The Impact of Risk Aversion and Stress on the Incentive Effect of Performance Pay

C. Bram Cadsby* Department of Economics and Finance University of Guelph Guelph, Ontario, Canada

Fei Song Ted Rogers School of Business Management Ryerson University Toronto, Ontario, Canada

Francis Tapon Department of Economics and Finance University of Guelph Guelph, Ontario, Canada

* Corresponding Author. Cadsby: Department of Economics and Finance, College of Business and Economics, University of Guelph, Guelph, Ontario, Canada, N1G 2W1. Email: <u>bcadsby@uoguelph.ca</u>. Phone: +1 519 824 4120 ext. 53320.

The Impact of Risk Aversion and Stress on the Incentive Effect of Performance Pay

Abstract

We demonstrate in a laboratory experiment that the effectiveness of performancecontingent incentives is inversely related to risk-aversion levels. For about 16.5% of participants, performance fails to improve under performance pay, and the probability of such deterioration increases with risk aversion. This phenomenon works in part through the reduced effort level of more risk-averse individuals when effort level is positively correlated with risk exposure. It is also associated with higher self-reported levels of stress by more risk-averse people working under performance-contingent pay. We find no evidence of such stress causing decrements in the quality of effort affecting performance after controlling for effort level. However, controlling for effort, more risk-averse participants perform better under a fixed salary, leaving less room for improvement under performance pay.

Keywords: risk aversion, performance pay, incentive, stress, productivity, experiment.

JEL Codes: C91, M52, J33.

The effect of financial incentives on human behavior is a central concern of economics (e.g., Baker et al., 1988; Bénabou and Tirole, 2003; Gibbons, 1998; Kreps, 1997; Lazear, 1986, 2000a, 2000b; Prendergast, 1999). Moreover, the relationship between such incentives and work performance has been an important focus of attention not only in economics, but also in management and psychology (e.g., Gerhart and Rynes, 2003; Gomez-Mejia and Welbourne, 1988; Rynes et al., 2005; Vroom, 1964). However, the impact of such incentives on behavior is viewed quite differently both within and across these disciplines. Many economists regard monetary incentives as a dominant and effective motivator of human behavior. Consequently, they argue that performance-contingent pay can be an effective incentive device to induce desired performance by mitigating principal-agent problems (e.g., Hart, 1989; Holmstrom, 1979; Jensen and Meckling, 1976; Jensen and Murphy, 1990; Kahn and Sherer, 1990; Kale et al., 2009; Milgrom and Roberts, 1992, pp. 206–247; Seiler, 1984). However, other economists and some psychologists argue, often with supporting experimental evidence, that under some circumstances extrinsic monetary incentives can crowd out intrinsic motivation, thereby adversely affecting motivation and hence performance (e.g., Deci and Ryan, 1985; Gneezy and Rustichini, 2000; see Frey and Jegen, 2001 for a review of this literature). Furthermore, a different strand in the psychology literature suggests that extrinsic incentives that increase the subjective importance of performing well on a task can result in "choking under pressure" (e.g., Baumeister, 1984; see Baumeister and Showers, 1986 for a review of this literature), thus hindering performance even when motivation is maintained or enhanced. Recently, economists have begun examining this issue as well (e.g., Ariely et al., 2009; Dohmen, 2008).

The incentive effect of pay-for-performance has been extensively examined in both the laboratory (e.g., Brase, 2009; Cadsby et al., 2007; Fessler, 2003) and the field (e.g., Abowd, 1990; Banker et al., 1996; Fehr and Goette, 2007; Gerhart and Milkovich, 1990; Lazear, 2000a, 2000b; Paarsch and Shearer, 1999, 2000, 2009; Shearer, 2004; Stajkovic and Luthans, 2001). Although a considerable number of studies show that monetary incentives are effective at improving performance, others indicate no effect on performance, while still others demonstrate that under some circumstances incentives can hinder performance (Bonner et al., 2000; Camerer and Hogarth, 1999; and Jenkins et al., 1998, all provide comprehensive reviews of this literature). All of these empirical studies are primarily concerned with the *average* effect of incentives on performance for a particular population in a specific context. This literature has focused on how

the characteristics of the work task, the relationships both between principals and agents and among agents, and the particular structure of the monetary incentives may affect the presence and magnitude of an incentive effect. However, to our knowledge, none of the earlier studies has explored whether individual differences among agents may also have an important and predictable impact on the effectiveness of financial incentives at improving task performance.¹ We focus on one such factor: differences in attitudes toward financial risk and uncertainty.

Whenever the link between effort and performance has a random component, the payoff from the exertion of effort involves financial uncertainty under performance-contingent pay. Indeed the trade-off between the provision of financial incentives and the transfer of risk from risk-neutral principals to risk-averse agents has been a central theme of agency theory (e.g. Milgrom and Roberts, 1992; Prendergast, 1999). This trade-off implies that the optimal level of incentive intensity is lower for higher levels of risk and lower levels of risk tolerance.² A lower level of incentive intensity is associated with less effort. However, in the standard linear model of agency theory, where the level of uncertainty is determined exogenously and is independent of the effort levels chosen by agents, the optimal amount of effort under a given level of incentive intensity is determined independently of either the amount of risk faced by agents or their attitudes toward it (e.g., Prendergast, 1999; Sloof and Praag, 2008).

Individual attitudes toward financial risk may nonetheless affect work performance under exogenously determined incentives. First, in contrast to the standard model, the amount of financial uncertainty faced by an agent may be positively correlated with his/her effort. In such circumstances, a more risk-averse agent may rationally choose to exert less effort than a less risk-averse counterpart in order to reduce exposure to risk. Less effort then translates into poorer average performance for the more risk-averse agent.

Second, there is empirical evidence that individual differences play an important role in the propensity of individuals to choke under pressure when faced with an academic test or sports competition (see Baumeister and Showers, 1986, pp. 373–375 for a review). Financial incentives may also enhance the perceived importance of performing well and have been associated with a similar choking phenomenon (Baumeister and Showers, 1986, pp. 368–369; Ariely et al., 2009). Since a higher level of risk aversion implies a greater discomfort with financial uncertainty, it

¹ Paarsch and Shearer (2007) do examine potential gender differences in the response to piece rates for workers from

² This result may break down in more complex models that permit multiple heterogeneous principals to be matched with multiple heterogeneous agents (e.g., Serfes, 2008).

may also be associated with a greater tendency for performance to be impaired by choking on incentivized tasks.

The purpose of this paper is to use experimental data to examine how an individual's attitude toward risk may influence the effectiveness of financial incentives at improving his/her performance on an assigned task. In Cadsby et al. (2007), we documented empirically the existence of a significant inverse relationship between risk aversion and performance improvement under a pay-for-performance scheme (henceforth *pfp*) using an anagram wordcreation game. In that study, however, which was largely concerned with how people choose between a fixed-salary scheme (henceforth fs) and pfp compensation schemes, we did not collect data on either effort levels or levels of stress. Thus, we could only speculate on possible reasons for the observed inverse relationship. Moreover, participants performed the real-effort task under each of the compensation schemes only twice. Thus, we did not have enough data to establish empirically whether the ex-post variance or riskiness of individual performances across periods increased with effort as required for the theory to predict a relationship between risk attitude and effort. In the current study, we examine the robustness of the previously reported result by employing a different task and subjects from a different country. Moreover, we examine its theoretical underpinnings by basing our new study on a theoretical economic model that predicts an inverse relationship between risk aversion and effort and on a psychological theory that suggests that individual levels of performance-related stress may adversely affect levels of performance for a given level of effort. Accordingly, we gather data on both effort and stress in order to better understand the mechanism through which individual attitudes toward risk may affect productivity improvement under *pfp*. We are thus able to consider both the amount of effort exerted under the financial uncertainty that inevitably accompanies performance pay and also the possibly choke-inducing stress experienced under the resultant pressure to perform well. We also have participants perform the real-effort task four times under each of the compensation systems. Thus, we are able to calculate the *ex-post* performance variance for each individual participant and determine whether effort and risk are positively correlated as required for the theoretical model's predictions regarding risk aversion and effort to be relevant to our experimental results. The next section provides some theoretical background. We then discuss the details of the experimental design, present some testable hypotheses based on the theory, and present the results of the experiment. Some conclusions follow.

1. THEORETICAL BACKGROUND

Under a fixed-pay compensation scheme where no performance-contingent incentives are offered, risk is borne by the firm. In contrast, under a performance-contingent incentive scheme, some of the risk is borne by the employees. In particular, a performance-based incentive scheme involves financial uncertainty, exposing the investment of effort by individual employees to financial risk. The reactions of employees to a change from a risk-free fixed-salary scheme (fs) to a riskier pay-for-performance scheme (pfp) may thus differ depending upon individual attitudes toward risk. This may occur for two reasons. First, under pfp, a more risk-averse employee may choose to exert less effort than a less risk-averse employee if effort is positively correlated not only with expected output, but also with risk. Second, a more risk-averse employee may react with greater stress than a less risk-averse employee to the financial uncertainty of pfp, and this increased stress may hinder his/her performance response to monetary incentives. We discuss both of these arguments below.

i. Optimal Effort and Risk Aversion when Risk Increases with Effort

To illustrate the relationship between optimal effort and individual levels of risk aversion under *fs* and *pfp*, we employ a simple agency model with a linear piece rate and both additive and multiplicative uncertainty (e.g. Bushman et al., 2000; Baker and Jorgensen, 2003; Sloof and van Praag, 2008). Consider an agent whose level of output, y, depends stochastically on the amount of effort s/he exerts, a, and two random terms, θ and ε :

(1) $y = \theta \cdot a + \varepsilon$,

where $\theta \sim N(\mu_{\theta}, \sigma_{\theta}^2)$ and $\varepsilon \sim N(0, \sigma_{\varepsilon}^2)$. Theta is multiplicative risk and can be thought of as representative of ability since it indicates how a certain amount of effort will be transformed into output. In the model, it is a random variable. For example, a person may be more or less productive from one day to the next or even from one hour to the next. This could depend on changes in mood or energy. For some tasks, it could depend on luck. For example, a sales person may have more positive responses to his/her identical efforts on some days than on others. Epsilon is additive risk or uncertainty that is independent of effort. For example, each day varying numbers of customers may either purchase or return goods regardless of a salesperson's effort or lack of effort.

The agent's pay, w, is determined by a fixed salary component, s, and a piece-rate component, b:

(2) $w = s + b \cdot y$.

The cost of effort is measurable in monetary terms and is assumed to be quadratic:

(3)
$$c(a) = k \cdot a^2/2 - f \cdot a$$

where a is effort, while k and f are constants with k > 0 and $f \ge 0$. The f term represents intrinsic motivation³, allowing for the possibility that effort exerted on the task may be enjoyable or satisfying up to a certain comfort level after which it becomes costly. This allows for some effort to be exerted under *fs*.⁴ Each agent is assumed to possess a Constant Absolute Risk Aversion (CARA) utility function:

(4) $U = -\exp[-r(w - k \cdot a^2/2 + f \cdot a)],$

where r is the Arrow-Pratt measure of risk aversion.^{5, 6} Since neither source of uncertainty in the relationship between effort and output is resolved prior to the exertion of effort, each agent must base his/her decisions on expected utility:

(5)
$$E(U) = -\exp\{-r[s - k \cdot a^2/2 + f \cdot a] - r \cdot b \cdot a \cdot \mu_{\theta} + [(r \cdot b)^2/2] \cdot [a^2 \cdot \sigma_{\theta}^2 + \sigma_{\varepsilon}^2]\}$$

Defining CE as the certainty equivalent in monetary terms of this expected utility and noting that U(CE) = E(U) by definition, we obtain:

A risk-averse agent maximizing his/her expected utility or its certainty equivalent under piece-rate *pfp* compensation will then determine his/her optimal effort as:

³ Like Frey and Jegen (2001) in their survey on motivation crowding out, we employ the definition of intrinsic motivation introduced by psychologist Deci (1971, p. 105) as follows: "one is said to be intrinsically motivated to perform an activity when one receives no apparent reward except the activity itself."

⁴ The standard assumption that effort always involves disutility is a special case of this assumption where f = 0. In fact, only one out of 85 participants in our experimental study exerted no effort at all in any of the four *fs* compensation periods of the experiment. Nine other participants exerted zero effort in some *fs* periods and positive effort in others. The remaining 75 participants exerted positive effort in all four *fs* periods. This may be due to a positive f in individual cost functions. However, there are alternative explanations such as a feeling of reciprocal obligation to do something in exchange for the fixed salary or a desire to practise for subsequent periods in which the participant may believe s/he could face *pfp* compensation.

⁵ The f·a term representing intrinsic motivation could be introduced as a direct positive component of the utility function rather than as a negative component of the c(a) function. Such an approach yields the same expression for equation (4) and identical theoretical predictions as the approach taken in the text.

⁶ We use CARA utility for illustrative purposes. In Cadsby, Song, and Zubanov (2016), we show that we obtain the same qualitative prediction for any utility function satisfying the usual neo-classical properties when rewards and costs are non-separable.

(7)
$$\tilde{a}_{pfp} = (f + b \cdot \mu_{\theta})/(k + r \cdot b^2 \cdot \sigma_{\theta}^2).$$

Note that r, the Arrow-Pratt measure of risk-aversion, is inversely related to the optimal amount of effort. This relationship depends on the variance of the multiplicative random term, σ_{θ}^2 . Since the multiplicative random term implies a positive correlation between effort and risk, a more risk-averse person will sacrifice more expected return than a less risk-averse person to mitigate risk. This can be thought of as analogous to a more risk-averse person investing less of his/her wealth in a risky portfolio of financial assets and correspondingly more in an asset that is risk-free. The variance of the additive random term, σ_{e}^2 , plays no role in the determination of optimal effort because additive risk is independent of the actions of the agent. Any part of income that is designated as fixed salary, s, and is thus paid regardless of performance, also plays no part in the determination of effort. For an agent who is paid solely by means of a fixed salary, the piece rate, b, is zero. For such an agent, optimal effort is:

(8)
$$\tilde{a}_{fs} = f/k$$

Thus, in the *fs* case, optimal effort is independent of the agent's attitude toward risk because there is no financial risk for the agent. Of course, if f = 0, there is no intrinsic motivation. In that case, any effort brings disutility to the agent and hence none will be exerted in the *fs* case.

The incentive effect of pfp is the difference between performance under pfp and performance under fs. This difference in performance is closely related to the difference in effort exerted under the two pay schemes as specified in equation (1). This difference in effort is:

(9)
$$\tilde{\mathbf{a}}_{pfp} - \tilde{\mathbf{a}}_{fs} = (\mathbf{f} + \mathbf{b} \cdot \boldsymbol{\mu}_{\theta})/(\mathbf{k} + \mathbf{r} \cdot \mathbf{b}^2 \cdot \boldsymbol{\sigma}_{\theta}^2) - \mathbf{f}/\mathbf{k}.$$

The most important implication of (9) for this paper is that, *ceteris paribus*, there is an inverse relationship between an individual agent's level of risk aversion and the difference in his/her motivation to exert effort under *pfp* relative to *fs*. This inverse relationship holds regardless of whether an agent exhibits intrinsic motivation (f > 0) or not (f = 0).

⁸ The first-order condition is $-\mathbf{k} \cdot \mathbf{a} + \mathbf{f} + \mathbf{b} \cdot \mathbf{\mu}_{\theta} - \mathbf{r} \cdot \mathbf{b}^{2} \cdot \mathbf{a} \cdot \mathbf{\sigma}_{\theta}^{2} = 0$. The second order condition is $-\mathbf{k} - \mathbf{r} \cdot \mathbf{b}^{2} \cdot \mathbf{\sigma}_{\theta}^{2} < 0$ whenever $\tilde{\mathbf{a}} > 0$. This encompasses all cases in which the agent is risk-averse or risk-neutral as well as those where a preference for risk is not too large, specifically where $-\mathbf{r} < \mathbf{k}/\mathbf{b}^{2} \cdot \mathbf{\sigma}_{\theta}^{2}$. A corner solution involving the exertion of the maximum possible level of effort is possible if an agent is sufficiently risk-loving, i.e if $-\mathbf{r} > \mathbf{k}/\mathbf{b}^{2} \cdot \mathbf{\sigma}_{\theta}^{2}$.

⁷ This expression was derived for the case where f = 0 by Bushman et al. (2000) and Baker and Jorgensen (2003). However, neither of these papers focused on the relationship between individual levels of risk aversion and the optimal level of effort.

Moreover, when f > 0 so that there is some intrinsic motivation, effort may either rise or fall when an agent moves from *fs* to *pfp* compensation. Whether it rises or falls also depends on the agent's degree of risk aversion. Define $\check{r} = k \cdot \mu_{\theta} / f \cdot b \cdot \sigma_{\theta}^2 > 0$. This is simply the value of r for which the difference in effort, $\tilde{a}_{pfp} - \tilde{a}_{fs} = 0$. If $r < \check{r}$, both effort and expected performance will be greater under *pfp* than under *fs*. This includes all agents who are either risk-neutral or risk-loving, together with those who possess levels of risk aversion below the critical value. However, if $r > \check{r}$, both effort and expected performance will be lower under *pfp* than under *fs*. Moreover, a higher level of multiplicative risk, a higher piece rate, and a higher intrinsic motivation term, f, all reduce \check{r} . This enlarges the range of risk-aversion levels resulting in less effort and expected performance under *pfp* than under *fs*.

Such a decline in effort under performance pay closely resembles the crowding out of intrinsic by extrinsic motivation. However, there is a subtle difference. Crowding out involves a reduction in intrinsic motivation through the introduction of *pfp*. While such crowding out may be an important phenomenon (e.g., Bénabou and Tirole, 2003; Frey and Jegen, 2001), we have demonstrated that even if intrinsic motivation, f, remains unchanged, effort may nonetheless fall in the presence of multiplicative uncertainty for a different reason. Despite unchanged intrinsic motivation, sufficiently risk-averse agents will choose to exert less effort under *pfp* than under *fs* in order to reduce the financial uncertainty associated with the *pfp* compensation scheme. In contrast to crowding out, this phenomenon should be associated empirically with individuals possessing higher levels of risk aversion. Of course, in the case of no intrinsic motivation (f = 0), an agent exerts no effort under *fs*. Thus, effort cannot fall under *pfp* in such a case.

ii. Choking under Pressure, Risk Aversion and Performance

Besides making a decision to exert less effort resulting in a lower expected level of output, a more risk-averse person working under *pfp* might well experience considerably more anxiety, stress or pressure than a less risk-averse person. Our argument for this relationship hinges on our understanding of financial risk as financial uncertainty. A risk-averse person is a person who dislikes or is averse to financial uncertainty. The more risk-averse a person is, the more that person dislikes financial uncertainty. Being placed in a situation that one dislikes may cause stress. The more one dislikes the situation, the more stress he/she may on average experience, recognizing of course that people may have differing abilities to cope with such situations. Thus,

when placed in the situation of financial uncertainty represented by *pfp*, more risk-averse people will on average experience a greater increase in stress than less risk-averse people.

In pioneering studies by Lazarus (1966) and Lazarus and Folkman (1984), stress is conceptualized to be relational and process-oriented in nature as "it is a particular relationship between the person and the environment that is appraised by the person as taxing or exceeding his or her resources and endangering his or her well-being" (Lazarus and Folkman, 1984, p. 19). A considerable literature addresses the relationship between stress and work performance (see LePine et al., 2005 and Muse et al., 2003, for reviews of this literature). Much of this literature suggests that stress, particularly "hindrance stress", is inversely related to performance.9 Furthermore, research on work-related stress has identified performancecontingent pay as one of these stressors (Schuler, 1980). If such stress impedes performance for a given level of motivation and effort, the incentive effect of *pfp* on performance may be weakened, eliminated or even reversed. Such "choking under pressure" is defined by Baumeister (1984) as "performance decrements under pressure circumstances." As outlined in Baumeister and Showers (1986), such "choking under pressure" may occur for a number of reasons. The payoff uncertainty created by the link between payoff and performance may cause a risk-averse employee to become distracted from the task at hand by thoughts irrelevant to the accomplishment of the task. For example, worry about whether or not one will perform well or the financial implications of performing poorly can seriously impede performance (Baumeister and Showers, 1986, pp. 366). Alternatively, the discomfort created by performance-dependent payoffs can make the risk-averse employee become more self-conscious about each step of the performance process, thereby hindering performance (Baumeister, 1984). Research by Beilock et al. (2004) and Beilock, Rydell, and McConnell (2007) suggests that a critical factor is the overconsumption of working memory, via worries about the situation and its consequences, when the working memory is the critical resource individuals need to perform at an optimal level. This effect is exacerbated for tasks that rely crucially on working memory, such as mathematical problem solving, the task we used in our experiment.

In an experimental study, Baumeister (1984) found that such choking occurs for a given level of effort. In particular, pressure worsened performance without reducing effort, i.e. effort

⁹ 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.

did not mediate the relationship between pressure and performance. Thus, for clarity, in this study we define "choking under pressure" as performance decrements under pressure circumstances for a fixed level of effort. We are then able to examine whether in addition to exerting less effort in order to reduce risk, a more risk-averse person might also have more difficulty successfully transforming his/her effort into output owing to excessive performance pressure arising from the financial uncertainty associated with *pfp*.

2. EXPERIMENTAL DESIGN

Participants were recruited at the University of Guelph in Guelph, Ontario by means of email solicitation through the Bachelor of Commerce program listserv. All 85 participants were undergraduates and majors in economics or other business subjects. There were 44 males and 41 females with an average age of 20.50 years and a standard deviation of 0.20 years. We employed a widely used arithmetic task involving the addition of sets of five double-digit integers (e.g. Niederle and Vesterlund, 2007; Cadsby et al., 2013). Specifically, participants were asked to play one practice and eight experimental three-minute periods using randomly generated sets of five double-digit integers. Subjects were not paid for the practice period. Its purpose was to familiarize the subjects with the assigned task. The experiment utilized two different compensation schemes, one representing *pfp* and the other *fs*. The *pfp* scheme paid \$0.25 Canadian¹⁰ for each correctly calculated sum. The *fs* scheme paid a fixed amount of \$1.75 Canadian for each three-minute work period, regardless of performance. This was based on a pre-test showing that seven was both the mean and median performance level under *pfp*.

Upon arrival, the instructions were read to the participants while they followed along on their own copies. Participants were provided with a prepared workbook containing the sets of numbers to be added in each period. The arithmetic tasks for each period were presented on a separate page of the workbook. Inserted just before each of these pages was another page used to explain which compensation scheme would apply in the subsequent period. 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 commencing work on that page. Participants were seated far enough apart so that they could not see the speed at which others were working.

¹⁰ At the time of the experiment, the Canadian and US dollars were approximately at par.

Half of the participants were paid according to the *fs* scheme in periods 1, 3, 5, and 7, while in periods 2, 4, 6, and 8, they were paid according to the *pfp* scheme. For the rest of the participants, the *pfp* scheme was used in periods 1, 3, 5, and 7, while the *fs* scheme was employed in periods 2, 4, 6, and 8. Participants were informed of the payment scheme immediately prior to the period. For all participants, regardless of payment scheme, the arithmetic questions were identical for each period. After every period, each participant's list of answers was collected by the experimenters and taken to another room where the number of correct answers was calculated. Participants did not receive feedback on the number of questions they had answered correctly until they were paid at the end of the session.

After participants completed the experimental task, they filled out a questionnaire in which they responded to a number of demographic questions such as age, gender, mother tongue and country of birth. In addition, the questionnaire was used to collect information on the stress experienced by participants during the experiment. There are three ways of measuring stress in the existing literature: psychological measures (self-reports), behavioral measures, and physiological measures (e.g., Cooper, Sloan and Williams, 1988; Fleming and Baum, 1987; Quick and Quick, 1984). In our study we follow the psychological perspective (Lazarus, 1966; Lazarus and Folkman, 1984; Parker and DeCotiis, 1983), in which stress is the result of an interaction between a person and the environment. Under this conceptualization, stress occurs when the environment imposes demands, which are perceived as being substantially out of balance with a focal person's capabilities. In other words, stress is a psychological experience and is thus best measured by a person's self-reported perceptions. Accordingly, we use a selfreport measure that asks participants to indicate how stressful they found working under the fs and *pfp* schemes, using a Likert scale of 1 (not at all) to 5 (extremely). Specifically, we asked the following questions: "How stressful was the task under the compensation method in which you earned \$0.25 for each arithmetic problem you solved correctly?" and "How stressful was the task under the compensation method in which you earned \$1.75 independent of your performance?" This method has the advantage of focusing the subject's attention on the stress associated with each of the two payment schemes rather than on other factors that may have an impact on physiological measures of stress. For example, a hot room may cause an increase in heart rate and/or perspiration, but has nothing to do with the stress that is the focus of our study. A disadvantage of self-reports is their subjective nature. Different people may assign different

numbers to similar feelings. This problem is partially mitigated by the fact that for each person, we focus on the difference between the measures reported for each pay scheme.¹¹

The final purpose of the questionnaire was to elicit risk preferences by asking participants to make ten lottery-choice decisions based on an instrument developed by Holt and Laury (2002). 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 are manipulated so that each decision involves 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. Following Holt and Laury (2002), we included those who switched more than once in the analysis, using the number of safe choices as a measure of risk aversion. In addition to being paid for the number of correct answers according to the compensation schemes outlined above, each participant was paid an additional sum based on the outcome of the lottery s/he chose from one of the ten lottery pairs selected at random by rolling a ten-sided die. We elicited risk preferences after the completion of the experimental task in order to avoid the possibility of biasing the behavioral decisions by priming participants to focus on risk. In our judgment, this was a more important consideration than the possibility that administering risk preferences after completion of the experimental task might lead to an impact of task performance or beliefs about the purpose of the experiment on the subsequent elicitation. To mitigate any impact that completing the experimental task might have on risk elicitation, we did not give any feedback on the number of correct solutions or the amount earned until the very end of the experiment after the risk data were collected. Nonetheless, subjects likely had some idea of how much they earned, implying the possibility of wealth effects on elicited risk attitudes. To examine this possibility, we performed a linear regression of our risk-attitude measure on earnings from the experimental task. It revealed no significant linear relationship between risk attitude and earnings.¹² 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. The numbers used

¹¹ Please also refer to Fried et al. (1984), which provides an extensive critique of using physiological measures, such as cardiovascular (heart rate and blood pressure) or biochemical (uric acid, blood sugar and steroid hormones) measures in work stress research. They provide evidence that factors such as room temperature, humidity, time of day, and consumption of caffeine, nicotine and alcohol can all influence the physiological measures drastically, but be unrelated to the experimental manipulation of stress per se.

¹² The details of this regression are not reported to save space, but are available from the authors upon request.

by Holt and Laury (2002) in Canadian-dollar units happened to correspond closely to the amount at stake in each real-effort work period. Accordingly, we used these numbers. At the end of the session, players were taken individually to another room, where they were paid privately in cash.

3. HYPOTHESES

Both the optimal-effort and the choking-under-pressure arguments outlined in the Theoretical Background section above suggest Hypotheses 1 and 2, which together form the central focus of this study:

H1: The effectiveness of pfp at improving performance is inversely related to individual levels of risk aversion.

H2: The probability of an improvement in performance under pfp relative to fs is inversely related to individual levels of risk aversion.

To examine the role played by the optimal-effort theory, it is necessary to find a reasonable representation for actual effort levels. We follow Eriksson et al. (2009) who examine the effects of different incentive and feedback schemes on effort using a real-effort task almost identical to the one used in our study.¹³ They point out that effort has both a quantitative and a qualitative dimension. The quantitative dimension is represented by the number of submissions or attempts within a specified time frame. This measure has a number of advantages. First, it is observable. Second, as discussed in detail immediately below, it is quite likely to be positively correlated with financial risk in a piece-rate environment. Third, this measure is relevant to many work tasks that involve repeated attempts at selling products, scoring goals or publishing articles within a specified period with some probability of success for each attempt that is greater than zero but less than one. Removing the time dimension would permit a large number of slow and careful attempts that could potentially result in a higher probability of success for each of them. This could potentially be the case for the arithmetic task used in this experiment. Thus, the time dimension is critical, making one's decision about the number of attempts equivalent to a decision about how fast to work, or alternatively on how much time to spend on each attempt including rest and relaxation time surrounding the attempt. Fourth, our within-person design allows us to compare the number of submissions for each person individually between the *pfp* and fs cases. Under fs, each person reveals the speed at which s/he prefer to work and the equivalent choice of the number of submissions. This may relate to such factors as a person's

¹³ Their subjects add sets of four two-digit numbers while ours add sets of five two-digit numbers.

skill at the task or intrinsic motivation at performing it, as well as the availability or lack thereof of other competing ways to spend one's time. We can use the *pfp-fs* comparison to examine the extent to which linking pay with performance changes the number of submissions from the utility maximizing choice made when pay is independent of both effort and performance for people with differing attitudes toward risk. Again following Eriksson et al. (2009), other factors such as the degree of concentration and care devoted to the task are deemed to affect the quality of effort reflected by the ratio of the number of correct answers to the number of submissions, or the *ex-post* probability of success, for each person in our experiment. Using the number of submissions as a proxy for the quantitative dimension of effort, we test two hypotheses related to the interplay between individual risk attitudes, number of submissions, and performance:

H3: If financial risk is positively correlated with the number of submissions, less risk-averse individuals will choose more submissions under pfp relative to their choice of submissions under fs than more risk-averse individuals.

H4: More submissions under pfp relative to fs in turn lead to higher performance in pfp relative to fs.

H3 is predicated on the condition that financial risk is positively correlated with the quantitative dimension of effort as proxied by the number of submissions, N. Let q = the probability of success for each submission. Under the simplifying assumption that q is independent of N, output follows a Binomial Distribution with mean = $q \cdot N$ and variance = $q(1 - q) \cdot N$.¹⁴ Since q(1 - q) > 0, the variance of output, proportional to financial risk under *pfp*, increases with the number of submissions, our proxy for effort level.¹⁵ We cannot directly observe the *ex ante* risk schedule for each participant at each potential submission level. However, since each participant played four periods under *pfp*, we can observe the *ex-post* performance variance for each participant under that payment scheme. This *ex-post* measure can be employed as a proxy for the *ex-ante* variance representing risk at the chosen submission level,

¹⁴ A discrete binomial distribution may be approximated by the continuous normal distribution employed to derive the theoretical predictions. The normal approximation possesses the same mean and variance as the binomial distribution it approximates with the approximation becoming more accurate as the number of attempts increases.

¹⁵ It might be thought that for each individual, q might fall with an increase in the number of that person's submissions or rise with a decrease. This would occur if for example working faster to submit more answers resulted in more careless mistakes. If this were the case, a higher number of submissions would lead to an even higher variance than if it q were constant as long as q > 0.5. In any case, this does not occur in our experiment. Regressions of q on the number of submissions with a fixed effect to control for the differing abilities of subjects show no systematic relationship between q and N for either the pfp or fs cases.

permitting us to examine whether or not financial risk was positively correlated with submissions in our experiment, a necessary condition for H3 to hold.

The roles played by stress and/or choking under pressure are examined by testing a final set of four hypotheses.

H5: Higher levels of risk aversion will generally be associated with larger differences between self-reported stress levels under pfp and self-reported stress levels under fs.

H6: Larger differences between self-reported stress levels under pfp and those under fs will in turn be associated with less performance improvement under pfp relative to fs.

H5 and H6 together imply a negative correlation between individual levels of risk aversion and performance improvement. This may occur through one or both of the channels summarized in H7 and H8 below.

H7: Relatively high levels of stress under pfp compared to fs are associated with less performance improvement through their correlation with relatively low submission levels under pfp compared to fs.

H8: Controlling for the number of submissions, higher reported stress levels in pfp relative to fs are correlated with decrements in performance or lower effort quality under pfp relative to fs, the choking under pressure hypothesis.

4. RESULTS

4.1 Overall Summary and Descriptive Results

All 85 participants completed the study. Table 1 reports means and standard errors of the variables. Performance refers to the number of problems correctly solved. Whether *pfp* (*fs*) was used in periods 1, 3, 5 and 7 (2, 4, 6 and 8) or in periods 2, 4, 6 and 8 (1, 3, 5 and 7) made no significant difference to any of the results. This was determined by running all of the statistical tests reported below with a dummy variable for ordering of compensation scheme, both as a main effect and as an interaction, with all of the other treatment variables. F-tests of the joint null hypothesis that both the main effect of ordering and its interactions with the other variables were equal to zero could never be rejected. Thus, we pooled the data for subsequent analysis.

An overview of the data summarized in Table 1 yields the following observations. First, most participants showed an improvement in productivity when moving from the *fs* periods to the *pfp* periods. On average, participants solved 6.90 more problems ($M_{pfp} = 26.01$ vs. $M_{fs} = 19.11$) over the four *pfp* periods than over the four *fs* periods, and the within-person difference

was significant (t = 8.04, df = 84, p < 0.0001). However, 14 out of the 85 participants, accounting for 16.5% of the sample, did not improve in the *pfp* periods. On average, those 14 participants solved 23.51 problems under fs (> 19.11 problems for all 85 participants) and 21.5 problems under *pfp*, (< 26.01 problems for all 85 participants), a reduction of 2.01 problems under pfp compared to fs. Second, the participants were quite risk-averse. The average riskaversion level was 6.47 safe lottery choices out of a possible 10 with 85.77% of participants exhibiting some degree of risk aversion. For comparison purposes, making 4 safe lottery choices was consistent with risk-neutrality. Only 5.89% of participants made this choice, while the remaining 8.24% of participants made risk-loving choices. The 14 subjects whose performance did not improve under *pfp* made 7.79 safe choices on average. All of them exhibited some degree of risk-aversion. Of the 85 participants, 75 (88.3%) made consistent choices, switching from the safer to the riskier lottery only once. Following Holt and Laury (2002), we elected to include all participants in the data analysis including the 10 who made inconsistent choices, and the numbers in Table 1 reflect their inclusion.¹⁶ Third, most participants (57 out of 85, i.e., 67.06%) reported a higher stress level working under pfp than under fs. On a scale of 1 to 5 for level of stress, the mean stress level under pfp was 4.44, while it was only 2.62 under fs. The mean within-person difference in stress levels between the two compensation schemes of 1.81 was significant (t = 8.99, df = 84, p < 0.0001). Fourth, quantitative effort level measured as the number of submissions over the four *pfp* periods averaged across participants was equal to 31.61, while the comparable number was 27.91 for the fs periods. The mean within-person difference of 3.70 was significant (t = 4.18, df = 84, p < 0.0001). The two demographic variables, gender and whether a participant was born in Canada or abroad, were both initially used as control variables in the analysis. Since the latter was never significant, it was ultimately dropped from the reported regressions. This made no difference to any of the qualitative results.

Insert Table 1 about here.

Figure 1 shows average *fs* and *pfp* performance for each individual subject. The data are sorted by risk category from the least to the most risk-averse. Within each risk category, the data are sorted by *fs* performance. Table 2 indicates the average *fs* performance and average *pfp* performance for all subjects in each risk category, and also presents the average difference

¹⁶ Our results are robust to the exclusion of those making inconsistent choices. If they are excluded, the mean number of safe choices for those making consistent choices is 6.45.

between *pfp* and *fs* performance. There are no consistent monotonic patterns in the levels. For *fs*, the lowest average performance levels are for the risk-loving subjects who made just 3 safe choices (4.04) and the risk-neutral (3.30) subjects who made 4 safe choices, while the highest is for those who made 6 safe choices (5.55) and the two subjects who irrationally made 10 safe choices (5.75).¹⁷ For *pfp*, the highest average level of performance is for the fairly risk-averse subjects who made 6 out of 10 safe choices (7.64) while the lowest average levels of performance are for those slightly more risk-averse subjects who made 7 (5.88) and the two subjects who made 10 safe choices (4.63). The differences in performance, while also not monotonically related to risk attitudes, appear to tell a somewhat clearer story. For those less risk-averse subjects who made 6 or fewer safe choices, the difference in performance between *pfp* and *fs* is greater than 2, while for those more risk-averse subjects who made more than 6 safe choices it is less than 2. Statistical analysis is of course required to determine whether this represents a systematic relationship between risk attitude and improvement in performance between *fs* and *pfp*.

Insert Figure 1 and Table 2 about here.

4.2 Hypothesis Tests

We employ two approaches to examine H1. First, we simply calculate the total performance of each participant over both the four *pfp* periods and the four *fs* periods. The difference is the performance improvement under *pfp*. There is one such observation per participant. The results of regressing performance improvement on risk aversion and a dummy variable that is one for males and zero for females are reported in Table 3 Column 1. They indicate a significant inverse relationship between individual levels of risk aversion and performance improvement as hypothesized (p = 0.011). There is no significant gender effect. This conservative approach has the advantage that it does not involve making any specific assumptions about the nature of the dependence between the observations on performance for each participant over the eight periods.

Could the inverse relationship between risk-aversion and the difference in performance between *pfp* and *fs* reflect a positive correlation between proficiency at the arithmetic task and risk aversion? If risk-averse people are more proficient, they may be able to improve more than others when given appropriate incentives to do so. There is no perfect measure of task

¹⁷ Making 10 "safe" choices is irrational because the tenth choice is between a certain return of \$2.85 in the safer column and a certain return of \$3.85 in the riskier column.

proficiency in our study. It is very challenging to identify proficiency because it is generally confounded with motivation when all we can observe is performance. The best proxy we can find is the highest of *fs*-period average performance and *pfp*-period average performance for each subject. This represents the most a subject produces under whichever scenario coaxes the most production out of him/her averaged over the four periods using that pay scheme to reduce random noise. To examine whether risk aversion is associated with task proficiency, we regress the risk-attitude measure on this proficiency proxy. It is not significant (p = 0.28), suggesting that positive correlation between risk aversion and proficiency is not an explanation for the observed inverse relationship between performance improvement under *pfp* versus *fs* and the level of risk aversion. This conclusion is strengthened when we insert the proficiency measure as a robustness control into the regression of performance improvement on risk aversion. While proficiency is significantly related to performance improvement, the risk aversion result remains intact.¹⁸

Insert Table 3 about here.

Assuming that performance incorporates an individually specific random effect that is not correlated with the explanatory variables permits the use of panel data methods. This has the advantage of permitting us to observe not only the relationship between performance improvement and risk aversion, but also how that relationship evolves over time. It will thus be our primary method of analysis in what follows.¹⁹ In Table 4 Column 1, we report the results of such a regression with random effects for each participant to account for individual data clustering.²⁰ The dependent variable is performance in each of the eight periods for each participant. The first independent variable is risk aversion (**ra**) centered on its mean of 6.47.²¹ The second independent variable is a dummy variable equal to one for the *pfp* periods and zero

¹⁸ The detailed regression results are available from the authors upon request.

¹⁹ Note that it is not possible to analyze the effect of risk aversion levels on performance under *pfp* and *fs* using an individual fixed-effects model because the risk aversion level for each participant is constant over the eight periods, making it impossible to separate its effect on performance from that represented by the individual fixed effects. ²⁰ As a robustness check, we also employed two alternative estimation techniques: the robust standard error

As a robustness check, we also employed two alternative estimation techniques: the robust standard error clustering of errors by individual participant and the combination of a random effect for each participant plus robust standard error clustering by participant. These different estimation techniques yield identical coefficients, but slightly different standard errors. We did the same robustness check for all estimations in this paper that use a random-effects panel-data approach. In no case were there any qualitative differences in statistical inferences. To save space, these results are not reported here, but are available from the authors upon request.

²¹ Centering simply involves subtracting the mean from each observation. Its purpose in the presence of interactions is to estimate coefficients that represent marginal effects at the mean rather than at zero. Cohen et al. (2003: 261-282, especially p. 281) contains an excellent discussion about centering and the proper interpretation of centered versus non-centered variables.

for the *fs* periods (**pfp**). The third independent variable is period number centered on its mean of 4.5 (**period**). The fourth independent variable is a dummy variable equal to one for males and zero for females (**male**). We also enter interactions between **ra** and **pfp**; **period** and **pfp**; **ra** and **period**; **ra**, **period** and **pfp**; and **male** and **pfp**.

Insert Table 4 about here.

The **pfp** dummy is positive and significant (p = 0.000), meaning that performance is significantly higher under *pfp* than under *fs*. The centering of **ra** means that this effect is evaluated at the mean level of risk aversion while the centering of **period** means it is evaluated halfway through the eight work periods. There is also a negative and significant interaction (p =0.000) between ra and pfp. The interpretation is that the higher the level of risk aversion, the lower is the difference between *pfp* and *fs* performance, corroborating H1. The main effect of period is not significant, meaning that under *fs* there is no significant improvement in performance from period to period evaluated at the mean level of risk aversion. However, the interaction between **period** and **pfp** is significant (p = 0.012), implying that under *pfp* there is significantly greater improvement in performance over time than in the *fs* periods again evaluated at the mean level of risk aversion. This suggests that *pfp* motivates participants to learn and improve, something not observed for the same participants during the *fs* periods. Under *fs*, both ra and the interaction between ra and period are positive and marginally significant (p =0.088 and p = 0.096 respectively). This suggests some tendency for more risk-averse participants to perform better under fs than participants who are less risk-averse, and for this tendency to be more pronounced over time. If true, this may represent a third channel, not anticipated in our theoretical discussion, through which risk aversion could be inversely related to performance improvement. Finally, the male dummy is not significant under fs, while its interaction with the **pfp** dummy is positive and marginally significant (p = 0.086).

To examine H2, we run a logit regression. The binary dependent variable is whether or not performance was higher under *pfp* than under *fs*. The independent variables are **ra** and **male**. The results are reported in Table 3 Column 2. They corroborate H2, showing that risk aversion is inversely and significantly related to the probability of performance improvement under *pfp* as predicted (p = 0.011). The control for **male** is also inversely related to the probability of performance improvement with marginal significance (p = 0.091).

H3 is predicated on the condition that financial risk is positively correlated with effort level

represented by the number of submissions. A regression of the *ex-post* performance variance for each participant on that participant's number of submissions under *pfp* controlling for gender is reported in Table 3 Column 4. The results confirm a significant positive relationship (p = 0.012).²² Thus, the required condition for H3 to hold is satisfied. To determine whether or not H3 does hold, we estimate a regression with the number of submissions of each participant in each period as the dependent variable and a random effect for each individual. The explanatory variables are identical to those in the performance regression employed to test H1. The results are reported in Table 4, Column 2. The positive and significant **pfp** dummy (p = 0.000) indicates that participants at the mean level of risk aversion submit more responses under *pfp* than under *fs*. The negative and significant interaction between **ra** and **pfp** (p = 0.037) demonstrates that the difference between the number of submissions under *pfp* and the corresponding number under *fs* is lower for individuals with higher levels of risk aversion, thus corroborating H3. The significant relationship between **period** and **pfp** (p = 0.001) points to more submissions under *pfp* relative to *fs* in later than in earlier periods.

H4 is examined by adding **number of submissions** centered at its mean in the *fs* treatment as an independent variable to the previously estimated random-effects regression with performance as the dependent variable. We also add interactions of **number of submissions** with **pfp**, **number of submissions** with **ra**, and **number of submissions** with **ra** and **pfp**. The results are reported in Table 4 Column 3. Not surprisingly, **number of submissions** has a strong effect on performance (p = 0.000), corroborating H4. The support for H3 and H4 taken together suggests that risk aversion affects performance at least in part through its differential effects on effort level represented by the number of submissions under *pfp* relative to *fs* as predicted by the optimal effort theory. This regression also makes it clear that the number of submissions, performance is significantly higher in the *pfp* periods (p = 0.000). Moreover, the effect of increasing the **number of submissions** on performance is significantly greater under *pfp* than under *fs* (p =0.005). It would seem that submissions involve more care and focus when success or failure affects one's financial payoff, i.e. not only the quantity but also the quality of effort is higher

 $^{^{22}}$ This result is robust to dropping the insignificant gender control from the regression. Not surprisingly, it also holds under *fs*. However, under *fs*, the relationship between *ex-post* performance variance and number of submissions is not central to our story because the performance variance is not associated with any variance in pay. These results are all available from the authors upon request.

under *pfp* than under *fs*.

Moreover, this analysis demonstrates that although risk aversion does affect performance via its effect on effort quantity represented by the number of submissions as predicted by the optimal effort theory, this is only part of the story of how risk aversion affects performance. In particular, after controlling for the **number of submissions**, ra is positive and significant (p = 0.000), while its interaction with **pfp** is negative and significant (p = 0.000). Thus, for a given number of submissions, a more risk-averse participant generally performs better than a less risk-averse person under fs. This effect is significantly mitigated under pfp. In addition, an increase in the number of submissions has a larger marginal effect on performance for participants who are more risk-averse (p = 0.000), but this effect is significantly reduced under *pfp* compared to *fs* (*p* = 0.018). This suggests that highly risk-averse people put more care into their attempts under both payment schemes, but that the relationship between risk aversion and effort quality, represented by the marginal effectiveness of effort, is particularly strong under fs. Together these results translate into less improvement in performance for more risk-averse participants when *pfp* is compared with fs through a channel that we did not anticipate. It would appear that more riskaverse participants make more serious attempts to achieve successful outcomes under fs than those who are less risk-averse. We conjecture that risk-aversion may be correlated with a more conscientious or self-motivated personality.²³ It is also possible more risk-averse participants are motivated by concern that the experimenter might impose unannounced ex-post penalties for poor performance under fs. However, we have no evidence that either of these conjectures is the case, and they remain a subject for further study.

We examine H5 by running a regression of the difference in reported stress levels under *pfp* versus *fs* for each participant on individual risk-aversion levels with the **male** dummy as a control variable. Table 3 Column 3 reports the results. Individual risk-aversion levels are positively and significantly related to the difference in reported stress levels (p = 0.002), thus corroborating H5. Thus, more risk-averse participants appear to be conscious of feeling a larger increase in stress when working under *pfp* versus *fs*.

Are these larger increases in stress levels associated with the effects of risk aversion levels

²³ Dohmen et al. (2010) show that risk-aversion is inversely related to cognitive ability as measured by two submodules of an IQ test. This would suggest that more risk-averse participants might, for a given amount of effort, perform more poorly and that the marginal effect of effort would be lower for those with higher levels of risk-aversion. However, this was not the case under either pay scheme.

on performance? To examine this question we test H6, initially by running a random-effects regression identical to the one used to test H1, but adding self-reported stress levels (stress) as an explanatory variable. For each participant, there are two self-reported stress levels: one for fs and one for *pfp*. We also try a second specification in which we add an interaction between stress and **pfp**. The results of both regressions are reported in Table 5 Columns 1 and 2. The results of the two regressions are very similar. However, adding the interaction suggests that it is only the self-reported stress numbers under *pfp* that have any effect on performance since the effect of stress under fs becomes insignificant once the interaction with **pfp** is taken into account. **Stress** is negative and significant in the initial regression (p = 0.023), while the interaction is negative and significant (p = 0.003) when it is added. Thus, higher stress levels under *pfp* are associated with decrements in performance.²⁴ In both specifications, the interaction of **ra** with **pfp** continues to be negative and significant (p = 0.002 in both cases), though the absolute value of the coefficient is lower than in the similar regression that did not contain stress as an explanatory variable. In Table 5 Column 3, we add a three-way interaction between ra, pfp, and stress. It is negative and significant (p = 0.003). However, the other coefficients of the regression change little. This suggests that the negative effect of risk aversion on performance improvement under *pfp* is magnified when reported stress levels are higher, or alternatively that the negative effect of reported stress levels on performance improvement under *pfp* is greater for more risk-averse participants. Since the two-way interaction between of ra and pfp remains significant in all three regressions, we can conclude that although risk aversion does appear to have a negative impact on performance in part through its association with greater increases in stress under *pfp*, this is only part of the explanation for the inverse relationship between risk aversion and performance improvement under pfp.

Insert Table 5 about here.

Does the pathway from higher levels of risk aversion through increased feelings of stress under *pfp* lead to reduced performance improvement under *pfp* by reducing the quantity of effort, the quality of effort, or through a combination of both? In Baumeister's (1984) classic article on choking under pressure, choking did not affect the quantity of effort, but caused decrements in

²⁴ A referee suggested that stress levels might be related to decrements in performance because they may be correlated to one's perceived proficiency at the task. We regress the difference in reported stress levels under *pfp* versus *fs* on our proficiency proxy defined in the second paragraph of section 4.2 to get at the idea that knowledge of one's own ability may influence stress. It is not significant (p=0.284). We perform an analogous regression for the stress level under *pfp*. Again it is not significant (p=0.651).

performance by a reduction in effort quality, making effort less effective. H7 puts forward the opposite hypothesis. It examines whether increased stress under *pfp* is associated with a reduced quantity of effort (number of submissions) by adding stress and an interaction of stress with pfp to the random-effects regression with number of submissions as a dependent variable used to test H3. A second specification adds a three-way interaction between ra, pfp, and stress to the regression. The results, reported in Table 5 Columna 4 and 5, show that as with its effect on performance, stress interacted with pfp is significant (p = 0.003), supporting H7. In contrast to the results of the performance regression, once the stress and interaction of stress with pfp are added to the submissions regression, the interaction of ra with pfp is no longer significant when evaluated at the mean reported stress level under pfp. However, the three-way interaction is negative and significant (p = 0.021), indicating that at very high stress levels risk aversion might still be associated with a reduced number of *pfp* submissions. Thus, higher levels of risk aversion are associated with both feelings of heightened stress when such a person is forced to work under the financial uncertainty of *pfp* and highly correlated behavior that mitigates risk through a reduction in submissions. This reduction in submissions might also occur because the more riskaverse people who are experiencing stress realize that unless they slow down, their accuracy rate may suffer. Thus, they could also be slowing down to mitigate a decrease in accuracy due to stress, further reducing the stressful risk they are experiencing.²⁵ This differs from Baumeister's (1984) choking under pressure, which led to decrements in performance without affecting effort levels.

H8 hypothesizes classic choking under pressure. We examine this by adding **stress** and the interaction of **stress** with **pfp** to the random-effects regression of performance used previously to test H4. Since this regression controls for the number of submissions, any further adverse effects of **stress** on *pfp* relative to *fs* performance could be interpreted as choking under pressure. However, as Table 5 Column 6 shows, no such effects are significant. Adding the three-way interaction between **ra**, **pfp**, and **stress** also indicates no significant effect on performance.²⁶ There appears to be no choking under pressure. H8 receives no support. Higher levels of stress for more risk-averse participants under *pfp* are associated with decrements in performance only insofar as they are associated with decrements in effort. However, it must be noted that the

²⁵ We thank an anonymous referee for this observation.

²⁶ We do not report the details of this regression here to save space, but the results are available from the authors upon request.

absence of choking under pressure in our experiment may be due to the moderate financial stakes involved coupled with the possibility noted in footnote 9 that small amounts of stress may aid performance while larger amounts may be required to impede it, known as the inverted U theory. Ariely et al. (2009) have shown that pay-for-performance compensation with very large stakes does impede performance on a fixed number of attempts for various real-effort tasks.

The results of this regression also reinforce the observations made previously in conjunction with the analysis of H4. Risk aversion affects performance improvement under pfp through a channel not anticipated by either the optimal-effort or choking-under-pressure hypotheses. In particular, for a given number of submissions, more risk-averse participants seem to exhibit a higher quality of effort and thus a higher marginal product of increased effort under fs. This leaves less room for improvement under pfp.

5. DISCUSSION AND CONCLUSIONS

In this paper, focusing on the role of performance-contingent incentives, we examined the interrelation among individual levels of risk aversion, effort level, effort quality, perceived stress and the effect of financial incentives on performance. To our knowledge, the discussion-paper version of this study presented the first theoretically-based laboratory experiment employing salient financial incentives to link attitudes toward risk with the effectiveness of performance pay at increasing productivity, and to examine the mechanism underlying this link.²⁷ Subsequent to the appearance of our paper, an interesting study by Zubanov (2015) used a quite different experimental design to compare the relationship between risk aversion and effort level under a piece rate with their relationship under a system that requires that a production target be met before any financial compensation is paid. Like us, Zubanov (2015) also finds evidence that higher levels of risk aversion reduce effort levels under piece rates relative to effort levels under fixed-salary compensation when risk is positively correlated with effort.

Our experiment examines the mechanisms through which risk aversion affects the improvement in performance under *pfp*. It presents evidence suggesting that more risk-averse people retreat from risk when it is correlated with effort level by reducing their number of attempts at a task even though by doing so they may lower their expected performance and hence

²⁷ Cadsby et al. (2009), the first discussion-paper version of this study used data from an earlier experiment. Unfortunately, some of the original data were lost because of a stolen laptop. Since it is important to be able to make one's data available to other researchers, we decided it was necessary to redo the experiments. The data from those new experiments are now reported in this revised version of the original study.

their expected pay. Furthermore, we show theoretically and empirically that performance may actually worsen under *pfp*, and that the probability of this occurring is higher, *ceteris paribus*, the more risk-averse a person is. While empirically resembling the crowding out of intrinsic by extrinsic motivation, the underlying reason for this phenomenon is not a loss of intrinsic motivation, but rather a reduction in effort levels or submissions by risk-averse people in order to reduce financial risk.

Our experiment also shows that more risk-averse people experience a greater increase in stress than less risk-averse people when moving from *fs* to *pfp*. Contrary to our expectations however, this increased stress did not lead to decrements in performance for a given level of effort as in Baumeister's (1984) seminal work on choking under pressure. Instead it was correlated with the reductions in submissions described above, which in turn led to decrements in performance. Thus, the economic theory of effort adjustment to reduce risk and the psychological theory suggesting that increased stress when stakes are high may be correlated with adverse effects on performance both receive support from our data. The economic theory is corroborated when one focuses attention on the number of submissions as utility-maximizing behavior toward risk, while the psychological theory is supported when one focuses on the emotions that accompany this behavior.

Finally and importantly, the results of our experiment suggest an unanticipated mechanism through which higher levels of risk aversion may be associated with less performance improvement under *pfp*. Higher levels of risk aversion are associated empirically with better performance for a given level of effort as well as a higher marginal product of increased effort under *fs* compensation. This leaves less room for performance to improve under *pfp*. Whether this effect is due to an association between a greater aversion to risk and a more conscientious personality, a desire by the more risk-averse to avoid any possibility of the experimenter reacting adversely to a poor performance, or some other factor is something to be explored in future work.

These results are important both in theory and in practice. Theoretically, they suggest that response to financial incentives depends not only on context, but also on individual heterogeneity, and in particular on individual attitudes toward risk. This warrants further theoretical and empirical study, focusing on four important issues.

The first issue concerns the generalizability beyond the lab of the results in this study. The experimental design involved each participant working first under one compensation scheme and

then under the other during periods 1 and 2. This pattern was repeated for periods 3 and 4, 5 and 6, and 7 and 8. This design was motivated by a number of factors. First, we needed the participants to experience several periods under each pay scheme and in particular under the *pfp* scheme so that we could get an *ex-post* measure of performance variance to examine whether more attempts were correlated with more risk as hypothesized. Second, it was important to be able to separate potential learning or boredom effects from the effects of changes in payment scheme as much as possible for each subject. If, for example, we had designed the experiment so that all four periods under one scheme preceded all four periods in the other, the treatment effects of interest would have been more difficult to isolate from the learning or boredom that could occur over the initial four periods of play. Third, given that we were in the lab and not able to alter a compensation scheme that would significantly affect a person's livelihood, we felt it necessary to move between pay schemes several times to ensure participants became conscious of the implications of the two schemes for their experimental earnings. The purpose of this design was not to create a real-world workplace in the lab, but to investigate whether the predictions from our economic model and the psychology literature on choking-under-pressure could be corroborated under the controlled conditions that we created for that purpose in the laboratory.²⁸

How robust are these results to different experimental designs and how generalizable are they to real-world workplaces and compensation schemes? In one such workplace, an employee might work for a long period of time under a fixed-salary scheme before a single once-and-for-all change to a pay-for-performance scheme. In another such workplace, an employee might work under a fixed salary much of the time, but at certain times of year have that salary supplemented by a sales contest using some form of performance pay. It is important to investigate the extent to which the results observed under our laboratory design with alternating treatments would hold up under such different circumstances both in the lab and more importantly in real-world workplaces.²⁹ For example, many of the less risk-averse among our

²⁸ Charness and Kuhn (2011) contains an excellent discussion of experimental design in general and designing principal-agent laboratory experiments in particular.

²⁹ We examine the robustness of our results by redoing the main analysis using only the data from periods 1 and 2, thus removing the subsequent data affected by the alternating-treatment design. While the direction of the interaction between risk-aversion level and the incentive effect of *pfp* remains negative, the coefficient is not significant (*p*=0.211). This reflects the fact that the inverse relationship between risk aversion and the treatment effect of *pfp* may strengthen somewhat over the course of the experiment. This can be seen by noting that in all of the panel-data specifications in Tables 4 and 5, the three-way interaction between risk-aversion, *pfp* treatment and

participants seemed to work more carelessly than their more risk-averse counterparts under *fs*, leaving more room for improvement under *pfp*. Was this due to a desire to conserve energy for the *pfp* period coming up next under our alternating-treatment design, or would a less risk-averse be more likely than a more risk-averse person to work more carelessly under a fixed salary that s/he expects to be long-lasting or permanent in a real-world workplace? Similarly, our design abstracted from peer effects and long-term experience and relationships on the job. The extent to which these factors might mitigate or alter our results is another important topic related to the generalizability of our results.

The second issue is to examine what mechanisms beyond the specific ones proposed and supported by the data in this paper may lead to a similar inverse relationship between risk aversion and productivity improvement under pfp. For example, our theoretical model only allows more risk-averse people to choose less risky strategies through a reduction in effort levels represented by the number of submissions. For this mechanism to work, the number of submissions and risk exposure must be positively correlated. However, such a correlation is not a necessary condition for higher levels of risk aversion to lead to less productivity improvement under pfp in real world settings. Any production process that allows each agent to select a level of risk exposure and also exhibits a positive correlation between that risk exposure and expected output can produce a similar result. For example, consider a situation where each agent must select from a set of available projects or work strategies. Assume those projects with greater risk exposures also have higher levels of expected net present value, as in Sung (1995). Effort is chosen independently of risk. Nonetheless, since more risk-averse agents can control risk exposure by choosing less risky projects with lower net present value, an inverse relationship between productivity improvement under pfp and risk aversion is quite likely.

Another example of how risk aversion could affect productivity improvement under *pfp* comes from an experimental paper by Oswald et al. (2015). Employing the same arithmetic task used in our study, it provides evidence that happier employees are more productive under a piece-rate system. If less risk-averse employees are happier than those who are more risk-averse

period (**ra_period_pfp**) is negative albeit not significant at conventional levels. Still this is suggestive of a growing understanding of the implications of the different payment schemes for these student subjects. The question is whether the effect of risk attitude on *pfp* versus *fs* performance improvement is a product solely of the alternating-treatment design, or would also appear with a once-and-for-all change in compensation design in a workplace where compensation represents the livelihood of the employees.

to work under the uncertainty of *pfp*, this may be another channel, related to but not necessarily the same as stress, through which risk-aversion may affect performance improvement under *pfp*.

A third issue worthy of study is to examine whether the themes examined in this paper may help elucidate and elaborate on other empirical findings. For example, an important study by Ariely et al. (2009) shows that very high monetary rewards can significantly reduce performance. The authors attribute this effect to stress and choking under pressure. Such choking may be related to increased risk aversion at higher stakes (Holt and Laury, 2002). It would be interesting to explore whether the extent of this kind of choking is also related to individual attitudes toward financial risk.

A fourth important issue is to explore how the effectiveness of different kinds of pay for performance may be influenced by the risk attitudes of those working under various compensation systems. Zubanov (2015) presents an interesting first attempt in this direction. More work exploring different kinds of bonus and tournament schemes along this dimension would provide a valuable new perspective for designers of compensation systems.

Does self-selection into jobs and compensation schemes lead to people selecting the scheme in which they would be most productive? Dohmen (2008), in a fascinating study of professional football (soccer) players in penalty kick situations finds that choking under pressure increases in front of home crowds, but does not increase in situations for which the stakes are higher. As he points out, the players that specialize in penalty kicks are likely to have self-selected into this task because of their ability to deal well with the resulting stress. However, the fact that many employees self-select into professions and compensation schemes does not imply that all or even most employees have selected the compensation schemes that best match their risk preferences. As Cadsby et al. (2007) demonstrate, while risk aversion is significantly and inversely related to the selection of a *pfp* compensation scheme, it is not the only factor that goes into such a choice. For example, a very risk-averse person may choose a *pfp* scheme if s/he expects to earn a lot more under *pfp* than *fs* due to his/her skill at the work task. However, his/her strong dislike of risk may nonetheless cause him/her to perform worse under *pfp* than under *fs*. S/he can make more money for him/herself under *pfp*, and yet produce less output for the company than under *fs*.

Practically, whether or not employees can self-select into compensation schemes, our results suggest that more risk-averse employees are likely to be less responsive, unresponsive, or even respond in the wrong direction when faced with performance-contingent pay. Since the stakes

are bound to be higher in the workplace than in the laboratory, and risk aversion is likely to be greater at higher stake levels (Holt and Laury, 2002), this phenomenon could well be even more pronounced in the workplace than in the laboratory. To determine whether or not this is so requires further research. However, our results in the lab suggest that different kinds of pay schemes may suit different kinds of workers, and risk attitudes may be a critical factor in determining the best employee-compensation fit.

Acknowledgements: We gratefully acknowledge funding from SSHRC grants #410-2007-1380 (Cadsby) and #2006-410-138 (Song). We are greatly indebted to J. Atsu Amegashie for many important and stimulating discussions and suggestions. We are also grateful to James Cox, Mike Hoy, Ling Liu, Alex Maynard, Thanasis Stengos and seminar audiences at the University of Guelph, University of Waterloo, University of Tasmania and University of Sydney for many helpful suggestions and comments. Thoughtful comments and suggestions by Sebastian Goerg and John Hamman, the Volume Editors of Research in Experimental Economics, vol. 19, and three anonymous referees helped us to improve the paper.

REFERENCES

- Abowd, John M. 1990. "Does Performance-Based Managerial Compensation Affect Corporate Performance?" *Industrial and Labor Relations Review*, 43(3): 52–74.
- Ariely, Dan, Uri Gneezy, George Loewenstein, and Nina Mazar. 2009. "Large Stakes and Big Mistakes." *Review of Economic Studies*, 76(2): 451–469.
- Baker, George P., Michael C. Jensen, and Kevin J. Murphy. 1988. "Compensation and Incentives: Practice vs. Theory." *Journal of Finance*, 43(3): 593–616.
- Baker, George P., and Bjorn Jorgensen. 2003. "Volatility, Noise, and Incentives." http://www.people.hbs.edu/gbaker/oes/papers/Baker_Jorgensen.pdf.
- Banker, Rajiv D., Seok-Young Lee, and Gordon Potter. 1996. "A Field Study of the Impact of a Performance-Based Incentive Plan." *Journal of Accounting and Economics*, 21(2): 195–226.
- Bénabou, Roland, and Jean Tirole. 2003. "Intrinsic and Extrinsic Motivation." *Review of Economic Studies*, 70(3): 489–520.
- Baumeister, Roy F. 1984. "Choking under Pressure: Self-Consciousness and Paradoxical Effects of Incentives on Skillful Performance." *Journal of Personality and Social Psychology*, 46(3): 610–620.
- Baumeister, Roy F., and Carolin J. Showers. 1986. "A Review of Paradoxical Performance Effects: Choking under Pressure in Sports and Mental Tests." *European Journal of Social Psychology*, 16(4): 361–383.
- Bonner, Sarah E., Reid Hastie, Geoffrey B. Sprinkle, and S. Mark Young. 2000. "A Review of the Effects of Financial Incentives on Performance in Laboratory Tasks: Implications for Management Accounting." *Journal of Management Accounting Research*, 12(1): 19–57.
- Brase, Gary L. 2009. "How Different Types of Participant Payments Alter Task Performance." *Judgment and Decision Making*, 4(5): 419–428.
- Bushman, Robert, Raffi Indjejikian, and Mark Penno. 2000. "Private Pre-Decision Information,

Performance Measure Congruity and the Value of Delegation." *Contemporary Accounting Research*, 17(4): 561–587.

- Cadsby, C. Bram, Fei Song, and Francis Tapon. 2007. "Sorting and Incentive Effects of Pay for Performance: An Experimental Investigation." *Academy of Management Journal*, 50(2): 387–405.
- Cadsby, C. Bram, Fei Song, and Francis Tapon. 2009. "The Impact of Risk Aversion and Stress on the Incentive Effect of Performance Pay." Working Paper, University of Guelph.
- Cadsby, C. Bram, Fei Song, and Zubanov, Nikolay, 2016. "The 'Sales Agent' Problem: Effort Choice under Performance Pay as Behavior Toward Risk." Unpublished Manuscript.
- Cadsby, C. Bram, Maros Servatka, and Fei Song. 2013. "How Competitive are Female Professionals? A Tale of Identity Conflict." *Journal of Economic Behavior and Organization*. 92: 284–303.
- Camerer, Colin, and Robin M. Hogarth. 1999. "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework." *Journal of Risk and Uncertainty*, 19(1-3): 7–42.
- Charness, Gary, and Peter Kuhn. 2011. "Lab-Labor: What Can Experimental Economists Learn from the Lab?" in *Handbook of Labor Economics*, Volume 4, Part A, edited by Orley Ashenfelter and David Card: 229–330.
- Cohen, Jacob, Patricia Cohen, Stephen Aiken, and Leona West. 2003. *Applied Multiple Regression/Correlation Analysis for the Behavior Sciences*. NJ: Lawrence Erlbaum Associates.
- Cooper, Cary, Stephen Sloan, and Stephen Williams. 1988. Occupational Stress Indicator Management Guide. Oxford, UK: NFER-Nelson.
- Deci, Edward L. 1971. "Effects of Externally Mediated Rewards on Intrinsic Motivation." *Journal of Personality and Social Psychology*, 18(1): 105–115.
- Deci, Edward L., and Richard M. Ryan. 1985. *Intrinsic Motivation and Self-Determination in Human Behavior*. New York: Plenum.
- Dohmen, Thomas. 2008. "Do Professionals Choke under Pressure?" *Journal of Economic Behavior and Organization*, 65(3-4): 636–653.
- Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde. 2010. "Are Risk Aversion and Impatience Related to Cognitive Ability?" *American Economic Review*, 120: 256-271.
- Eriksson, Tor, Anders Poulsen, and Marie Claire Villeval. 2009. "Feedback and Incentives: Experimental Evidence." *Labor Economics*, 16: 679–688.
- Fehr, Ernst, and Lorenz Goette. 2007. "Do Workers Work More if Wages are High? Evidence from a Randomized Field Experiment." *American Economic Review*, 97(1): 298–317
- Fessler, Nicholas J. 2003. "Experimental Evidence on the Links among Monetary Incentives, Task Attractiveness, and Task Performance." *Journal of Management Accounting Research*, 15(1): 161–182.
- Fleming, India, and Andrew Baum. 1987. "Stress: Psychological assessment." Journal of Organizational Behavior Management, 8 (2), 117-140.
- Frey, Bruno S., and Reto Jegen. 2001. "Motivation Crowding Theory: A Survey of Empirical Evidence." *Journal of Economic Surveys*, 15(5): 589–611.
- Fried, Yitzhak, Kendrith Rowland, and Gerald Ferris. 1984. "The physiological measures of work stress: A critique." *Personnel Psychology*, 37(4): 583-615.
- Gerhart, Barry, and George T. Milkovich. 1990. "Organizational Differences in Managerial Compensation and Financial Performance." *Academy of Management Journal*, 33(4): 663–

691.

- Gerhart, Barry, and Sara Rynes. 2003. *Compensation: Theory, Evidence, and Strategic Implications*. Thousand Oaks, CA: Sage.
- Gibbons, Robert. 1998. "Incentives in Organizations." *Journal of Economic Perspectives*, 12(4): 115–132.
- Gneezy, Uri, and Aldo Rustichini. 2000. "Pay Enough or Don't Pay at all." *Quarterly Journal of Economics*, 115(3): 791–810.
- Gomez–Mejia, Luis R., and Theresa M. Welbourne. 1988. "Compensation Strategy: An Overview and Future Steps." *Human Resource Planning*, 11(3): 173–190.
- Hart, Oliver 1989. "An Economist's Perspective on the Theory of the Firm." *Columbia Law Review*, 89(7): 1757–1774.
- Holmstrom, Bengt. 1979. "Moral Hazard and Observability." *Bell Journal of Economics*, 10(1): 74–91.
- Holt, Charles and Susan Laury. 2002. "Risk Aversion and Incentive Effects." *American Economic Review*, 92(5): 1644–1655.
- Jenkins, Douglas. G. Jr., Atul Mitra, Nina Gupta, and Jason D. Shaw. 1998. "Are Financial Incentives Related to Performance? A Meta-Analytic Review of Empirical Research." *Journal of Applied Psychology*, 83(5): 777–787.
- Jensen, Michael C., and William H. Meckling. 1976. "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure." *Journal of Financial Economics*, 3(4): 305–360.
- Jensen, Michael C., and Kevin J. Murphy. 1990. "Performance Pay and Top Management Incentives." *Journal of Political Economy*, 98(2): 225–264.
- Kahn, Lawrence M., and Peter D. Sherer. 1990. "Contingent Pay and Managerial Performance." *Industrial and Labor Relations Review*, 43(3): 107–20.
- Kale, Jayant R., Ebru Reis, and Anand Venkateswaran. 2009. "Rank Order Tournaments and Incentive Alignment: The Effect on Firm Performance." *Journal of Finance*, 64(3): 1479– 1512.
- Kreps, David. 1997. "Intrinsic Motivation and Extrinsic Incentives." *American Economic Review*, 87(2): 359–64.

Lazarus, Richard. 1966. Psychological Stress and the Coping Process. New York: McGraw Hill.

- Lazarus, Richard, and Susan Folkman. 1984. Stress Appraisal and Coping, 1st Edition. New York: Springer
- Lazear, Edward P. 1986. "Salaries and Piece Rates." Journal of Business, 59: 405-432.
- Lazear, Edward P. 2000a. "The Power of Incentives." *American Economic Review*, 90(2): 410–415.
- Lazear, Edward P. 2000b. "Performance Pay and Productivity." *American Economic Review*, 90(5): 1346–1361.
- LePine, Jeffrey A., Nathan P. Podsakoff, and Marcie A. LePine. 2005. "A Meta-Analytic Test of the Challenge Stressor–Hindrance Stressor Framework: An Explanation for Inconsistent Relationships Among Stressors and Performance." *Academy of Management Journal*, 48(5): 764–775.
- Milgrom, Paul, and John Roberts. 1992. *Economics, Organization and Management*. Englewood Cliffs, NJ: Prentice Hall.
- Muse, Lori A., Stanley G. Harris, and Hubert S. Feild. 2003. "Has the Inverted-U Theory of Stress and Job Performance Had a Fair Test?" *Human Performance*, 16(4): 349–364.
- Niederle, Murial, and Lise Vesterlund. 2007. "Do Women Shy Away from Competition? Do

Men Compete Too Much?" The Quarterly Journal of Economics, 122: 1067–1101.

- Oswald, Andrew J., Eugenio Proto, and Daniel Sgroi. 2015. "Happiness and Productivity." *Journal of Labor Economics*, 33(4): 789-822.
- Paarsch, Harry J., and Bruce S. Shearer. 2009. "The Response to Incentives and Contractual Efficiency: Evidence from a Field Experiment." *European Economic Review*, 53(5): 481-494.
- Paarsch, Harry J., and Bruce S. Shearer. 2007. "Do Women React Differently to Incentives? Evidence from Experimental Data and Payroll Records." *European Economic Review*, 51(7): 1682–1707.
- Paarsch, Harry J., and Bruce S. Shearer. 2000. "Piece Rates, Fixed Wages, and Incentive Effects: Statistical Evidence from Payroll Records." *International Economic Review*, 41(1): 59–92.
- Paarsch, Harry J., and Bruce S. Shearer. 1999. "The Response of Worker Effort to Piece Rates: Evidence from the British Columbia Tree-Planting Industry." *Journal of Human Resources*, 34(4): 643–667.
- Parker, Donald F., and Thomas A. DeCotiis. 1983. "Organizational Determinants of Job Stress." *Organizational Behavior and Human Performance*, 32(2): 160–177.
- Prendergast, Canice. 1999. "The Provision of Incentives in Firms." *Journal of Economic Literature*, 37(1): 7–63.
- Quick, James, and Jonathan Quick. 1984. Organizational Stress and Preventive Management. New York: McGraw-Hill.
- Rynes, Sara, Barry Gerhart, and Laura Parks. 2005. "Personnel Psychology: Performance Evaluation and Pay for Performance." *Annual Review of Psychology*, 56: 571–600.
- Seiler, Eric. 1984. "Piece Rate versus Time Rate: The Effects of Incentives on Earnings." *Review of Economics and Statistics*, 66(3): 363–376.
- Serfes, Konstantinos. 2008. "Endogenous Matching in a Market with Heterogeneous Principals and Agents." *International Journal of Game Theory*, 36: 587–619.
- Shearer, Bruce S. 2004. "Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment." *Review of Economic Studies*, 71(2): 513–534.
- Schuler, Randall S. 1980. "Definition and Conceptualization of Stress in Organizations." *Organizational Behavior and Human Decision Processes*, 25(2): 184–215.
- Sloof, Randolph and C. Mirjam van Praag. 2008. "Performance Measurement, Expectancy and Agency Theory: An Experimental Study." *Journal of Economic Behavior and Organization*, 67(3-4): 794–809.
- Stajkovic, Alexander D., and Fred Luthans. 2001. "Differential Effects of Incentive Motivators on Work Performance." *Academy of Management Journal*, 44(3): 580–590.

Sung, Jaeyoung. 1995. "Linearity with Project Selection and Controllable Diffusion Rate in Continuous-Time Principal-Agent Problems." *RAND Journal of Economics*, 26(4): 720–743.

- Vroom, Victor H. 1964. Work and Motivation. New York: Wiley.
- Zubanov, Nikolay. 2015. "Risk Aversion and Effort in an Incentive Pay Scheme with Multiplicative Noise: Theory and Experimental Evidence." *Evidence-based HRM: A Global Forum for Empirical Scholarship,* 3: 130-144.



Figure 1 Average performance in *fs* and *pfp* periods (fvag and pavg) on risk-aversion

Tuble I Duta Summary Theuns and Standard El	1015 been een	Subjects
Variable	Mean	Standard Error
<i>pfp</i> periods performance	26.01	0.96
fs periods performance	19.11	1.01
<i>pfp-fs</i> performance difference	6.90	0.86
ra (risk-aversion)	6.47	0.19
<i>pfp</i> periods stress experienced	4.44	0.19
fs periods stress experienced	2.62	0.18
<i>Pfp-fs</i> stress experienced difference	1.81	0.20
<i>pfp</i> periods number of submissions (effort level)	31.61	1.06
<i>fs</i> periods number of submissions (effort level)	27.91	1.33
Male	0.52	0.05
Born-abroad	0.19	0.43

Table 1 Data	Summary:	Means an	d Standard	l Errors	between	Subjects
	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	1.1.0.000000000000000000000000000000000			~~~~~	

Note: These statistics are for all subjects including the 10 subjects who made inconsistent risk choices. Following Holt and Laury (2002), we use the number of safer lottery choices as our measure of risk aversion. Its mean is reported here.

ra	3	4	5	6	7	8	9	10
performance								
pfp	6.14	6.15	6.82	7.64	5.88	6.76	6.18	4.63
fs	4.04	3.30	4.45	5.55	4.63	5.29	4.79	5.75
pfp-fs	2.10	2.85	2.37	2.09	1.25	1.47	1.39	-1.12

Note: 85 observations.

Table	3 OLS	S and	Logit	Models:	Coeff	ïcients	with	Standard	Errors	in I	Parentheses

Table 5 OLS and Logit Models. Coefficients with Standard Errors in Latentineses										
	DV: pfp-fs	DV: Probability	DV: pfp-fs	DV: <i>pfp</i> performance						
	performance	of positive	stress	va	riance					
	difference	performance	difference							
		improvement								
		(Logit Model)								
ra (risk	-1.23**	-0.59**	0.34***	pfp	0.05**					
aversion)	(0.47)	(0.23)	(0.10)	submissions	(0.02)					
Male	1.93	-1.22*	0.73**	Male	0.35					
	(1.68)	(0.72)	(0.37)		(0.37)					
Constant	5.91***	2.74***	1.41***	Constant	0.19					
	(1.21)	(0.64)	(0.26)		(0.64)					
Obs.	85	85	85		85					
F-stat.	3.67		8.22		3.88					
$R^2$	0.08	0.18	0.17		0.09					

Note: ***, ** and * denote 1%, 5% and 10% respectively. Risk-aversion is centered at its mean. Male is a dummy variable equal to 1 for males and 0 for females.

	DV: performance	DV: submissions	DV: performance
ra (risk aversion)	0.23*	0.06	0.23***
	(0.13)	(0.15)	(0.06)
<i>pfp</i> treatment	1.48***	1.44***	0.45***
	(0.20)	(0.22)	(0.15)
period	0.03	-0.06	0.04
-	(0.04)	(0.05)	(0.03)
male	-0.11	0.46	-0.38*
	(0.48)	(0.54)	(0.22)
ra pfp	-0.31***	-0.18**	-0.28***
	(0.08)	(0.09)	(0.06)
period pfp	0.16**	0.24***	0.003
	(0.06)	(0.07)	(0.04)
ra period	0.04*	0.06**	0.01
	(0.02)	(0.03)	(0.02)
ra_period_pfp	-0.04	-0.06	-0.02
	(0.03)	(0.04)	(0.03)
male_pfp	0.48*	0.37	0.15
	(0.28)	(0.31)	(0.20)
# of submissions			0.65***
			(0.03)
submissions_pfp			0.11***
			(0.04)
submissions_ra			0.10***
			(0.02)
submissions_pfp_ra			-0.06**
			(0.02)
Constant	4.83***	5.98***	4.95***
	(0.34)	(0.39)	(0.16)
Obs.	680	680	680
$R^2$	0.12	0.09	0.72
Wald Chi ²	195.66	137.58	1273.90

Table 4 Random-Effects Models: Coefficients and Standard Errors in Parentheses

Note: ***, ** and * denote 1%, 5% and 10% significance respectively. Risk-aversion is centered at its mean. Period is centered at its mean. Submissions are centered at their mean in the *fs* treatment.

Table 5 Kandolii Elik				DI	7	DU
	DV:	D	V:		V :	DV:
	performance	perform	nance	submis	ssions	performance
ra (risk aversion)	0.20	0.23*	0.24*	0.07	0.08	0.23***
	(0.13)	(0.13)	(0.13)	(0.15)	(0.15)	(0.06)
<i>pfp</i> treatment	1.73***	1.47***	1.44***	1.33***	1.30***	0.34*
	(0.23)	(0.24)	(0.24)	(0.27)	(0.27)	(0.18)
period	0.03	0.03	0.03	-0.06	-0.06	0.05
	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.03)
male	-0.19	-0.11	-0.07	0.49	0.52	-0.35
	(0.48)	(0.48)	(0.48)	(0.54)	(0.54)	(0.22)
ra_pfp	-0.25***	-0.25***	-0.25***	-0.13	-0.13	-0.29***
	(0.08)	(0.08)	(0.08)	(0.09)	(0.09)	(0.07)
period pfp	0.16***	0.16***	0.15***	0.24***	0.24***	0.002
	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)	(0.05)
ra period	0.04*	0.04*	0.04*	0.06**	0.06**	0.01
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)
ra period pfp	-0.04	-0.04	-0.04	-0.06	-0.06	-0.02
	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)
male pfp	0.59**	0.52*	0.48*	0.38	0.35	0.11
	(0.28)	(0.28)	(0.28)	(0.31)	(0.31)	(0.21)
stress	-0.17**	-0.01	0.07	-0.06	0.12	0.06
	(0.07)	(0.09)	(0.09)	(0.10)	(0.11)	(0.05)
stress nfn	(****)	-0 29***	-0 36***	-0 39***	-0 46***	-0.04
put the put		(0.10)	(0.10)	(0.11)	(0.11)	(0.07)
-		(0.10)	(0.10)	(0.11)	(0.11)	(0.07)
stress_pfp_ra			-0.12***		-0.11**	
			(0.04)		(0.04)	
# of submissions						0.65***
						(0.03)
submissions_pfp						0.11***
						(0.04)
submissions_ra						0.10***
						(0.02)
submissions_pfp_ra						-0.06**
						(0.02)
Constant	4.56***	4.82***	4.93***	6.07***	6.16***	5.04***
	(0.34)	(0.37)	(0.37)	(0.42)	(0.42)	(0.18)
Obs.	680	680	680	680	680	680
$R^2$	0.10	0.11	0.14	0.10	0.11	0.72
Wald Chi ²	202.11	213.55	225.41	158.78	165.33	1289.78

Table 5 Random-Effects Models with Stress: Coefficients and Standard Errors in Parentheses

Note: ***, ** and * denote 1%, 5% and 10% significance respectively. Risk-aversion is centered at its mean. Period is centered at its mean. Submissions are centered at their mean in the *fs* treatment. Stress is centered at its mean in the *pfp* treatment.

# SUPPLEMENARY MATERIALS

# FOR REVIEWERS ONLY - NOT FOR PUBLICATION

# Instructions

Thank you for participating today. All of your responses in this study will remain completely anonymous. It is important that during this experiment you do not talk or make any noise that might disrupt others around you. If you have any questions, please raise your hand and the experimenter will answer your questions individually. During this experiment you will be asked to add up sets of five double-digit integers such as the following.

98	42	69	50	78	

The first round is a trial round, which will last for two minutes, for you to get familiar with the task while the rest of the rounds will be experimental rounds, which will be used to calculate your earnings as explained below. Each experimental round will last for three minutes. You are not allowed to use a calculator, but may write numbers down on scratch paper provided by us. The numbers are randomly drawn and each problem is presented as above.

You will have a *Workbook* that will contain all of your work. Your job is to solve as many problems as you can in each round. Your earnings in this experiment will depend on your performance and/or the specific compensation method applied to each of the experimental rounds. All of the experimental rounds will be used for payment. Once we begin the experiment, you will not be able to look ahead to future pages or to go back to previous pages. To ensure confidentiality, just write down your participant number on the cover of the *Workbook*. Please do not write your name on any of these materials.

Please make sure that you completely understand the instructions for the experiment. Once again, remember not to make any noises that might disturb others around you. If you have any questions, raise your hand and we will answer your questions individually.

**Fixed Salary:** You will earn \$1.75 in this round, regardless of the number of arithmetic problems you solve in this round. Thus, your total earnings for Round 2 will be: \$1.75.

**Piece-rate:** You will earn \$0.25 for each arithmetic problem you solve correctly. For example, if you correctly solve P problems in Round 1, your total earnings for Round 1 will be:  $$0.25 \times P$ .

# Sample Workbook Page

						Answer							Answer
Line 1	47	79	17	23	40		Line 21	97	26	51	71	17	
Line 2	78	16	11	97	80		Line 22	20	79	29	11	25	
Line 3	62	82	23	71	40		Line 23	81	83	94	52	31	
Line 4	55	39	57	53	62		Line 24	55	20	13	81	36	
Line 5	13	44	29	41	13		Line 25	81	63	33	11	62	
Line 6	60	38	79	18	47		Line 26	49	63	74	77	78	
Line 7	78	18	63	92	62		Line 27	22	19	59	47	13	
Line 8	75	40	18	35	90		Line 28	99	12	47	14	82	
Line 9	31	98	79	22	26		Line 29	72	42	96	57	99	
Line 10	35	68	59	15	72		Line 30	98	51	63	71	26	
Line 11	80	47	52	82	33		Line 31	52	95	89	99	84	
Line 12	42	72	43	20	49		Line 32	86	46	46	31	80	
Line 13	68	22	61	35	45		Line 33	75	78	34	68	89	
Line 14	73	30	25	75	34		Line 34	61	22	71	99	61	
Line 15	50	50	27	50	40		Line 35	18	29	27	54	31	
Line 16	49	39	66	91	30		Line 36	76	61	85	94	37	
Line 17	11	98	73	26	10		Line 37	72	17	68	30	75	
Line 18	22	15	19	91	87		Line 38	51	97	47	77	60	
Line 19	43	59	99	48	80		Line 39	40	64	21	30	86	
Line 20	45	38	78	19	13		Line 40	16	84	16	93	41	