Revised August 19, 2005

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 alphabetical ordering of the authors' names does not reflect the relative contribution of each author.

**Corresponding Author

Acknowledgements: We would like to thank Professor Allan Layton and the School of Economics and Finance at the Queensland University of Technology for their warm hospitality and support for this research. We are also grateful to Sara Rynes, the editor of the *Academy of Management Journal*, and three anonymous referees for many important suggestions that helped improve this paper substantially. In addition, we thank Louis Christofides, Rekha Karambayya, John Laurence Miller, Christine Oliver, Miana Plesca, and Mary-Anne Sillamaa as well as participants at the 2004 Economic Science Association meeting and the 2004 Canadian Experimental and Behavioral Economics conference for their helpful comments and suggestions. Cadsby would also like to acknowledge generous funding from the Social Sciences and Humanities Research Council of Canada, grant #410–2001–159.

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

Abstract

Using real-effort laboratory experiments with salient incentives, we examine the impact of pay-for-performance (PFP) versus fixed-salary (FS) compensation on productivity. PFP achieved significantly higher 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. Risk preferences were also important. Less risk-averse individuals were both more likely to select PFP and also more responsive to PFP incentives.

In a recently published review of the extensive literature on compensation, Gerhart and Rynes (2003) noted that compensation costs comprise, on average, 65 to 70% of total production costs in the U.S. economy (U.S. Bureau of Labor Statistics, 2001) and are similarly substantial elsewhere (European Parliament, 1999). Not surprisingly then, the relationship between pay and both employee and firm performance has long 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, 2000; 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 also to motivate those employees to maximize their output. An important early literature focused on the latter objective, and noted the principal-agent problem that arises when the objectives of employees are not perfectly aligned with those of firms (e.g., Coase, 1937; Jensen & Meckling, 1976; Lawler, 1971; Williamson, 1975). Such problems can arise, for example, when employees are more concerned with their remuneration and/or maintaining a comfortable level of effort 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 the principal-agent problem because it aligns the preferences of firms and employees (e.g. Jensen & Murphy, 1990).

More recently, some organizational theorists and labor economists have begun to focus on the former objective, arguing that PFP can also act as a sorting device to identify and attract the most capable employees (Choping & Schulman, 1997; Gerhart & Milkovich, 1992; Gerhart & Rynes, 2003; Jensen, 2003; Lazear, 1986, 1999, 2000; Zenger, 1992). For example, Michael Jensen (2003) observed that the shape of a firm's compensation scheme, i.e. whether people are paid a fixed salary or on the basis of performance, leads job applicants to self-select into a preferred compensation scheme based on their productivity. In particular, he argues that more

productive workers will choose the PFP scheme, while less productive workers will opt for the fixed-salary (FS hereafter) scheme. The notion of sorting based on compensation scheme is consistent with the idea of person-organization fit, a widely held idea that job applicants make job-choice decisions based on their perceptions of the match between their own dispositions and organizational characteristics (Bretz & Judge, 1994; Schneider, 1987). Rynes (1987: 190) emphasized that "compensation systems are capable of attracting (or repelling) the right kinds of people because they communicate so much about an organization's philosophy, values, and practice." If this is true, PFP may help a firm both to attract and to retain productive employees.

The incentive property of PFP has been extensively examined theoretically (Baker, Jensen, & Murphy, 1988; Lazear, 2000; O'Dell & McAdams, 1987) and empirically in both the laboratory (e.g., Fessler, 2003; Kuhn & Yockey, 2003; see also Bonner, Hastie, Sprinkle, & Young, 2000; Camerer & Hogarth, 1999, for reviews of the literature) and the field (e.g., Abowd, 1990; Gerhart & Milkovich, 1990; Stajkovic & Luthans, 2001). Indeed, several meta-analysis studies of behavioral responses to pay have concluded that individual pay incentives significantly improve productivity (Guzzo, Jette, & Katzell, 1985; Jenkins, Mitra, Gupta, & Shaw, 1998; Locke, Feren, McCaleb, Shaw, & Denny, 1980; Stajkovic & Luthans, 1997).

In contrast, empirical support for the sorting property is less abundant. Evidence based on self-report survey data suggests that high performers are most likely to leave an organization if performance does not lead to sufficient financial rewards (Trevor, Gerhart, & Boudreau, 1997), and that low performers are more likely to stay with an organization when relationships between pay and performance are weak (Harrison, Virick, & William, 1996). An exceptionally interesting and important study by Lazear (1999, 2000) 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. However, the nature of Lazear's data makes it difficult to determine whether the observed effects were primarily the result of an increase in payment level or a change in payment system. 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, as illustrated

in Figure 1 of Lazear (2000), 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 had to result in a wage level that was greater than or equal to the previous hourly wage for all of Safelite's workers. In fact, Lazear (2000) reported that pay for a given worker rose by an average of 10.6% under the new regime.

Higher pay levels may have effects similar to PFP. Cable and Judge (1994: 320) argued that it is "generally accepted that … high pay levels will attract greater quantities of higher quality applicants." Yellen (1984), Weiss (1980), and Greenwald (1979) have all developed theories in the efficiency wage tradition to support this view. Higher pay levels may also lead a given employee to produce more (Akerlof & Yellen, 1986; see also Gerhart & Rynes, 2003, for a critical review of the theoretical and empirical literature on this subject). Thus, it is possible that both the sorting and incentive effects observed in Lazear's field study could have resulted from higher real wages rather than from the move to piece rates per se.

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. While this methodology has its inherent weaknesses, it also has significant strengths. An important advantage of the laboratory methodology is that it enables us to examine the hypothesized sorting and incentive effects of PFP in isolation from those of higher pay. In our setup there was no minimum pay guarantee associated with the PFP system. Thus, in contrast to Lazear (1999, 2000) there was no presumption on the part of any participant that he/she would definitely earn at least as much under this scheme as under the FS scheme. In fact, the average cost per unit of output and the average pay per participant both turned out to be *lower* under the assigned PFP than under the assigned FS treatment ($.20 versus $.23 and $2.11 versus $2.20 respectively). Thus, the increases in productivity due to both sorting and incentives observed in our study can be attributed to the difference in compensation schemes rather than differences in pay level. This is an important advance on Lazear's pioneering study. It also has important practical implications, suggesting that it is not necessary for a movement towards PFP

to be coupled with higher pay for a firm to experience the hypothesized sorting and incentive effects.

The laboratory setup also enables us to investigate how individual risk preferences might influence both the sorting and incentive effects of PFP. Cable and Judge (1994), using a self-report questionnaire-based methodology, 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. Corroboration of this finding in a behavioral setting represents significant additional empirical support for the risk-aversion hypothesis. In addition, we are able to say more about the important role played by risk aversion. In particular, 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.

Finally, Lazear (2000: 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 confounding effects.

THEORETICAL SYNTHESIS

The Sorting Effect of PFP: "Rational" Expectations versus the "Lake Wobegone Effect"

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, Ash, and Dreher (1989) have suggested that

the job attraction literature 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".

In order to decide whether to choose a PFP or FS firm, a job applicant must formulate forecasts or expectations about his/her own future job-specific productivity. 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 and unbiased manner. Many economic models assume that people have "rational" expectations, which means that people make the best possible statistical forecast based on all the information at their disposal (Muth, 1961; Samuelson, 1965; Sargent, 1973). Employees who choose a compensation scheme formulate their expectations rationally, but may make mistakes. However, their mistakes should not be biased towards either over- or under-confidence in their own abilities. Thus, in aggregate, their erroneous choices should not be biased towards either the PFP or FS scheme. In addition, we might also expect more experienced employees with more information about their own abilities and the nuances of the two compensation schemes to choose more accurately than their less experienced counterparts.

However, an increasing number of studies have reported that in many contexts people are over-confident about their own abilities (e.g., Dunning, Meyerowitz, & Holzberg, 1989; Taylor, 1989; Taylor & Brown, 1988). This literature suggests that such optimism might be particularly prevalent among people who place a high value on a particular characteristic such as their own productivity at work, but have little concrete evidence on how good they actually are (Alicke, Klotz, Breitenbecher, Yurak, & Vredenburg, 1995; Epley & Dunning, 2000; Miller & Ratner,

1998). To the extent that people are overly self-confident, the so-called "Lake Wobegone effect" (Meyer, 1975), they may frequently self-select into PFP systems when they would have done better under a FS scheme, leading to a self-selection bias, contrary to the rational-expectations hypothesis. Camerer and Lovallo (1999) explored whether such an over-confidence bias could plausibly and predictably influence economic behavior in one particular setting: entry into competitive games or markets. In particular, they examined and tested the hypothesis that new businesses frequently fail as a result of entrepreneurs/managers acting on the basis of excessive optimism about their own relative skills. Camerer and Lovallo's findings are consistent with the prediction that over-confidence leads to excessive business entry. Similar studies have also demonstrated that over-confidence about one's own abilities can lead to sub-optimal decisions even in the face of significant consequences, monetary or otherwise (e.g., Heath & Tversky, 1990). Of course, it is also conceivable that to the extent that some people are lacking self-confidence and thus overly pessimistic, they may self-select into the FS scheme when they would have done better under the PFP scheme. However, the literature on the over-confidence effect suggests that such people would be in the minority, and that any observed bias within a population of employees would, at least in the absence of experience, be in the direction of incorrectly choosing PFP over FS. In accordance with an untested conjecture of Camerer and Lovallo (1999), we predict that such a bias would dissipate with experience working at the task. The following hypotheses summarize the discussion above.

H1: For the initial self-selection, due to the over-confidence bias, participants will be optimistically biased toward the PFP scheme.

H2: For the final self-selection, the over-confidence bias will dissipate.

Determinants of Self-Selection into PFP versus FS Schemes

The effectiveness of self-selection as a sorting device depends critically on the hypothesized close relationship discussed above between individual productivity level and self-selection into either a FS or a PFP scheme (Lazear, 1986). However, 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. Thus, standard economic theory would predict that in addition to the expected earnings from the different compensation schemes, the level of risk aversion should also affect self-selection. This prediction is consistent with the central argument of the classic agency model: contracts between principals and agents involve a tradeoff between incentives and insurance (see Gibbons, 1998, for an excellent review of agency theory).

H3: Controlling for individual attitudes toward risk, more productive individuals are more likely to select the PFP compensation scheme.

H4: Controlling for individual productivity differences, more risk-averse individuals are less likely to select the PFP compensation scheme.

Studies of individual attitudes toward financial risk have consistently found that most people are risk-averse (e.g. Holt & Laury, 2002). For such people, the uncertain nature of PFP may tempt them to opt for the FS scheme even though they may be quite confident that they are likely to produce enough to earn more under the former scheme. Since far fewer people are generally risk-loving, the reverse will rarely be true, i.e. we would rarely find low-productivity individuals opting for PFP because they are attracted by the uncertainty associated with it. 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. In contrast, they are likely to play 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:

H5: The level of productivity will moderate the relationship between risk aversion and self-selection. Specifically, this relationship will be stronger for higher-productivity participants than for those with lower productivity.

Besides individual productivity levels and attitudes toward risk, we also examine whether gender affects self-selection. Outside the lab, females earn less on average than males (e.g., Sorensen, 1989). Some researchers have suggested that this may be due to females possessing less salary information than males (e.g., Gehart & Rynes, 1991), while others have argued that it may stem from females setting lower reservation wages for themselves (e.g., Stevens, Bavetta, & Gist, 1993). Subich, Barrett, Doverspike, and Alexander (1989: 98–99) argued that the male tendency toward "risk taking correlates with salary expectations and attitudes regarding salary compromise". Most relevant to our research, Chauvin and Ash (1994) found that, after controlling for job and other factors, most of the unexplained pay advantage male business graduates had over their female counterparts was attributable to men receiving a higher proportion of performance-based pay than women. Chauvin and Ash (1994) suggested that a likely explanation for this observed difference was that women tend to prefer less risk than do men, and thus choose jobs with less variable compensation. However, the evidence on the relationship between gender and risk preference has been mixed and inconclusive (e.g., Dufwenberg, Gneezy, & Rustichini, 2004; Eckel & Grossman, forthcoming; Shupp & Williams, 2003; Subich et al., 1989). It is thus important to examine whether there is a relationship between gender and self-selection of compensation schemes mediated by risk preferences as suggested by Chauvin and Ash (1994).

In addition, Hollenbeck, Ilgen, Ostroff, and Vancouver (1987) utilized expectancy theory to partially explain male-female wage differentials through differences in occupational choice. Although they showed that differing perceptions influence job choice and hence salary differentials, they were unable to determine whether these perceptions were accurate or not. Is there any such relationship, apart from possible gender differences in risk preferences, between gender and choice of payment scheme in our laboratory setting? If so, is that relationship based

on accurate or inaccurate perceptions about the linkage between one's effort and one's output? A recent review of the empirical literature on sex differences in cognitive abilities suggests that, on average, females may have better verbal abilities than males (Halpern, 2000). Thus, females may have been better at making words from anagrams, the production task in our experiment. If expectancies accurately reflect reality, gender should affect self-selection only through the mediation of such potential productivity differences and/or differences in risk aversion. If, in contrast, gender-based expectancies are inaccurate, but nonetheless affect self-selection, gender should have additional explanatory power, after controlling for both productivity and risk aversion. This discussion is summarized in H6.

H6: The relationship between gender and self-selection is mediated by risk attitudes and/or productivity, and has no additional impact on self-selection.

The ability to solve anagrams would likely be strongly influenced by verbal ability. It is thus interesting to consider whether initial self-selection decisions might be informed by the years of experience our participants have had with language-based tasks and their success therein. The idea here is that general knowledge of their own verbal ability might help participants make accurate forecasts of their level of productivity at the anagram task. This in turn should help them to self-select into the most lucrative compensation scheme even before they have any experience with the actual task.[1] Unfortunately, we were unable to access the GPAs or standardized test scores of our participants, which might have been good indicators of ability. However, we were able to identify whether or not a participant was a native English speaker and use this as an admittedly rough proxy for verbal ability. Specifically, we predict that native English speakers would be more likely to self-select into PFP, and this relationship would be mediated by productivity. This hypothesis is most interesting at the initial self-selection, where it indicates whether participants could translate a general notion of their verbal ability into

[1] We thank an anonymous referee for suggesting this line of reasoning.

an accurate enough prediction of their task-specific ability to choose the best payment scheme for themselves prior to experiencing the task. Hypothesis 7 summarizes the preceding discussion.

H7: The relationship between native language and self-selection is mediated by productivity.

The Incentive Effect of PFP and its Differential Effect According to Risk-Preference

Expectancy theory (Vroom, 1964) identifies three factors, the effort-performance expectancy, the performance-outcome expectancy, and valence, all of which play an interactive role in motivation. Under a pure FS system, there is no link between performance and outcome. In contrast, 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. 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).

H8: Overall, controlling for individual differences in average productivity and self-selection, 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 risk-averse individuals. Since this weakens one of the three important links in expectancy theory,

one might expect that the incentive effects of PFP might be less pronounced for more risk-averse participants.

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, for a critical review of this literature). Most of the empirical literature suggests that stress is negatively related to performance (see the references in Muse et al, 2003).[2] If this is the case, the higher level of stress experienced by a more risk-averse participant assigned to the PFP scheme along with the lower level of reward valence found by Cable and Judge (1994) might together reduce the strength of the hypothesized incentive effect of PFP.

H9: The effectiveness of PFP at improving productivity is inversely related to individual levels of risk aversion.

METHODS

Participants

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. 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 undergraduate students, most but not all were majors in economics or other subjects taught within the business faculty.

Experimental Design and Procedure

A widely used anagram word-creation game (Locke & Latham, 1990; 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

[2] 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.

anagram games using prescribed sets of seven letters. This task is particularly appropriate for investigating whether people sort into different compensation schemes according to their productivity because it is a task where productivity depends on both ability and effort. The experiment utilized two different compensation schemes, representing PFP and FS. The first compensation scheme paid $0.20 per correct word created. The second scheme paid a fixed salary of $2.20, independent of performance. Since a person creating 11 words under the productivity-based scheme would earn $2.20, 11 words was the break-even point between the two schemes. Figure 1 illustrates the two compensation schemes.

Insert Figure 1 about here.

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, 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, for rounds 3 and 5, all participants were paid according to the FS scheme and for rounds 4 and 6, all participants were paid according to the PFP scheme. 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 by the experimenters 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, in which they responded to a number of demographic questions such as age, gender, and native language. Besides collective demographic data, another primary purpose of the questionnaire was 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 words they created according to the compensation schemes 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 in order to avoid biasing the behavioral decisions by priming participants to pay undue attention to risk. 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 self-selection into 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. On average, participants earned $21.20 AUD (about $15.11 US) for a session lasting approximately one hour and 15 minutes. This was significantly higher than the average wage a student could earn working typical jobs on or off campus.

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 and the productivity statistics under those treatment conditions. Table 2 also reports the average productivity for each anagram in Vance and Colella (1990) and in a pre-test we ran prior to the current study. Vance and Colella (1990) used psychology undergraduates from Ohio State University. In their study, participants were given performance targets, but were not paid on the basis of performance. Our pre-test employed 99 business undergraduates at a large Canadian business school, 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. Table 3 reports means, standard deviations, and correlations of the variables.

Compensation Schemes as Sorting Mechanisms: Hypotheses 1 and 2

For rounds 1 and 2, 57 participants (50%) chose the PFP scheme, while 58 selected the FS scheme. For rounds 7 and 8, 60 (52%) chose the PFP scheme, while 55 chose the FS scheme. We first discuss the extent to which participants chose the payment scheme that maximized their earnings. In Table 4, we report the percentage of participants who gained, broke even and lost under their selected compensation scheme compared to the alternative scheme. Recall that 11 words was the break-even point. As reported in Panel A, for those who chose the PFP scheme, 52.6% produced an average of fewer than 11 words in rounds 1 and 2, thus earning less than they would have under the FS scheme. This suggests that over half of those selecting the PFP scheme prior to playing the game were overly optimistic, consistent with the over-confidence effect.[3] In contrast, as reported in Panel B, only 22.4% of those who chose the FS scheme produced an average of more than 11 words, thus earning less than they would have if they had produced the same number under the PFP scheme. This result may have occurred because pessimism about one's own ability was less prevalent than optimism as we predict in Hypothesis 1, or alternatively because some of those who were able to produce more than 11 words with effort were not motivated to exert such effort under the FS scheme. To distinguish between these two possibilities, we also calculated the productivity of those who initially selected the FS scheme (in rounds 1 and 2) but were compelled to produce under the PFP scheme in rounds 4 and 6. As reported in Panel C, increased effort under the PFP scheme made only a small difference. Only two more participants, resulting in a total of 25.9%, made more money under the PFP scheme than they had made under the FS scheme in rounds 1 and 2. It would thus seem that as predicted in Hypothesis 1, a bias towards overly optimistic self-confidence was primarily responsible for the higher percentage of participants who made the wrong choice, financially speaking, at the first self-selection.

[3] An anonymous referee suggested that rather than reflecting overconfidence, this result might have arisen because of low levels of loss aversion at small laboratory stakes. Both our behavioral experiment and our risk-elicitation instrument were framed in the gain rather than in the loss domain. Only 3.5% of our participants appeared to be risk loving in this domain. Only risk-loving participants should choose the PFP scheme when they expect to produce fewer than 11 words. However, we cannot rule out the possibility that some people nonetheless perceived the game in the loss domain and were more risk-loving in that domain than our risk-elicitation technique suggested.

Insert Table 4 about here.

By the beginning of round 7, participants had gained experience with the game. Although they had not received explicit feedback on their performance from the experimenters, they knew how many words they had submitted for each anagram and presumably had a reasonable idea of how many were likely to be correct. Thus, we might expect that they would have more realistic expectations of their subsequent performance. This indeed appeared to be the case for those who chose the PFP scheme for rounds 7 and 8, only 28.3% of whom made less money than they would have done under the FS scheme compared to 52.6% for rounds 1 and 2. Thus, the excessive unrealistic optimism exhibited initially seemed to dissipate with experience, consistent with Hypothesis 2. This experience effect was much less dramatic for those choosing the FS scheme for rounds 7 and 8, 18.2% of whom made less money than they would have done had they produced the same number of words under the PFP scheme, compared to 22.4% for rounds 1 and 2. Applying the rounds 7 and 8 selection criteria to rounds 4 and 6 when everyone played under the PFP scheme, the corresponding number was 14.5%.

To test Hypotheses 1 and 2 more formally, we investigated whether the probability of at least breaking even for those optimistically choosing the PFP scheme was significantly greater than the probability of at least breaking even for those pessimistically choosing the FS scheme. In the former case, this implies producing 11 words or more, while in the latter case it implies producing 11 words or less. We examined this question by running the following logistic regression separately for the initial and final self-selection cases:

$$\ln [p_e/(1-p_e)] = \beta_0 + \beta_1 \cdot SS_i \qquad (1)$$

where p_e is the probability of earning at least as much under the self-selected compensation scheme as under the alternative scheme and SS_i is a dummy variable equal to 1 if the PFP scheme is selected and 0 if the FS scheme is selected. The i subscript refers to the initial and final selections. The results are reported in Table 5.

Insert Table 5 about here.

Initially, those selecting the PFP scheme exhibited a significantly lower probability of at least breaking even than those selecting the FS scheme ($p = .000$). However, this significance disappeared with the final self-selection ($p = .201$). Further, we ran a test to determine whether the overconfidence effect becomes significantly smaller from the initial to the final self-selection. The *p*-value was a marginal 0.10. Thus, the predicted bias towards more rather than less self-confidence was present initially, consistent with previous evidence in favor of an overconfidence effect. However, there is some though not conclusive evidence that it dissipated with experience, as suggested by Camerer and Lovallo's (1999) conjecture. Thus, Hypotheses 1 and 2 were both corroborated, though support for hypothesis 2 was only marginal.

To further examine how participants improve their self-selections with experience, we next compared the average productivity in rounds 3, 4, 5, and 6 for those who maintained, with those who altered, their choice of compensation scheme. The tests indicated a significant difference between those who maintained a choice of the FS scheme through both self-selection decisions and those who changed their choice from FS to PFP ($M_{(FS, FS)} = 8.47$ vs. $M_{(FS, PFP)} = 10.71$, $p = .001$). Similarly, there was a significant difference between those who maintained a choice of the PFP scheme and those who changed their choice from PFP to FS ($M_{(PFP, PFP)} = 11.51$ vs. $M_{(PFP, FS)} = 8.97$, $p = .001$). Thus, despite not receiving explicit feedback on the number of correct words they had produced during the game, participants seemed to have had a good idea of how well they were doing and to have responded accordingly.

Risk, Productivity and other Factors Influencing Self-Selection: Hypotheses 3 to 6

In order to assess the role of attitude toward risk together with productivity in the selection of compensation scheme, we elicited risk preferences using an instrument developed by Holt and Laury (2002) as described in the discussion of the experimental design above. Our participants were highly risk-averse with 93% exhibiting some degree of risk aversion[4]. Notice that this was consistent with the reasoning used in the development of Hypothesis 5, which relied

[4] 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.

on the previously reported findings that most people are risk-averse in the financial domain. Of the remaining participants, 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.[5] Having established the risk attitudes measure, we then 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 regression with a two-by-two factorial design as follows:

$$\ln [p_{ssi}/(1-p_{ssi})] = \beta_0 + \beta_1 \cdot \text{Productivity} + \beta_2 \cdot \text{Risk aversion} + \beta_3 \cdot \text{Interaction} \qquad (2)$$

where p_{ssi} = the probability of self-selecting the PFP scheme and i 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 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 Hypothesis 3 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. The null hypothesis was contrasted with the alternative suggested by Hypothesis 4 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 Hypothesis 5 that higher levels of productivity would be associated with a more 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

[5] 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.

significant impact[6]. Risk aversion was conventionally centered at its mean so that the coefficient on productivity was estimated at the mean level of risk aversion.

Insert Table 6 about here.

Table 6 reports the results. 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 was not significant (p = .123). Thus, despite the excessive amounts of self-confidence noted above, those who subsequently performed better did have a significantly higher probability of self-selecting themselves into PFP, controlling for their differing attitudes toward risk. At the same time, controlling for productivity, those who exhibited a higher degree of risk aversion in their lottery choices were more likely to choose the risk-free FS compensation scheme. The lack of significance of the interaction effect suggests that the unobservable expected earnings of most participants were higher under PFP than under FS so that risk aversion mattered significantly and in a quantitatively similar manner even for those who expected to perform relatively poorly. This is consistent with the excessive self-confidence documented earlier. 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. The coefficients were larger and the p-values smaller than in the initial selection, suggesting that participants learned from experience. The null hypothesis for the interaction was also rejected in the direction of the predicted negative effect (p = .048). This was consistent with more participants expecting to produce fewer than 11 words than under the initial selection. Risk aversion mattered less for those producing less because many of them did not expect higher earnings under the risky PFP and hence opted for FS regardless of their level of risk aversion.

[6] Cohen, Cohen, West and Aiken (2003: 261–282, especially p. 281) contains an excellent discussion on the importance of centering the covariate. If productivity were not centered, the coefficients on the other explanatory variables would indicate its effect at a fictional productivity level of zero that is nonsensical and hence never utilized in the experiments.

Clearly, self-selection into the two compensation schemes depended on both risk aversion and productivity in the manner predicted by Hypotheses 3, 4, and 5.[7]

Next, we examined whether gender affected self-selection into a compensation scheme, and whether the relationship was mediated by risk aversion and/or productivity as predicted in H6. Following Baron and Kenny (1986), we began by examining whether or not gender was related to one of the hypothesized mediator variables, namely risk aversion. The results revealed that there was no gender effect on risk attitudes. In fact, the mean levels of risk aversion were almost identical for males and females ($M_{Female} = 6.92$ vs. $M_{Male} = 6.86$). Since gender was not related to risk aversion, no further test was necessary. The prediction that risk aversion would mediate the relationship between gender and self-selection, was not supported by the data. However, gender was significantly related to productivity ($M_{Female} = 10.64$ vs. $M_{Male} = 9.60$, $p = .02$, one-tail test). Further analysis revealed that, controlling for risk aversion, both gender ($p = .05$, one-tail test) and productivity ($p = .001$, one-tail test) significantly predicted for initial self-selection in the absence of the other variable. When both gender and productivity were utilized in the logistic regression for initial self-selection, the results showed that the coefficient for gender became insignificant, while the coefficient for productivity remained significant ($p = .02$, one-tail test). This pattern supported the hypothesized mediated relationship between gender and self-selection through productivity, suggesting that gender differences in expectancies concerning the relationship between effort and productivity reflected real gender differences in task-specific cognitive capital. Surprisingly, gender did not affect the final self-

[7] 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 referee 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 either the initial or 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).

selection either through the mediation of productivity or in any other manner.

In regard to native language, 85 of our participants were native speakers of English, while 30 were not. We constructed a dummy variable (0= English, 1= Non-English) for first language. The analysis yielded the following results. As expected, the ability to solve anagrams was strongly influenced by first language (p = .002, one-tail test). However, surprisingly, first language was not related to the initial self-selection. Thus, although first language was strongly related to task-specific ability, the existence of this relationship did not seem to inform the initial self-selection made by participants. At the final self-selection, the initial regression analysis revealed that first language significantly influenced the final self-selection (p = .015, one-tail test). When first language was replaced by productivity, the coefficient on productivity was likewise significant (p = .000, one-tail test). Using both first language and productivity together in the logistic regression, we found that the relationship between first language and final selection became insignificant (p = .212) while the coefficient on productivity remained significant (p = .000). This pattern was consistent with H7, which predicted that productivity mediated the relationship between first language and self-selection. In sum, the results suggest that participants were unable to use the information contained in the knowledge of their first language to help them make a wise initial self-selection. However, once they were placed in the work environment and obtained some task-specific experience, they were in a better position to make an informed decision.

The Incentive Effect of PFP and Its Differential Effect According to Risk Preference: Hypotheses 8 and 9

H8 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 order to isolate this effect from the effect of self-selection. Recall that everyone was compensated in rounds 3 and 5 using the FS scheme, and in rounds 4 and 6 using the PFP scheme. We were thus able to perform a within-person pairwise comparison of productivity under the two schemes. The results indicated

that, as predicted, participants performed significantly better under PFP than under FS (M_{PFP} = 10.56 vs. M_{FS}= 9.43, $p < .001$). Note that this was the case even though the mean levels of productivity were slightly higher for the anagrams used in the FS compensation scheme than for those used in the PFP scheme in both Vance and Colella's (1990) study (M_{PFP} = 9.24 vs. M_{FS}= 9.37) and our pre-test (M_{PFP} = 12.05 vs. M_{FS}= 12.12). There was however another possible confounding factor because the FS rounds (rounds 3 and 5) were run before the PFP rounds (rounds 4 and 6). Participants may have improved with practice. To remove this confound, we compared productivity in the earlier PFP round 4 with that in the later FS round 5. Although the difference was lower and the *p*-value higher, PFP still resulted in significantly higher productivity than FS (M_{PFP} = 10.46 vs. M_{FS}= 9.86, $p = .042$). Therefore, the data supported H8.

We tested H9, which predicted that the effectiveness of PFP at improving productivity was inversely related to the level of risk aversion, by regressing the difference in average productivity in the imposed PFP rounds (rounds 4 and 6) and that in the imposed FS rounds (rounds 3 and 5) on risk aversion. We found a significant inverse relationship between productivity improvement under PFP and the level of risk aversion ($p = .017$), corroborating H9. Thus, the opportunity to earn more money through better performance was a less effective motivator for more risk-averse individuals. In fact, for 25.2% of our participants, PFP actually caused a decline in productivity.

We have now established that people who self-selected into the PFP scheme performed better than those who self-selected into the FS scheme 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% ($p = .013$) in the initial self-selection period and by 38.09% ($p < .001$) in the final self-selection period.

Insert Figure 2 about here.

To get a sense of how much of the PFP "firm's" increased productivity was due to the sorting effect (i.e. its ability to attract higher-quality employees) versus the incentive effect of the

payment scheme, we decomposed the higher productivity of the PFP "firm's" employees relative to the FS "firm's" employees into these two components. This decomposition is illustrated in Figure 2. Those selecting and hence working under PFP were significantly more productive than those selecting and hence working under the FS scheme (For rounds 1 and 2, $P_P = 10.87$ vs. $F_F = 9.33, p = .013$). While it is not possible to decompose this difference, it is possible to examine the productivity of those initially selecting each scheme during rounds 3 and 5, in which everyone was assigned to FS, versus rounds 4 and 6, in which everyone was assigned to PFP. For those selecting the PFP scheme initially, average productivity in rounds 4 and 6 under PFP was 11.43 words. For those selecting the FS scheme initially, average productivity in rounds 3 and 5 under FS was 8.71 words. The difference of 2.72 words was significant ($p < .001$) just as the comparable difference was significant in rounds 1 and 2. This difference can be decomposed into one component due to sorting differentials between persons who had made different self-selections measured when they were assigned to a common scheme and another due to the incentive effects of working under the different schemes. The sorting differential under the assigned PFP was 1.73 words (for PFP rounds 4 and 6, $P_P = 11.43$ vs. $L_F = 9.70, p = .001$), while the incentive effect for those selecting the FS scheme was .99 words (for PFP rounds 4 and 6 versus FS rounds 3 and 5, $P_F = 9.70$ vs. $F_F = 8.71, p < .001$). The incentive effect for those selecting the PFP scheme was 1.26 words (for PFP rounds 4 and 6 versus FS rounds 3 and 5, $L_L = 11.43$ vs. $F_L = 10.17, p < .001$), while the sorting differential under the assigned FS scheme was 1.46 words (for FS rounds 3 and 5, $F_P = 10.17$ vs. $F_F = 8.71, p = .006$). Thus, even at the initial self-selection, both sorting and incentive differences between the two schemes contributed significantly to productivity differences, and interestingly, sorting differences were quantitatively the more important of the two.

A comparison between the mean productivity of those selecting the PFP scheme in rounds 7 and 8 and those selecting it in rounds 1 and 2 revealed a significant increase of 1.89 words (12.76 vs. 10.87, $p = .002$). The slight decrease in the average productivity of those selecting the FS scheme was not significant (9.24 vs. 9.33, $p = .533$). However, the differential

between those selecting and working under PFP and those selecting and working under FS increased to 3.52 words, which continued to be significant (For rounds 7 and 8, $P_P = 12.76$ vs. $F_F = 9.24, p < .001$). When the data in the middle rounds were reorganized based on the final self-selection decisions, there was little change in the incentive differences. However, the sorting differentials increased, 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, although both sorting and incentive differences between the two schemes continued to contribute significantly to productivity differences, the former were now more than two and a half times as important as the latter in their contributions to the observed differences in productivity between the two compensation schemes.

CONCLUSION AND DISCUSSION

Key Findings and Discussion

Rynes, Gerhart, and Parks (2005), have called 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, 2000; 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 this paper, focusing on the role of performance incentives, operationalized as the contrast between PFP versus FS compensation schemes, we examined the interrelations between levels of risk aversion, employee quality, job-choice decisions, and subsequent performance levels. PFP has two advantages for firms relative to FS pay: first, they attract higher-quality employees, and second, they motivate employees to put forward more effort. The importance of these sorting and incentive effects has been proposed and examined before, but the separate presence and importance of the two effects in isolation from higher pay levels have not to our knowledge been disentangled and directly tested. Key findings are summarized below.

First, our data showed strong support for the "Lake Wobegone Effect" (Meyer, 1975), the notion that more people are overly optimistic than are overly pessimistic about their abilities, and that this creates an initial bias toward many less capable but over-confident participants selecting the PFP scheme. However, by the final self-selection, participants, who were then experienced at the task, re-sorted themselves, apparently correcting the bias, though it must be cautioned that a marginal *p*-value of 10% was associated with the test of whether this bias had significantly dissipated. Thus, experience appears to have attenuated the initial over-confidence, making the sorting property more efficient at selecting the higher-quality participants into the PFP compensation scheme.

Second, we found that individual productivity level and risk attitudes both significantly affected the selection of compensation scheme: in particular, the more skilled one was and/or the less risk-averse one was, the more likely one would select the PFP scheme. Furthermore, we demonstrated that risk aversion interacts with individual productivity in determining the choice of payment system. In particular, risk aversion affects decisions of high-productivity individuals significantly more than it affects the decisions of low-productivity individuals once they have gained some experience at the anagram task used in our study.

Methodologically, Cable and Judge (1994) collected their data using survey questionnaires and commented on the self-report bias that may have influenced their results. 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. Such bias can cause a gap between what people say and do with respect to pay since people often hold the view that money is a "less noble source of motivation," and may be reluctant or unable to report accurately about its impact on their behavior. We employed a behavioral experiment. Both our risk-aversion measures and our payment-scheme selection measures are based on decisions with salient financial consequences. The behavioral methodology also has significant limitations. It is not in any way superior to that used by Cable and Judge (1994). However, given that all methodologies have

their strengths and weaknesses, we believe that our finding that risk aversion is an important determinant of sorting behavior using a methodology different from Cable and Judge (1994) represents significant additional empirical support for the risk-aversion hypothesis.

Third, disentangling the incentive effect from the sorting effect by conducting a within-person analysis to examine productivity differences under the two compensation schemes, we found strong support for the incentive effect of PFP. Moreover, we showed that the effectiveness of PFP at improving productivity is inversely related to individual levels of risk aversion. 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 contrasts with the results of Lazear's (1999, 2000) field-study results, where sorting and incentive effects made roughly equal contributions to PFP productivity gains. This is not surprising. In Lazear's study, all employees were removed from fixed hourly wages and placed into PFP with a guaranteed minimum equal to the previous hourly wage. It was costly for workers to leave and look for fixed hourly wages elsewhere. Besides, there was little incentive to do so because of the guaranteed minimum. In our laboratory context, there was no cost at all in 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 costs of choosing one scheme over another.

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 very useful to treat the same individual with different compensation schemes in order to examine the effect of each compensation scheme on individual behavior. By examining how behavior changed for each individual as he/she was exposed to the two different compensation schemes, we were able to control for the various unidentified factors specific to each individual that may well affect 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 in future work. However, they could not have emerged under the between-person approach. Third, under our within-person design, the productivity data from the middle rounds, in which participants were assigned to identical treatments, 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 should be acknowledged. First, 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? As argued in Sackett and Larson (1990), 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 certainly possible that different results would have been obtained by sampling from other populations. Indeed it is important that behavioral theories in general be tested using samples from a variety of populations.

Second, the money at stake in our experiment was far less than a year's worth of pay for a typical employee. We acknowledge that this may affect risk preferences, with people

exhibiting less risk aversion when less is at stake. However, 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. Thus, we used the Holt-Laury instrument to elicit levels of risk aversion 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 will be less likely to choose variable compensation schemes regardless of the stakes involved. 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., Cameron, 1999; Fehr & Tougareva, 1995; Slonim & Roth, 1998). Of course, it is always possible that our particular context is an exception to these general findings. This warrants further examination.

Third, switching compensation schemes as a design feature of our experiment is quite unlike the real world. However, 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, such differences between a laboratory study like ours and the real world may cause unexpected problems and biases. 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 could do alone.

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, contextual/cultural factors, and time frames. 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 did 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.

Managerial Implications and Directions for Future Research

This paper offers some important insights for managers. However, those insights must be qualified and interpreted with caution. First, we have demonstrated that it is possible for a firm to attract higher-quality applicants by implementing PFP without having to simultaneously increase its level of pay. However, it is important to understand that, in practice, there are factors that could potentially impede the implementation or effectiveness of PFP. For example, PFP schemes can have a destructive effect on intrinsic motivation, self-esteem, teamwork, and creativity (Amabile, 1988; Beer & Katz, 2003; Deci & Ryan, 1985). 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. In addition, increased pay dispersion, resulting from PFP, may trigger negative comparison effects to the extent that they magnify the variance in pay differences among employees (Pfeffer & Langton, 1993; Zenger, 1992). In particular, according to distributive justice theory (Deutsch, 1985), pay variance under PFP schemes may dampen their incentive power, lead to high voluntary turnover, and even lead to company sabotage if employees perceive pay differences as inequitable or unjust. Others have pointed out further implementation difficulties such as the availability, accuracy, and cost of performance measures (Holmstrom & Milgrom, 1991), and the inter-temporal problem of incentive ratcheting (Jensen, 2003). PFP is especially problematic when team-based production makes it difficult, if not impossible, to disentangle individual relative contributions. As argued by Milgrom and Roberts (1988), incentive systems can encourage counterproductive organizational activities due to a narrow focus on individual outcomes to the exclusion of socially efficient and value-enhancing cooperation with colleagues. 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 motivating high levels of performance through other means as well. Future research should be devoted to exploring the effectiveness of such mechanisms and their interactions with various forms of PFP in different organizational contexts.

Second, we have found that employees who self-select into performance-based compensation schemes may be more productive but also tend to be less risk-averse, and as a result possibly more prone to bending or even breaking rules. This may create enormous problems for their firms, possibly leading all the way to bankruptcy. One example is the bond trading scandal at Salomon Brothers in the early 1990s, which almost led to the firm being shut down by regulators. Jensen (2003) discussed at length how compensation systems with bonuses linked to the attainment of performance thresholds are a virtual guarantee of illegal and/or value-destroying behavior. Examples abound in the pharmaceutical and consumer goods industries.

Thus, organizations using PFP schemes may also require more vigilance on the part of senior management and company directors. The cost of such vigilance must be weighed against the productivity benefits of performance-based compensation.

Third, we have shown that the incentive effects of PFP are neither uniform nor universal. In particular, 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 (2000, p. 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 economic theory 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. These findings represent an important caveat as 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 is very stressful, and that such stress impairs performance. Individual differences are important, and managers must not lose sight of such differences in designing effective compensation systems for their organizations.

REFERENCES

Abowd, J. 1990. Does performance-based managerial compensation affect corporate performance? ***Industrial and Labor Relations Review***, 43: 52–74.

Akerlof, G. A. 1982. Labor contracts as partial gift exchange. ***Quarterly Journal of Economics,*** 97: 543–569.

Akerlof, G., & Yellen, J. 1986. ***Efficiency Wage Models of the Labor Market.*** UK: Cambridge University Press.

Alicke, M. D., Klotz, M. L., Breitenbecher, D., Yurak, T., & Vredenburg, D. 1995. Personal contact, individuation, and the better-than-average effect. ***Journal of Personality and Social Psychology***, 68: 805–825.

Amabile, T. 1988. A model of creativity and innovation. In B. M. Staw, & L. L. Cummings (Eds.), ***Research in Organizational Behavior***, 10. Greenwich, CT: JAI Press.

Baker, G., Jensen, M., & Murphy, K. 1988. Compensation and incentives: Practice versus theory. ***Journal of Finance***, 43: 593–617.

Baron, R. M., & Kenny, D. A. 1986. The moderator-mediator distinction in social psychological research: Conceptual, strategic, and statistical considerations. ***Journal of Personality and Social Psychology***, 51: 1173–1182.

Beer, M., & Katz, N. 2003. Do incentives work? The perceptions of a worldwide sample of senior executives. ***Human Resource Planning,*** 26: 30–44.

Bonner, S. E., Hastie, R., Sprinkle, G. B., & Young, M. 2000. A review of the effects of financial incentives on performance in laboratory tasks: Implications for management accounting. ***Journal of Management Accounting Research***, 12: 19–57.

Bretz, R. D., Ash, R. A., & Dreher, G. F. 1989. Do people make the place? An examination of the attraction-selection-attrition hypothesis. ***Personnel Psychology,*** 42: 561–581.

Bretz, R. D., & Judge, T. A. 1994. The role of human resource systems in job applicant decision processes. ***Journal of Management***, 20: 531–551.

Cable, D. M., & Judge, T. A. 1994. Pay preferences and job search decisions: A person-organization fit perspective. ***Personnel Psychology,*** 47: 317–348.

Camerer, C., & Hogarth, R. M. 1999. The effects of financial incentives in experiments: A review and capital-labor-production. ***Journal of Risk and Uncertainty***, 19: 7–42.

Camerer, C., & Lovallo, D. 1999. Overconfidence and excess entry: An experimental approach. ***American Economic Review***, 89: 307–318.

Cameron, L. 1999. Raising the stakes in the ultimatum game: Experimental Evidence from Indonesia. ***Economic Inquiry,*** 37: 47–59.

Chauvin, K. W., & Ash, R. A. 1994. Gender earnings differentials in total pay, base pay, and contingent pay. ***Industrial and Labor Relations Review***, 47: 634–649

Choping, M. C., & Schulman, C. T. 1997. Performance pay as a screening device. ***Studies in Economic and Finance***, 18: 94–109.

Coase, R. 1937. The Nature of the Firm. ***Economica***, 4: 386–405.

Cohen, J., Cohen, P., West, S., & Aiken, L. 2003. ***Applied Multiple Regression /Correlation Analysis for the Behavioral Sciences*** (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Deci, E. L. 1971. Effects of externally mediated rewards on intrinsic motivation. ***Journal of Personality and Social Psychology***, 18: 105–115.

Deci, E. L, & Ryan, R. M. 1985. ***Intrinsic Motivation and Self-determination in Human Behavior.*** New York: The Free Press.

Deutsch, M. 1985. ***Distributive Justice: A Social Psychological Perspective.*** New Haven, CT: Yale University Press.

Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. 1989. Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving appraisals of ability. ***Journal of Personality and Social Psychology***, 57: 1082–1090.

Dufwenberg, M., Gneezy, U., & Rustichini, A. 2004. Price competition: The role of gender and education. In S. Huck (ed.) ***Advances in Understanding Strategic Behaviour: Game Theory,***

Experiments, and Bounded Rationality: Essays in Honour of Werner Güth: 286–297. Basingstoke: Palgrave

Eckel, C., & Grossman, P. forthcoming. Men, women, and risk aversion. In Plott and Smith (Eds.), ***Handbook of Results in Experimental Economics***. New York: Elsevier Science.

Epley, N., & Dunning, D. 2000. Feeling 'holier than thou': Are self-serving assessments produced by errors in self- or social prediction? ***Journal of Personality and Social Psychology***, 79: 861–875.

European Parliament. 1999. ***Labour Costs and Wage Policy with EMU.*** Luxembourg, Belgium: Directorate–General for Research, Economic Affairs Series, ECON 111 EN.

Fehr, E., & Tougareva, E. 1995. ***Do high stakes remove reciprocal fairness - Evidence from Russia.*** Discussion Paper, Univ. of Zurich.

Fessler, N. J. 2003. Experimental evidence on the links among monetary incentives, task attractiveness, and task performance. ***Journal of Management Accounting Research***, 15: 161–182.

Gerhart, B., & Milkovich, G. T. 1990. Organizational differencces in managerial compensation and financial performance. ***Academy of Management Journal***, 33: 663–691.

Gerhart, B., & Milkovich, G. T. 1992. Employee compensation: Research and practice. In M. D. Dunnette, & L. M. Hough (Eds.), ***Handbook of Industrial and Organizational Psychology*** (2^{nd} ed.), Vol. 3: 481– 570. Palo Alto, CA: Consult. Psychol. Press.

Gerhart, B., & Rynes, S. 1991. Determinants and consequences of salary negotiations by male and female MBA graduates. ***Journal of Applied Psychology***, 76: 256–263.

Gerhart, B., & Rynes, S. 2003. ***Compensation: Theory, Evidence, and Strategic Implications***. Thousand Oaks, CA: Sage.

Gibbons, R. 1998. Incentives in organizations. ***Journal of Economic Perspectives***, 12: 115–132.

Gomez–Mejia, L. R., & Balkin, D. 1989. Effectiveness of individual and aggregate compensation strategies. ***Industrial Relations,*** 28: 431–445.

Gomez–Mejia, L. R., & Welbourne, T. M. 1988. Compensation strategy: An overview and future steps. ***Human Resource Planning***, 11: 173–190.

Greenwald, B. C. N. 1979. ***The Labor Market as a Market for Lemons***. New York: Garland.

Guzzo, R. A., Jette, R. D., Katzell, R. A. 1985. The effects of psychologically based intervention programs on worker productivity: A meta-analysis. ***Personnel Psychology,*** 38: 275–291.

Halpern, D. 2000. ***Sex Differences in Cognitive Abilities.*** New York: Lawrence Erlbaum.

Harrison, D. A., Virick, M., & William, S. 1996. Working without a net: Time, performance, and turnover under maximally contingent rewards. ***Journal of Applied Psychology,*** 81: 331–345.

Heath, C., & Tversky, A. 1990. Preference and belief: Ambiguity and competence in choice under uncertainty. In K. Borcherding, O. I. Larchev, and D. M. Messick (Eds.), ***Contemporary Issues in Decision Making***: 93–121. Amsterdam: Elsevier Science.

Heneman, R. L., Ledford, G. E., & Gresham, M. T. 2000. The changing nature of work and its effects on compensation design and delivery. In S. Rynes & B. Gerhart (Eds.), ***Compensation in Organizations: Current Research and Practice***: 195–240. San Francisco: Jossey–Bass.

Hollenbeck, J. R., Ilgen, D. R., Ostroff, C., & Vancouver, J. B. 1987. Sex differences in occupational choice, pay, and worth: a supply-side approach to understanding the male-female wage gap. **Personnel Psychology** 40: 715–743.

Holmstrom, B., & Milgrom, P. 1991. Multitask principal-agent analyses: Incentive contract, asset ownership, and job design. ***Journal of Law, Economics, and Organization***, 146: 85–105.

Holt, C. & Laury, S. 2002. Risk aversion and incentive effects. ***American Economic Review***, 92: 1644–1655.

Jenkins, D. G. Jr., Mitra, A., Gupta, N., Shaw, J. D. 1998. Are financial incentives related to performance? A meta-analytic review of empirical research. ***Journal of Applied Psychology,*** 83(5): 777–787.

Jensen, M. C. 2003. Paying people to lie: The truth about the budgeting process. ***European Financial Management,*** 9: 379–406.

Jensen, M. C. & Meckling, W. H. 1976. Theory of the firm: Managerial behavior, agency costs and ownership structure. ***Journal of Financial Economics***, 3: 305–360.

Jensen, M. C., & Murphy, K. J. 1990. Performance pay and top-management incentives. ***Journal of Political Economy***, 98: 25–264.

Kennedy, P. 2003. ***A Guide to Econometrics*** (5th ed.). Cambridge, MA: The MIT Press.

Keppel, G. 1991. ***Design and Analysis: A Researcher's Handbook*** (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Kuhn, K. M., & Yockey, M. D. 2003. Variable pay as a risky choices: Determinants of the relative attractiveness of incentive plans. ***Organizational Behavior and Human Decision Processes***, 90: 323–345.

Lawler, E. E., III. 1971. ***Pay and Organizational Effectiveness: A Psychological View.*** New York: McGraw–Hill.

Lawler, E. E., III. 1990. ***Strategic Pay: Aligning Organizational Strategies and Pay Systems***. San Francisco, CA: Jossey–Bass.

Lazear, E. P. 1986. Salaries and piece rates. ***Journal of Business***, 59: 405–432.

Lazear, E. P. 1999. Performance pay and productivity. ***American Economic Review***, 90: 1346–1361.

Lazear, E. P. 2000. The power of incentives. ***American Economic Review***, 91: 410–415.

Locke, E. A., Feren, D. B., McCaleb, V. M., Shaw, K. N., Denny, A. T. 1980. The relative effectiveness of four ways of motivating employee performance. In K. D. Duncan, M. M. Gruenberg, D. Wallis (Eds.), ***Changes in Working Life***: 363–388. New York: Wiley.

Locke, E. A., & Latham, G. P. 1990. ***A theory of goal setting and task performance***. Englewood Cliffs, NJ: Prentice Hall.

Meyer, H. H. 1975. The pay for performance dilemma. ***Organizational Dynamics,*** 3: 39–50.

Milgrom, P., & Roberts, J. 1988. ***Economics, Organization and Management***. Englewood Cliffs, NJ: Prentice Hall.

Milkovich, G., & Newman, J. 2002. ***Compensation*** (7th ed.). New York: Irwin / McGraw-Hill.

Miller, D. T., & Ratner, R. K. 1998. The disparity between the actual and assumed power of self–interest. ***Journal of Personality and Social Psychology***, 74: 53–62.

Muse, L. A., Harris, S. G., & Field, H. S. 2003. Has the inverted-U theory of stress and job performance had a fair test? ***Human Performance***, 16: 349–364.

Muth, J. F. 1961. Rational expectations and the theory of price movements. ***Econometrica***, 29: 315–335.

O'Dell, C., & McAdams, J. 1987. The revolution in employee rewards. ***Management Review,*** 76: 30–34.

Pfeffer, J., & Langton, N. 1993. The effect of wage dispersion on satisfaction, productivity, and working collaboratively: Evidence from college and university faculty. ***Administrative Science Quarterly,*** 38: 382–407.

Rynes, S. 1987. Compensation strategies for recruiting. ***Topics in Total Compensation,*** 2: 185–196.

Rynes, S., Gerhart, B., & Minette, K. 2004. The importance of pay in employee motivation: Discrepancies between what people say and what they do. ***Human Resource Management,*** 43: 381–394.

Rynes, S., Gerhart, B., & Parks, L. 2005. Personnel psychology: Performance evaluation and pay for performance. ***Annual Review of Psychology,*** 56: 571–600.

Rynes, S., & Lawler, J. 1983. A policy-capturing investigation of the role of expectancies in decisions to pursue job alternatives. ***Journal of Applied Psychology***, 68: 620–631.

Rynes, S., Schwab, D., & Heneman III, H. 1983. The role of pay and market pay variability in job application decisions. ***Organizational Behavior and Human Performance***, 31: 353–364.

Samuelson, P. A. 1965. Proof that properly anticipated prices fluctuate randomly. ***Industrial Management Review,*** 6: 41–49.

Sackett, P., & Larson, J. 1990. Research strategies and tactics in industrial and organizational psychology. In Dunnette M. D., Hough L. M. (Eds.), ***Handbook of Industrial and Organizational Psychology*** (2nd ed.), 1: 419–489. Palo Alto, CA: Consulting Psychologists Press.

Sargent, T. J. 1973. Rational expectations, the real rate of interest, and the natural rate of unemployment. In Arthur M. Okun and George Perry (Eds.), ***Brookings Papers on Economic Activity,*** Vol. 2.

Schneider, B. 1987. The people make the place. ***Personnel Psychology,*** 40: 437–453.

Schweitzer, M., Ordóñez, L., & Dumaz, B. 2004. Goal-setting as a motivator of unethical behavior. ***Academy of Management Journal,*** 47: 422–433.

Shupp, R. S., & Williams, A. 2003. ***Risk Preference Differentials of Small Groups and Individuals.*** Working Paper, Indiana University.

Slonim, R., & Roth, A. Learning in high stakes ultimatum games: An experiment in the Slovak Republic. ***Econometrica***. 66: 569–597.

Sorensen, E. 1989. Measuring the Effect of Occupational Sex and Race Composition on Earnings. In Michael, R. Hartmann, H., O'Farrel, B. (Eds.), ***Pay equity: Empirical Inquires.*** Washington D.C.: National Academic Press.

Stajkovic, A., & Luthans, F. 1997. A meta-analysis of the effects of organizational behavior modifdication on task performance, 1975–1995. ***Academy of Management Journal***, 40: 1122–1149.

Stajkovic, A. D., & Luthans, F. 2001. Differential effects of incentive motivators on work performance. ***Academy of Management Journal,*** 44: 580–601.

Stevens, C. K., Bavetta, A. G., & Gist, M. E. 1993. Gender differences in the acquisition of salary negotiation skills: The role of goals, self-efficacy, and perceived control. ***Journal of Applied Psychology,*** 78: 723–735.

Subich, L., Barrett, G., Doverspike, D., & Alexander, R. 1989. The effects of sex-role-related factors on occupational choice and salary. In Michael, R. Hartmann, H., O'Farrel, B. (Eds.), ***Pay equity: Empirical Inquires.*** Washington D.C.: National Academic Press.

Taylor, S. E. 1989. ***Positive Illusions: Creative Self–Deception and the Healthy Mind.*** New York: Basic Books, Inc.

Taylor, S. E., & Brown, J. D. 1988. Illusion and well-being: A social psychological perspective on mental health. ***Psychological Bulletin,*** 103: 193–210.

Trank, C. Q., Rynes, S. L., & Bretz, R. D. Jr. 2002. Attracting applicants in the war for talent: Differences in work preferences among high achievers. ***Journal of Business Psychology,*** 17: 331–345.

Trevor, C. O., Gerhart, B., & Boudreau, J. W. 1997. Voluntary turnover and job performance: Curvilinearity and the moderating influences of salary growth and promotions. ***Journal of Applied Psychology,*** 82: 44–61.

U. S. Bureau of Labor Statistics. 2001. ***Productivity and Costs.*** Retrieved from the World Wide Web January 2003 at: http://www.bls.gov/lpc/peoplebox.htm.

Vance, R., & Colella, A. 1990. Effects of two types of feedback on goal acceptance. ***Journal of Applied Psychology,*** 75: 68–77.

Vroom, V. H. 1964. **Work and Motivation.** New York: Wiley.

Weiss, A. 1980. Job queues and layoffs in labor markets with flexible wages***. Journal of Political Economy***, 88: 526–538.

Williamson, O. E. 1975. ***Markets and Hierarchies: Analysis and Antitrust Implications.*** New York: Free Press.

Yellen, J. 1984. Efficiency wage models of unemployment. ***American Economic Review Proceedings***, 74: 200–205.

Zenger, T. R. 1992. Why do employees only reward extreme performance? Examining the relationship among performance pay and turnover. ***Administrative Science Quarterly,*** 37: 198–219.

Table 1
Risk-Aversion Measure: Holt-Laury Instrument

Option A	Option B	MY CHOICE (Please Write A or B)
10% chance of $4.40, 90% chance of $3.52	10% chance of $8.47, 90% chance of $.22	
20% chance of $4.40, 80% chance of $3.52	20% chance of $8.47, 80% chance of $.22	
30% chance of $4.40, 70% chance of $3.52	30% chance of $8.47, 70% chance of $.22	
40% chance of $4.40, 60% chance of $3.52	40% chance of $8.47, 60% chance of $.22	
50% chance of $4.40, 50% chance of $3.52	50% chance of $8.47, 50% chance of $.22	
60% chance of $4.40, 40% chance of $3.52	60% chance of $8.47, 40% chance of $.22	
70% chance of $4.40, 30% chance of $3.52	70% chance of $8.47, 30% chance of $.22	
80% chance of $4.40, 20% chance of $3.52	80% chance of $8.47, 20% chance of $.22	
90% chance of $4.40, 10% chance of $3.52	90% chance of $8.47, 10% chance of $.22	
100% chance of $4.40, 0% chance of $3.52	100% chance of $8.47, 0% chance of $.22	

Table 2
Productivity Results and Treatment Conditions
(Standard Deviations in Parentheses)

Round &Anagram	Treatment Condition	FS	PFP	VC (1990)	Pre-Test
1. OADMHUP	SS1 (FS: n=58; PFP: n=57)	10.41 (3.05)	12.14 (3.48)	11.16	14.58
2. AEDBKUG	SS1 (FS: n=58; PFP: n=57)	8.66 (2.75)	9.60 (3.37)	8.95	10.98
3. OELBJAM	FS	9.00 (3.47)	-	9.42	12.15
4. UADQWER	PFP	-	10.46 (3.47)	8.84	11.89
5. EASCKIY	FS	9.86 (3.81)	-	9.32	12.09
6. AODJGIP	PFP	-	10.65 (3.81)	9.63	12.20
7. UONHMEY	SS2 (FS: n=55; PFP: n=60)	9.51 (2.76)	12.77 (3.09)	10.21	13.06
8. OELHMAZ	SS2 (FS: n=55; PFP: n=60)	8.96 (3.09)	12.75 (4.04)	9.63	13.62
1& 2 (SS1) Average	SS1 (FS: n=58; PFP: n=57)	9.53 (2.58)	10.87 (3.05)	10.06	12.78
3 & 5 (FS) Average	N=115	9.43 (2.85)	-	9.37	12.12
4 & 6 (PFP) Average	N=115	-	10.56 (2.89)	9.24	12.05
7 & 8 (SS2) Average	SS2 (FS: n=55; PFP: n=60)	9.24 (2.48)	12.76 (3.23)	9.92	13.34

Table 3
Means, Standard Deviations, and Correlations

	Variable	*M*	*SD*	1	2	3	4	5	6	7
1.	Gender (female =0, male=1)	.38	.49							
2.	Native Language (English=0, Non-English=1)	.26	.44	.10						
3.	Risk aversion	6.77	1.92	.06	-.07					
4.	Productivity under the FS Scheme (Round 3 & 5)	9.43	2.85	**.21**	**-.21**	.04				
5.	Productivity under the PFP Scheme (Round 4 & 6)	10.56	2.89	.15	**-.30**	-.11	**.78**			
6.	Productivity in Middle 4 rounds (Rounds 3 to 6)	9.99	2.71	**.19**	**-.27**	-.04	**.94**	**.94**		
7.	Initial Self-Selection (FS = 0, PFP = 1)	.50	.50	.15	.01	-.12	**.26**	**.30**	**.30**	
8.	Final Self-Selection (FS = 0, PFP = 1)	.52	.50	.04	**-.18**	-.17	**.45**	**.48**	**.49**	**.39**

Note. $N = 115$. 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
Number of People who Gained, Broke Even, and Lost Each Pay Scheme

Panel A: Those who Chose the PFP Scheme

	Rounds 1 and 2	Rounds 7 and 8
>11 (gained)	22 (38.6%)	39 (65%)
=11 (broke even)	5 (8.8%)	4 (6.7%)
<11 (lost)	30 (52.6%)	17 (28.3%)
Total	57	60

Panel B: Those who Chose the FS Scheme

	Rounds 1 and 2	Rounds 7 and 8
<11 (gained)	43 (74.1%)	41 (74.5%)
=11 (broke even)	2 (3.4%)	4 (7.3%)
>11 (lost)	13 (22.4%)	10 (18.2%)
Total	58	55

Panel C: Productivity in Rounds 4 and 6 (PFP) for Those Who Chose the FS Scheme Initially (rounds 1 and 2) and Finally (rounds 7 and 8)

	Chose the FS Scheme Initially	Chose the FS Scheme Finally
<11 (gained)	41 (70.7%)	43 (78.2%)
=11 (broke even)	2 (3.4%)	4 (7.3%)
>11 (lost)	15 (25.9%)	8 (14.5%)
Total	58	55

Table 5
Logistic Regression of Probability of Breaking Even or Better on Self-Selection of Compensation Schemes (Two-Tail Test *P*-values in Parentheses)

Independent Variable	Initial Self-Selection (Rounds 1 and 2)	Final Self-Selection (Rounds 7 and 8)
Constant	1.242	1.504
	(.001)	(.000)
Self-Selection	-1.347	-.576
(FS=0, PFP=1)	(.000)	(.201)

Table 6
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)

Independent Variable	Initial Self-Selection	Final Self-Selection
Constant	.280	.921
	(.226)	(.004)
Productivity	.241	.574
(Centered at 11)	(.003)	(.000)
Risk aversion	-.264	-.450
(Centered at Mean)	(.040)	(.015)
Interaction	-.064	-.136
	(.123)	(.048)

Figure 1
PFP vs. FS Compensation Schemes

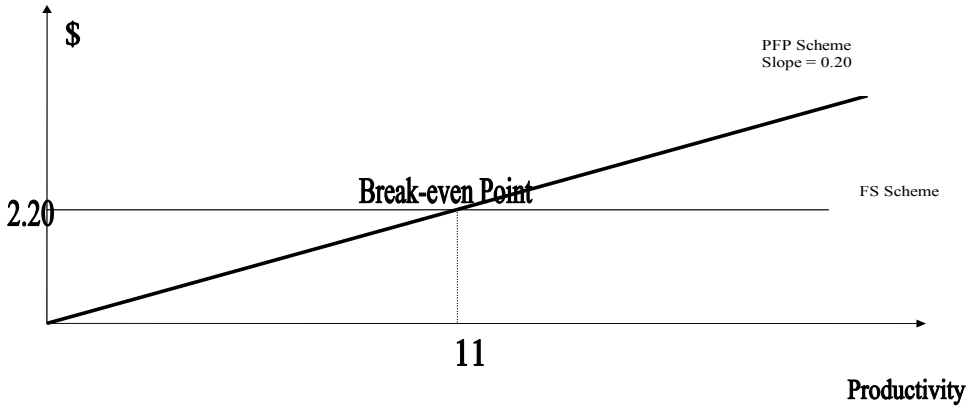
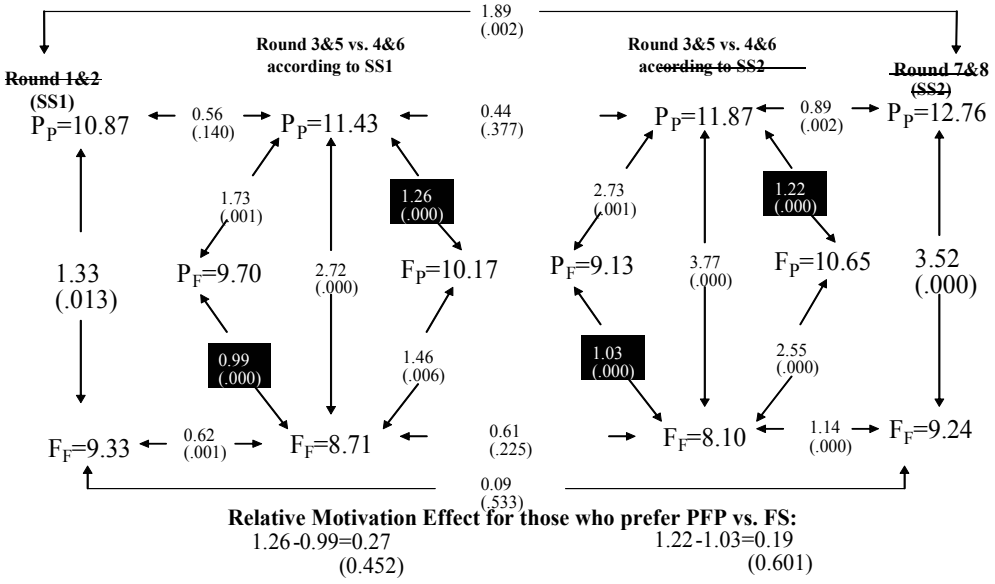


Figure 2
Decomposing Sorting and Incentive Effect of Pay-for-performance

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 linear scheme who also chose the linear scheme; F_F represents those working under the flat scheme who also chose the flat scheme; P_F represents those working under the linear scheme who chose the flat scheme; and F_P represents those working under the flat scheme who chose the linear scheme. Differences are indicated between the arrows and p-values associated with those differences are presented in parentheses.



Relative Motivation Effect for those who prefer PFP vs. FS:
1.26-0.99=0.27 (0.452)
1.22-1.03=0.19 (0.601)