June 14, 2006

**Sorting and Incentive Effects of Pay-for-Performance: An Experimental Investigation**

C. Bram Cadsby*  
Department of Economics  
University of Guelph  
Guelph, Ontario, Canada  
Tel: (519) 824-4120  
Fax: (519) 763-8497  
e-mail: bcadsby@uoguelph.ca

Fei Song**  
School of Business Management  
Ryerson University  
Toronto, Ontario, Canada  
Tel: (416) 979-5000  
Fax: (416) 979-5266  
e-mail: fsong@ryerson.ca

Francis Tapon  
Department of Economics  
University of Guelph  
Guelph, Ontario, Canada  
Tel: (519) 824-4120  
Fax: (519) 763-8497  
e-mail: ftapon@uoguelph.ca

*The two first authors made equal contributions to the paper.

**Corresponding Author

Acknowledgements: We thank the Economics Department at the Queensland University of Technology for their warm hospitality and support for this research and Rekha Karambayya, Christine Oliver, and Mary-Anne Sillamaa as well as participants at the 2004 Canadian Economics Association meeting, the 2004 Economic Science Association meeting and the 2004 Canadian Experimental and Behavioral Economics conference for their helpful comments. We are especially grateful to the editor of the *Academy of Management Journal*, Sara L. Rynes, and three anonymous reviewers, each of whom offered many interesting and stimulating suggestions, which substantially improved the paper. We also thank David Prescott, Marion Steele, Yiguo Sun, and Henry Thille for statistical advice. Cadsby would also like to acknowledge generous funding from the Social Sciences and Humanities Research Council of Canada, grant #410-2001-1590.
Abstract

Agency theory highlights potential losses in productivity that can occur when the interests of owners and employees are imperfectly aligned. Pay-for-performance (PFP) has been proposed as a potential solution to this problem. Using a real-effort laboratory experiment with salient incentives, we examine the impact of PFP versus fixed-salary (FS) compensation. PFP achieved significantly higher firm productivity through both sorting and incentive effects. In particular, more productive employees selected PFP, and employees on average, regardless of their preferred compensation scheme, produced more under PFP. However, more risk-averse individuals were both less likely to select PFP and less responsive to PFP incentives.
Compensation costs comprise 65% to 70% of total production costs in the U.S. economy and are similarly substantial elsewhere (Gerhart & Rynes, 2003). Accordingly, the relationship between pay and both employee and firm performance has been a focus of attention in management, economics, organizational studies, and sociology (e.g., Gomez-Mejia & Welbourne, 1988; Jensen, 2003; Lawler, 1971, 1990; Lazear, 1986, 1999, 2000a; Rynes & Lawler, 1983; Rynes, Schwab, & Heneman, 1983; Zenger, 1992). The last two decades have witnessed an increase in the prevalence of different systems of pay-for-performance in many organizations (Heneman, Ledford, & Gresham, 2000; Milkovich & Newman, 2002).

Firms seeking to maximize productivity on behalf of their shareholders will seek both to hire the most highly skilled employees and to motivate those employees to maximize their output. An important early literature focused on the latter objective, noting the agency problem that arises when the objectives of employees are not well aligned with those of firms (e.g., Coase, 1937; Jensen & Meckling, 1976; Lawler, 1971; Williamson, 1975). Such problems occur, for example, when employees are more concerned with their remuneration and/or maintaining a comfortable effort level than with exerting their maximum effort to produce as much as possible for their organization. Pay-for-performance (PFP hereafter) has been justified as an effective contracting mechanism to resolve agency problems because it aligns the preferences of firms and employees (e.g., Jensen & Murphy, 1990). More recently, some organizational theorists and labor economists have focused on the former objective, arguing that PFP can also act as a sorting device to identify and attract the most capable employees (Gerhart & Milkovich, 1992; Gerhart & Rynes, 2003; Jensen, 2003; Lazear, 1986, 1999, 2000a; Zenger, 1992). For example, Jensen (2003) argued that more productive workers would choose PFP, while less productive workers would opt for a fixed-salary (FS hereafter). Sorting based on compensation scheme is consistent with the idea of person-organization fit, which asserts that job applicants choose organizations based on their perceptions of the match between their own dispositions and organizational characteristics (Bretz & Judge, 1994; Rynes, 1987; Schneider, 1987).

The incentive property of PFP has been extensively examined theoretically (Baker,
Jensen, & Murphy, 1988; Lazear, 2000a) and empirically in both the laboratory (e.g., Fessler, 2003; Kuhn & Yockey, 2003; see Camerer & Hogarth, 1999 for a review of the literature) and the field (e.g., Abowd, 1990; Gerhart & Milkovich, 1990; Stajkovic & Luthans, 1997). In contrast, empirical support for sorting is less abundant. Data from human resource records suggest that high performers are most likely to leave an organization if performance does not lead to sufficient financial rewards (Trevor, Gerhart, & Boudreau, 1997), while low performers are more likely to stay put if relationships between pay and performance are weak (Harrison, Virick, & William, 1996). An exceptionally important study by Lazear (1999, 2000a, 2000b) used field data to examine both sorting and incentive effects and found that PFP resulted in a 44% increase in productivity, divided roughly equally between the two effects. The Safelite Glass Corporation, whose workforce was the object of the study, moved from a system of hourly wages to that of piece rates. However, the piece rates were introduced with a minimum wage guarantee, which was equal to the hourly wage under the previous regime. Thus, the piece-rate system with guaranteed minimum wage introduced by Safelite implied that employees had no “pay at risk” relative to the previous hourly-wage system, where we define pay at risk as the guaranteed pay foregone to have a chance at higher earnings through PFP. Pay at risk is often considered an important characteristic of PFP systems (e.g., Jensen and Murphy, 1990). Moreover, Safelite’s guaranteed wage meant that each employee would have earned either the same or more per hour under the new scheme even if he/she had produced exactly the same amount as under the old scheme. According to Lazear (2000b), this is why the union, which had initially opposed the introduction of piece rates at Safelite, agreed to the new scheme. Thus, many employees may well have perceived the combination of a generous pay guarantee along with the opportunity of earning more through PFP as an improvement in pay. Such a perception may have had an impact on productivity as predicted by reciprocity-based efficiency-wage theories (e.g. Akerlof, 1982) in addition to the predicted impact of PFP alone.

1 We thank an anonymous reviewer for suggesting this definition of “pay at risk” and the accompanying argument.
Rynes, Gerhart, and Parks (2005) have stressed the need for more work investigating the distinction between the incentive and sorting effects of compensation systems, arguing that the latter appears to be very important based on the limited work available (e.g., Lazear, 1999, 2000a; Trevor et al., 1997). Furthermore, they pointed out that, despite the fact that risk is a central factor in many current forms of pay (e.g., incentives, gain sharing, profit sharing, and stock options), little attention has been given to its measurement. Hence, they advocated more attention to individual differences in how people perceive risk. In our paper, we focus on the role of performance incentives, operationalized as the contrast between simple PFP versus FS compensation schemes, to examine the interrelations between levels of risk aversion, employee quality, job-choice decisions, and subsequent performance levels.

The primary goal of this paper is to provide insights into both the sorting and incentive effects of PFP versus FS by investigating the presence and importance of such effects in a controlled laboratory environment. An important advantage of the laboratory methodology is that it enables us to examine the hypothesized sorting and incentive effects of PFP in the absence of a minimum pay guarantee that removes all pay at risk relative to the corresponding FS system. Thus, in contrast to Lazear (1999, 2000a, 2000b), none of our participants could be certain that he/she would earn at least as much under our PFP scheme as under our FS scheme. In fact, the average pay per participant would have been lower under our assigned PFP than under our assigned FS treatment ($1.87 versus $2.20 respectively) if productivity had remained unchanged. Thus, the increases in productivity due to both sorting and incentives observed in our study can be unambiguously attributed to the relationship between pay and performance per se rather than to PFP coupled with a generous minimum pay guarantee that ensures equal or higher earnings as at Safelite. This is an important advance on Lazear’s pioneering study. It also has important practical implications, suggesting that PFP need not be accompanied by a minimum pay guarantee that removes all pay at risk relative to available FS opportunities for a firm to experience the hypothesized sorting and incentive effects. Moreover, Lazear (2000a: 1359) pointed out that the method by which Safelite’s new payment scheme was implemented and the
fact that its introduction followed shortly after the firm was placed under new management could have been at least partially responsible for the observed productivity gain. Laboratory-generated data allow us to avoid such potential confounding effects.

The laboratory setup also enables us to investigate how risk preferences might influence both the sorting and incentive effects of PFP. Cable and Judge (1994), using self-report questionnaire-based data, demonstrated that risk aversion was a significant predictor of a preference for fixed versus contingent pay. By eliciting behavioral risk-preference measures in the laboratory, we were able to corroborate this result, which represents significant additional empirical support for the risk-aversion hypothesis. In addition, we show that the level of risk aversion interacts with individual productivity in determining which payment scheme will be selected, affecting decisions of high-productivity individuals significantly more than it affects those of low-productivity individuals once they have gained some experience at the anagram task used in our study. We also demonstrate that the effectiveness of PFP at improving productivity is inversely related to individual levels of risk aversion. For about a quarter of our participants, PFP actually lowers productivity.

**THEORETICAL SYNTHESIS**

**Determinants of Self-Selection into PFP versus FS Schemes**

An individual selects a compensation contract based on its perceived desirability compared to other available compensation options. Psychologists have noted that PFP is more attractive to those higher in cognitive ability (Trank, Rynes, & Bretz, 2002) and need for achievement (Bretz, Ash, & Dreher, 1989). Bretz et al. (1989) have suggested that the job attraction literature should focus on how such individual differences are associated with important organizational outcomes. In a world of asymmetric information, where a job applicant knows more about his/her own ability and need for achievement than a potential employer, implementing a compensation system that attracts high-quality applicants is just such an important issue. A high-quality applicant is one who has the ability and disposition to produce more than a low-quality applicant. Camerer and Hogarth (1999: 10) refer to the combination of
“heuristics, personal skills and traits, domain-specific procedures and so forth” that determine employee quality as “cognitive capital”. An organization is interested in attracting high-quality applicants, i.e. people with large amounts of “cognitive capital”.

If PFP is to be an efficient sorting device, it is necessary that employees be able to forecast their own future productivity in a reasonably accurate manner. Many economic models assume that people have “rational” expectations, which means that people make the best possible forecast based on all the information at their disposal (Muth, 1961). Employees who choose a compensation scheme may formulate their expectations rationally, and yet sometimes make mistakes due to lack of complete information. However, to the extent that employees have some information about their job-related cognitive capital, more productive employees should exhibit a higher probability of sorting into the PFP scheme. Since information increases with relevant experience, more experienced employees with more information about their own quality and the nuances of the compensation schemes should choose more accurately than their less experienced counterparts.

In addition to individual productivity differences, there are other factors that could potentially influence self-selection into a particular compensation scheme. Cable and Judge (1994) hypothesized that a high level of risk aversion, a low level of self-efficacy, and an external locus of control would all be associated with a preference for a fixed-pay versus a contingent-pay system. However, they found that only the level of risk aversion was in fact a significant predictor of a preference for fixed pay. Similarly, Gomez-Mejia and Balkin (1989) found that employees under a variable compensation scheme were more likely to contemplate leaving their jobs if they were more averse to risk. Since a FS scheme offers a certain financial outcome, while PFP involves financial uncertainty, the latter does expose an agent to risk. Following Milgrom and Roberts (1992, p. 187), we define risk-averse employees as those who “would rather have a smaller income whose magnitude is certain than an uncertain income that is somewhat larger on average but is subject to unpredictable and uncontrollable variability.” Agency theory predicts that such employees will require a risk premium to select PFP over FS.
Empirical studies have largely supported this proposition (e.g. Aggarwal & Samwick, 1999; Bloom & Milkovich, 1998). Following this logic, more risk-averse individuals should require a higher risk-premium to select PFP. A higher level of expected productivity implies higher expected PFP earnings and hence a higher expected risk premium. The rational expectations theory implies a positive correlation between expected and actual productivity. Thus, for a given level of productivity, and hence a given expected risk premium, a more risk-averse individual is less likely to choose PFP.

H1: Controlling for individual attitudes toward risk, more productive individuals are more likely to select the PFP compensation scheme.

H2: Controlling for individual productivity differences, more risk-averse individuals are less likely to select the PFP compensation scheme.

Individuals who expect to perform poorly are almost certain to choose FS, regardless of their risk-aversion level. In contrast, those who expect to perform well will be more subject to risk-aversion effects, as they have a rational basis for choosing PFP that could be undercut by fear of failure under such a plan. To the extent that performance expectations are rational, low-productivity people will be in the former, while high-productivity people will be in the latter category. For this reason, risk attitudes are likely to play little discernable role at low levels of productivity, where risk aversion merely reinforces the attractiveness of the FS scheme, but a substantially more important role at high levels of productivity, where risk aversion mitigates the attractiveness of PFP. Hence, we expect to find an interaction effect between the degree of risk aversion and the level of productivity as follows:

H3: The level of productivity will moderate the relationship between risk aversion and self-selection, such that the inverse relationship between individual levels of risk aversion and the probability of choosing PFP will be stronger for more productive participants.

2 It is theoretically possible that a low-productivity individual would select PFP if he/she were highly risk-loving. However, studies of individual attitudes toward financial risk have consistently found that most people are risk-averse to a greater or lesser degree. For example, Holt and Laury (2002) found that only 8% of their subjects made risk-loving choices at smaller stakes than in our experiment, while just 6% made risk-loving choices at stakes higher than in our experiment.

3 We thank an anonymous reviewer for suggesting this line of reasoning.
The Incentive Effect of PFP and its Differential Effect According to Risk-Preference

Expectancy theory (Vroom, 1964) identifies three factors that play an interactive role in motivation: effort-performance expectancy, instrumentality (the link between performance and outcome), and valence. In contrast to a pure FS system, where there is no link between performance and outcome, PFP provides a very direct and explicit link. Thus, the theory implies that PFP can induce higher performance if two conditions hold. First, employees must place value on the monetary reward, i.e. the monetary reward must possess sufficient valence to render its achievement worth the effort. Second, employees must perceive that greater effort will lead to better performance, and thus to the valued higher reward. Under these circumstances, the possibility of higher pay contingent on performance should motivate employees to work hard, thereby discouraging shirking.

As discussed in the introduction, the incentive property of PFP has been extensively examined both theoretically and empirically. Most of this literature, which includes several important meta-analyses of pay and productivity (e.g., Jenkins, Mitra, Gupta, & Shaw, 1998; Locke, Feren, McCaleb, Shaw, & Denny, 1980; Stajkovic & Luthans, 1997), suggests that individual pay incentives significantly improve performance. Thus, we expect to observe this incentive effect in our study. We are able to examine the incentive effect of PFP on productivity in isolation from the hypothesized self-selection effects by assigning all participants to each of the two compensation treatments and observing the differences in productivity for each individual under the FS versus the PFP scheme. This within-person design controls for differing productivity levels across individuals (Keppel, 1991: 323-324).

\( H4: \) On average, people are more productive under PFP than under FS.

Cable and Judge (1994: 341) reported an interesting unhypothesized finding, namely that “risk-averse individuals placed less emphasis on pay level as a criterion in their job pursuit process”. This suggests that the valence associated with reward may be lower on average for more risk-averse individuals. Pappas and Flaherty (2006) recently corroborated this result by means of a questionnaire administered to 214 business-to-business salespeople. In addition to
showing that more risk-averse individuals exhibited significantly lower levels of valence, they also found that such individuals had lower expectancy levels (Pappas & Flaherty, 2006, Table III, p. 30). Since this weakens two of the three important links in expectancy theory, one might expect the incentive effects of PFP to be less pronounced for more risk-averse individuals. Pappas and Flaherty (2006) also showed that individual risk attitudes moderate the relationship between pay mix and valence. In particular, they found an increase in valence associated with more variable pay, which was greater for less than for more risk-averse individuals. This result also suggests weaker PFP incentive effects for those who are more risk-averse.

In addition, a more risk-averse person assigned to a PFP system with its uncertain payoffs might well experience considerable discomfort and stress when compared to a less risk-averse person. If such stress impedes performance, the hypothesized incentive effect of PFP may be weakened, eliminated or even reversed. A considerable literature exists concerning the relationship between stress and job performance (see Muse, Harris, & Field, 2003, and Lepine, Podskoff, & Lepine, 2005, for critical reviews of this literature). Much of the empirical literature suggests that stress, particularly “hindrance stress”, is negatively related to performance. Lepine et al. (p. 765–766) link hindrance stress with low expectancy. As argued above, there is evidence that higher levels of risk aversion may be correlated with lower levels of expectancy, which may adversely affect motivation in a variable-pay environment. If this is the case, the higher level of hindrance stress and the associated lower level of expectancy experienced by more risk-averse participants assigned to PFP might, together with the lower level of reward valence, reduce the strength of the hypothesized incentive effect of PFP.

H5: The effectiveness of PFP at improving productivity is inversely related to individual levels of risk aversion.

METHODS

Experimental Design and Procedure

4 However, Muse et al. (2003) argue that the inverted-U theory, which suggests that small amounts of stress aid performance while larger amounts impede it, has not yet been fairly tested.
Participants were recruited at a large, urban Australian university by means of both announcements in economics classes and random recruitment in the lounge area of the business school. During recruiting, participants were told that they were participating in a study about workplace issues, and that they would be paid. They were not given any other details prior to the experiment. All 115 participants (71 males and 44 females with an average age of 20.9 years and a standard deviation of 4.51 years) were undergraduates, and most but not all were majors in economics or other business subjects. A widely used anagram word-creation game (e.g., Schweitzer, Ordóñez, & Dumaz, 2004; Vance & Colella, 1990) was employed as the experimental task. Specifically, participants were asked to play one practice and eight experimental three-minute anagram games using prescribed sets of seven letters. This task is particularly appropriate to investigate sorting because it is a task where performance depends on both ability and effort. Vance and Colella (1990) used psychology undergraduates from Ohio State University in an experiment employing these anagrams. In their study, participants were given performance targets, but not paid on the basis of performance. We conducted a pre-test at a large Canadian business school employing 99 business undergraduates, who were given salient performance-based financial incentives. The average productivity in our pre-test was higher than in the Vance and Colella (1990) study in eight of the nine anagram games. The overall mean excluding the practice round was 9.65 words per round in Vance and Colella (1990) and 12.57 in our pre-test. Since the current study involved both PFP and FS compensation schemes, we chose the average of these two numbers, approximately 11 words, as our break-even point in order to make the expected earnings of both compensation schemes equivalent for an average performer. Our aim was to induce approximately equal numbers of participants to choose each scheme. The first compensation scheme paid $0.20 per correct word created. The second scheme paid a fixed salary of $2.20, independent of performance. We chose $2.20 so that over the eight rounds participants could earn an amount comparable or higher to what they could earn elsewhere. The experimental economics literature argues that this is important to ensure that participants are facing salient economic incentives (e.g., Friedman and Cassar, 2004). On average, participants
earned $21.20 AUD (about $15.11 US) for a session lasting about one hour and 15 minutes. This was substantially higher than the average wage a student could earn working on or off campus.

Upon arrival, the experimental instructions were read to the participants while they followed along on their own copies. Participants were provided with a prepared workbook containing the anagrams. Each anagram was presented on a separate page of the workbook. Other pages were used for participants to record their choices of compensation scheme or devoted to explaining which compensation scheme would apply in a subsequent anagram round. Participants were not permitted to look ahead to future pages or to go back to previous pages. They were allowed to tear off one page and look at the next only when instructed to do so by the experimenter. To ensure anonymity, participants wrote their assigned participant numbers, but not their names, on each page of the workbook immediately prior to beginning work on that page. After the instructions of the experiment were read, but before they played the practice round, participants chose which one of the two compensation schemes they would like to adopt for calculating their earnings for rounds 1 and 2. The middle four rounds were non-self-selection rounds in which all participants were assigned to identical compensation schemes. Specifically, all participants were paid according to the FS scheme for rounds 3 and 5 and the PFP scheme for rounds 4 and 6. In each case, they were informed of the payment scheme immediately prior to the round. Switching back and forth between the two payment schemes allowed us to separate experience or learning effects from the effects of changing the compensation scheme. For rounds 7 and 8, participants were again given a choice between the two compensation schemes. After each round, each participant’s list of words was collected and taken to another room where the number of correct words was calculated. Participants did not receive feedback on the number of correct words they had produced until they were paid at the end of the session.

Insert Table 1 about here.

After participants completed the eight rounds of anagrams, they filled out a questionnaire. Besides collecting demographic data, the questionnaire was used to elicit risk preferences. This
was accomplished by asking participants to make ten lottery-choice decisions based on an instrument developed by Holt and Laury (2002). As presented in Table 1, each of the ten lottery decisions presented to the participants involved a relatively safe choice (option A) versus a relatively risky choice (option B). The probabilities of each lottery outcome were manipulated so that each decision involved progressively higher expected earnings for the risky choice relative to the safe choice. Accordingly, everyone should have a switching point, above which safer choices are selected and below which riskier choices are selected. In addition to being paid for the anagram task as outlined above, participants were paid an additional sum based on the outcome of their chosen lottery from the pair of randomly-selected lotteries.

We elicited risk preferences after the completion of the self-selection game because to do otherwise might prime participants to pay undue attention to risk, thus biasing their behavioral decisions. To mitigate any impact that playing the self-selection game might have on risk elicitation, we did not give any feedback on how many correct words were created or how much had been earned until the very end of the experiment after the risk data were collected. The purpose of eliciting risk preferences was to examine the role of such preferences in the selection of payment schemes. Holt and Laury (2002) found that risk preferences were affected by the amount of money at stake. In particular, larger stakes were associated with a higher level of risk aversion. We therefore adjusted the stakes used by Holt and Laury (2002) to correspond as closely as possible to the amount at stake in the two rounds of the anagram game affected by each self-selection decision. This involved multiplying Holt and Laury’s (2002) lottery numbers by 2.2 to obtain the appropriate amounts in Australian dollars. At the end of the session, players were taken individually to another room, where they were paid privately in cash.

**Measures**

For H1, H2, and H3, the dependent variable is the choice of compensation scheme. Each participant makes an initial choice for rounds 1 and 2 and a final choice for rounds 7 and 8, each of which is analyzed separately. There are two independent variables associated with these hypotheses. The first is the number of correct words per anagram averaged over the assigned
rounds 3, 4, 5, and 6 for each participant, which we call productivity. The second is a measure of risk-aversion for each participant using the Holt-Laury (2002) instrument described above and presented in Table 1. Specifically, the number of gambles above the switching point from safe to risky gambles represents our measure of risk-aversion. For H4 and H5, the dependent variable is the number of correct words for each of the assigned rounds 3, 4, 5, and 6 for each participant. For H4, the independent variables are a treatment dummy for the assigned compensation scheme with FS = 0 and PFP = 1 and a linear time trend to take account of possible learning effects. For H5, we add the participant-specific levels of risk-aversion described above as a potential moderator of the treatment effect.

DATA ANALYSIS AND RESULTS

Data Overview

Insert Tables 2 & 3 about here.

All 115 participants completed the study. Table 2 presents the anagrams used by round along with the treatment conditions employed, as well as the productivity, earnings, and per-unit cost statistics under those treatment conditions. Table 2 also reports the average productivity for each anagram in Vance and Colella (1990) and in the pre-test we ran prior to the current study. In accordance with agency theory, earnings were higher under PFP ($2.55) than under FS ($2.20) in rounds 7 and 8. In round 1, PFP earnings were also higher ($2.43), while in round 2, PFP earnings were lower ($1.92) than FS earnings. The lower earnings in round 2 were likely due to the fact that participants were still learning to sort themselves appropriately and that, according to our pre-test, the anagram in round 2 was particularly difficult. Table 3 reports means, standard deviations, and correlations of the variables.

Factors Influencing Self-Selection: H1 to H3

We investigated the roles played by risk aversion and productivity on the compensation scheme selected, both for the initial two rounds and for the final two rounds, utilizing a logistic

5 Of the 115 participants, 99 had one switching point, consistent with expected utility theory. In our analysis, involving attitudes toward risk, we followed the cautious approach of discarding data from those who exhibited more than one switching point, leaving us with 99 usable data points.
regression based on our two-by-two factorial design as follows:

\[ \ln \left( \frac{p_{ssj}}{1 - p_{ssj}} \right) = \beta_0 + \beta_1 \cdot \text{Productivity} + \beta_2 \cdot \text{Risk aversion} + \beta_3 \cdot \text{Interaction} \quad (1) \]

where \( p_{ssj} \) = the probability of self-selecting the PFP scheme and \( j \) refers to the initial versus the final selection. A separate estimation was run for each of the two selection decisions.

The first factor was the average productivity of each participant as measured by the data from the four middle rounds when all players were compensated in the same manner. The null hypothesis was contrasted with the alternative suggested by H1 that higher levels of productivity would be associated with a higher probability of selecting the PFP scheme. The second factor was each participant’s level of risk aversion as measured by the lottery mechanism. Our participants were highly risk-averse with 93% exhibiting some degree of risk aversion. Of those remaining, 3.5% were risk-neutral, while another 3.5% were risk-loving. These levels of risk aversion were somewhat higher than those found by Holt and Laury (2002) in their lower-stakes setting and roughly comparable to those found in their higher-stakes setting.\(^6\) The null hypothesis regarding this factor was contrasted with the alternative suggested by H2 that higher levels of risk aversion would be associated with a lower probability of selecting the PFP scheme. Finally, we tested the null hypothesis on the interaction of risk-attitude and productivity against the alternative suggested by H3 that higher levels of productivity would be associated with a stronger negative impact of the level of risk aversion, resulting in a predicted negative interaction effect. We centered productivity at the break-even point of 11 words so that the coefficient on risk aversion was estimated at the lowest point where it was likely to have a significant impact.\(^7\) Risk aversion was conventionally centered at its mean so that the coefficient on productivity was estimated at the mean level of risk aversion.

---

\(^6\) Recall that we multiplied Holt and Laury’s (2002) lower stakes setting by 2.2 to approximate the monetary stakes in two rounds of our anagram game. Hence our stakes were in between their lower and higher stakes settings.

\(^7\) Cohen, Cohen, West and Aiken (2003: 261–282, especially p. 281) contain an excellent discussion on the importance of centering the covariate. If productivity were not centered, the coefficient on the risk-aversion variable would indicate its effect at a fictional productivity level of zero that is never observed in the experiments. Note that we did not center the interaction term, but rather created the interaction term as a product of the two centered main effect terms.
Table 4 reports the results with coefficients and p-values in Panel A and some examples of effect sizes in Panel B. For the initial self-selection, the null hypotheses for both productivity ($p = .003$) and risk aversion ($p = .040$) were both rejected in the direction of the specified alternatives. However, the interaction coefficient, though of the predicted negative sign, was not significant ($p = .123$). Thus, those who subsequently performed better did have a significantly higher probability of selecting PFP, controlling for their differing attitudes toward risk. For example, at the mean level of risk aversion, the probability of selecting PFP rose by 0.31 from 0.41 for productivity one standard deviation (henceforth SD) below to 0.72 for productivity one SD above the breakeven point of 11 words. At the same time, controlling for productivity, those who exhibited a higher degree of risk aversion in their lottery choices were less likely to choose the risky PFP compensation scheme. For example, for a productivity level of 11 words, the probability of selecting PFP fell by 0.13 from 0.57 for a participant at the mean level of risk aversion to 0.44 for a participant one SD above the mean risk-aversion level. For the final self-selection, the null hypotheses for both productivity ($p = .000$) and risk aversion ($p = .015$) were also both rejected in the direction of the specified alternatives. With more experience, at the mean level of risk aversion, the probability of selecting PFP was 0.57 higher ($0.92 - 0.35$) for a participant whose productivity was one SD above compared to one whose productivity was one SD below the 11-word breakeven point. The null hypothesis for the interaction was also rejected in the direction of the predicted negative effect ($p = .048$). Figure 1 illustrates the relationship between the logistic function, $\ln \left[ \frac{p}{1 - p} \right]$, and the level of risk aversion for productivity levels one SD above and one SD below 11 words for both the initial and final cases. Examples of how this interaction affected the relationship between the probability of selecting PFP and the level of risk aversion for different levels of productivity are presented in Table 4, Panel B.

Clearly, self-selection into the two compensation schemes depended on both risk aversion and productivity in the manner predicted by H1 and H2 in both self-selections, and by H3 in the
The coefficients were larger and the $p$-values smaller in the final than in the initial selection. In addition, the model correctly predicted the choice of compensation scheme 77.8% of the time in the final compared to 62.6% of the time in the initial selection. This suggests that participants were learning from experience both to forecast their expected earnings under the two available schemes, and to adjust their choices based on those forecasts and on their individual levels of risk aversion.

**PFP’s Incentive Effect and its Relationship to Risk Preference: H4 and H5**

Insert Table 5 about here.

H4 predicted that, besides attracting higher-quality employees, PFP would induce more effort and hence higher productivity from employees, regardless of their quality, than FS. We tested this hypothesis with the data from the middle four rounds in which participants were assigned to payment schemes, thus isolating this effect from the effect of self-selection. We employed a two-level hierarchical linear regression model to deal with the repeated measures for each participant as follows:

$$\text{Prod}_{it} = \beta_0 + \beta_1 \cdot \text{Round} + \beta_2 \cdot \text{Treatment} + \eta_i + \epsilon_{it}$$

where $\text{Prod}_{it}$ is the number of correct words made by participant $i$ in round $t$, Round is a linear time trend with round $3 = 0$, Treatment is a dummy variable equal to 1 for PFP rounds and 0 for FS rounds, $\eta_i$ is an participant-specific random effect, and $\epsilon_{it}$ is a randomly-distributed disturbance term. Initially, we included an anagram-difficulty-score variable based on the Vance

---

8 An explained in the experimental design section, the Holt-Laury instrument was administered after the behavioral experiment was completed, but before participants received any feedback or payment. Nonetheless, as an anonymous reviewer has pointed out, it is possible that the self-selection decisions in the game affected the elicited measures of risk aversion. This would result in biased estimates of the coefficients in equation (2). We examined this possibility by using average individual productivity in the four assigned rounds as an instrumental variable to estimate the effect of self-selection on risk aversion. Since productivity is correlated with self-selection, but not with the disturbance term of a regression of risk aversion on self-selection, such a technique yields a consistent estimate of the impact of self-selection on risk aversion even if risk aversion also influences self-selection as specified in equation (2). Kennedy (2003): 159–160, 162–163, 167–168 and 174–176 contains an excellent explanation of why this procedure is necessary and proves that it yields consistent estimates in a case like ours. We fail to reject the null hypothesis of no effect of self-selection on risk-aversion level for both the initial and final self-selection ($p = .751$ in the initial case and $p = .745$ in the final case). Of course, there is always a possibility of type-two error. However, we can conclude that there is no evidence in the data that either the initial or final self-selection had an impact on the risk-aversion measure, and thus no evidence of bias in the estimated coefficients of equation (2). Furthermore, there is no significant correlation between individual levels of productivity and individual risk-aversion measures, whether productivity is measured under the assigned FS scheme, the assigned PFP scheme, or based on an average under both schemes as indicated in Table 3.
and Collela (1990) study to control for differences in difficulty. However, this variable was not
significant. Thus, we dropped it from the analysis. We also initially attached individual-level
random effects to the coefficients on both round and treatment, allowing these coefficients to
vary randomly between participants. Since these random effects were not significant, we also
dropped them. None of the dropped variables made any difference to the coefficients or
significance levels of the included independent variables.

The results are reported in the first column of Table 5. There was clearly learning going
on from round 3 through round 6 as evidenced by the significant coefficient on the round
variable ($p = .006$). Treatment was significant in the predicted positive direction with a treatment
effect size of 0.86 words ($p = .000$). Therefore, the data supported H4. We tested H5, which
predicted that the effectiveness of PFP at improving productivity was inversely related to the
level of risk aversion, by adding an interaction between the treatment variable and centered risk
aversion to the above model as reported in the last column of Table 5. The coefficient of -0.21
was significant in the predicted negative direction ($p = .020$). While the treatment effect size was
0.86 words at the mean level of risk aversion, it fell to 0.46 words one SD above the mean and
rose to 1.26 words one SD below the mean.\(^9\) Thus, the opportunity to earn more money through
better performance was a less effective motivator for more risk-averse individuals as predicted
by H5. In fact, for 25.2% of our participants, PFP was actually associated with a decline in
productivity.

Since risk aversion significantly affected both choice of pay scheme and the incentive
effect of PFP, it is interesting to examine whether the H4 incentive effect differed between those
selecting each scheme. The results are also reported in Table 5. For both the initial (SS1) and
final (SS2) selections, the treatment effect was significantly positive for both the FS (SS1: $p =
.050$, SS2: $p = .039$) and PFP groups (SS1: $p = .001$, SS2: $p = .001$). Although the treatment
coefficients were higher for those selecting PFP, the null hypothesis that the coefficients were

\(^9\) Note that H5 does not predict a significant main effect for risk aversion. Including a main effect in the regression yields an
insignificant p-value. Thus, we dropped the main effect of risk aversion from the analysis.
identical could not be rejected using a two-tail test for either SS1 ($p = .152$) or SS2 ($p = .347$).

Choice of payment scheme was affected not only by risk aversion, but also by expected productivity and resultant earnings. Thus, highly risk-averse people who expect a high enough risk premium under PFP will choose that scheme, while less risk-averse people who expect to earn less under PFP than under FS choose FS. Hence allowing employees to choose between FS and PFP cannot guarantee that those with low (high) levels of risk aversion for whom the incentive effect is predicted to be strongest (weakest) will actually choose the PFP (FS) scheme.

**Supplementary Analysis: Decomposing Sorting and Incentive Effects**

We have now established that people who selected PFP performed better than those who selected FS for two reasons: sorting and incentives. For both of these reasons, a laboratory “firm” offering PFP achieved significantly higher productivity than an identical “firm” offering FS by 14.25% ($10.87 - 9.53, p = .013$) in the initial self-selection rounds and by 38.09% ($12.76 - 9.24, p < .001$) in the final self-selection rounds. It is impossible to decompose these differences into sorting and incentive components, using data from self-selection rounds 1 and 2 or 7 and 8 alone. However, it is possible to combine the compensation-scheme choices from the self-selection rounds with the productivity data from rounds 3 and 5, in which everyone was assigned to FS, versus rounds 4 and 6, in which everyone was assigned to PFP, to separate the two effects. This decomposition is illustrated in Figure 2. For those selecting PFP initially, average productivity in rounds 4 and 6 under PFP was 11.43 words. For those selecting FS initially, average productivity in rounds 3 and 5 under FS was 8.71 words. The difference of 2.72 words (31.23%) was significant ($p < .001$), just as was the comparable difference in rounds 1 and 2. This difference can be decomposed into sorting effects between persons making different self-selections measured under a common assigned scheme and incentive effects on individuals of working under the different schemes.\(^{10}\) The sorting effect under the assigned PFP scheme for

\(^{10}\) For simplicity, in the analysis that follows, we ignore any learning that takes place between adjacent rounds. The regression coefficient on the ‘Round’ variable in equation (2) was 0.26 ($p = .013$), suggesting that participants are generally making 0.26 more words with each new round of experience. Thus, the incentive effects discussed in the text are likely about 0.26 words lower than the decompositions suggest since those effects are calculated by comparing productivity numbers between adjacent rounds.
rounds 4 and 6 was 1.73 words (11.43 – 9.70, \( p = .001 \)), while the incentive effect for those selecting FS was .99 words (9.70 – 8.71, \( p < .001 \)) for PFP rounds 4 and 6 versus FS rounds 3 and 5. The incentive effect for those initially selecting PFP was 1.26 words (11.43 – 10.17, \( p < .001 \)) for PFP rounds 4 and 6 versus FS rounds 3 and 5, while the sorting effect under the assigned FS scheme for rounds 3 and 5 was 1.46 words (10.17 – 8.71, \( p = .006 \)). Thus, even at the initial self-selection, both sorting and incentive effects contributed significantly to productivity differences, with the sorting effect quantitatively the more important of the two.

When the middle rounds data were reorganized based on the final self-selection decisions, there was little change in the incentive effects, which were 1.03 words (9.13 – 8.10, \( p = .000 \)) for those selecting FS and 1.22 words (11.87 – 10.65, \( p = .000 \)) for those selecting PFP in the final rounds. However, the sorting effects increased to 2.73 words (11.87 – 9.13, \( p = .001 \)) under the assigned PFP scheme and 2.55 words (10.65 – 8.10, \( p = .000 \)) under the assigned FS scheme, indicating that participants made self-selections more in tune with their actual quality once having gained experience with the experimental task. Thus, at the final self-selection, while both sorting and incentive effects continued to contribute significantly to productivity differences, the former was now more than twice as important as the latter.

**CONCLUSION AND DISCUSSION**

**Key Findings and Discussion**

The importance of the sorting and incentive effects of PFP has been proposed and examined before. However, the separate presence and significance of the two effects in isolation from a pay guarantee ensuring that PFP earnings cannot fall below FS earnings as in Lazear’s (1999, 2000a, 2000b) Safelite study have not to our knowledge been disentangled and directly tested. Our study has demonstrated that PFP has two advantages for firms relative to FS: first, it attracts higher-quality employees, and second, it motivates employees to exert more effort.

Both attraction and attrition facilitated sorting. In fact, 19 (16) participants who had chosen FS (PFP) initially changed their choice to PFP (FS) in the final rounds. A closer look reveals that those who changed from FS to PFP were more productive in the assigned rounds
than those who stayed in FS (10.71 vs. 8.47, *t*(57) = 3.34, *p* < 0.001), while those who changed from PFP to FS were less productive in the assigned rounds than those who stayed in PFP (8.97 vs. 11.51, *t*(56) = 3.68, *p* < 0.001). Thus, experience seemed to help participants sort themselves toward higher earnings whether that meant a move toward or away from PFP. However, in Lazear (2000a), most sorting occurred through hiring, i.e. attraction. The major reason for this discrepancy is likely that nobody made less under Safelite’s PFP scheme than under hourly wages because of the guarantee Safelite set at the level of the previous hourly wages. Thus, there was little incentive to leave Safelite after the change in compensation system. In contrast, in our laboratory setup, less productive employees could earn more under the FS scheme.

Disentangling the incentive from the sorting effect by a within-person analysis to examine productivity differences under the two compensation schemes, we found strong support for the incentive effect of PFP. On average, the incentive effect on productivity was however quantitatively less important than the sorting effect, particularly once participants had gained experience with the task and made the final self-selection. Interestingly, this differs from the results of Lazear’s (2000a) field-study results, where sorting and incentive effects made roughly equal contributions to PFP productivity gains. This is not surprising. It is costly for employees to leave and look for fixed hourly wages elsewhere in the real world. Besides, there was little incentive to do so because of the guaranteed minimum wage accompanying Safelite’s PFP scheme. In our laboratory context, there was no cost at all to choosing one scheme over the other or moving from one scheme to the other at the time of the second selection. Thus, there were fewer impediments to sorting. The importance of sorting outside the lab is clearly related to the precise details of the PFP scheme and the costs of moving from one scheme to another.

We have also shown that risk attitudes have an important impact on both the compensation scheme chosen and the incentive effect of PFP. In particular, high-productivity individuals are less likely to select PFP when they are highly risk-averse, and the incentive effects of PFP are weaker for more risk-averse individuals. Agency theory stresses the important role played by risk attitudes in sorting decisions. By showing that risk attitudes can play a role in
sorting choices in a behavioral experiment with salient financial incentives, we demonstrate that the theory has some explanatory power albeit in a laboratory rather than a workplace setting. We believe this represents an advance in our knowledge relative to theory supported by self-report questionnaire data alone (Cable & Judge, 1994). By showing that the incentive effect of PFP relative to FS is weaker for more risk-averse people in a laboratory setting, we draw attention to the possibility that such heterogeneity in the responsiveness of individuals to PFP incentives in the workplace may be an important issue. This result is consistent with Deckop, Merriman and Blau (2004), who demonstrated by means of a survey study that individual levels of risk aversion moderate the effects of their “control-by-pay” measure on withdrawal cognitions, contingent pay satisfaction, and organizational citizenship behavior. The importance of risk aversion under the many varieties of PFP schemes and workplace settings of course requires further research.

**Methodological Advantages, Caveats, and Limitations**

This experimental design, incorporating both between- and within-person analyses of the effects of making pay contingent on individual performance, offers several unique advantages. First, it is useful to treat the same individual with different compensation schemes in order to examine the effect of each scheme on individual behavior. The within-person approach allowed for pairwise comparison between treatments, thus controlling for subject variability to a greater extent than random assignment. For a given sample size this resulted in greater statistical power. Second, the within-person design also allowed us to examine the diversity of individual responses to the different payment schemes. For example, we were able to test and corroborate the hypothesis that the effectiveness of PFP at improving performance was inversely related to individual levels of risk aversion, and show that more than a quarter of the participants performed better under FS than under PFP. We believe these are interesting and important findings, which should be investigated further. However, they could not have emerged under a between-person approach. Third, under our within-person design, the productivity data from the

---

11 Rynes, Gerhart, and Minette (2004) recently argued that results of many attitudinal studies on the relative importance of various types of pay systems may have been subject to social desirability bias, a potential threat to both internal and external validity.
middle imposed-scheme rounds acted as a measure of productivity in the self-selection regressions. This permitted us to test our central sorting hypothesis that people sort into compensation schemes based on an exogenous person-specific measure of average productivity. It is just such a task-specific measure of employee quality that is of primary concern to managers. Finally, comparing the performance of participants who have chosen between the two compensation schemes may have more external validity than a comparison based on random assignment since, as argued by Schneider (1987), people make precisely such choices when applying for jobs in real life.

However, this study has a number of limitations that must be acknowledged. First, there are many forms of PFP. We used piece rates to represent PFP in our laboratory study. The incidence of piece rate varies among industrialized countries from 0.9% based on the Panel Study of Income Dynamics to 3.6% based on the National Longitudinal Survey of Youth (Parent, 2002: 19) of employees in the US, to 13% of manufacturing establishments in Germany (Jirjahn, 2002: 152). Considerably more employees participate in other forms of PFP. We could not study all forms of PFP in this study. We concentrated on piece rates because they are the simplest and most direct form of PFP, and thus a good place to begin examining the sorting and incentive effects of PFP predicted by theory. Tournament schemes, merit increases, and group-based incentives are all relevant and interesting to study. However, the incentives embedded in these schemes are more complex both in terms of the economic incentives they may provide and in the psychological effects they may induce. We hope to examine the effects of other forms of PFP in future research. The results of this study corroborate the theoretical predictions regarding PFP, and illustrate the possibility of both sorting and incentive effects. However, their applicability to the many other forms of PFP that exist in the workplace is obviously limited.

Second, we sampled from a population of university students. To what extent is it legitimate to generalize from the behavior of such a sample to a broader population of job-seekers? Sackett and Larson (1990) argued that sampling from a student population in this context is appropriate because most of the students were actually working at part-time jobs,
seeking jobs at the time of the study, or would soon enter the job market. Nevertheless, it is possible that different results would have been obtained by sampling from other populations. Indeed it is important to test behavioral theories using samples from a variety of populations.

Third, the money at stake in our experiment was far less than a year’s worth of pay for a typical employee. Of course, we cannot create a situation identical to a real-world workplace in the laboratory. We acknowledge that this may affect risk preferences, with people exhibiting less risk aversion when less is at stake. Thus, we caution the reader not to extrapolate the measures of risk aversion in our study to situations in which the stakes are substantially higher. Our aim in this study was not to identify the level of risk aversion typically brought to an individual compensation-scheme decision per se; rather, our purpose was to examine the role played by differing levels of risk aversion (along with expected productivity) on the compensation-scheme decision. Our risk-aversion elicitation was based on choices involving financial stakes of the same magnitude as those in the behavioral experiment. We do not claim that these elicited levels of risk aversion are similar to those that would be elicited if much higher stakes were involved. Rather, we hypothesize and test that, based on agency theory, individuals with higher levels of risk aversion are less likely to choose PFP schemes regardless of the stakes involved12.

Fourth, there was no minimum standard below which an employee would be fired in our experimental design. In contrast, in the workplace, such a minimum generally exists both in FS and PFP environments (e.g. Lazear, 2000a) and the possibility of being fired and losing one’s salary could motivate higher productivity. Inserting such a feature into this experimental design would have likely had little effect if the minimum standard had been set substantially below the 11-word break-even point for FS and PFP. For example, in rounds 3 and 5, when all participants were assigned to the FS scheme, 6.1% (7 participants) made 5 or fewer words. In rounds 4 and 6, when all participants were assigned to the PFP scheme, 2.6% (3 participants) made 5 or fewer

12 In regard to the stake effect on behavioral responses in experiments, a number of studies examine the extent to which stakes matter to behavioral decisions in a variety of contexts. Over a broad range of contexts, they demonstrate that as long as the financial stakes in a behavioral experiment are equal to or greater than the opportunity cost of a person’s time (i.e. the amount they could earn elsewhere), there are no significant effects on behavior (e.g., Slonim & Roth, 1998). Of course, it is always possible that our particular context is an exception to these general findings. This warrants further examination.
words. Thus, if anyone making fewer than 6 words had been subject to termination and cessation of salary, it is likely that 4 participants would have performed marginally better under the FS scheme. This would have had little impact on the overall results. Of course, if the minimum standard were set close to 11 words under an FS scheme, it might motivate higher productivity over a broader range of participants. It would be interesting to compare the effects of such a scheme with the effects of piece rates or other types of PFP in future work.

Fifth, the laboratory nature of the task and the starkness of the manipulation, i.e. switching compensation schemes and providing no feedback in between rounds, necessarily limits the generalizability of the results to complex organizational settings. In addition, the time frame in the lab was minimal compared to the workplace. This factor may also affect behavior and decisions. Another limitation of our experimental design was that its within-person nature made it impossible to make certain comparisons of interest. For example, randomly assigning subjects to one of a PFP condition, a FS condition, or a condition in which they could choose a pay scheme would have allowed us to compare, the productivity of those choosing PFP with those assigned to that approach. Unfortunately, such a design coupled with the salient financial incentives central to our methodological approach, would have been very costly, and would have precluded the within-person comparisons facilitated by our design. Nonetheless, such an examination could yield additional new insights.

Finally, laboratory findings such as ours should not be generalized to complex organizational settings without taking into account other potentially important factors such as task characteristics, and contextual/cultural factors. For example, it is possible that in an occupational setting, the constant stress of PFP might ultimately result in lower productivity, while a generous FS scheme might engender effort motivated by positive reciprocity. Akerlof (1982) discussed an interesting example of the latter in which more talented employees in a tightly-knit community exerted effort out of gratitude to an employer who set standards that were low enough for both lower- and higher-quality employees to meet. While our experiment does not capture such potentially important social factors, it nonetheless reveals that a firm’s
compensation scheme can significantly affect not only the motivation of employees, but also the types of employees it attracts with respect to both productivity and risk attitudes.

Notwithstanding these limitations, in our view, the point of the laboratory is not to recreate the real world, but to control important factors that may be difficult to control using field data. Of course, differences between a laboratory and workplace environment may cause unexpected problems. However, every methodology has its strengths and its weaknesses. For this reason, we strongly advocate that a variety of methodologies including field studies, hypothetical self-report questionnaires and also laboratory behavioral studies be employed to examine such important questions as the effects of different compensation systems. Together such studies can help us learn more about the proposed theories than any one methodological approach alone.

Managerial Implications and Directions for Future Research

This paper offers some important insights for managers. First, we have demonstrated that it is possible for a firm to attract higher-quality applicants by implementing PFP while simultaneously decreasing its unit cost of production. Of course, under PFP the per-unit cost is always $0.20. In Table 2, we report per-unit costs under FS of $0.23 for the middle rounds, $0.23 for the initial rounds and $0.24 for the final rounds. Thus, our study demonstrates that the labor cost per-unit of production can be lower under PFP, consistent with Sturman et al.’s (2003) hypothetical scenario study. Of course, whether this will actually occur depends on the particular parameters of the PFP and FS schemes being compared.

Second, the implications of PFP for organizations and their employees are quite different when PFP is chosen versus when it is imposed. As reported above, when employees were permitted to select their preferred compensation scheme, those who selected PFP earned higher wages on average than under FS, corroborating the importance of a PFP risk premium as predicted by agency theory. In contrast, as reported in Table 2, in the middle rounds in which participants were assigned to pay schemes, average earnings were lower under PFP ($2.11 averaged over rounds 4 and 6) than under FS ($2.20). The risk premium that must be paid to compensate for PFP risk is based on the notion that employees willingly choose such a scheme
over a less risky alternative. This is the case in our experimental setting when participants have a choice of payment scheme in the initial and final rounds, and is the case in workplace settings when people select a job from a number of possibilities with different payment characteristics. In such cases, those who expect that their productivity will be high enough to earn enough under PFP to compensate them for the risk involved will choose PFP, while others will choose FS. This is not however the case when participants have no such choice during the middle rounds of our experiment. The middle rounds are analogous to a situation in which employees have little choice but to accept a piece-rate job because that is the only type of job available. *The Guardian* (2003) reported that a legal loophole that had enabled employers to “pay homeworkers and others on piece rates 20% less than the minimum wage” in the U.K had just been closed. Previously, employers were permitted “to set piece rates so that only the most proficient or speediest worker” received the minimum wage of £4.20, “while the average worker would be on 80% of this.” Such low rates of piece-rate pay with no financial compensation for pay risk are indicative of a piece-rate ghetto, in which work alternatives are not available. The assigned PFP treatment in rounds 4 and 6 resembles such a situation.

Third, we have shown that the incentive effects of PFP are neither uniform nor universal. More risk-averse employees are less responsive to performance-based financial incentives, and may even suffer a decline in productivity under PFP. This somewhat mitigates Lazear’s (2000a: 1347) conclusion based on his important field study that: “Workers respond to prices just as economic theory predicts. Claims by sociologists and others that monetizing incentives may actually reduce output are unambiguously refuted by the data.” It would appear that the economic theory of financial or price incentives predicts better for less risk-averse persons, whereas the “sociologists and others” (Lazear cites Deci, 1971) make better predictions for those who are more risk-averse. Indeed our results suggest that more risk-averse employees are likely to be less responsive, unresponsive, or even respond in the wrong direction when faced with PFP. Since the stakes are likely to be higher in the workplace than in the laboratory, and risk aversion is likely to be greater at higher stake levels (Holt & Laury, 2002), this phenomenon could well be
even more pronounced in the workplace than in the lab. These findings represent an important 
caveat to the effectiveness of the incentive effects of PFP not evident in Lazear’s field study. It 
may be that for many highly risk-averse people, PFP causes hindrance stress, and that such stress 
impairs performance. Individual differences are important, and managers must not lose sight of 
them in designing effective compensation systems.

Fourth, we have corroborated the prediction of agency theory that those selecting PFP 
tend to be less risk-averse than those selecting FS. Following Schneider’s (1987) “people-make-
the-place” argument, this may affect the culture of the organization, leading to less risk-averse 
organizational decisions. Managers should be cognizant of this possibility. Finally, the potential 
benefits of PFP must be assessed in conjunction with the understanding that in practice the 
design of payment policies to support the provision of sustainable high quality work is a complex 
undertaking. As argued by Sturman et al. (2003), designers of PFP programs must determine the 
size of the necessary financial rewards and whether the potential return on investment in terms of 
additional productivity outweighs the financial and human costs involved. It is important to 
recognize as well that non-financial incentives can also drive positive behavior changes. In 
designing PFP schemes, it is also vital to attend to and minimize unintended negative 
consequences. For example, PFP schemes can have a destructive effect on the actual 
performance of some employees as evidenced by the 25.2% of our participants who were 
actually less productive under PFP than under FS. Furthermore, PFP may dysfunctionally 
motivate employees to focus excessively on tasks that lead to individual financial rewards, 
sometimes at the expense of other equally important tasks. Moreover, PFP can lead to 
considerable deception, which can have devastating consequences for organizations (Jensen, 
2003; Schweitzer, et al., 2004). Thus, notwithstanding the merit of PFP, managers must pay 
special attention to designing PFP programs that avoid these drawbacks and to motivating high 
levels of performance through other means as well. Future research should explore the 
effectiveness of such mechanisms and their interactions with various forms of PFP in different 
organizational settings.
REFERENCES


TABLE 1
Risk-Aversion Measure: Holt-Laury Instrument

<table>
<thead>
<tr>
<th>Option A</th>
<th>Option B</th>
<th>MY CHOICE</th>
</tr>
</thead>
<tbody>
<tr>
<td>10% chance of $4.40, 90% chance of $3.52</td>
<td>10% chance of $8.47, 90% chance of $.22</td>
<td></td>
</tr>
<tr>
<td>20% chance of $4.40, 80% chance of $3.52</td>
<td>20% chance of $8.47, 80% chance of $.22</td>
<td></td>
</tr>
<tr>
<td>30% chance of $4.40, 70% chance of $3.52</td>
<td>30% chance of $8.47, 70% chance of $.22</td>
<td></td>
</tr>
<tr>
<td>40% chance of $4.40, 60% chance of $3.52</td>
<td>40% chance of $8.47, 60% chance of $.22</td>
<td></td>
</tr>
<tr>
<td>50% chance of $4.40, 50% chance of $3.52</td>
<td>50% chance of $8.47, 50% chance of $.22</td>
<td></td>
</tr>
<tr>
<td>60% chance of $4.40, 40% chance of $3.52</td>
<td>60% chance of $8.47, 40% chance of $.22</td>
<td></td>
</tr>
<tr>
<td>70% chance of $4.40, 30% chance of $3.52</td>
<td>70% chance of $8.47, 30% chance of $.22</td>
<td></td>
</tr>
<tr>
<td>80% chance of $4.40, 20% chance of $3.52</td>
<td>80% chance of $8.47, 20% chance of $.22</td>
<td></td>
</tr>
<tr>
<td>90% chance of $4.40, 10% chance of $3.52</td>
<td>90% chance of $8.47, 10% chance of $.22</td>
<td></td>
</tr>
<tr>
<td>100% chance of $4.40, 0% chance of $3.52</td>
<td>100% chance of $8.47, 0% chance of $.22</td>
<td></td>
</tr>
</tbody>
</table>
### TABLE 2

**Productivity, Earnings, and Cost-per-Unit by Round and Treatment**

(Standard Deviations in Parentheses)

<table>
<thead>
<tr>
<th>Anagram</th>
<th>Treatment</th>
<th>FS</th>
<th>PFP</th>
<th>VC (1990)(^a)</th>
<th>Pre-Test</th>
<th>Earnings</th>
<th>Cost-per-Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. OADMHUP</td>
<td>SS1</td>
<td>10.41 (3.05)</td>
<td>12.14 (3.48)</td>
<td>11.16</td>
<td>14.58</td>
<td>FS: 2.20 / PFP: 2.43</td>
<td>FS: 0.21 / PFP: 0.20</td>
</tr>
<tr>
<td>2. AEDBKUG</td>
<td>SS1</td>
<td>8.66 (2.75)</td>
<td>9.60 (3.37)</td>
<td>8.95</td>
<td>10.98</td>
<td>FS: 2.20 / PFP: 1.92</td>
<td>FS: 0.25 / PFP: 0.20</td>
</tr>
<tr>
<td>3. OELBJAM</td>
<td>FS</td>
<td>9.00 (3.47)</td>
<td>N/A</td>
<td>9.42</td>
<td>12.15</td>
<td>FS: 2.20 / PFP: -</td>
<td>FS: 0.24 / PFP: -</td>
</tr>
<tr>
<td>4. UADQWER</td>
<td>PFP</td>
<td>N/A</td>
<td>10.46 (3.47)</td>
<td>8.84</td>
<td>11.89</td>
<td>FS: - / PFP: 2.09</td>
<td>FS: - / PFP: 0.20</td>
</tr>
<tr>
<td>5. EASCKIY</td>
<td>FS</td>
<td>9.86 (3.81)</td>
<td>N/A</td>
<td>9.32</td>
<td>12.09</td>
<td>FS: 2.20 / PFP: -</td>
<td>FS: 0.22 / PFP: -</td>
</tr>
<tr>
<td>6. AODJGIP</td>
<td>PFP</td>
<td>N/A</td>
<td>10.65 (3.81)</td>
<td>9.63</td>
<td>12.20</td>
<td>FS: - / PFP: 2.13</td>
<td>FS: - / PFP: 0.20</td>
</tr>
<tr>
<td>7. UONHMEY</td>
<td>SS2</td>
<td>9.51 (2.76)</td>
<td>12.77 (3.09)</td>
<td>10.21</td>
<td>13.06</td>
<td>FS: 2.20 / PFP: 2.55</td>
<td>FS: 0.23 / PFP: 0.20</td>
</tr>
<tr>
<td>8. OELHMAZ</td>
<td>SS2</td>
<td>8.96 (3.09)</td>
<td>12.75 (4.04)</td>
<td>9.63</td>
<td>13.62</td>
<td>FS: 2.20 / PFP: 2.55</td>
<td>FS: 0.25 / PFP: 0.20</td>
</tr>
<tr>
<td>1 &amp; 2 Average</td>
<td>SS1</td>
<td>9.53 (2.58)</td>
<td>10.87 (3.05)</td>
<td>10.06</td>
<td>12.78</td>
<td>FS: 2.20 / PFP: 2.17</td>
<td>FS: 0.23 / PFP: 0.20</td>
</tr>
<tr>
<td>3 &amp; 5 Average</td>
<td>FS</td>
<td>9.43 (2.85)</td>
<td>N/A</td>
<td>9.37</td>
<td>12.12</td>
<td>FS: 2.20 / PFP: -</td>
<td>FS: 0.23 / PFP: -</td>
</tr>
<tr>
<td>4 &amp; 6 Average</td>
<td>PFP</td>
<td>N/A</td>
<td>10.56 (2.89)</td>
<td>9.24</td>
<td>12.05</td>
<td>FS: - / PFP: 2.11</td>
<td>FS: - / PFP: 0.20</td>
</tr>
<tr>
<td>7 &amp; 8 Average</td>
<td>SS2</td>
<td>9.24 (2.48)</td>
<td>12.76 (3.23)</td>
<td>9.92</td>
<td>13.34</td>
<td>FS: 2.20 / PFP: 2.55</td>
<td>FS: 0.24 / PFP: 0.20</td>
</tr>
</tbody>
</table>

\(^a\)VC (1990): Results from Vance and Colella (1990).
TABLE 3
Means, Standard Deviations, and Correlations\textsuperscript{a}

<table>
<thead>
<tr>
<th>Variable</th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Risk aversion</td>
<td>6.77</td>
<td>1.92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Productivity under the FS Scheme (Round 3 &amp; 5)</td>
<td>9.43</td>
<td>2.85</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Productivity under the PFP Scheme (Round 4 &amp; 6)</td>
<td>10.56</td>
<td>2.89</td>
<td>−0.11</td>
<td>0.78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Productivity in Middle 4 rounds (Rounds 3 to 6)</td>
<td>9.99</td>
<td>2.71</td>
<td>−0.04</td>
<td>0.94</td>
<td>0.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Initial Self-Selection (FS = 0, PFP = 1)</td>
<td>0.50</td>
<td>0.50</td>
<td>−0.12</td>
<td>0.26</td>
<td>0.30</td>
<td>0.30</td>
<td></td>
</tr>
<tr>
<td>6. Final Self-Selection (FS = 0, PFP = 1)</td>
<td>0.52</td>
<td>0.50</td>
<td>−0.17</td>
<td>0.45</td>
<td>0.48</td>
<td>0.49</td>
<td>0.39</td>
</tr>
</tbody>
</table>

\textsuperscript{a}N = 115. Correlations ≥ |.30| are significant at \( p < .001 \); correlations ≥ |.25| are significant at \( p < .01 \); and correlations ≥ |.18| are significant at \( p < .05 \) (two-tail test).
TABLE 4
Logistic Regression of the Probability of Choosing the PFP Compensation Scheme as a Function of Productivity and Attitude Toward Risk (Two-Tail p-values for Constants and One-Tail p-values for Explanatory Variables in Parentheses)

Panel A: Logistic Regression

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Initial Self-Selection</th>
<th>Final Self-Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>.28 ( .226)</td>
<td>.92 (.004)</td>
</tr>
<tr>
<td>Productivity</td>
<td>.24 (.003)</td>
<td>.57 (.000)</td>
</tr>
<tr>
<td>(Centered at 11)</td>
<td>(.003)</td>
<td>(.000)</td>
</tr>
<tr>
<td>Risk aversion</td>
<td>−.26 (.123)</td>
<td>−.45 (.048)</td>
</tr>
<tr>
<td>(Centered at Mean)</td>
<td>(.040)</td>
<td>(.015)</td>
</tr>
<tr>
<td>Interaction</td>
<td>−.06</td>
<td>−.14</td>
</tr>
</tbody>
</table>

Panel B: Effect Sizes

<table>
<thead>
<tr>
<th>Risk Aversion</th>
<th>Productivity</th>
<th>Initial Probability</th>
<th>Final Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Level</td>
<td>Effect</td>
</tr>
<tr>
<td>6.77 (mean)</td>
<td>8.29 (−1 SD)</td>
<td>0.41</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>11 (breakeven point)</td>
<td>0.57</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>13.71 (+1 SD)</td>
<td>0.72</td>
<td></td>
</tr>
<tr>
<td>6.77 (mean)</td>
<td>11 (breakeven point)</td>
<td>0.57</td>
<td>0.13</td>
</tr>
<tr>
<td>8.69 (+1 SD)</td>
<td>0.44</td>
<td>0.51</td>
<td></td>
</tr>
<tr>
<td>6.77 (mean)</td>
<td>13.71 (+1 SD)</td>
<td>0.72</td>
<td>0.36</td>
</tr>
<tr>
<td>8.69 (+1 SD)</td>
<td>0.36</td>
<td>0.71</td>
<td></td>
</tr>
<tr>
<td>6.77 (mean)</td>
<td>8.29 (−1 SD)</td>
<td>0.41</td>
<td>0.02</td>
</tr>
<tr>
<td>8.69 (+1 SD)</td>
<td>0.39</td>
<td>0.31</td>
<td></td>
</tr>
</tbody>
</table>
TABLE 5
Repeated Measures Two-Level Hierarchical Linear Model of Productivity with Individual Random Effects using Restricted Maximum Likelihood (Two-Tail \( p \)-values for Constants and One-Tail \( p \)-values for Explanatory Variables in Parentheses)

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>H4 (All Data)</th>
<th>H4 (SS1=FS, n=58)</th>
<th>H4 (SS1=PFP, n=57)</th>
<th>H4 (SS2=FS, n=55)</th>
<th>H4 (SS2=PFP, n=60)</th>
<th>H5 (All Data)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>9.17</td>
<td>8.25</td>
<td>10.11</td>
<td>7.71</td>
<td>10.51</td>
<td>9.17</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Round</td>
<td>0.26</td>
<td>0.46</td>
<td>0.06</td>
<td>0.40</td>
<td>0.14</td>
<td>0.26</td>
</tr>
<tr>
<td>(Round 3=0)</td>
<td>(0.006)</td>
<td>(0.001)</td>
<td>(0.342)</td>
<td>(0.007)</td>
<td>(0.154)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>Treatment</td>
<td>0.86</td>
<td>0.53</td>
<td>1.20</td>
<td>0.63</td>
<td>1.075</td>
<td>0.86</td>
</tr>
<tr>
<td>(PFP=1, FS=0)</td>
<td>(0.000)</td>
<td>(0.050)</td>
<td>(0.001)</td>
<td>(0.039)</td>
<td>(0.001)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Centered Risk Aversion</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>-0.21</td>
</tr>
<tr>
<td>interacted with Treatment</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>(0.020)</td>
</tr>
</tbody>
</table>
FIGURE 1
Interaction of RA and Productivity on Probability of Choosing PFP

Panel A: Initial Self-Selection

Panel B: Final Self-Selection
FIGURE 2
Decomposing Sorting and Incentive Effects of Pay-for-Performance
The left half of the figure deals with rounds 1 and 2, labeled as Self-Selection 1 (SS1). The right half of the figure deals with rounds 7 and 8, labeled as Self-Selection 2 (SS2). Each number represents the average level of productivity for those working under a particular compensation scheme in the indicated rounds, given the compensation scheme selected. P_p represents those working under the PFP scheme who also chose the PFP scheme; F_F represents those working under the FS scheme who also chose the FS scheme; P_F represents those working under the PFP scheme who chose the FS scheme; and F_p represents those working under the FS scheme who chose the PFP scheme. Differences are indicated between the arrows and p-values associated with those differences are presented in parentheses.
BIOGRAPHICAL NOTE FOR C. BRAM CADSBY

C. Bram Cadsby (Ph.D., MIT) is a professor of Economics at the University of Guelph, Guelph, Ontario, Canada. He has published numerous articles in the fields of Experimental Economics and Finance. His current research is focused on social preferences, compensation systems, coordination and equilibrium selection in the context of multiple game-theoretic Nash equilibria, and psychological and economic deterrents to cheating and tax evasion.

BIOGRAPHICAL NOTE FOR FEI SONG

Fei Song (Ph.D., York University) is an assistant professor of Organizational Behavior / Human Resources in the School of Business Management at Ryerson University in Toronto, Canada. Her research interests include trust and reciprocity, individual and group decision-making in strategic interactions, social preferences, compensation systems and productivity.

BIOGRAPHICAL NOTE FOR FRANCIS TAPON

Francis Tapon (Ph.D., Duke University) is a professor of Economics at the University of Guelph, Ontario, Canada. He has published numerous articles in the fields of industrial organization, corporate strategy and investment finance. This is his first article using lab experimental methodology. His current research interests are in the area of organizational economics particularly the efficiency of team work and compensation policy, and stock market anomalies.