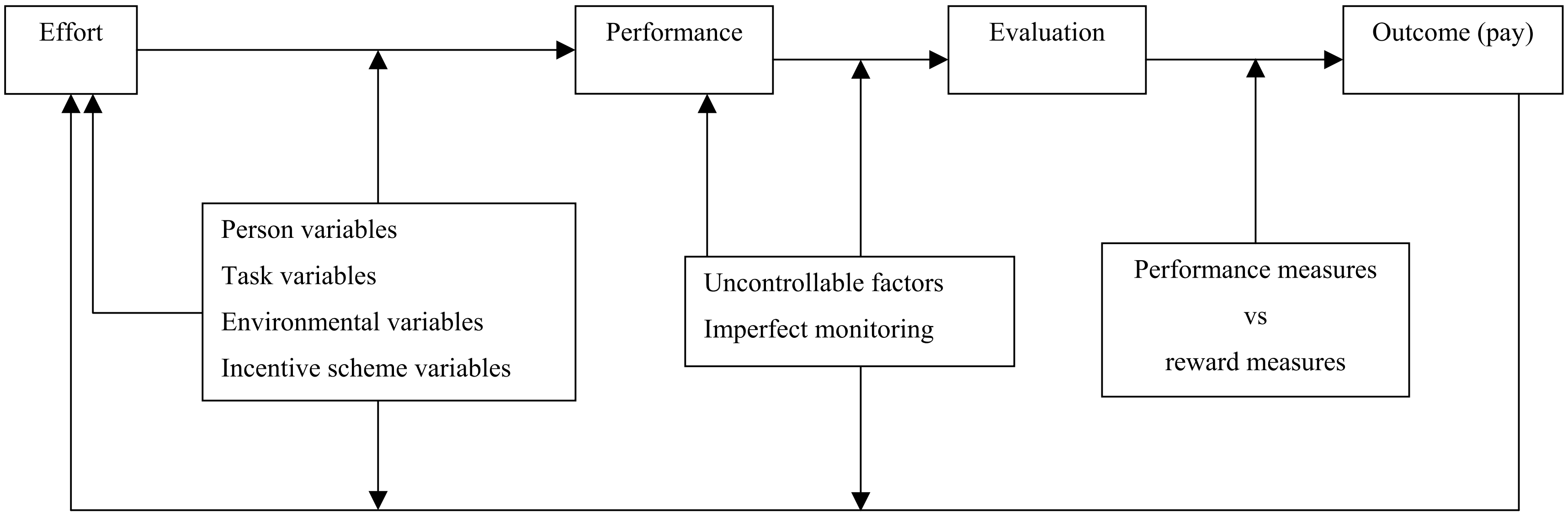

Figure 2: Uncertainty framework relevant for this paper after controlling for the uninvestigated sources of uncertainty

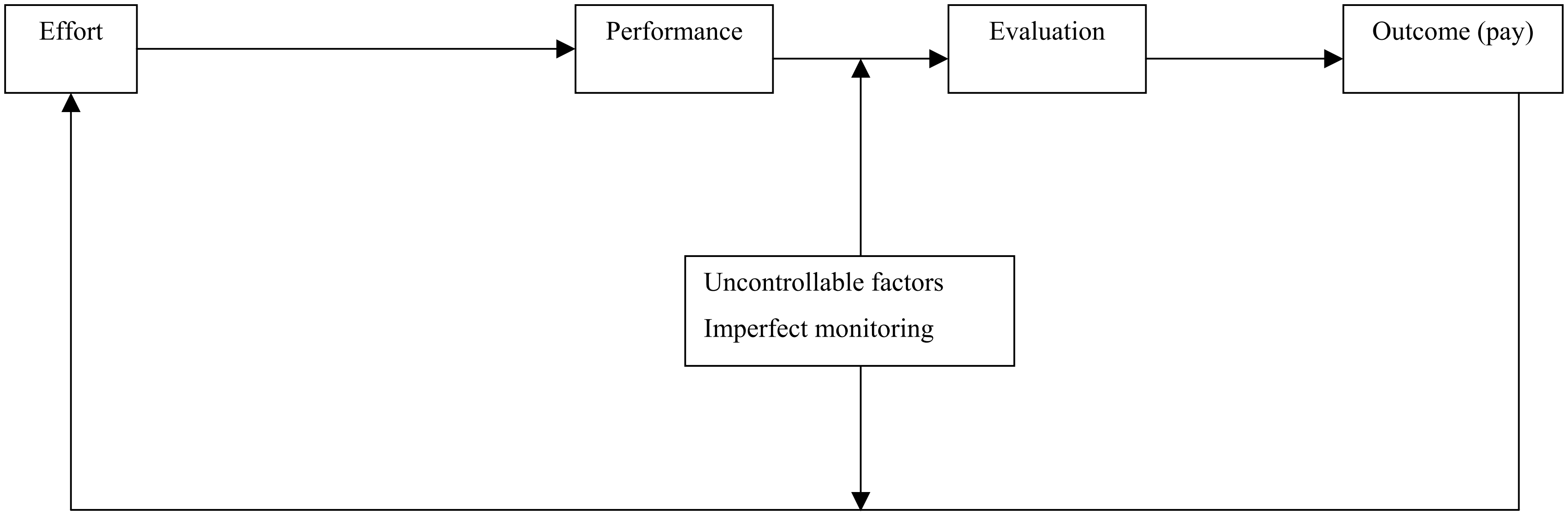

Figure 3: Performance means for Incentive Framing and Uncertainty (Certainty versus UF)

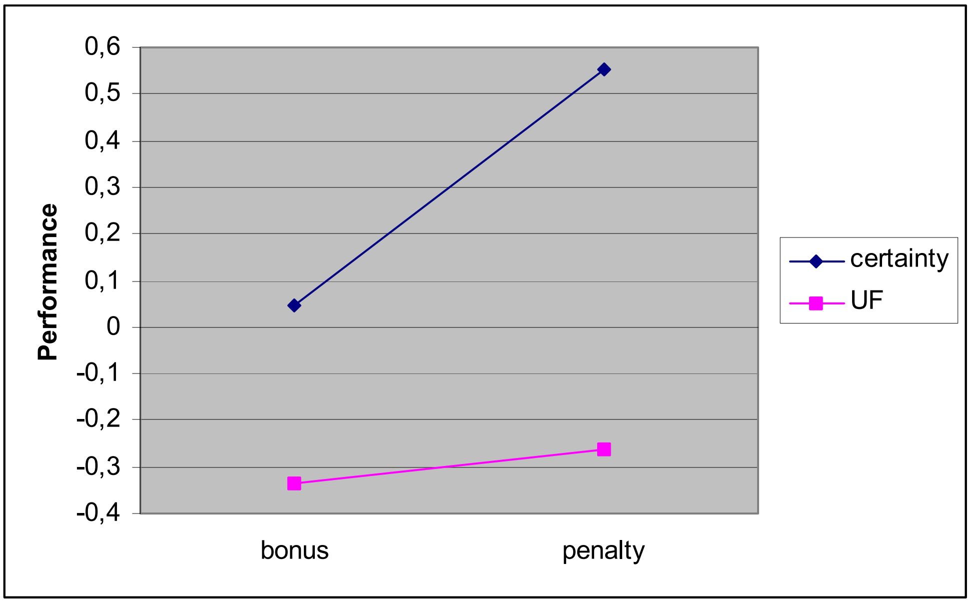

Figure 4: Performance means for Incentive Framing and Uncertainty (Certainty versus IM)