Incentive Schemes, Sorting and Behavioral Biases of Employees: Experimental Evidence

Ian Larkin, Harvard Business School

Stephen Leider, Ross School of Business, University of Michigan

April, 2011

Abstract

We investigate how the convexity of a firm's incentives interacts with worker overconfidence to affect sorting decisions and performance. We demonstrate experimentally that overconfident employees are more likely to sort into a non-linear incentive scheme over a linear one, even though this reduces pay for many subjects and despite the presence of clear feedback. Additionally, the linear scheme attracts demotivated, underconfident workers who perform below their ability. Our findings suggest that firms may design incentive schemes that adapt to the behavioral biases of employees to "sort in" ("sort away") attractive (unattractive) employees; such schemes may also reduce a firm's wage bill.

Introduction

As economists' understanding of behavioral biases exhibited by individuals has deepened, an emerging literature has investigated how firms can best adapt their pricing, incentive and contract offerings in light of these biases. Several papers have studied how consumer biases affect the optimal pricing and product terms in settings such as health clubs (DellaVigna and Malmendier 2006), credit cards (Ausubel 1999), payday loans (Skiba and Tobacman 2008), magazine subscriptions (Oster and Scott-Morton 2005) and cellular phone services (Grubb 2009). For example, DellaVigna and Malmendier (2006) use data from three health club establishments to demonstrate that customers purchasing a monthly contract offering unlimited use of the gym pay more on average than they would if they bought a series of pay-per-visit contracts. This finding suggests some customers have commitment or estimation problems around their visits, and that gyms offer contracts which capitalize on these behavioral problems.

In employment settings several experiments have studied how firms can adapt incentives for their workers in response to reciprocity, including whether firms should offer gift exchange wages (Fehr and Falk 1999), whether incentives should be framed as a bonus or a fine (Fehr and Gächter 2002), and whether firms should rely on long-run employment relationships (Brown et al. 2004). Furthermore, Dohmen and Falk (forthcoming) investigate how several different forms of incentive pay interact with a variety of individual characteristics (such as risk tolerance, social preferences and overconfidence) to affect not just performance, but also worker sorting. Understanding the role of behavioral biases in employee sorting may be particularly important to firms for two reasons: first, employment contracts that take advantage of employee biases could better attract and retain employees; and second, these contracts might allow firms to hire employees at a lower cost compared to contracts that do not address the behavioral biases of employees.

In this research we use a laboratory experiment to investigate the relationship between sorting, the use of non-linear incentive schemes and overconfidence. The convex incentive scheme studied in this paper is commonly used by firms, and in particular our experiment models the form of incentives often seen in sales environments and in the compensation contracts offered by venture capital companies to entrepreneurs. For example, Larkin (2007) reports that at the software vendor investigated in the paper, the commission on a sale increased with a salesperson's quarterly sales performance, and could range from 2% to 25%, depending on the total sales the salesperson had made in that quarter[1]. Larkin reports that these "kinked" systems are common not only in enterprise software, but also in hardware, telecom equipment, aerospace and other industries. Wasserman (2006) shows that similar structures are often used for executives in early-stage companies.

We argue that overconfidence by employees – systematic upward bias in beliefs about one's ability (DellaVigna 2009) – may help explain why firms offer convex incentive schemes, and why some employees seem to prefer them. Overconfidence potentially interacts with convex incentive schemes in several ways. First and foremost, a convex scheme is arguably more attractive to potential employees who are overconfident, in that they overestimate the likelihood of receiving a commission rate from the steep portions of the curve. For some industries (over)confidence may be an important skill for employees to have. For example, a typical salesperson is rejected "hundreds if not thousands of times before he hears 'yes'" (Zoltners et al. 2006), and successful salespeople must therefore be highly confident of their abilities in the face of persistent negative feedback. In these settings workers with an overconfident disposition may be more productive.[2] In this case a convex incentive scheme could help attract and retain these desirable employees. Secondly, firms may reduce their wage bill by proposing convex schemes to biased employees, particularly if these employees fail to update their beliefs about their ability or future performance. For example, a survey of enterprise software salespeople discussed in

[1] The industry term for these pay schemes is an "accelerating" incentive system, because the piece rate paid accelerates as total performance increases.
[2] Englmaier (forthcoming) similarly argues that it may be desirable to hire overconfident (specifically overoptimistic) managers to commit to an aggressive R&D policy.

Gino et al. (2009) reports that the average salesperson at the company surveyed expected to make commissions over the course of a year that were nearly eight times larger than the realized average commission level. If these salespeople accurately reported their beliefs in the survey[3], the firm saves on the wage bill by sorting in employees who expect to make more than they actually do.

The attractiveness of convex pay schemes to overconfident employees is clear: by overestimating their likely performance, these employees believe they will earn more than they would in a typical linear pay scheme where pay does not accelerate as performance increases. For this argument to be robust, overconfident employees must not update their beliefs as to their ability, or must do so slowly. However, scholars know little about the extent to which overconfident employees update their beliefs in the face of either persistent feedback about actual performance, or lower wages earned by choosing a job environment with a convex incentive plan.

In our study, we allowed experimental subjects facing a standard multiplication task to choose how they wanted to be compensated: via a linear piece rate, or via a convex payout schedule that paid them more per correct answer as their productivity increased.[4] The choice continued over nine rounds. We tracked the choices of the subjects, whether their choice of scheme was correct (in that it maximized their pay), and their level of confidence measured in a number of different ways. We used two initial periods where all subjects were assigned the linear piece rate in order to gauge subject skill at multiplication, and to set individual convex payoff functions so that subjects expecting to perform at a similar level as the initial periods would rationally have a slight preference for the linear pay scheme. Additionally, in order to control for incentive effects, subjects in our control treatment were randomly

[3] This survey was, of course, non-incentivized and therefore the validity of responses is questionable. As we will discuss, our experimental design is somewhat similar to this survey approach, but with subjects incentivized to report their expectations accurately.

[4] We considered introducing the choice of a flat payment to the experiment, since this scheme might be most attractive to highly underconfident subjects. However, Dohmen and Falk (forthcoming) looks expressly at the choice between a flat wage and a piece rate and finds no correlation with confidence, and the introduction of a flat wage would introduce additional confounds and complexity. We therefore decided to look only at performance-based incentives.

assigned a pay scheme in each non-practice round, and their performance was compared to that of subjects who were given a choice of pay schemes.

Our results are striking. Subjects who are highly overconfident, as measured by their ex-ante performance expectations on the task, are 45 percentage points more likely to incorrectly choose the convex scheme. These mistakes cost the subjects 15% of their payoff in these rounds, suggesting that a firm could save on the wage bill by offering non-linear incentives and "sorting in" overconfident workers. We also find qualitatively similar results for other measures of overconfidence, such as results from an independent self-reported measure of a subject's propensity for relative overconfidence. Incentive effects do not explain our results; subjects who chose the convex scheme did not perform any differently from subjects who were randomly assigned the convex scheme. However, subjects who chose the linear scheme over the convex scheme performed 8% worse than subjects randomly assigned to the linear scheme, suggesting that offering only convex schemes in some work environments could help "sort away" employees who lack motivation.[5] Furthermore, the subjects who choose the linear scheme in all six periods, and perform the worst relative to their initial ability, are also somewhat more likely to be underconfident. Therefore, overconfident subjects may be more likely to maintain motivation throughout the task. We examined a large number of demographic or personality predictors of overconfidence with limited results; only being male, extroverted, predicting a high future salary, or desiring a career in management predicted overconfidence on the task.

Interestingly, this propensity to wrongly sort into the convex pay scheme does not go down with experience; overconfident subjects are just as likely to wrongly pick the convex scheme in round 9 as they are in early rounds. We reported performance at the end of each round, and it was therefore clear to subjects whether they chose the right scheme. In one treatment we gave even clearer feedback to subjects after the third choice period, indicating how much they would have earned under each scheme given their performance and telling them exactly how much money they lost by choosing the wrong scheme. Even

[5] It is very uncommon for firms to offer the choice of incentive scheme to employees for a certain job function. Our experiment is better thought of as two separate firms offering different incentive schemes to a potential employee.

this stronger feedback made no statistical difference to the level of overconfidence, nor to the propensity of overconfident subjects to mistakenly choose the convex scheme. Finally, we analyzed subjects' rate of learning about their ability as measured by changes in their reported expected performance. While subjects change their beliefs in early periods, a significant number of employees fail to update their beliefs in the last several periods, despite continuing to make substantial prediction errors.

We also investigated the extent to which overconfidence persisted across tasks by introducing a second, separate task. We gave subjects a set of trivia questions and measured overconfidence on this task by asking subjects to guess the number of questions they expected to get right, and to rate their degree of confidence in their answers to each trivia question. Overconfidence on overall performance in the trivia task was *not* predictive of choosing (correctly or incorrectly) the convex scheme in the multiplication task; however individuals who were highly confident about their answers to many questions that they actually answered incorrectly *were* more likely to mistakenly choose the convex scheme in the multiplication task. This is consistent with the social psychology literature, which suggests that overconfidence is influenced both by individual characteristics and by the specific task at hand[6].

We believe that our paper contributes to the extensive theoretical, experimental and empirical literature on the effects of overconfidence on economic decision-making. While we believe our results suggest future directions for research using archival data from firms, our experimental approach has several important strengths. First, we are able to directly elicit several measures of task-specific and general overconfidence, rather than having to infer overconfidence from proxy variables. Furthermore, because we observe individuals repeatedly we can show that this overconfidence is persistent, rather than transitory. Additionally, we can cleanly observe sorting decisions because we know all of the options each individual faces. Moreover, by isolating on a specific work task, we are able to rule out other forms of incentives (such as career concerns or reputational motivations) that might otherwise complicate worker behavior. Lastly, we are able to directly measure and control for alternative explanations, such as

[6] We briefly review the social psychology literature on overconfidence in the next section.

risk attitudes or individual ability, to more conclusively show that sorting decisions (and mistakes) are being driven by overconfidence. Of course, our study does not conclusively demonstrate that convex schemes are optimal for certain firms or job tasks. Under the task and payoff structure used in our experiment, the average payoff for subjects in the convex scheme was higher than that in the linear scheme. Our experiment does demonstrate, however, that there are potential benefits to firms that offer convex wage schemes to employees. In future research we hope to show the firm, employee and/or task conditions under which the "benefit [to firms] of [convex] contracts that outweighs [their] apparent costs" (Lazear and Oyer 2009).

Previous Literature

Pricing/Incentive Contracts and Sorting

The idea that firms use employment contracts as a mechanism to screen for worker type is an old one in economics, stretching back at least as far as Spence's seminal paper on job market signaling (Spence 1973). Indeed, the literature on agency, reviewed in Hart and Holmstrom (1987) and elsewhere, is fundamentally concerned with the principal's use of incentive contracts to align employee effort with desired outcomes. While the traditional argument in the literature holds that sorting mechanisms are important when skills are heterogeneous, a number of studies have examined how incentives interact with social preferences to influence employee performance and sorting (e.g. Fehr and Falk 1999; Fehr and Gächter 2002; Brown et al. 2004). In this paper we build on recent work such as Dohmen and Falk (forthcoming) in examining how sorting between incentive scheme type can be driven by other behavioral factors, such as individual overconfidence.

There is a similar emerging literature that links overconfidence and other behavioral biases of customers to pricing and contract terms offered by firms. DellaVigna and Malmendier (2006) show that many customers choosing an unlimited monthly contract appear to pay more for health club services than they would if they chose a contract that charged a per-visit fee. One interpretation of this finding is that customers are overconfident about their future likely use of the health club. Ausubel (1999) presents evidence that credit card companies offer “teaser” rates that appear to successfully take advantage of customer overconfidence about ability to restrain their spending. Skiba and Tobacman (2008) show data from payday loan companies that corroborate a model of customer overconfidence about their ability to borrow less in the future. Oster and Scott-Morton (2005) find evidence that subscription discounts vary systematically, with high-brow intellectual having lower discounts to extract rents from consumers who want to subscribe as a commitment device. Similarly, Grubb (2009) shows that data on three-part cell phone contract terms closely match a theoretical model on consumer overconfidence about future use.

Our study also contributes to the literature on non-linear incentive schemes. Empirical work has documented that these schemes are very costly (Oyer 1998; Larkin 2007). However, there has been little theoretical or empirical work on the benefits of such schemes, leading some scholars of incentives to suggest that there widespread use presents a puzzle to economists (Lazear and Oyer 2009)[7]. In this paper we examine how such non-linear (convex) schemes can influence employee sorting in the presence of overconfidence. In particular, we study convex *individual* incentives, such as those created by non-linear commissions, stock options or bonuses with performance hurdles; many previous studies of non-linear incentives have focused on incentives where the non-linearity is driven by *relative* incentives, such as tournaments.

Overconfidence

Studies in social psychology and experimental economics have demonstrated that individuals are overconfident on a wide variety of topics, including answering general knowledge questions (Fischhoff et al. 1977), predicting horse races (Fischhoff and Slovic 1980), diagnosing the malignancy of ulcers (Fischhoff and Slovic 1980), being a good driver (Svenson 1981), and the likelihood of encountering a variety of positive and negative life events (Weinstein 1980). The literature has generally considered

[7] There is a large experimental literature on the positive motivational effects of goal setting; however, many of these studies did not involve incentive differences if a goal was reached (Locke and Latham 2002). Also, many instances of non-linear incentives do not use explicit goals. There are many sensible explanations for non-linear incentive schemes that are not centered on employee overconfidence, but there is little theoretical or empirical work on the benefits of highly non-linear schemes. This is in contrast to the many studies documenting the costly behavior engendered by non-linear incentive schemes.

three distinct forms of overconfidence (Healy and Moore 2007): *absolute overconfidence*, which consists of overly positive beliefs about one's future performance (sometimes called "overestimation" or "self-efficacy"); *relative overconfidence*, which consists of overly positive beliefs about one's performance in comparison to the performance of others (aka "overplacement", "better-than-average" or "the Lake Wobegon effect"); and *overprecision*, which consists of having excessive certainty or precision in one's beliefs (often measured as having confidence intervals that are too narrow). In this paper we are largely concerned with absolute overconfidence, but also include some measures to elicit relative overconfidence and overprecision.

A number of mechanisms have been suggested to explain overconfidence. One theory argues that overconfident beliefs allow individuals to maintain an unrealistically positive view of themselves (Alicke et al. 1995; Kwan et al. 2004). Accordingly, individuals tend to be most overconfident about positive traits, as well as traits that are perceived to be under their control (Alicke 1985). The quality of information about performance and ability can also be important, as overconfidence tends to be greatest in settings with infrequent and noisy feedback (Nisbett and Lee 1980). Healy and Moore (2007), however, argue that many patterns in overconfidence can be reconciled if individuals correctly update their beliefs according to Bayes' Rule, and have more precise information about their own performance than the performance of others. Möbius et al. (2010) demonstrate theoretically and experimentally that overconfidence can occur because individuals' adjust beliefs too slowly given new information, and because they incorporate positive information more than negative information. In our experiment, subjects will perform the same task repeatedly and get accurate feedback on their actual and predicted score in each period. In principle this should allow them converge to accurate beliefs about their own ability, even if they begin with over- or underconfident beliefs. We also introduce an even stronger feedback condition for some subjects.

Overconfidence appears to depend both on general traits of the individual, such as personality (Schaefer et al. 2004), narcissism (Campbell et al. 2004), genetics and developmental environment

(Cesarini et al. 2009), as well as specific characteristics of the task. That is, while some individuals tend to be more overconfident than others, someone can be highly overconfident on some tasks and highly underconfident on others. A number of studies have reported that people are typically overconfident about common activities such as driving, getting along with others, or other tasks they find easy, and are underconfident on tasks they find difficult (Klayman et al. 1999; Kruger 1999; Moore and Cain 2007). We take these results into account in our experiment by basing pay on absolute performance only and by designing individual pay scheme options based on each participant's skill level. Additionally, we introduce a second, non-correlated task to examine the extent to which overconfidence in this second task predicts overconfidence in the experiment's main task.

Overconfidence and Sorting

Previous experimental studies have documented the link between confidence and pay scheme choice.[8] For example, Niederle and Vesterlund (2007) demonstrate that men are much more likely to select to be compensated via a tournament than women, who prefer to be paid via a piece rate. Using a task-specific measure of confidence, the paper demonstrates that men tend to be more overconfident than women, which helps explain their preference for the tournament. Dohmen and Falk (forthcoming) allow individuals to choose between receiving a fixed payoff, or sorting into either an individual linear piece rate, a tournament, or a group revenue-sharing scheme. They find that overestimating one's productivity predicted sorting into the tournament, but not sorting into an individual piece-rate or a group revenue sharing scheme. Similarly, in a laboratory study, Sautmann (2009) finds that "employers" offering a binary wage scheme to "employees" based on high or low output would lower wage levels for overconfident subjects, since overconfident subjects overestimated their likelihood of reaching the high output wage. Our paper builds on these results in four ways: it investigates convex piece rates, which are

[8] Cadsby et al. (2007) also examine experimentally sorting between fixed pay and performance pay schemes, but do not measure overconfidence.

very commonly used to determine wages, instead of a tournament[9] or binary wage scale, which are less commonly used to set pay; it adds a number of different confidence measures to the measure commonly used; it investigates a "hard" feedback condition where subjects were clearly told whether they were choosing the wrong scheme; and it examines the interaction of overconfidence across tasks.

Camerer and Lovallo's (1999) experimental study on excess entry into markets is also highly relevant. To imitate market entry, participants were asked to decide whether to opt into a pay-scheme that diminished in pay as more participants enter, but also pays by rank of performance on a trivia task. Those who were overconfident on the trivia task were the most prone to opting in and decreasing their net profits. However, participants were not provided with feedback on their own performance in the trivia task. Our paper differs in the following ways: we provide clear feedback on absolute and relative performance; our payoffs are completely based on individual performance; and we measure confidence using both absolute and relative terms.

The implications of overconfidence demonstrated by these laboratory experiments have also been shown to be remarkably robust in studies using archival field data. Managers across a number of disciplines have been shown to be substantially overconfident in their work-related judgments (Russo and Schoemaker 1992). Entrepreneurs (Busenitz and Barney 1997) and venture capitalists (Zacharakis and Shepherd 2001), who face substantially convex returns to the success of startup ventures, are both characteristically overconfident. Furthermore, overconfidence has been shown to distort financial trading decisions (Barber and Odean 2001) by leading overconfident men to trade 45% more often than women, significantly decreasing their net returns. Similarly, overconfident CEOs make distorted investment decisions and make more (and worse) acquisitions (Malmendier and Tate 2005, 2008).

The effect of overconfidence on performance incentives and competition identified by the experiments of Niederle and Vesterlund (2007) and Dohmen and Falk (forthcoming) have also found empirical support. Employee overconfidence provides one of the best explanations for why many

[9] Although promotion decisions can be usefully thought of as a tournament, there are very few actual incentive systems in companies that base pay on explicit tournaments (Lazear and Oyer 2009).

companies grant stock options to all employees and not just managers, because overconfident employees overvalue these stock grants (Oyer and Schaefer 2005; Bergman and Jenter 2007). Additionally, overconfidence about ones' own performance has been shown to explain the sorting between different competition groups of long distance runners (Nekby et al. 2008). Finally, Wu and Knott (2006) find empirically that overconfidence among entrepreneurs leads to the excess market entry demonstrated experimentally by Camerer and Lovallo (1999), resolving a major contradiction in the literature as considerable evidence shows most entrepreneurs are risk averse, despite their apparently "risky" choices.

Experimental Design

Our experiment consisted of nine periods of multiplication problems, one period of trivia questions, a risk preference elicitation mechanism, and a concluding questionnaire that included several psychological measures. The experimental tasks were programmed using the software z-Tree (Fischbacher 2007). Subjects were paid for their earnings in two randomly selected math/trivia periods, one randomly selected risk decision, as well as a $10 show-up fee. Screen shots of the experiment, including instructions, are included in the appendix.

Multiplication Task

In each 150 second period, subjects solved randomly generated multiplication problems; all problems consisted of multiplying a two-digit number by a one-digit number.[10] Their score on the task was the number of correct answers minus half the number of incorrect answers. In the first two periods a subject's payoff was based on a linear piece rate equal to $0.30*score. We then calculated the subject's baseline ability (B) as the maximum score in the first two periods. We used each subject's baseline ability to calibrate a subject-specific convex piece rate that remained fixed throughout the rest of the experiment. We used a subject-specific rather than a universal convex payoff function because there is

[10] We used this task both because it was used previously by Dohmen and Falk (forthcoming) to study overconfidence and sorting, and because it is a simple task. While a more complicated task may be more realistic, we believe a simple task is a stronger test of the persistence of overconfidence. In a complicated task it will generally be more difficult to identify the causes of performance (e.g. ability, luck, task strategy), and therefore it may be easier for an individual to maintain incorrect beliefs about her own ability over time.

substantial variance in ability between subjects, and we wanted to avoid low- and high-ability subjects having a trivial choice over pay schemes.[11] The convex piece rate was set according to the table below, where each cell shows the total piece rate.

Score	(B-3) or less	(B-2.5) to (B+1)	(B+1.5) to (B+5)	(B+5.5) or more
Basic Convex[12]	$0.20	$0.25	$0.35	$0.50
Escalated Convex	$0.20	$0.25	$0.35	$0.75

Therefore, the linear scheme yields a higher payoff for scores equal to (B+1) or less, while the convex scheme yields a higher payoff for scores of (B+1.5) or higher. We chose the convex piece rates so that ex-post one choice would always yield a strictly higher payoff. In the third period all subjects were paid under the basic convex piece rate, so that subjects became familiar with it.

In periods four through nine the pay scheme used varied between treatments. In the control **No Choice** treatment, each subject was randomly assigned to either the linear or basic convex scheme in each period. This treatment provides an important control for testing the effect of the introduction of choice on performance, confidence and other important variables. In the **Choice** treatment subjects were given a choice between the linear or basic convex scheme at the beginning of each round. In the **Escalation** treatment, subjects chose between the linear and basic convex pay schemes for periods four through six, and between the linear and escalated convex pay schemes in periods seven through nine. We used this treatment to test whether a highly convex scheme affected overconfidence, pay scheme choice, or their

[11] Having individual-specific payment schemes is also realistic. Many firms have individualized performance hurdles and bonus points. For example, sales quotas are usually set based on a salesperson's individual performance in previous years.

[12] In an early pilot, we tried using a payoff function with varying marginal piece rates (e.g. $0.20 per point for the first 10 points, then $0.25 per point for the next 5 points…) to avoid discontinuities in the payoff function and more closely match typical convex commission schemes. However, subjects found this pay scheme too confusing. Additionally, by applying one piece rate to the total score it is very easy for subjects to identify the minimum score needed to make the convex scheme optimal. While convex schemes with varying marginal piece rates are more common than schemes with varying absolute piece rates, the incentive effects of the two schemes are always in the same direction.

interaction. As discussed earlier, in the **Comparative Feedback** treatment subjects received a report after the sixth period detailing how much they would have earned under each scheme given their performance, and the amount of money they lost if they chose the wrong scheme.

In each period we also elicited two measures of subjects' beliefs about their performance. Before the multiplication task (but after they chose the pay scheme for that round) subjects predicted what score they would achieve. Subjects had an incentive to predict correctly, since an additional 50 cents was added to their round earnings if they were correct, or an additional 25 cents if their guess was within one point.[13] After the task, subjects were asked to guess which quartile their score fell in for that period compared to the other experimental subjects in the room, earning an additional 50 cents for the period if they were correct. We used these two beliefs to create measures of both absolute and relative overconfidence in each period.

After each Multiplication Task, subjects were informed of their score on the task, their payoff given the pay scheme, whether their guesses were correct, and the actual quartile of their score. Subjects were required to click a button indicating they had read this information before moving on to the next round.

Trivia Task

In order to obtain a measure of overconfidence in an entirely separate domain, we had subjects answer trivia questions. Subjects had two and a half minutes to answer 10 multiple choice trivia questions drawn from Nelson and Narens (1980). Subjects earned $0.75 for each question they answered correctly. As in the multiplication task, subjects predicted their score before performing the task and guessed their relative ranking after the task. As before, subjects earned additional payoffs for accurate predictions. We also asked subjects to rate their confidence in their answer for each question on a scale of one to five.

Risk Measure

[13] However, during the task subjects were not able to see their score, so this incentive should not have distorted performance.

Because risk averse (loving) subjects may be more likely to choose the linear (convex) scheme, we assessed subject risk tolerance using the risk elicitation mechanism from Dohmen and Falk (forthcoming). Subjects made fifteen choices between a lottery and a fixed payment. In each case the lottery had a 50% chance of paying $4 and a 50% chance of paying $0, while the fixed payment increased in $0.25 increments from $0.25 to $3.75. Subjects were paid for one randomly selected decision.

Psychological and Demographic Questionnaire

At the end of the experiment subjects filled out a questionnaire that elicited basic demographic information, a measure of the Big 5 personality traits,[14] and different measures of overconfidence and optimism. We built an independent self-reported measure of a subject's propensity towards relative overconfidence by asking them to rate the accuracy of several statements reflecting overconfidence on relative performance (e.g. "I am a better driver than most people").[15] We also asked subjects to rate statements designed to measure optimism (e.g. "In uncertain times, I usually expect the best"). Finally, we asked subjects to rate the likelihood of a variety of events happening to them over the course of their life compared to other people they know of the same age and gender. Half the events were largely out of one's control (e.g. being selected randomly for an IRS audit), and half were events the subject could potentially influence (e.g. getting a job in the next 5 years that pays more than $75,000/year). We relied on typical examples in the social psychology literatures on confidence and optimism to develop these questions.

Results

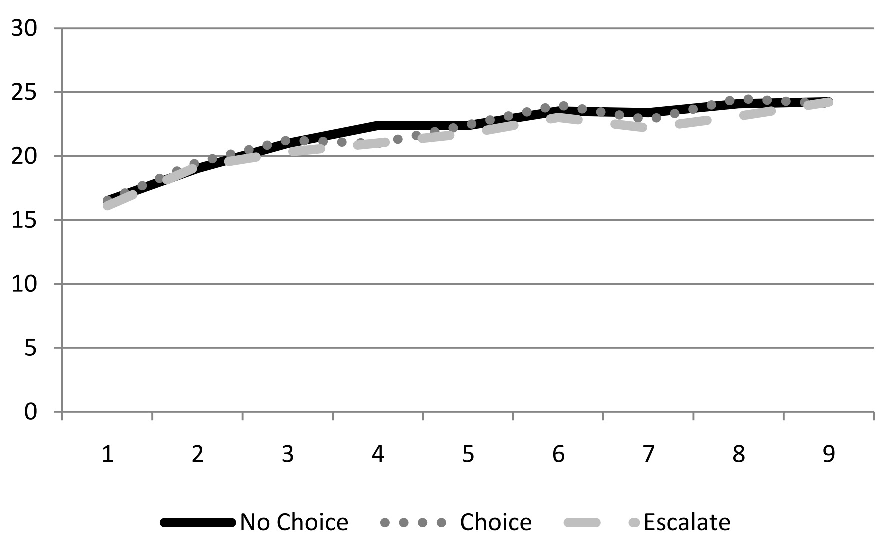
Our results are presented as follows. In the first section, we review our measures of confidence and evaluate our subject pool for over- and underconfidence, including measuring for learning over time and response to feedback. This includes identifying the group of subjects that are most likely to be overconfident based on demographics and other factors. In the second section, we extend our analysis to

[14] Questions were taken from Gosling et al. (2003), Rammstedt and John (2007) and Saucier (1994).
[15] In what follows we term this measure "independent overconfidence," only to differentiate it from the task-specific measure of "relative overconfidence" discussed above. As noted in the "Related Literature" section, there is no general measure of overconfidence.

pay scheme choice, and evaluate the impact that over- and underconfidence have on pay scheme choice and total wages earned. In the third section, we focus on incentive effects by comparing subject performance under random assignment of pay scheme to those who were able to choose their pay scheme.

Sessions were run at the Harvard Business School CLER lab using the standard subject pool. A total of 179 subjects participated, with 41 in the No Choice control, 68 in the Choice treatment and 70 in the Escalation treatment. The average subject earned $28.78. Figure 1 reports average subject performance in the Multiplication Task in each treatment. Subjects earned a score of approximately 16 in the first period, increasing to roughly 21 in period 4, and to 24 in period 9. The treatments do not substantially differ.

Figure 1: Multiplication Task Performance



Confidence Measurements

A large advantage of our experimental approach is that it collects a number of independent measures of subject overconfidence, across the three overconfidence types suggested by Healy and Moore (2007). These are presented in Figure 2 below.

Figure 2: Measures of Overconfidence Used in Study

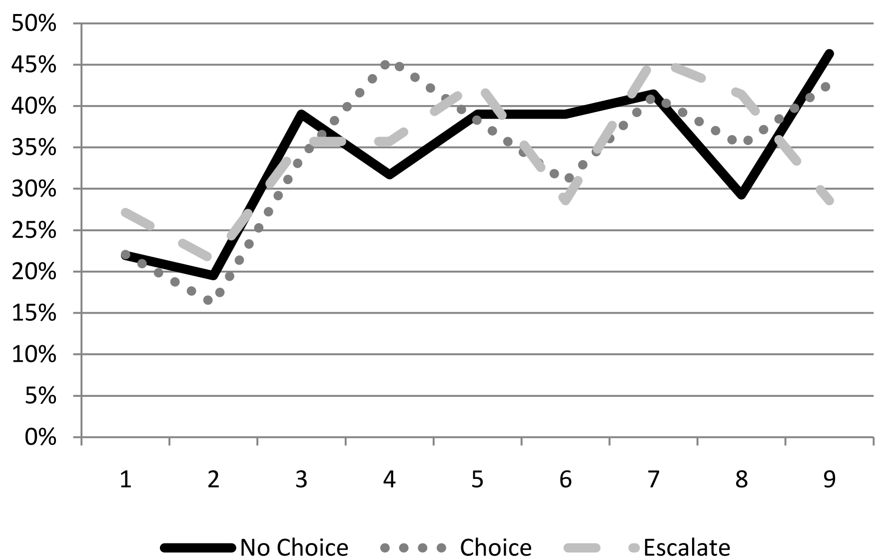
Measure name	Definition	Overconfidence type	Notes
Overconfidence (Difference in belief)	A subject's predicted performance in a given multiplication round minus his actual performance	Task-specific Absolute	Primary definition used in the paper; belief elicitation was incentivized
Relative overconfidence	Dummy variable =1 if subject overestimated his performance quartile in a given round	Task-specific Relative	Belief elicitation was incentivized; measured for both multiplication and trivia tasks, but only multiplication task reported in the paper
Trivia overconfidence	Same as overconfidence above, but for the trivia task	Alternate Task Absolute	Belief elicitation was incentivized
Question accuracy	The total number of incorrect trivia answers (out of 10) in which the subject reported being "Extremely Confident" in her answer	Alternate Task Precision	Subjective report of confidence in answer on 5 point scale, Non-incentivized
Independent measure of relative overconfidence	Z-score of 10 question survey on questions about general relative overconfidence (e.g. "I am better at board games than most people")	General Relative	Non-incentivized, but subjects were told survey was an important part of the experiment. Survey was given at the end of the experiment, so it did not bias results.

Aggregate over- and underconfidence

As noted in Figure 2, our primary measure of confidence is the difference between the predicted score and the achieved score. Thus a subject whose prediction was higher than her actual score is called overconfident, and a subject whose prediction was below her actual score is called underconfident. Figure 3 displays the frequency of overconfident subjects over time. As would be expected, subjects were rarely able to exactly predict their score. During the 3 practice rounds, only 4.8% of scores were predicted correctly, while 26.3% of predictions were too high and 68.9% of predictions were too low. Similarly, during the experimental rounds 4-9, subjects predicted their score correctly only 5.7% of the time, while 38.0% over-predicted and 56.3% under-predicted. These numbers are quite consistent across all three treatments and do not differ substantially across periods. Averaging across treatments and periods, subjects are slightly underconfident: the mean difference between prediction and performance is -1.36.

To show these measurements of over- and under confidence are not a result of small, random within-subject variation in task performance, we also look at large errors in prediction (when subject predictions are incorrect by more than +/- 10 or 20 percent of their actual score)[16]. This creates a measure of confidence that excludes consideration of small, random variations in task performance. These large errors occurred quite often: 20.9% of predictions overestimated the subject's score by at least 10 percent, and 12.5% of predictions overestimated performance by at least 20 percent. Similarly, 41.2% of predictions underestimated the subject's score by at least 10 percent, and 24.3% underestimated by at least 20 percent. Overall the average absolute difference between guess and score is 24.7%.

Figure 3: Frequency of Overconfident Subjects



To compare these trends in overconfidence to random variation in performance, we fit a logarithmic trend for each subject and looked at the deviation between actual performance and trend-fitted

[16] We consider percentage deviations because of the wide variation in performance. However, an error of at least 10% implies that the guess fell outside of the payment range of +/- 1 for 92% of observations.

performance.[17] Actual performance deviates from trend by at least 10 percent in 45.4% of multiplication tasks, significantly less often than the 62.1% of subject predictions with errors of at least 10 percent (test of proportions: $p < 0.01$). Similarly, actual performance deviates by at least 20 percent in 16.4% of the multiplication tasks, significantly less often than 36.7% of subject predictions with errors of at least 20 percent ($p < 0.01$). Therefore, the prediction errors are much larger than the variance in subject performance, suggesting that random shocks to performance cannot explain the observed levels of substantial over- and underconfidence.

Trends in aggregate over- and underconfidence

After an initial learning period there is a stable and persistent level of overconfidence throughout the experiment. Notably, the average number of over- and underconfident subjects did not change significantly over time, even though subjects received feedback on their prediction and actual performance after every period. Across all treatments, 38.5% of the subjects in the first treatment round (period 4) were overconfident, while 38.0% in the last experimental round (period 9) were overconfident. The distribution of overconfidence was not significantly different between periods 4 and 9 (Kolmogorov-Smirnov test: $p = 0.69$). Furthermore, many subjects do not appear to be engaging in sufficient learning, given their prediction errors, after the first few periods. In periods 6 through 9 more than 25% of subjects made the same prediction in consecutive periods (indicating they placed zero weight on the information revealed by their performance) despite these subjects on average making absolute prediction errors of 15 percent. By contrast, only 7% of subjects had the same prediction in periods 1 and 2.[18]

We find similar trends when focusing only on rounds exhibiting large magnitudes of over- and underconfidence. 26.3% of subjects in period 4 are overconfident by at least 10 percent, and this figure

[17] Just using the within-subject standard deviation of performance would overstate the random component of performance because of the increase in performance over time due to learning.

[18] Similarly, if we assume all our subjects are Bayesian, and using the fact that performance is approximately normal, we find that learning overall slows down quite rapidly. The median subject (among those with well defined weights) places three times more weight on performance in period 1 than on her prior in period 1 when forming her posterior, but by periods 7-9 the median subject is placing two to three times more weight on her prior than on her performance. That is, beliefs converge quite rapidly during the experiment despite the fact that many subjects continue to make large prediction errors.

decreases only slightly to 19.0% in period 9. Similarly, 15.1% of subjects are overconfident by at least 20 percent in period four, decreasing by less than half to 8.4% in period 9. Absolute prediction errors are also stable across periods. The mean error in each treatment is between 2 and 4.5 in periods 4 through 9 with no significant time trend in any treatment (the distribution of absolute errors is not different between periods 4 and 9: $p = 0.89$). Taken together, these results suggest that many overconfident subjects did not update their belief of their ability over the course of the experiment.

Individual heterogeneity in over- and underconfidence

We find that specific individuals were more prone to being overconfident in any given period. First, significantly more subjects were over- or underconfident in multiple periods than would be expected if prediction errors were randomly distributed across subjects in each period.[19] Furthermore, subjects in the fourth quartile of overconfidence in one period are significantly more likely to be in the fourth quartile of overconfidence in the next period (32%) and less likely to be in the first quartile (14%) than others (chi-square test: $p < 0.01$). These results show that specific individuals are predisposed to be over- or underconfident.

Subjects who were initially overconfident are also more likely to be overconfident over the course of the experiment. To show this, we build a measure of "initial overconfidence," defined as the average difference between predicted and actual score for the first three periods. We then compare the average difference between predicted and actual score for the group of subjects who were initially overconfident and the group of subjects that were not initially overconfident. The average overconfidence in the experimental rounds for the subjects who were initially overconfident was 0.40, suggesting the average subject in this group remained overconfident. Conversely, the average confidence for the latter group was -1.12, suggesting underconfidence also persists. The difference between these

[19] We use a non-parametric permutation test where we randomly permute being overconfident or underconfident between subjects in each period. In our data 15.64% of subjects were overconfident in 4 of 6 periods, 2.2% in 5 of 6 periods, and 1.7% in all six periods. The probability of seeing at least that many subjects who were overconfident in at least 4, 5 or 6 periods are $p = 0.02$, $p = 0.34$, $p = 0.01$. Similarly, we observe 27.4% of subjects who were underconfident in 4 of 6 periods, 14.0% in 5 of 6 periods, and 5.6% in all 6 periods. The probability of seeing at least that many subjects who were underconfident in at least 4, 5 or 6 periods are $p = 0.54$, $p = 0.26$, $p = 0.04$.

two measures is statistically significant (using a nonparametric permutation test $p < 0.01$). Similarly, initially overconfident subjects were overconfident an average of 2.80 times in the last six periods, while the other subjects were overconfident 2.16 times; this difference is also statistically significant ($p = 0.02$)[20].

More formally, we regress the difference between predicted and actual score on initial overconfidence, as well as a dummy variable indicating relative overconfidence (guessing a higher quartile of performance than the subject achieved) and corresponding subject-level overconfidence measures from the trivia task and the independent questionnaire on relative overconfidence.[21,22,23] We also control for subject skill level in the task by including the subject's baseline ability and including subject random effects. The results are presented in model (1) in Table 1, which confirms that subjects who were more overconfident in the first three periods are significantly more overconfident in the last six periods.

We examine the relationship between predicted and actual performance on several alternative measures of overconfidence in models (2) – (5) of Table 1. Model (2) shows that absolute and relative overconfidence are positively correlated. Interestingly, this differs from the results of Healy and Moore (2007), where absolute and relative overconfidence are negatively correlated. Model (5) shows that subjects who were more overconfident according to the independent self-reported measure of relative overconfidence were more likely to be overconfident in their predictions. Therefore the observed overconfidence is coming disproportionately from a specific set of subjects, rather than simply random error. In model (3), we see that subjects who were overconfident on the Trivia Task as measured by the

[20] We find similar differences (albeit less significant) between subjects who were overconfident and underconfident during periods 4 to 6. During periods 7 to 9, these overconfident subjects were on average more overconfident (difference = 0.68, $p = 0.08$), and were overconfident somewhat more often (difference = 0.21, $p = 0.17$).

[21] The confidence measures from the trivia task include the difference between predicted and actual performance, as well as a count of the number of times a subject claimed they were "extremely confident" on a trivia question that they answered incorrectly.

[22] The overconfidence measure from the independent questionnaire is a standardized scale of overconfidence based on the subject's accuracy ratings of various statements reflecting overconfidence (e.g. "I am an above average driver"). Subjects were excluded whose answers on the questionnaire were nonresponsive.

[23] We also examined the psychological measures of optimism; however none of the measures were significant.

difference between their guess and their score were not more likely to be overconfident on the Multiplication Task. This suggests that overconfidence may be specific to the task at hand, which is in accord with much of the psychological literature on the domain-specificity of overconfidence (see Klayman et al. 1999). Finally, model (4) shows that subjects who claimed to be extremely confident in trivia answers that were actually incorrect were also more overconfident on the Multiplication Task. One interpretation of this finding is

Table 1: Overconfidence Measures

VARIABLES	(1)	(2)	(3)	(4)	(5)
Mean Initial Overconfidence (Pds 1-3)	0.119***				
	(0.0240)				
Relative Overconfidence		3.909***			
		(0.282)			
Trivia Overconfidence (Total)			-0.0201		
			(0.0619)		
Trivia Overconf. (Question Rating)				0.428*	
				(0.220)	
Independent Overconfidence					0.430**
					(0.210)
Choice Treatment	0.0592	-0.0944	-0.176	-0.264	-0.162
	(0.469)	(0.504)	(0.511)	(0.497)	(0.506)
Escalation Treatment	0.0435	-0.195	-0.196	-0.286	-0.135
	(0.443)	(0.469)	(0.488)	(0.477)	(0.487)
Period	0.0740**	0.0679**	0.0581*	0.0520	0.0426
	(0.0310)	(0.0344)	(0.0328)	(0.0322)	(0.0340)
Baseline Ability	0.0203	0.0215	0.0203	0.0203	0.0203
	(0.0708)	(0.0659)	(0.0710)	(0.0712)	(0.0708)
Constant	-2.217***	-3.079***	-1.988**	-2.035**	-1.712**
	(0.777)	(0.824)	(0.820)	(0.811)	(0.808)
Observations	1074	1074	1074	1074	1074
Number of Subjects	179	179	179	179	179

Robust standard errors in parentheses. Significant levels: *** p<0.01, ** p<0.05, * p<0.1
The dependent variable is the difference between the subject's prediction of their score and the actual score for each subject in each round.

that extreme overconfidence by an individual may correlate across several tasks. Based on these results, it is clear that much of the overconfidence within the sample is from a distinct subgroup of subjects. Additionally, note again that there are no significant differences between treatments, and that there is either no time trend or a slight positive time trend (i.e. more overconfidence over time). Furthermore, baseline ability in the math task does not predict overconfidence.[24] Finally, for no measure of overconfidence did the escalated pay scheme affect overconfidence.

Demographics

We also look at demographic predictors of overconfidence. In the first model in Table 2, we regress our measure of overconfidence on gender and the Big 5 personality factors. We find that extraversion is predictive of overconfidence while being male is also marginally predictive. This is consistent with the common finding that men are more overconfident than women (e.g. Bengtsson et al. 2005; Beyer 1990; Niederle and Vesterlund 2007).

In model (2), we look at predictions including future job type (e.g. doctor, teacher) and estimated salary in 5 years. A higher predicted salary in 5 years is very significantly associated with overconfidence, but for the most part choice of career path is not (although choosing a career path in management is marginally predictive).[25] In general, the demographic factors reaching significance are traits often associated with an overconfident individual: male, extroverted, predicting high future salaries, and desiring a career in management.[26]

[24] The speed of learning (e.g. the increase in performance in the third period over baseline ability) does not correlate with overconfidence either.

[25] The omitted job type in regression 2 was "other" and "don't know." We also grouped sales and administrative/clerical into this category, because each had only one subject indicating they expected to end up in this job type. Finally, we ran a version of regression 2 using company type (e.g. education, government, finance etc.) instead of job type, but did not find any company type to be predictive of overconfidence.)

[26] We also ran the same regression using the independent measure of relative overconfidence as the dependent variable. Male, predicted salary, and management were still predictive using this alternate measure of overconfidence.

Table 2: Demographics

	(1)		(2)	
VARIABLES	Coefficient	Std. Error	Coefficient	Std. Error
Male	0.5804*	(0.329)		
Extraversion	0.3156**	(0.138)		
Agreeableness	0.0083	(0.209)		
Conscientiousness	-0.1280	(0.190)		
Emotional Stability	-0.0298	(0.186)		
Openness	0.2640	(0.232)		
Salary in 5 years ('000s)			0.0050***	(0.0017)
Job: Media/Artist/Writer			0.467	(0.494)
Job: Doctor/Healthcare			0.203	(0.529)
Job: Engineering/Tech			1.227	(1.031)
Job: Management			1.806*	(0.963)
Job: Marketing			0.519	(0.609)
Job: Teacher/Professor			1.304	(0.936)
Job: Researcher			0.0579	(1.046)
Job: White Collar			-0.0910	(0.545)
Choice	-0.0956	(0.429)	-0.226	(0.428)
Escalate	-0.0920	(0.420)	0.162	(0.392)
Period	-0.0027	(0.0816)	0.00493	(0.0840)
Baseline Ability	0.0424	(0.0287)	0.0454	(0.0326)
Constant	-1.8387***	(0.616)	-2.261***	(0.743)
Observations	978		936	
Number of Subjects	163		156	

Robust standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1
The dependent variable is the difference between the predicted score and actual score at the subject-round unit of observation.

Pay Scheme Choices

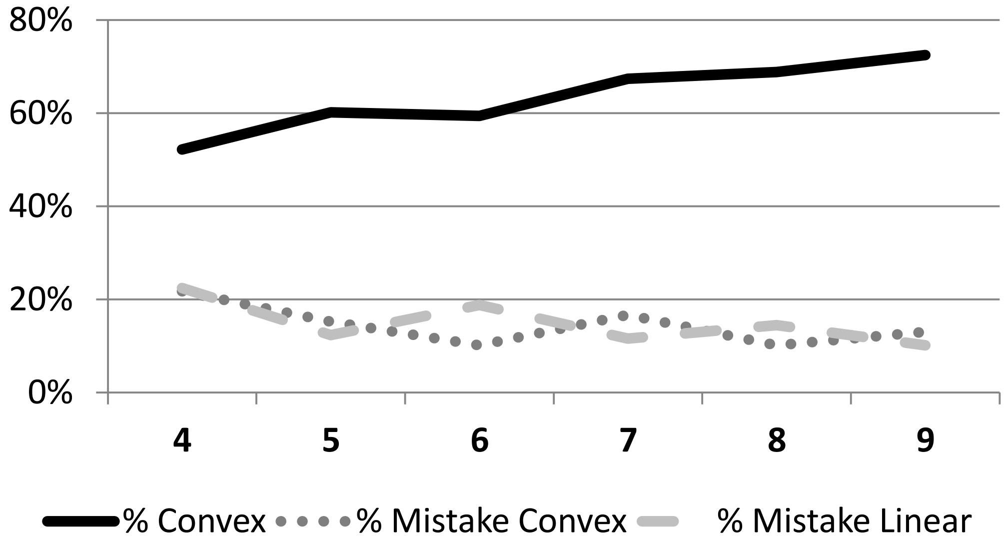
We now examine subjects' choices over the linear and convex pay schemes. In this section and subsequently, we refer to choices that do not maximize a subject's pay in the experiment as "mistakes." It is important to note that this assumes that subjects do not gain utility from choosing the linear or convex scheme in ways not reflected in pay. For example, it is well known that many people face disutility from making risky choices. Our regressions in this and other sections of the paper fully control for risk preferences, but there may be some other unobserved factor that gives subjects utility from making a pay scheme choice, even if it does not maximize pay. In the absence of a theoretical or empirical literature suggesting what these factors might be, we will call these choices "mistakes," with the caveat that subjects may have chosen to enter into the "wrong" contract even had they known it would result in less pay[27]. We note, however, that in nearly 89% of rounds subjects chose the scheme which would maximize their earnings given their performance expectations. This suggests that subjects largely made their incentive scheme choice based on expected earnings (given their beliefs), not other factors.

Figure 4 reports the fraction of subjects who chose the convex scheme in each period (pooling between the Choice and Escalation treatments), as well as the fraction of subjects who made a mistake in choosing linear and convex schemes. In period 4, the first choice period, 52% of subjects chose the convex scheme; this increased to 72% by period 9. This increase in the choice of the convex scheme is not surprising, given the increase in subject scores seen in Figure 2. In each period, between 20 and 30 percent of subjects selected a pay scheme that paid them less than the alternative; a similar number of mistakes came from those who mistakenly chose the linear scheme and those who mistakenly chose the convex scheme. While subjects are certainly doing better than random choice of pay scheme, it is notable that even after several periods (including clear feedback after each period) a substantial number of subjects continue to make mistakes in choosing their compensation scheme: 23% of subjects make a mistake in period 9. Furthermore, we find that the same subjects are persistently making mistakes: significantly more subjects mistakenly choose convex many times (or mistakenly choose linear many times) than if mistakes were randomly allocated across individuals.[28] We also find that subjects who had mistakenly chosen convex in one period are more likely to do so in the next period (39%) than subjects who had mistakenly chosen linear (12%) or correctly chosen their pay scheme (10%). Similarly, subjects

[27] There was nothing in our post-survey questionnaire that suggests subjects derived utility from incentive scheme choice per se. In contrast, a number of subjects complained about not being paid as much as they would have had they not made a "mistake."

[28] In our experiment 27% of subjects mistakenly choose convex in one period, 17% in two periods, 4% in three periods, 2% in four periods, and 1% in five periods. Using a non-parametric permutation test, we find significantly more subjects made this mistake at least two, three, four and five times than if mistakes were randomly allocated across individuals ($p = 0.05$, $p = 0.03$, $p < 0.01$ and $p < 0.01$ respectively). Similarly, 22% of subjects mistakenly chose linear in one period, 12% in two periods, 8% in three periods, 4% in four periods, and 1% in five periods. Significantly more subjects made this mistake at least three, four and five times than if mistakes were random ($p < 0.01$, $p < 0.01$, $p = 0.05$ respectively).

Figure 4: Pay Scheme Choices



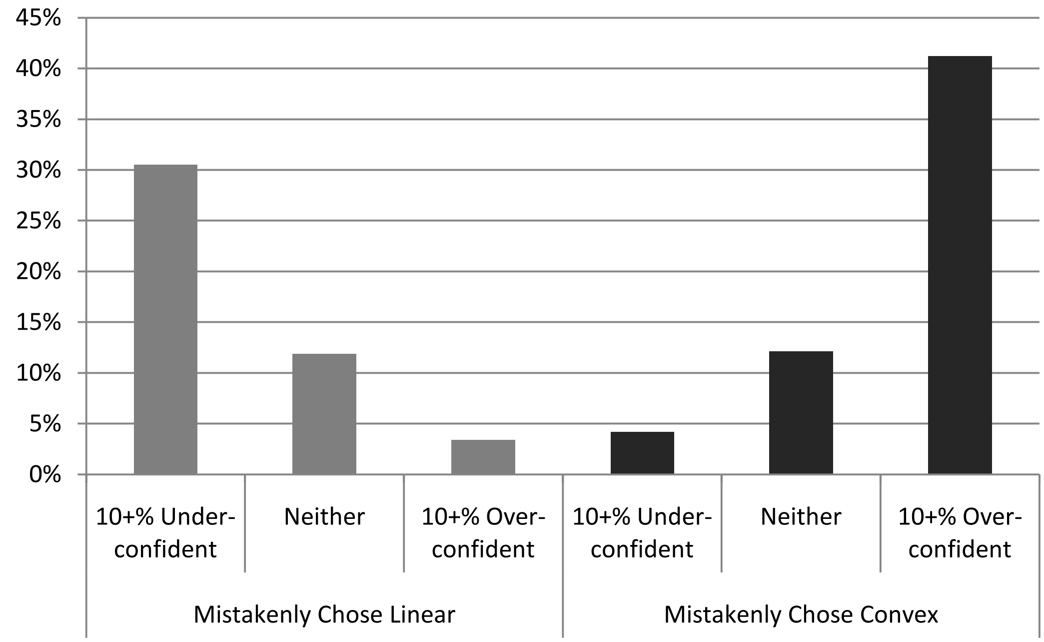
who had mistakenly chosen linear are more likely to do so again (35%) than subjects who had mistakenly chosen convex (10%) or correctly chosen their pay scheme (9%).[29]

We can also look at how these mistakes vary with subject over- and underconfidence. Figure 5 shows the fraction of over- and underconfident subjects who mistakenly choose the linear piece rate and who mistakenly choose the convex piece rate. For this figure we identify a subject as underconfident if their guess is 10 percent or more below their predicted performance, and similarly we identify a subject as overconfident if their guess is 10 percent or more above their predicted performance.[30] These results indicate that subjects are substantially more likely to choose the linear scheme when they would have earned more under convex if they are underconfident, and conversely are substantially more likely to choose the convex scheme when they would have earned more under linear if they are overconfident.

[29] The distributions are significantly different (chi-square test: $p < 0.01$).

[30] We find similar results using our main measure comparing guesses to realized performance, and from using the independent measure of relative overconfidence.

Figure 5: Mistakes and Overconfidence



To further analyze the role of overconfidence in sorting between the piece rate options, we regress payment scheme choices on overconfidence, a dummy for the escalated convex piece rate in the last three periods of the Escalation treatment, an interaction term, the subject's maximum practice score, and the certainty equivalent from the lottery choices.[31] Results of the random effects regressions are presented in Table 3. Model (1) uses pay scheme choice as the dependent variable, model (2) uses an indicator for mistakenly choosing the convex scheme, model (3) uses the amount (in percentage terms) of the foregone payoff from this mistake, and models (4) and (5) use the corresponding variables for mistakenly choosing the linear scheme.

From model (1) in Table 3, we can again see that overconfidence strongly predicts choosing the convex pay scheme. For example, subjects in the 90th percentile of overconfidence (henceforth "highly overconfident subjects") are 23 percentage points more likely to choose convex than subjects in the 10th

[31] The median subject is risk neutral by this measure. Two subjects were excluded because they did not have a unique switching point between the lottery and the certain payoff. We obtain quantitatively similar results for overconfidence if we run the regressions for all subjects excluding the risk measure.

Table 3: Pay Scheme Choices

VARIABLES	(1) Chose Convex	(2) Mistakenly Chose Convex	(3) % Lost Pay From Convex Mistake	(4) Mistakenly Chose Linear	(5) % Lost Pay From Linear Mistake
Overconfidence	0.0184***	0.0362***	0.0124***	-0.0266***	-0.0109***
	(0.0035)	(0.0049)	(0.0018)	(0.0044)	(0.0022)
Escalated Convex Available	0.0288	-0.0425	-0.0124	-0.0647*	-0.0168
	(0.0665)	(0.0276)	(0.0084)	(0.0335)	(0.0140)
Escalated x Overconfidence	-0.0387	0.0866	0.0083	0.0535	0.0020
	(0.0371)	(0.0645)	(0.0215)	(0.0407)	(0.0334)
Baseline Ability	-0.0059	-0.0018	-0.0001	-0.0013	-0.0003
	(0.0043)	(0.0020)	(0.0007)	(0.0021)	(0.0008)
Lottery Certainty Equivalent	0.2806**	0.1224**	0.0328**	-0.1724***	-0.0436**
	(0.1158)	(0.0489)	(0.0151)	(0.0592)	(0.0215)
Period	0.0369***	-0.0155**	-0.0046**	-0.0177**	-0.0021
	(0.0082)	(0.0069)	(0.0021)	(0.0072)	(0.0031)
Constant	0.2394	0.2083***	0.0564**	0.4665***	0.1073***
	(0.1754)	(0.0767)	(0.0226)	(0.0965)	(0.0277)
Observations	816	816	816	816	816
Number of Subjects	136	136	136	136	136

Robust standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1

percentile. For comparison, this is roughly the same magnitude as the effect of risk aversion: subjects in the 90th percentile of risk attitudes (most risk loving) are 21 percentage points more likely to choose the convex scheme than subjects in the 10th percentile.

Highly overconfident subjects are also 45 percentage points more likely to mistakenly choose the convex scheme than subjects who are not highly overconfident, with the mistake costing them 16% of their payoff. Risk-loving subjects are similarly more likely to mistakenly choose the convex scheme. The frequency of these mistakes decreases by only 8 percentage points between period 4 and period 9. Conversely, underconfident subjects are more likely to mistakenly choose the linear scheme. Subjects at the 10th percentile of overconfidence are 33 percentage points more likely to mistakenly choose the linear scheme than subjects at the 90th percentile, with these mistakes costing them 14% of their payoff. Risk averse subjects are also more likely to make this mistake, though the effect is much smaller: subjects at the 10th percentile of certainty equivalent are 14 percentage points more likely to make this mistake than subjects at the 90th percentile. The frequency of mistakenly choosing the linear scheme decreases by 7 percentage points between period 4 and period 9.

The effect of overconfidence on mistaken pay scheme choice is persistent throughout the experiment. Running the same regression specification for only the last three periods results in essentially the same estimates. For example, the estimated coefficient on overconfidence for mistakenly choosing convex is $\beta = .0389$ (s.e. = .00588, $p < 0.001$). Similarly, the estimated coefficient on overconfidence for mistakenly choosing linear is $\beta = -.0249$ (s.e. = .00556, $p < 0.001$). Therefore even though the overall frequency of the mistakes decreases slightly over time, overconfidence still causes the mistakes to occur even at the end of the experiment. Note also that in the Escalation treatment increasing the highest piece rate did not have a main effect, nor an interaction with overconfidence, on any of these behaviors.

Because both our main measure of overconfidence and ex-post pay scheme mistakes depend on realized performance, it could be argued that our measures represent a high number of unexpected mistakes, a random shock in the difficulty of questions, or some other random factor affecting realized performance instead of overconfidence. As a robustness check we replicate our analysis using an alternate measure of overconfidence by replacing actual performance in each period with predicted performance.[32] This measure removes the effect of any random shocks like those mentioned above. Table 4 replicates our analysis with this alternate measure.

Overall we find very similar results to our main specification. Subjects who are highly overconfident by this measure are 44 percentage points more likely to choose the convex scheme, and are 26 percentage points more likely to mistakenly choose the convex scheme (costing them 9% of their potential payoff). Highly underconfident subjects are 22 percentage points more likely to mistakenly choose the linear scheme (costing them 7% of their potential payoff). Our main measure of

[32] As before we estimate performance using a logarithmic trend for each subject.

overconfidence, therefore, does not seem to represent mistakes or other random factors leading to decreased performance.

Table 4: Overconfidence with Predicted Performance

VARIABLES	(1) Chose Convex	(2) Mistakenly Chose Convex	(3) % Lost Pay From Convex Mistake	(4) Mistakenly Chose Linear	(5) % Lost Pay From Linear Mistake
Overconfidence (fitted perf.)	0.0503***	0.0297***	0.0100***	-0.0251***	-0.00848***
	(0.00804)	(0.00742)	(0.00274)	(0.00517)	(0.00204)
Escalated Convex Available	0.0394	0.0387	0.00480	-0.0384	-0.0124
	(0.0334)	(0.0372)	(0.0115)	(0.0281)	(0.0129)
Escalated x Overconfidence	-0.00963	0.0160	0.00146	0.00193	-0.000364
	(0.00802)	(0.0139)	(0.00448)	(0.00896)	(0.00466)
Baseline Ability	0.326***	0.121***	0.0313**	-0.162***	-0.0380*
	(0.107)	(0.0450)	(0.0142)	(0.0542)	(0.0198)
Lottery Certainty Equivalent	-0.00766*	-0.00123	6.08e-05	-0.00135	-0.000456
	(0.00453)	(0.00192)	(0.000674)	(0.00241)	(0.000929)
Period	0.0303***	-0.0190**	-0.00534**	-0.0115	-0.000189
	(0.00766)	(0.00864)	(0.00265)	(0.00801)	(0.00298)
Constant	0.303**	0.194***	0.0506**	0.396***	0.0892***
	(0.142)	(0.0743)	(0.0240)	(0.0922)	(0.0309)
Observations	816	816	816	816	816
Number of Subjects	136	136	136	136	136

Robust standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1

As an additional robustness check, we regress mistakenly choosing the convex scheme on the same explanatory variables as model (2) in Table 3, but with the alternate measures of overconfidence used earlier in Table 1. We report these findings in Table 5. All of the alternate measures of overconfidence are at least marginally correlated with mistakenly choosing the convex scheme, except for overconfidence on the Trivia task. This suggests that a quite broad notion of overconfidence underlies the propensity of overconfident subjects to (mistakenly) prefer the convex piece rate. Additionally, the fact that subject-level measures of overconfidence (such as initial overconfidence and the independent test of

Table 5: Alternate Measures of Overconfidence

VARIABLES	(1)	(2)	(3)	(4)	(5)
Mean Overconfidence (Pds 1-3)	0.00715**				
	(0.00361)				
Escalated x Overconfidence (Pds 1-3)	-0.00633				
	(0.00501)				
Relative Overconfidence		0.216***			
		(0.0428)			
Escalated x Relative Overconfidence		0.0314			
		(0.0816)			
Trivia Overconfidence (Total)			0.00242		
			(0.00444)		
Escalated x Trivia (Total)			-0.00326		
			(0.0124)		
Trivia Overconf. (Question Rating)				0.0301	
				(0.0187)	
Escalated x Trivia (Question)				-0.0217	
				(0.0300)	
Independent Overconfidence					0.0388**
					(0.0165)
Escalated x Indep. Overconfidence					-0.0569**
					(0.0268)
Escalated Convex Available	0.0166	0.0320	0.0379	0.0474	0.0380
	(0.0386)	(0.0324)	(0.0363)	(0.0382)	(0.0354)
Baseline Ability	0.00139	0.00152	0.000683	0.000300	-0.000699
	(0.00221)	(0.00227)	(0.00222)	(0.00219)	(0.00244)
Lottery Certainty Equivalent	0.0822*	0.0756	0.0757	0.0875*	0.0858
	(0.0480)	(0.0501)	(0.0492)	(0.0492)	(0.0584)
Period	-0.0206**	-0.0201**	-0.0208**	-0.0206**	-0.0222**
	(0.00884)	(0.00856)	(0.00885)	(0.00885)	(0.00891)
Constant	0.181**	0.115	0.181**	0.160**	0.203**
	(0.0791)	(0.0803)	(0.0808)	(0.0794)	(0.0838)
Observations	816	816	816	816	738
Number of Subjects	136	136	136	136	123

Robust standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1
The dependent variable is an indicator for mistakenly choosing the convex scheme at the subject-round unit of observation.

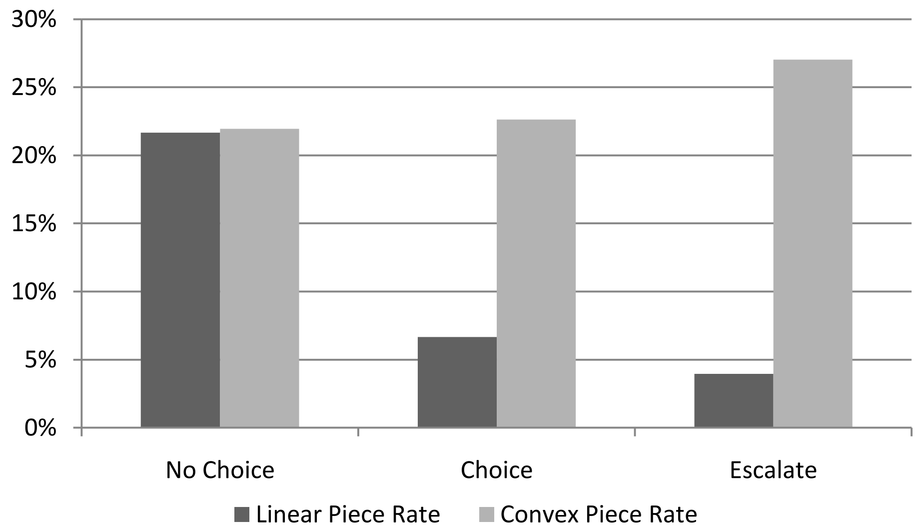
relative overconfidence) are correlated to overconfidence in the multiplication task corroborates our finding that a specific subgroup of people are overconfident, and are not learning to correct their expectations over time.

Incentive and Payoff Effects

We next consider whether the two payment schemes differ in terms of their incentive effects. We begin by comparing the average percent increase over baseline ability under the linear and convex piece rates in each of the three treatments (presented in Figure 6). In the No Choice control the convex piece rate appears to have no differential incentive effect compared to the linear piece rate; subjects in this condition who were randomly assigned the convex piece rate do not perform any better than subjects randomly assigned the linear piece rate. However, when subjects are allowed to choose their payment scheme, subjects who choose the linear scheme perform much worse (while subjects who choose the convex scheme perform similar to the control treatment). This suggests that subjects who want to exert low effort sort out of the convex scheme and into the linear scheme.

A substantial share of this lack of motivation comes from the 17% of subjects in the Choice and Escalate treatments who choose the linear piece rate in all six periods. These subjects perform significantly worse under the linear scheme than the subjects who choose a mix of linear and convex – subjects who always choose linear have an average score of 2% below their baseline ability, compared to 13% above baseline for the other subjects (Mann-Whitney U test on subject averages: $p < 0.001$). Additionally, these unmotivated subjects who always choose linear are somewhat more likely to be underconfident, measured both by overconfidence in the first three periods of the math task (test of proportions for being initially overconfident: $p = 0.098$) and by the self-reported measure of relative overconfidence (test of proportions: $p = 0.065$). Hence overconfidence may be related to motivation when subjects are allowed to sort.

Figure 6: Performance Relative to Baseline



We confirm this result by regressing performance on ability, treatment and pay scheme. Table 6 reports the results of the random effects regression of subject score in periods 4 through 9 on baseline ability, period, and dummy variables for the linear and convex piece rates in each treatment (with random assignment of the linear piece rate in the No Choice treatment being the omitted category). We find no difference in performance in the control treatment between the linear and convex piece rates. In the two choice treatments, however, subjects who opt into the linear piece rate perform significantly worse than subjects with the linear piece rate in the control (a difference of approximately 7.5% for a subject with average baseline ability). There is no significant difference between subjects who choose a convex piece rate and those who were randomly assigned a convex piece rate ($p = 0.88$). There is also no significant difference between the basic and escalated convex piece rate ($p = 0.46$).[33] Thus while it appears that a linear piece rate does not have a deleterious incentive effect in general, when subjects are free to sort into the linear scheme it induces less effort. This could explain why firms almost never give employees the

[33] We find the same results if we focus only on period 6 to 9, where the base and escalate convex piece rate are different.

choice of a linear pay scheme when a convex scheme is available; firms would prefer employees that are attracted to the linear scheme "sort away" to another firm.

Table 6: Incentive Effects

VARIABLES	Coeff.	Std. Err.
No Choice & Convex	0.148	(0.436)
Choice & Linear	-1.402**	(0.719)
Choice & Convex	0.044	(0.698)
Escalate & Linear	-1.558***	(0.649)
Escalate & Convex	-0.395	(0.635)
Baseline Ability	0.896***	(0.0361)
Period	0.494***	(0.0582)
Constant	2.422**	(0.933)
Observations	1074	
Number of Subjects	179	

Robust standard errors in parentheses.
Significance levels: *** p<0.01, ** p<0.05, * p<0.1
The dependent variable is total score on the multiplication task at the subject-round unit of observation.

Subject Payoffs

How costly are these mistakes to subjects, and how much would a hypothetical firm save in wage payments due to employee mistakes in wage scheme choice? Table 7 reports the average total earnings for the six choice periods in the Multiplication Task for each treatment,[34] as well as the minimum earnings subjects would earn if they were always placed in the lower payoff scheme and the maximum earnings subjects would earn if they were always placed in the higher payoff scheme. In the No Choice condition subjects' earnings are, not surprisingly, roughly halfway between their worst possible earnings and their best possible earnings. Subjects in the Choice treatment certainly do better than random choice of pay scheme, but they make mistakes one third of the time, costing them on average $3.64 over the six

[34] Recall, however, that we only paid subjects for two randomly selected periods from the nine Multiplication Task periods and the one Trivia Task period.

periods, or 31% of their potential improvement in payoffs from choosing correctly. In the Escalated condition subjects make mistakes one fourth of the time, losing out on $2.49, or 18% of their potential payoff improvement. Despite getting clear feedback on their performance every period, subjects who can select their pay scheme still leave money on the table, earning considerably less than they would if they always picked the right scheme.

Moreover, subjects in the choice treatments who were overconfident in the first three periods only improved their payoffs by 53% relative to the worst possible decisions. Essentially, these subjects' earnings were comparable to the randomly assigned subjects, suggesting overconfident subjects do not

Table 7: Subject Payoffs

Treatment	Actual Payoff	Min Payoff	Max Payoff	Payoff Improvement	Mistake Pct
No Choice	$47.62	$39.60	$55.55	50%	52%
Choice	$50.66	$39.04	$54.30	69%	34%
Escalate	$55.47	$37.75	$57.96	82%	25%

benefit by having choice of pay scheme, since they so often choose the wrong pay scheme. By contrast initially underconfident subjects earned $52.65 on average, improving their payoff by 73% compared to the worst possible decisions. In contrast to those that are overconfident, underconfident subjects do benefit from having a choice of pay scheme.

While our experiment focused on employee decisions and confidence levels, it is worthwhile to analyze the results from the point of view of a hypothetical firm that sets compensation. Table 8 breaks down average subject-round and per-question payments across treatments. In all three treatments, average subject-round payments were substantially higher under the convex scheme than the linear one. However, because of the performance differences across treatments discussed in the previous section, the per-question payments were more similar. Still, under the task and payoff schedules used in our experiment, a hypothetical firm would prefer to use only a linear compensation scheme.

Table 8: Average Total and Per-Question Payments

Incentive Scheme	No Choice: Average Subject-Round Payoff	No Choice: Average Payoff per Question	Choice: Average Subject-Round Payoff	Choice: Average Payoff per Question	Escalate: Average Subject-Round Payoff	Escalate: Average Payoff per Question
Linear	$7.17	$0.30	$6.17	$0.30	$6.54	$0.30
Convex	$8.69	$0.36	$9.77	$0.38	$10.80	$0.45

However, our results suggest a firm may be able to find circumstances under which the convex scheme is optimal, and point to several potential avenues for future research on this question. First, it is clear that the performance "break" point where the convex scheme pays more than the linear scheme is important. We set the same "break" point for all subjects – initial performance plus 1.5 questions – and did not adjust it over the course of the experiment. Firms commonly adjust these "break" points; for example, sales quotas and goals based on metrics like revenue and market share are typically adjusted at least annually. Second, as demonstrated by the "Escalate" treatment, optimality is clearly affected by the degree of convexity. Future research could investigate this in much greater detail. Finally, convex schemes may be most attractive for firms when overconfidence has a direct effect of task performance. This was not true of our simple multiplication task, but has been shown to be true for certain real-world job activities such as sales and entrepreneurship.

Direct Comparative Feedback

In order to further examine the persistence of overconfidence and mistaken pay scheme choice, we conducted an additional **Comparative Feedback** treatment. This treatment was the same as the Choice treatment; however, in addition to the feedback subjects receive on their score, guess and payoff after each period, they also received an additional screen of feedback after period 6 (the third choice period) comparing what their payoff would have been under each pay scheme given their performance in period 6. This makes it even easier for subjects to see what the optimal pay scheme choice would have been, and makes the exact cost of their mistake completely clear.

Results

An additional 76 subjects participated in these sessions. Average earnings were $25.63. Figure 7 shows the frequency of convex choices, as well as mistake choices, in each period. As in our other treatments, subjects increasingly choose the convex piece rate over the course of the experiment; similarly, they also continue to mistakenly choose linear and convex between 10 and 20 percent of the time each. Additionally, there does not appear to be a decrease in the frequency of these mistakes in the last three periods despite having received the comparative feedback.

Figure 7: Comparative Feedback Treatment

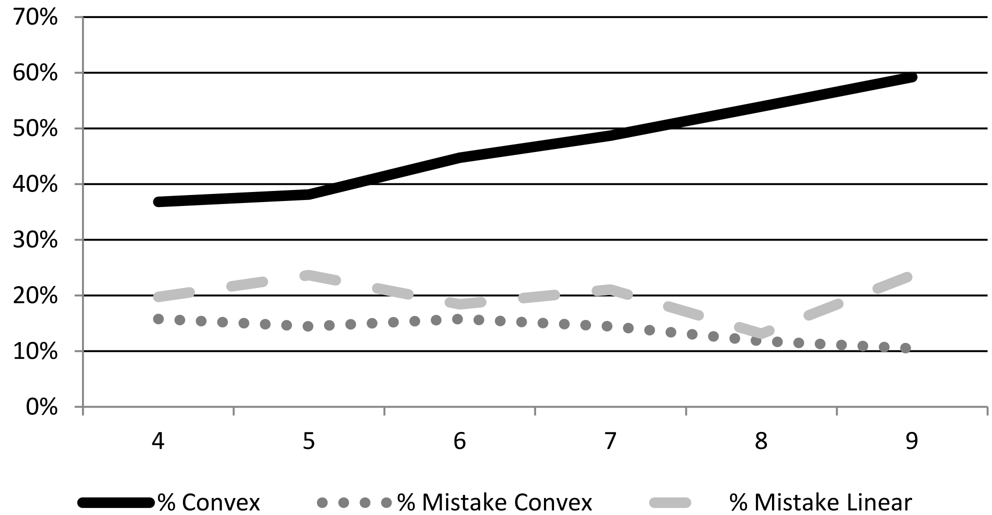


Table 9 reports the results of random effects regressions with either mistakenly choosing convex or mistakenly choosing linear as the dependent variables.[35] In models (1) and (3) we include a dummy variable for the periods after the comparative feedback was given (i.e. periods 7 to 9), while in models (2) and (4) we have a dummy for the last three periods for the subjects who made a mistake during the comparative feedback period. Providing comparative feedback does not significantly decrease the frequency of mistakes, nor does it affect the relationship between overconfidence and mistakes, even for

[35] One subject is excluded because he did not have a unique switching point in the risk measure. The same results are obtained if we run the specifications without the risk measure and include this subject.

the subjects who were shown that they were making a mistake by the comparative feedback. Thus the propensity for overconfident and underconfident individuals to choose the wrong pay scheme appears robust, even in the presence of very clear feedback showing the optimal pay scheme given their performance.

Table 9: Comparative Feedback Treatment

	(1)	(2)	(3)	(4)
VARIABLES	Mistakenly Chose Convex		Mistakenly Chose Linear	
Overconfidence	0.0308***	0.0317***	-0.0378***	-0.0426***
	(0.0075)	(0.0067)	(0.0085)	(0.0079)
After Feedback Given	-0.0287		-0.0646	
	(0.0345)		(0.0560)	
After Feedback x Overconfidence	0.0048		-0.0038	
	(0.0091)		(0.0112)	
After Feedback & Mistake at Feedback		-0.0020		-0.0281
		(0.0583)		(0.0495)
Feedback Mistake x Overconfidence		0.0023		0.0166
		(0.0104)		(0.0116)
Lottery Certainty Equivalent	0.0381	0.0381	-0.0819	-0.0882
	(0.0867)	(0.0876)	(0.0674)	(0.0654)
Max Practice Score	-0.0032	-0.0031	-0.0048	-0.0049
	(0.0035)	(0.0035)	(0.0031)	(0.0031)
Constant	0.2023*	0.1851*	0.3422***	0.3478***
	(0.1088)	(0.1096)	(0.0807)	(0.0781)
Observations	450	450	450	450
Number of Subjects	75	75	75	75

Robust standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1

Conclusion and next steps

Our findings suggest that non-linear incentive schemes “sort in” overconfident workers, and “sort out” workers who may become demotivated. We find that many subjects are persistently over- or underconfident despite receiving clear feedback, and that these belief biases affect incentive scheme choices. Overconfident subjects are more likely to choose the convex scheme, and are more likely to make a mistake in doing so – overconfident subjects do no better than chance in picking the correct pay scheme – and their mistakes cost them up to 15% of their potential payoff. We also find that, while the linear and convex schemes have no difference in their direct incentive effects for the task in the experiment, subjects who *choose* the linear scheme significantly underperform compared to subjects who choose the convex scheme.

Extrapolating our experimental results, overconfident employees facing a choice between a linear pay system and a convex one are, ceteris paribus, significantly more likely to choose a job setting with convex pay and continue to do so even if it pays them less. The fact that these subjects appear to persist in their mistaken pay scheme choices may help explain why convex schemes have lasted so long in corporate environments. Similarly, our results suggest that underconfident employees may shy away from pay that is more dependent on high performance. In short, the linearity of the incentive system could play an important role in sorting employees by their level of confidence. Additionally, unmotivated employees who wish to exert lower effort may also prefer a linear incentive scheme. Importantly, our findings suggest this sorting does not occur along the skill dimension, which is the standard argument in the literature on the benefits of sorting. Rather, as with the literature on social preferences and sorting, our results suggest using an incentive system which sorts by a behavioral trait – overconfidence in the case of our study – may be beneficial to firms.

It is important to note that our study, like any experimental study, abstracted away from many important factors. In our experiment, the two "firms" are identical except for their offered incentive schemes; in reality, firms differ across many dimensions which likely affect sorting decisions made by employees. Interestingly, a leading survey of MBAs suggests that compensation is only the second most important factor driving job decisions, after lifestyle (Hudson Group 2008), so our study is obviously incomplete. Similarly, there may be unobserved differences in employee type that are correlated with confidence that better explain why non-linear incentive schemes are so prevalent. Still, we find comfort in the fact that our sorting results were so highly correlated with overconfidence measured in five separate ways.

The experimental design in this paper was focused on understanding the choices and performance of workers in the context of two commonly-observed incentive schemes, and how these employee choices could benefit firms. It did not directly address the firm's problem of determining the optimal menu of incentive schemes given the sorting behavior of workers. A natural and interesting direction for future research would be to directly examine firm behavior and identify cases where exploiting worker sorting increases the profit of the firm. The existing literature on overconfidence as well as our experimental results suggests that the benefit of identifying overconfident workers will vary between tasks – depending on both the production process and the value of output to the firm.

In future research, we hope to further explore the several plausible benefits to firms of offering contracts that appeal to overconfident employees. The first is that overconfidence is likely a beneficial worker trait for certain job functions, particularly tasks like sales that require persistence in the face of repeated failure (Zoltners et al. 2006). Compte and Postlewaite (2004) argue that for tasks where nervousness or stress are detrimental to performance, e.g. for a lawyer arguing a high-profile case, overconfidence can increase performance. For these jobs overconfident workers may be more productive. On the other hand, underconfidence may be desirable in environments where the cost of mistakes is very high. These are ripe subjects for future research.

Finally, the non-linearity of the incentive system may allow firms to lower their wage bill. A convex scheme could allow firms to take advantage of overconfident employees' systematic and persistent bias towards believing they will perform well, while a linear scheme may exploit the underconfidence of employees. Indeed, highly overconfident subjects in our experiment lost approximately 15% of their potential payment because they made mistakes in their choice of incentive scheme.

References

Alicke, Mark D. (1985). "Global Self-Evaluation as Determined by the Desireability and Controllability of Trait Adjectives." *Journal of Personality and Social Psychology*, 49(6), 1621-1630.

Alicke, Mark D., M.L. Klotz, David L. Breitenbecher, Tricia J. Yurak, and Debbie S. Vredenburg (1995). "Personal Contact, Individuation and the Better-Than-Average Effect." *Journal of Personality and Social Psychology*, 68(5), 804-825.

Ausubel, Lawrence M. (1999). "Adverse Selection in the Credit Card Market." Working Paper, University of Maryland.

Barber, Brad and Terrance Odean (2001). "Boys Will be Boys: Gender, Overconfidence, and Common Stock Investment." *The Quarterly Journal of Economics*, 116 (1), 261-292.

Bengtsson, Claes, Persson, Mats, and Peter Willenhag (2005). "Gender and overconfidence." *Economics Letters*, 86 (2), 199-203.

Bergman, Nittai K. and Dirk Jenter (2007). "Employee Sentiment and Stock Option Compensation," *Journal of Financial Econometrics*, 84(3), 667-712.

Beyer, Sylvia (1990). "Gender differences in the accuracy of self-evaluations of performance." *Journal of Personality and Social Psychology*, 59 (5), 960-970.

Brown, Martin, Armin Falk and Ernst Fehr (2004) "Relational Contracts and the Nature of Market Interactions," *Econometrica*, 72, 747-780.

Busenitz, Lowell W. and Jay B. Barney (1997). "Differences Between Entrepreneurs and Managers In Large Organizations: Biases and Heuristics In Strategic Decision Making," *Journal of Business Ventures*, 12(1), 9-30.

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.

Camerer, Colin and Dan Lovallo (1999). "Overconfidence and Excess Entry: An Experimental Approach," *American Economic Review*, 89(1), 306-318.

Campbell, W. Keith, Adam S. Goodie, and Joshua D. Foster. (2004). "Narcissism, Confidence, and Risk Attitude." *Journal of Behavioral Decision Making*, 17(4), 481-502.

Cesarini, David, Magnus Johannesson, Paul Lichtenstein, and Björn Wallace (2009). "Heritability of Overconfidence." *Journal of the European Economic Association*, 7(2), 617-627.

Compte, Olivier and Andrew Postlewaite (2004). "Confidence-Enhanced Performance." *American Economic Review*, 94(5), 1536-1557.

DellaVigna, Stefano (2009). "Psychology and Economics: Evidence from the Field." *Journal of Economic Literature*, 47, 315-372.

DellaVigna, Stefano and Ulrike Malmendier (2006). "Paying Not to Go to the Gym." *The American Economic Review*, 96(3), 694-719.

Dohmen, Thomas and Armin Falk (forthcoming). "Performance Pay and Multidimensional Sorting: Productivity, Preferences and Gender." *American Economic Review*.

Englmaier, Florian (forthcoming) "Managerial Optimism and Investment Choice," *Managerial and Decision Economics*.

Fehr, Ernst and Armin Falk (1999). "Wage Rigidity in a Competitive Incomplete Labor Market," *Journal of Political Economy*, 107, 106-134.

Fehr, Ernst and Simon Gächter (2002). "Do incentive contracts crowd out voluntary cooperation?" *working paper*.

Fischbacher, Urs (2007). "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Economics*, 10(2), 171-178.

Fischoff, Baruch and Paul Slovic (1980). "A Little Learning: Confidence in Multicue Judgement." in *Attention and Performance VIII*, R. Nickerson (ed.), Hillsdale: Erlbaum.

Fischhoff, Baruch, Paul Slovic and Sarah Lichtenstein (1977). "Knowing with Certainty: The Appropriateness of Extreme Confidence." *Journal of Experimental Psychology: Human Perception and Performance*, 3, 552-564.

Gino, Francesca, Ian Larkin and J. Lamar Pierce (2009). "The Psychological Costs of Pay-For-Performance: Implications for Strategic Compensation." Mimeo.

Gosling, Samuel D., Peter J. Rentfrow and William B. Swann Jr. (2003). "A Very Brief Measure of the Big Five Personality Domains." *Journal of Research in Personality*, 37(6), p. 504-528.

Grubb, Michael D. (2009)."Selling to Overconfident Consumers." *American Economic Review*. 99(5), 1770-1807.

Hart, Oliver and Bengt Holmstrom (1987). "The Theory of Contracts". In T.F. Bewley (Ed.), *Advances in Economic Theory: Fifth World Congress*. Cambridge University Press.

Healy, Paul and Done Moore (2007). "Bayesian Overconfidence." Working paper. Available at http://cess.nyu.edu/exp_seminar/fall_08_papers/Healy_Moore-Overconfidence.pdf. Retrieved on July 30, 2009.

Klayman, Joshua, Jack Soll, Claudia Gonzalez-Vallejo and Sema Barlas (1999). "Overconfidence: It Depends on How, What, and Whom You Ask." *Organizational Behavior and Human Decision Processes*, 79(3), 216-247.

Kruger, Justin (1999). "Lake Wobegon be gone! The 'Below-Average Effect' and the Egocentic Nature of Comparative Ability Judgments." *Journal of Personality and Social Psychology*, 77, 221-232.

Kwan, Virginia S.Y., Oliver P. John, David A. Kenny, Michael H. Bond, and Richard W. Robins (2004). "Reconceptualizing Individaul Differences in Self-Enhancement Bias: An Interpersonal Approach." *Psychological Review*, 111(1), 94-110.

Larkin, Ian. (2007). "The Cost of High-Powered Incentives: Employee Gaming in Enterprise Software Sales." Working paper, Harvard Business School.

Lazear, Edward P., and Paul Oyer (2009). "Personnel Economics." Forthcoming chapter in *Handbook of Organizational Economics*, Robert Gibbons and John Roberts, editors. Princeton University Press. Retrieved at http://faculty-gsb.stanford.edu/oyer/wp/handbook.pdf. Last accessed March 9, 2010.

Locke, Edwin and Gary Latham (2002). "Building a Practically Useful Theory of Goal Setting and Task Motivation: A 35-Year Odyssey," *American Psychologist*, 57 (9): 705-717.

Malmendier, Ulrike and Geoffrey Tate (2005). "CEO Overconfidence and Corporate Investment." *The Journal of Finance*, 60 (6), 2661-2700.

Malmendier, Ulrike and Geoffrey Tate (2008). "Who Makes Acquisitions? CEO Overconfidence and the Market's Reaction." *Journal of Financial Economics*, 89(1), 20-43.

Möbius, Markus M., Muriel Niederle, Paul Niehaus and Tanya Rosenblat (2010). "Self-Confidence Management: Theory and Experimental Evidence", *working paper*.

Moore, Don and Daylian Cain (2007). "Overconfidence and underconfidence: When and why people underestimate (and overestimate) the competition." *Organizational Behavior and Human Decision Processes,* 103(2), 197-213.

Nekby, Lena, Peter S. Thoursie and Lars Vahtrik (2008). "Gender and Self-Selection Into a Competitive Environment: Are Women More Overconfident Than Men?" *Economic Letters*, 100(3), 405-407.

Nelson, Thomas O., and Louis Narens (1980). "Norms of 300 general-information questions: Accuracy of recall, latency of recall, and feeling-of-knowing ratings." *Journal of Verbal Learning and Verbal Behavior*, 19, 338-368.

Niederle, Muriel and Lise Vesterlund (2007). "Do Women Shy Away from Competition? Do Men Compete Too Much?" *The Quarterly Journal of Economics*, 122(3), 1067-1101.

Nisbett, Richard E. and Ross Lee (1980). *Human Inference: Strategies and Shortcomings of Social Judgement*, Prentice Hall: Englewood Cliffs, NJ.

Oster, Sharon M. and Fiona M. Scott-Morton (2005). "Behavioral Biases Meet the Market: The Case of Magazine Subscription Prices," *Advances in Economic Analysis & Policy*, 5(1), 1-30.

Oyer, Paul. (1998). "Fiscal Year Ends and Nonlinear Incentive Contracts: The Effect on Business Seasonality," *Quarterly Journal of Economics*, 113(1), 149-185.

Oyer, Paul and Scott Schaefer (2005). "Why Do Some Firms Give Stock Options to All Employees? An Empirical Examination of Alternative Theories," *Journal of Financial Economics*, 76(1), 99-133.

Rammstedt, Beatrice and Oliver P. John (2007). "Measuring Personality in One Minute or Less: A 10 Item Short Version of the Big Five Inventory in English and German." *Journal of Research in Personality*, 41(1), 203-212.

Russo, J. Edward and Paul J.H. Schoemaker (1992). "Managing Overconfidence," *Sloan Management Review*, 33(2), 7-17.

Saucier, Gerard (1994). "Mini-Markers: A Brief Version of Goldberg's Unipolar Big-Five Markers." *Journal of Personality Assessment*, 63(3), 506-516.

Sautmann, Anja (2009). "Contracts for Agents with Belief Biases: Some Theory and an Experiment." New York University, Working Paper. Retrieved at https://files.nyu.edu/as3304/public/Sautmann-SConf-0910.pdf. Last accessed June 28, 2010.

Schaefer, Peter S., Cristina C. Williams, Adam S. Goodie, and W. Keith Campbell. (2004). "Overconfidence and the Big Five." *Journal of Research in Personality*, 38(5), 473-480.

Skiba, Paige M. and Jeremy Tobacman (2008). "Payday Loans, Uncertainty, and Discounting: Explaining Patterns of Borrowing, Repayment and Default", Working Paper. Retrieved at http://www.economics.ox.ac.uk/members/jeremy.tobacman/papers/payday.pdf. Last accessed Nov. 17, 2010.

Spence, A.M. (1973). "Job Market Signaling". *Quarterly Journal of Economics*. 87(3): 355-374.

Svenson, Ola (1981). "Are We All Less Risky and More Skillful than Our Fellow Drivers?" *Acta Psychologica*, 47, 143-151.

Wasserman, Noam (2006). "Stewards, Agents, and the Founder Discount: Executive Compensation in New Ventures." *Academy of Management Journal*, 49(5), 960-976.

Weinstein, Neil D. (1980). "Unrealistic Optimism About Future Life Events." *Journal of Personality and Social Psychology*, 39(5), 806-820.

Wu, Brian and Anne Marie Knott (2006). "Entrepreneurial Risk and Market Entry," *Management Science*, 52(9), 1315-1330.

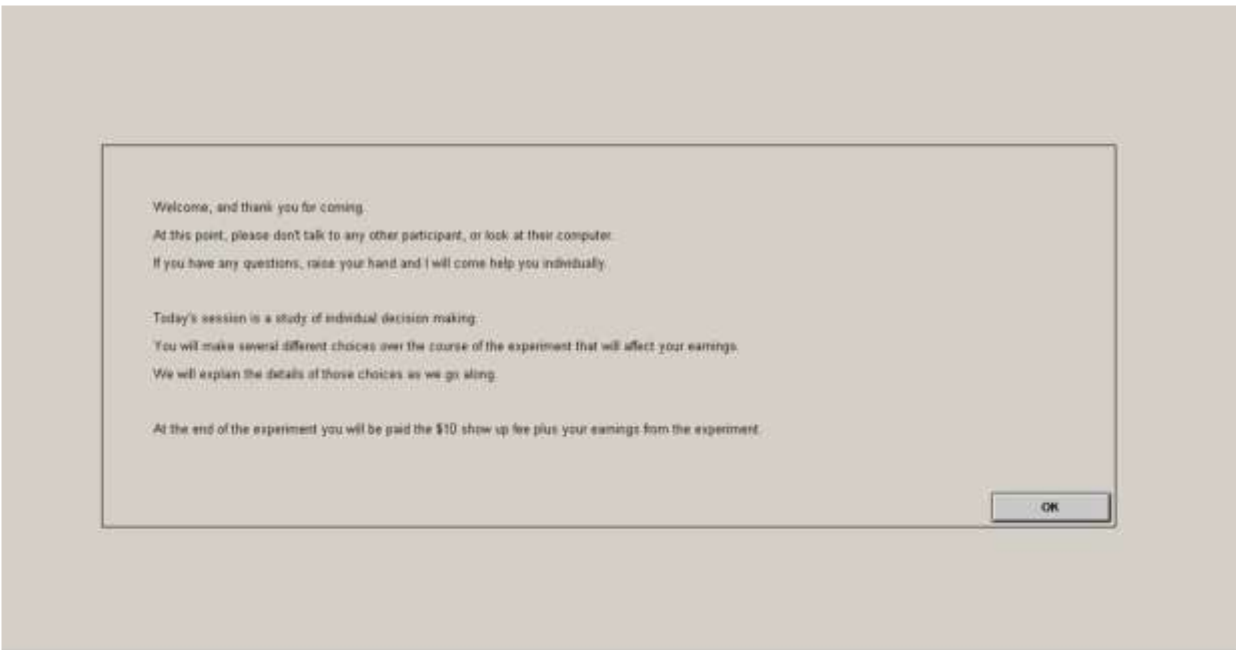
Zacharakis, Andrew L. and Dean A. Shepherd (2001). "The Nature of Information and Overconfidence On Venture Capitalists' Decision Making," *Journal of Business Ventures*, 16 (4), 311-332.

Zoltners, Andris A., Prabhakant Sinha, and Sally E. Lorimer (2006). *The Complete Guide to Sales Force Incentive Compensation: How to Design and Implement Plans That Work.* AMACOM, USA.

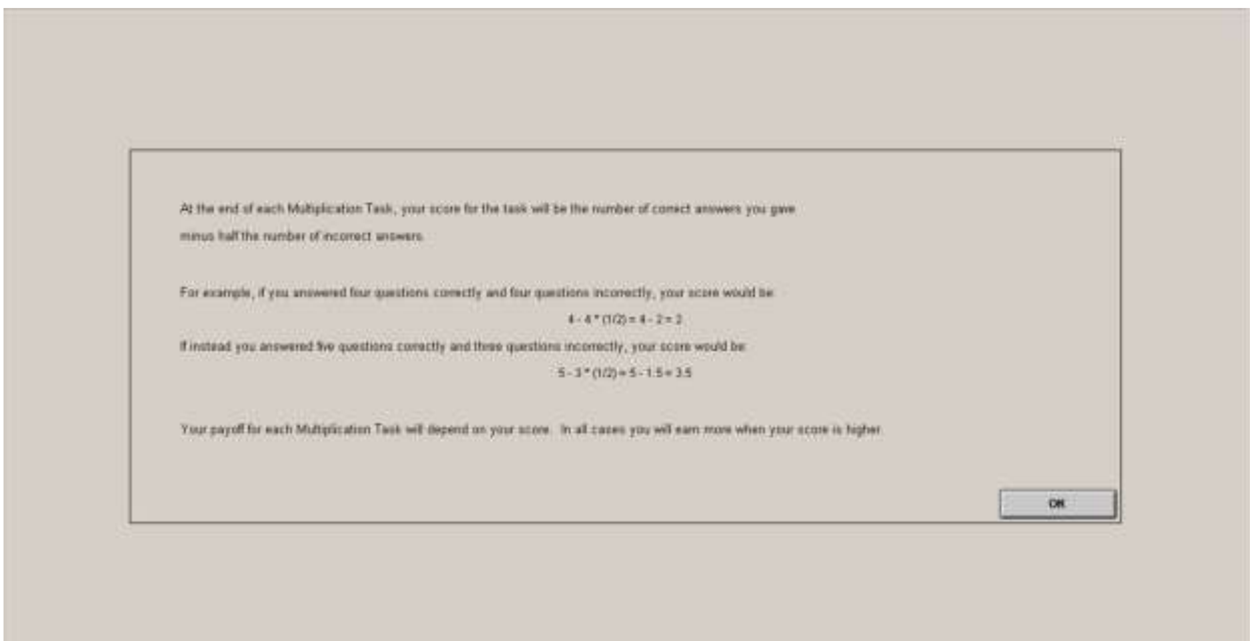
APPENDIX: EXPERIMENT INSTRUCTIONS AND SCREEN SHOTS*

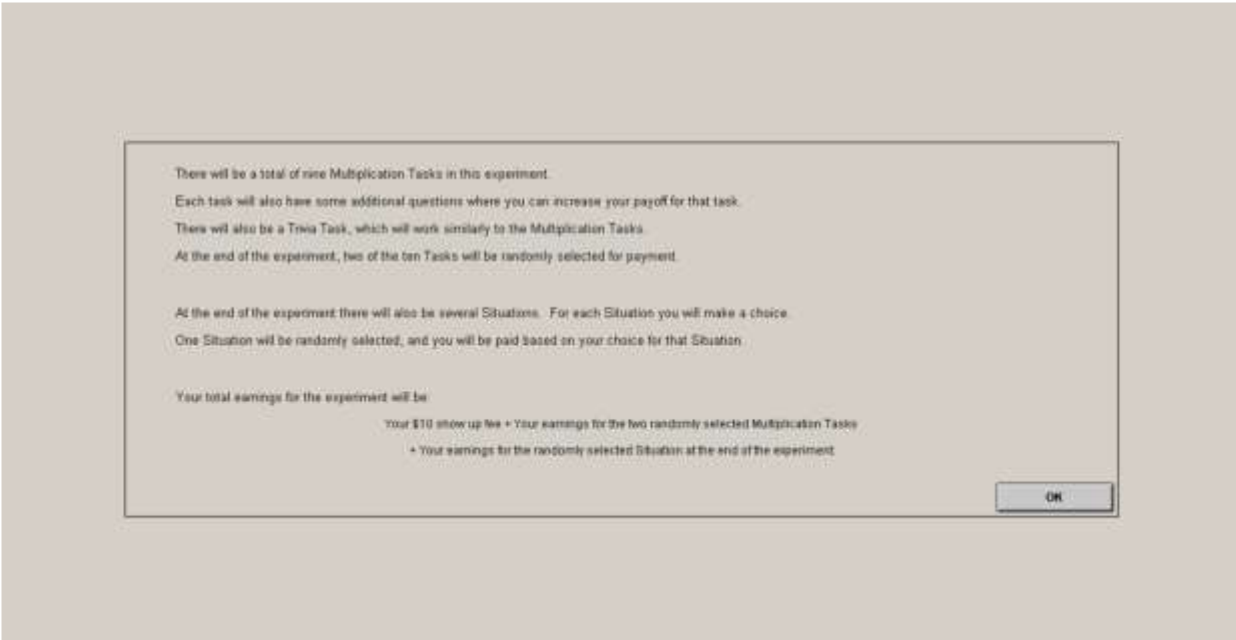
*To be posted online. Not intended for publication.

GENERAL INSTRUCTIONS:

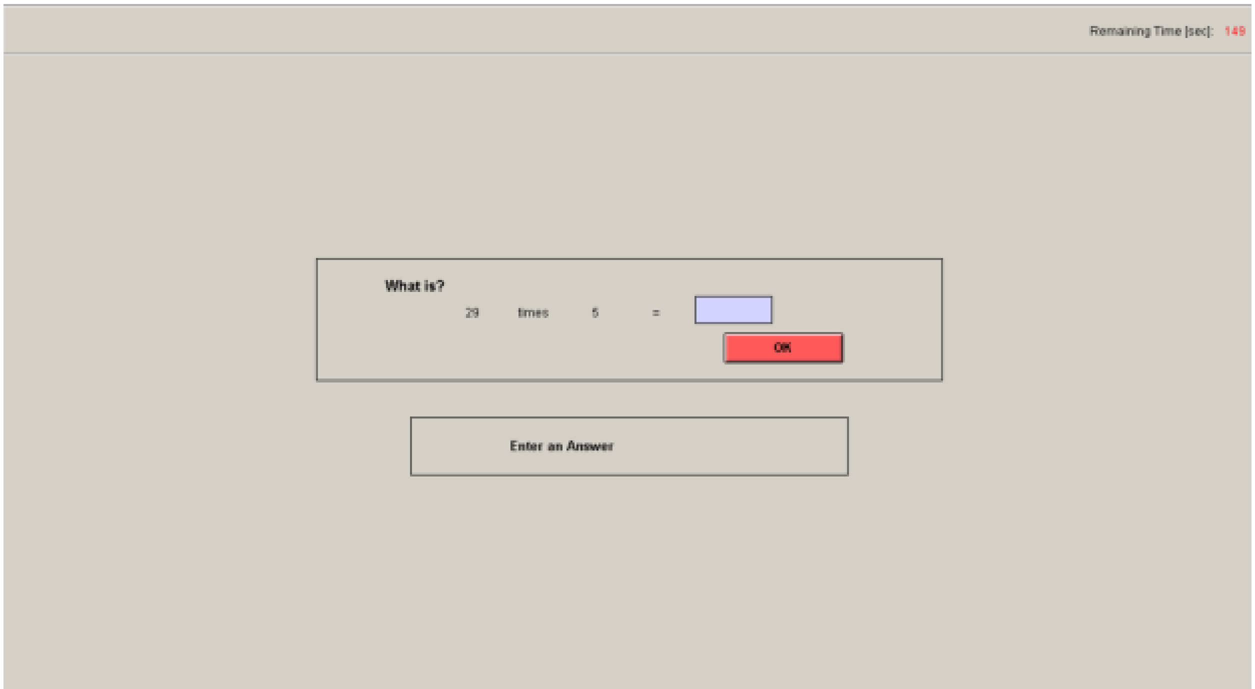



Throughout the experiment, you will be performing the Multiplication Task.
In the Multiplication Task, you will try to solve correctly as many multiplication problems as possible within two and a half minutes.
You may not use a calculator, but you may use scrap paper to answer the questions.
There are five different difficulty levels for multiplication problems in the Multiplication Task.
Here are examples of each difficulty level:
Difficulty 1: What is 24 times 3?
Difficulty 2: What is 37 times 4?
Difficulty 3: What is 45 times 8?
Difficulty 4: What is 73 times 9?
Difficulty 5: What is 89 times 8?
OK

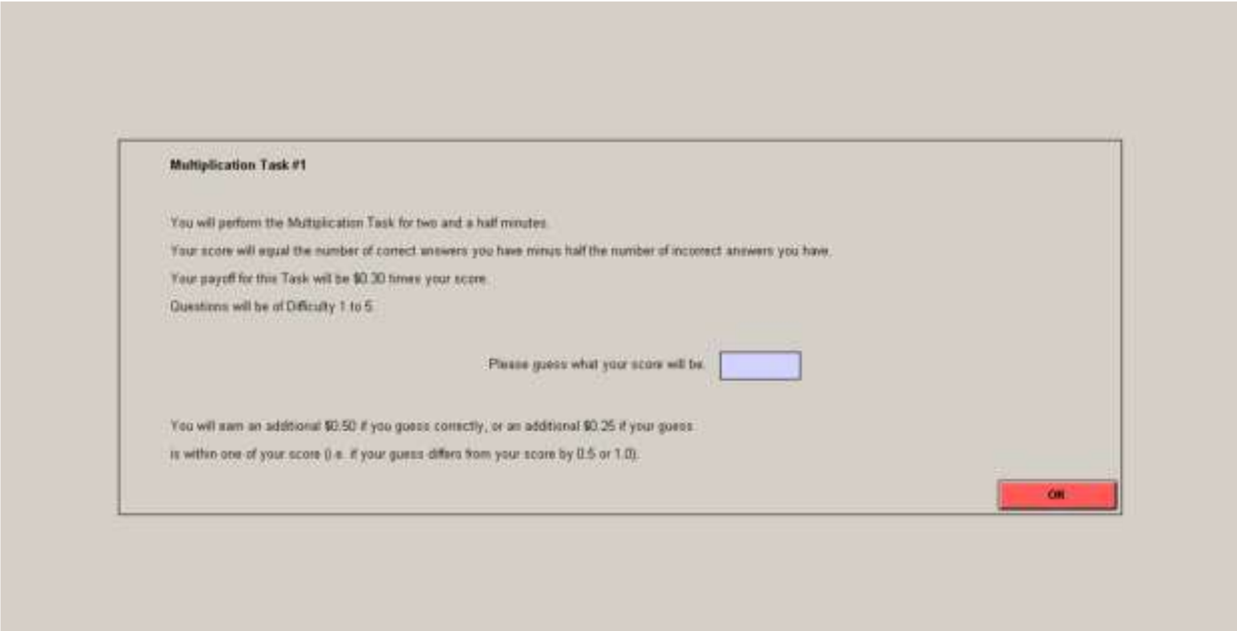

At the end of each Multiplication Task, your score for the task will be the number of correct answers you gave
minus half the number of incorrect answers.
For example, if you answered four questions correctly and four questions incorrectly, your score would be:
4 - 4 * (1/2) = 4 - 2 = 2
If instead you answered five questions correctly and three questions incorrectly, your score would be:
5 - 3 * (1/2) = 5 - 1.5 = 3.5
Your payoff for each Multiplication Task will depend on your score. In all cases you will earn more when your score is higher.
OK


There will be a total of nine Multiplication Tasks in this experiment.
Each task will also have some additional questions where you can increase your payoff for that task.
There will also be a Trivia Task, which will work similarly to the Multiplication Tasks.
At the end of the experiment, two of the ten Tasks will be randomly selected for payment.
At the end of the experiment there will also be several Situations. For each Situation you will make a choice.
One Situation will be randomly selected, and you will be paid based on your choice for that Situation.
Your total earnings for the experiment will be:
Your $10 show up fee + Your earnings for the two randomly selected Multiplication Tasks
+ Your earnings for the randomly selected Situation at the end of the experiment.
OK

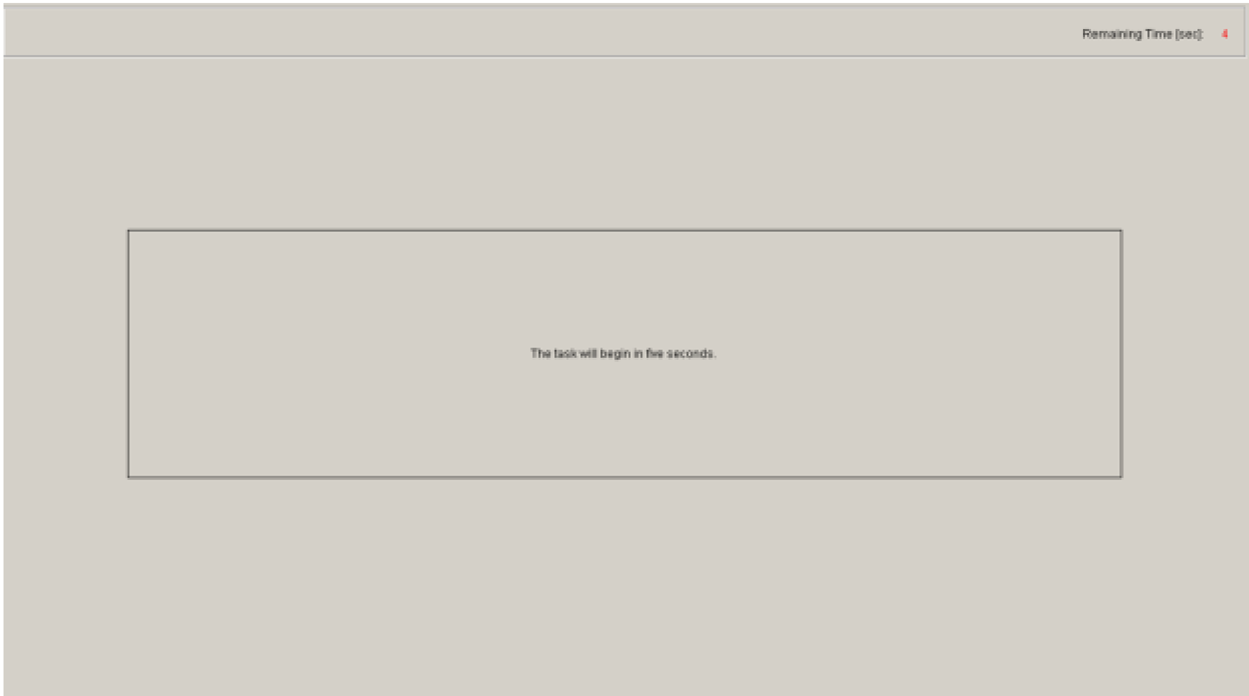
SAMPLE QUESTION (same for all multiplication rounds and all treatments):



PRACTICE ROUNDS 1-2:


Multiplication Task #1
You will perform the Multiplication Task for two and a half minutes.
Your score will equal the number of correct answers you have minus half the number of incorrect answers you have.
Your payoff for this Task will be $0.30 times your score.
Questions will be of Difficulty 1 to 5.
Please guess what your score will be.
You will earn an additional $0.50 if you guess correctly, or an additional $0.25 if your guess
is within one of your score (i.e. if your guess differs from your score by 0.5 or 1.0).
OK

(Presented prior to every task):



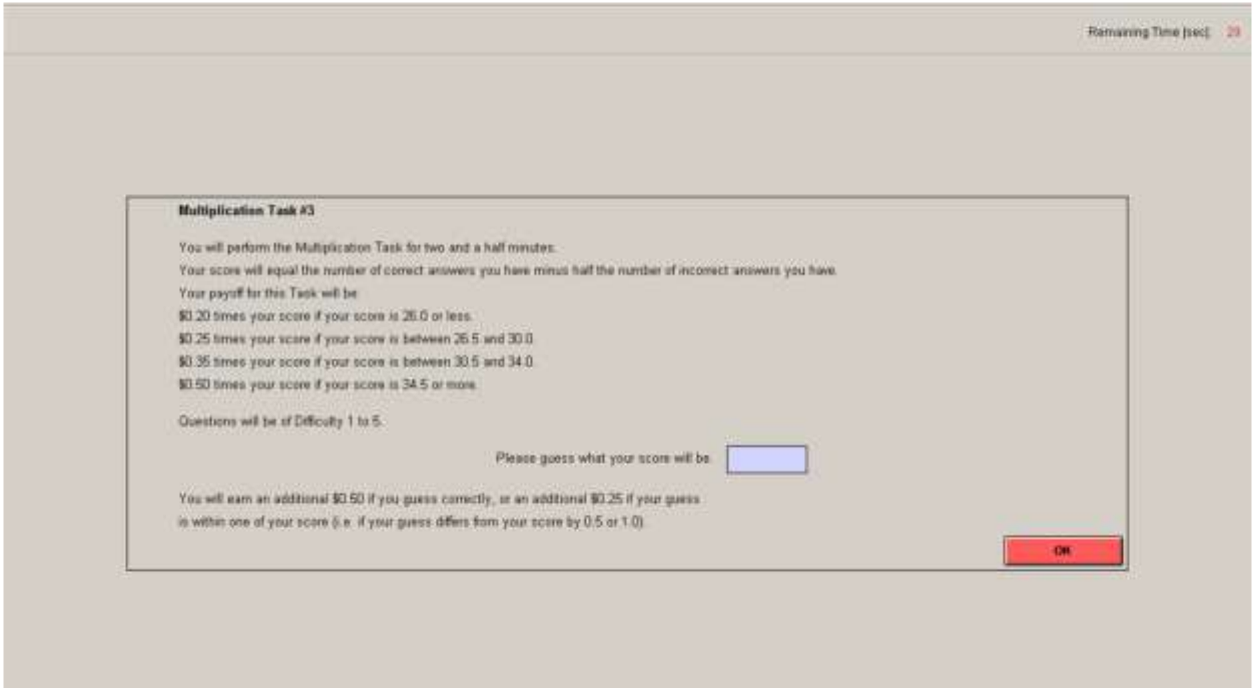
(This screenshot is presented after task completion in all ten rounds and for every treatment):


Remaining Time [sec]:
Number of other subjects: 20
Please guess how many other subjects had a lower score than you.
fewer than 5
at least 5, but fewer than 10
at least 10, but fewer than 15
at least 15
You will receive an additional $0.50 if you are correct.
OK

(This screenshot is also presented after task completion in all ten rounds and for every treatment):

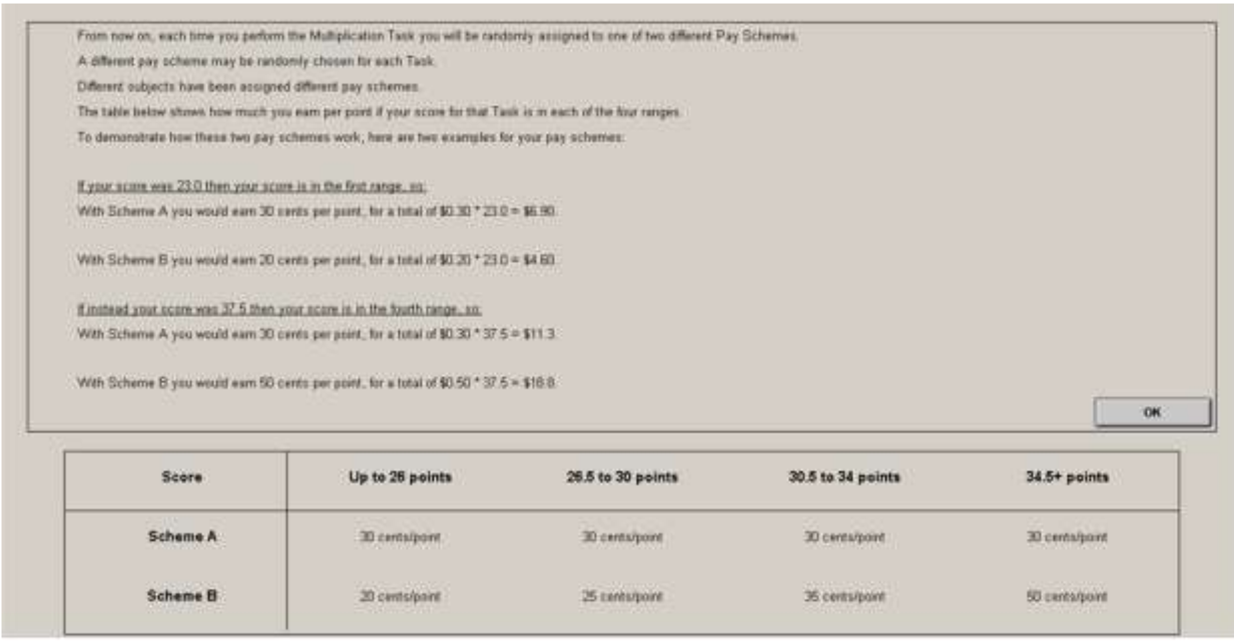
PRACTICE ROUND 3:

(Cutoffs adjust based on the maximum score achieved in practice rounds 1 & 2. The example seen here is based on a maximum practice score of 28.5).



Remaining Time [sec]: 29
Multiplication Task #3
You will perform the Multiplication Task for two and a half minutes.
Your score will equal the number of correct answers you have minus half the number of incorrect answers you have.
Your payoff for this Task will be:
$0.20 times your score if your score is 26.0 or less.
$0.25 times your score if your score is between 26.5 and 30.0.
$0.35 times your score if your score is between 30.5 and 34.0.
$0.50 times your score if your score is 34.5 or more.
Questions will be of Difficulty 1 to 5.
Please guess what your score will be:
You will earn an additional $0.50 if you guess correctly, or an additional $0.25 if your guess
is within one of your score (i.e. if your guess differs from your score by 0.5 or 1.0).
OK

ROUNDS 4-9 No-Choice Treatment:


From now on, each time you perform the Multiplication Task you will be randomly assigned to one of two different Pay Schemes.
A different pay scheme may be randomly chosen for each Task.
Different subjects have been assigned different pay schemes.
The table below shows how much you earn per point if your score for that Task is in each of the four ranges.
To demonstrate how these two pay schemes work, here are two examples for your pay schemes:

If your score was 23.0 then your score is in the first range, so:
With Scheme A you would earn 30 cents per point, for a total of $0.30 * 23.0 = $6.90.

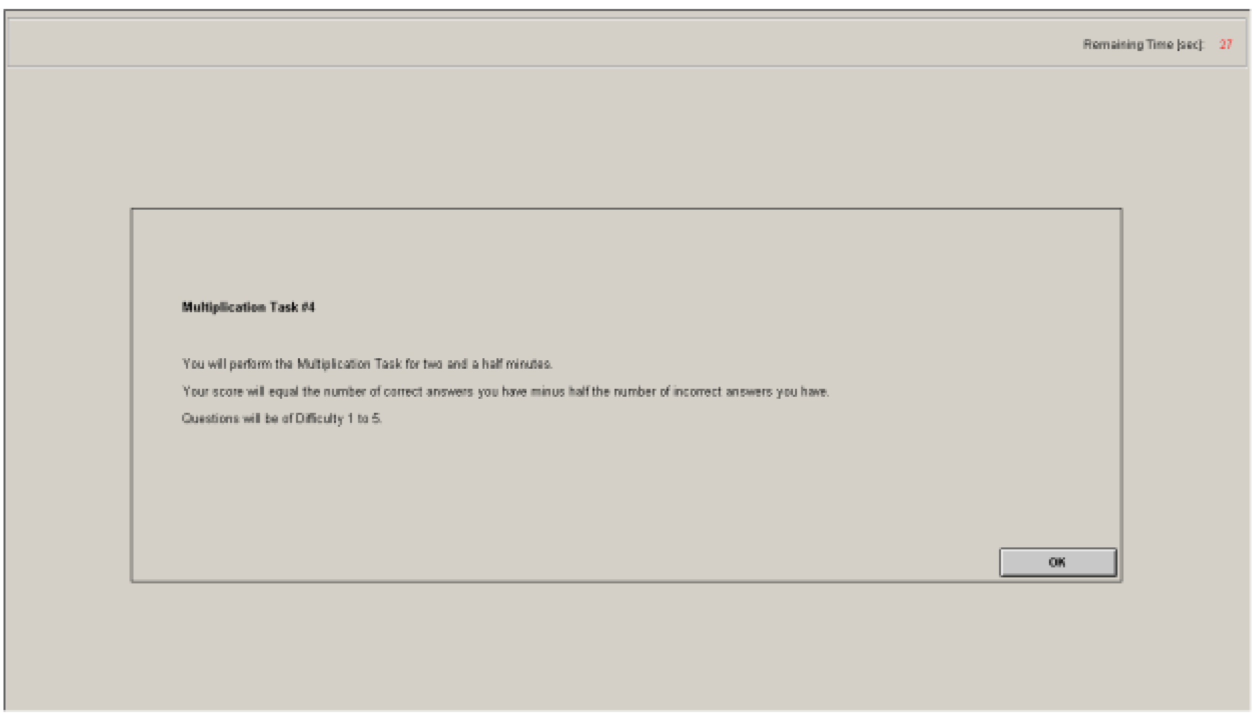
With Scheme B you would earn 20 cents per point, for a total of $0.20 * 23.0 = $4.60.

If instead your score was 37.5 then your score is in the fourth range, so:
With Scheme A you would earn 30 cents per point, for a total of $0.30 * 37.5 = $11.3.

With Scheme B you would earn 50 cents per point, for a total of $0.50 * 37.5 = $18.8.

OK

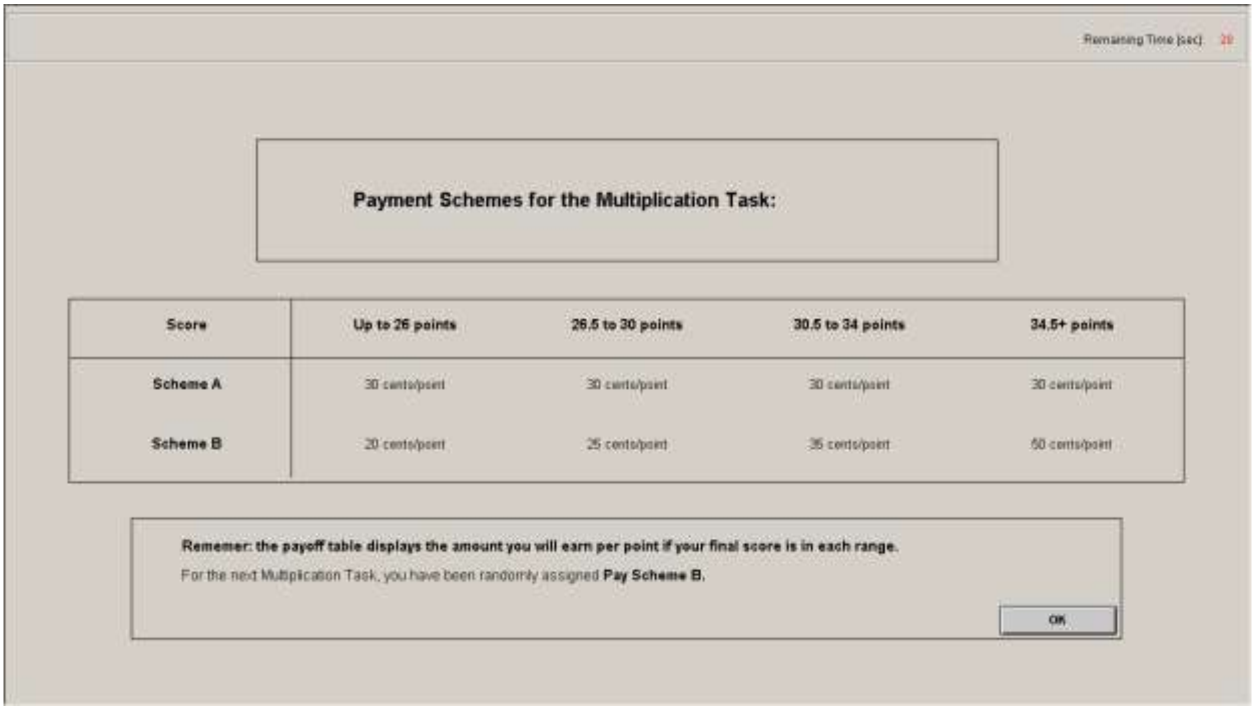
Score	Up to 26 points	26.5 to 30 points	30.5 to 34 points	34.5+ points
Scheme A	30 cents/point	30 cents/point	30 cents/point	30 cents/point
Scheme B	20 cents/point	25 cents/point	35 cents/point	50 cents/point


Remaining Time [sec]: 27

Multiplication Task #4

You will perform the Multiplication Task for two and a half minutes.
Your score will equal the number of correct answers you have minus half the number of incorrect answers you have.
Questions will be of Difficulty 1 to 5.

OK



Remaining Time [sec]

Payment Schemes for the Multiplication Task:

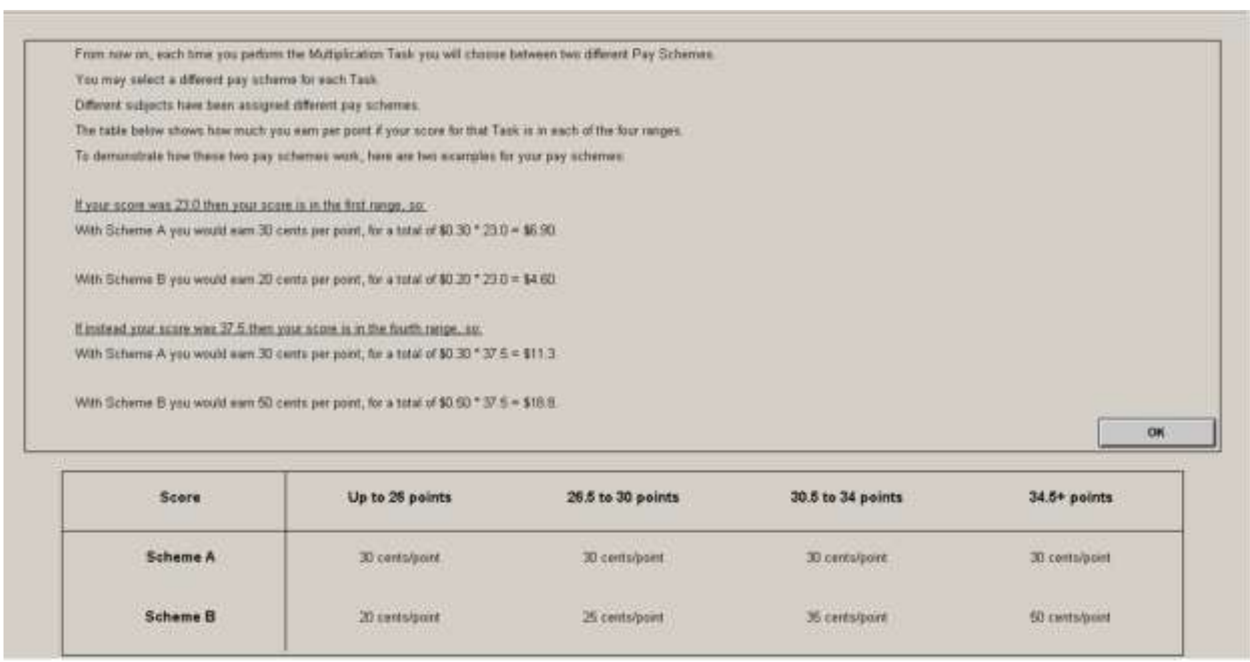
Score	Up to 26 points	26.5 to 30 points	30.5 to 34 points	34.5+ points
Scheme A	30 cents/point	30 cents/point	30 cents/point	30 cents/point
Scheme B	20 cents/point	25 cents/point	35 cents/point	50 cents/point

Rememer: the payoff table displays the amount you will earn per point if your final score is in each range.

For the next Multiplication Task, you have been randomly assigned **Pay Scheme B.**

OK

ROUNDS 4-9 Choice Treatment / ROUNDS 4-6 Escalate Treatment:



From now on, each time you perform the Multiplication Task you will choose between two different Pay Schemes.

You may select a different pay scheme for each Task.

Different subjects have been assigned different pay schemes.

The table below shows how much you earn per point if your score for that Task is in each of the four ranges.

To demonstrate how these two pay schemes work, here are two examples for your pay schemes:

If your score was 23.0 then your score is in the first range, so:

With Scheme A you would earn 30 cents per point, for a total of \$0.30 * 23.0 = \$6.90

With Scheme B you would earn 20 cents per point, for a total of \$0.20 * 23.0 = \$4.60

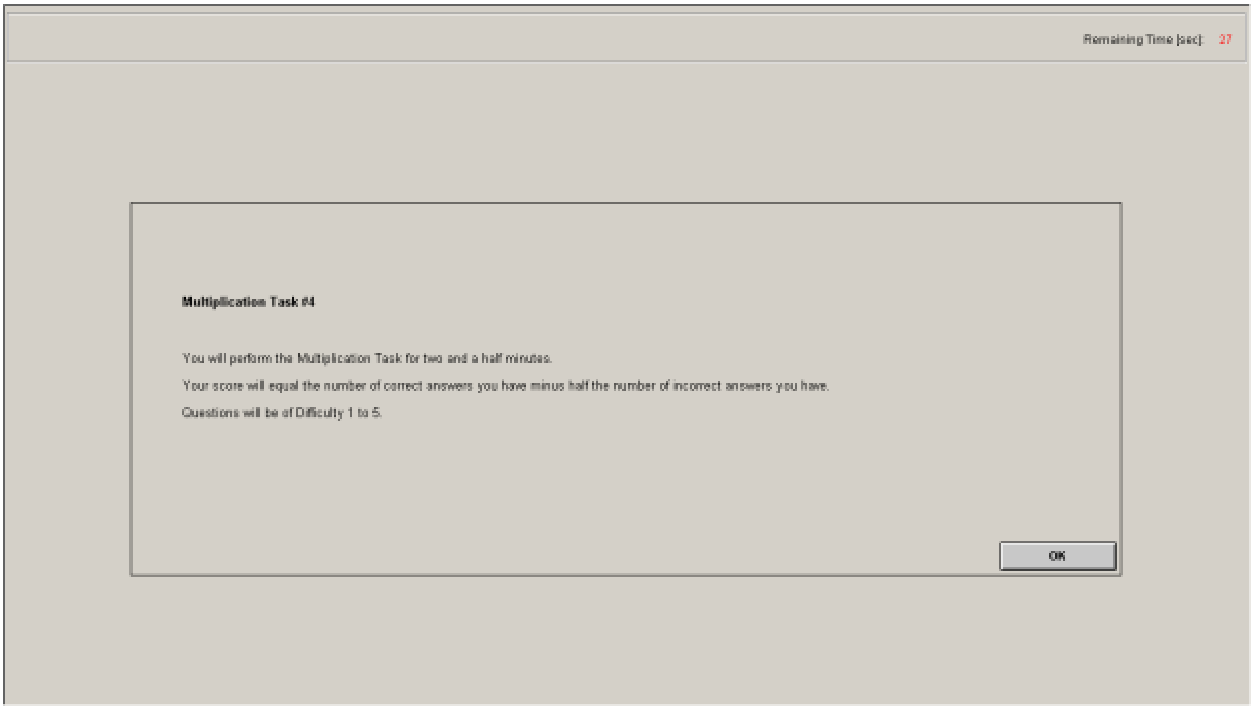
If instead your score was 37.5 then your score is in the fourth range, so:

With Scheme A you would earn 30 cents per point, for a total of \$0.30 * 37.5 = \$11.3

With Scheme B you would earn 50 cents per point, for a total of \$0.50 * 37.5 = \$18.8

OK

Score	Up to 26 points	26.5 to 30 points	30.5 to 34 points	34.5+ points
Scheme A	30 cents/point	30 cents/point	30 cents/point	30 cents/point
Scheme B	20 cents/point	25 cents/point	35 cents/point	50 cents/point



Remaining Time [sec]: 27

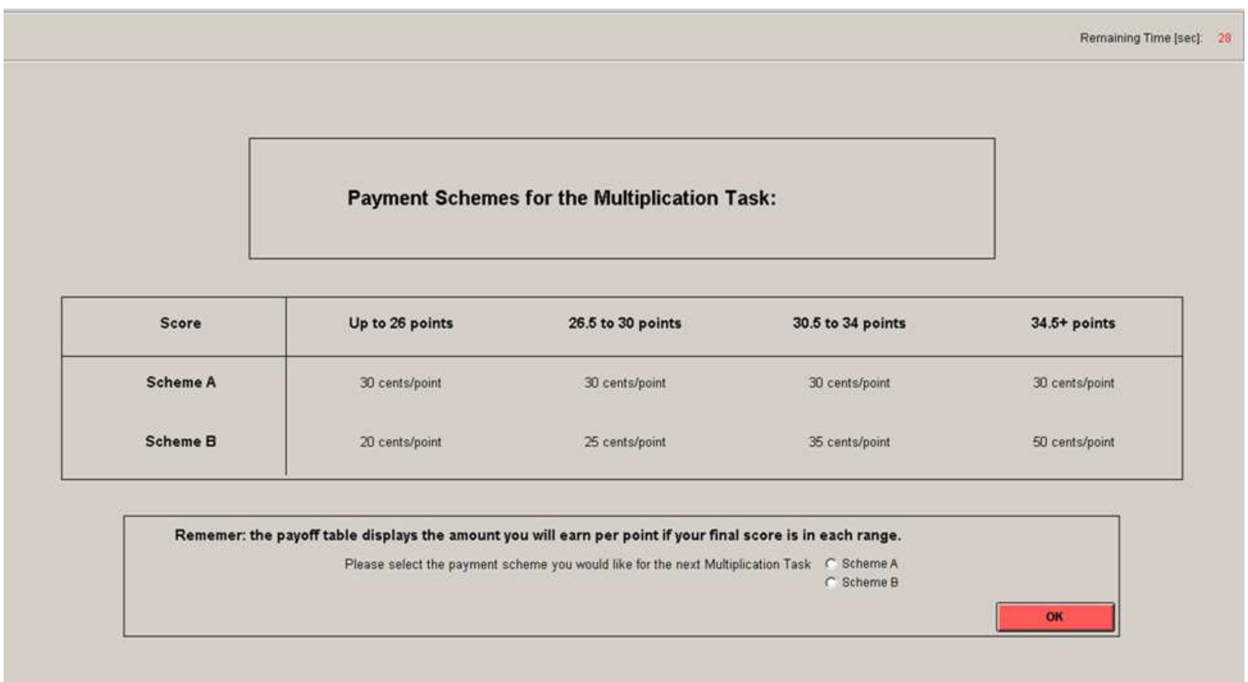
Multiplication Task #4

You will perform the Multiplication Task for two and a half minutes.

Your score will equal the number of correct answers you have minus half the number of incorrect answers you have.

Questions will be of Difficulty 1 to 5.

OK



Remaining Time [sec]: 28

Payment Schemes for the Multiplication Task:

Score	Up to 26 points	26.5 to 30 points	30.5 to 34 points	34.5+ points
Scheme A	30 cents/point	30 cents/point	30 cents/point	30 cents/point
Scheme B	20 cents/point	25 cents/point	35 cents/point	50 cents/point

Rememer: the payoff table displays the amount you will earn per point if your final score is in each range.

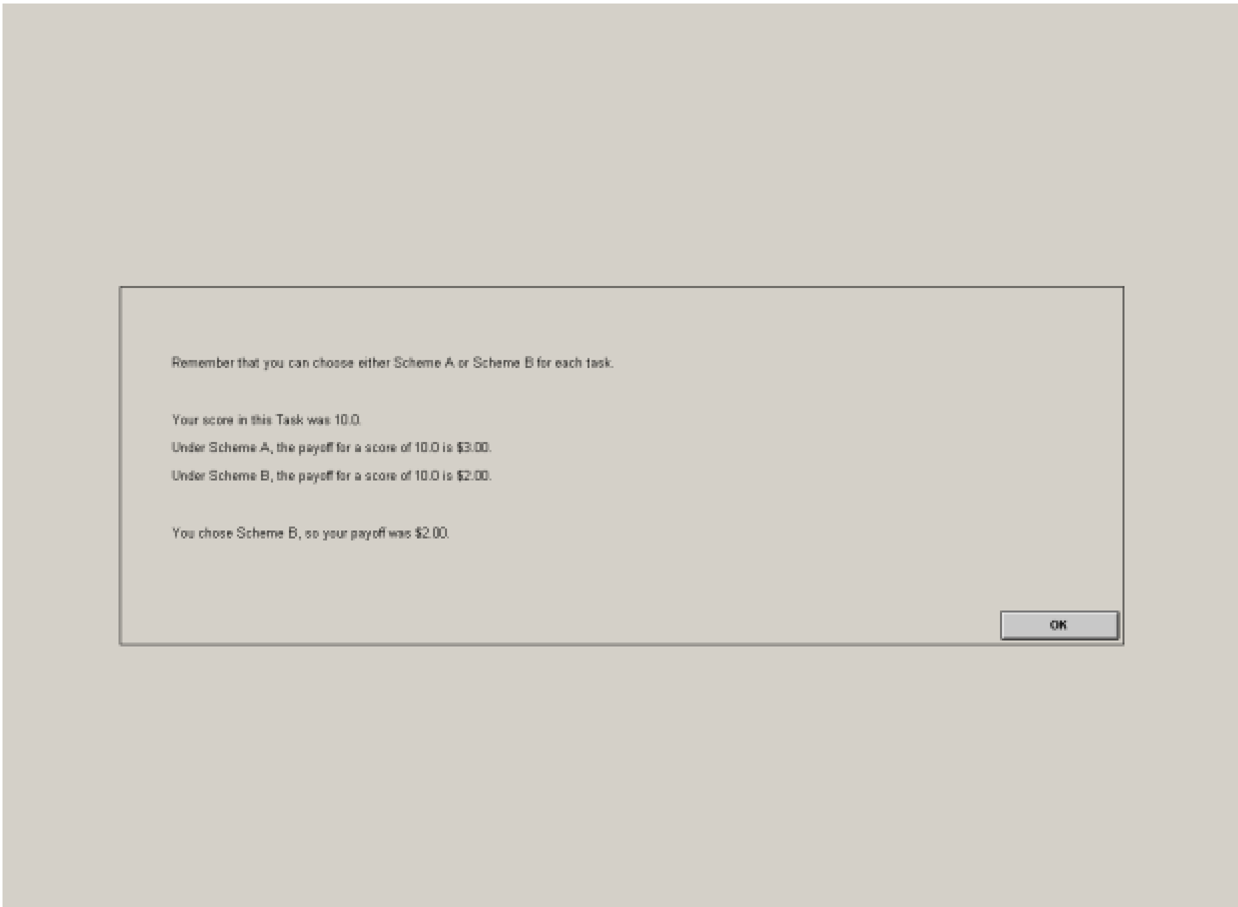
Please select the payment scheme you would like for the next Multiplication Task
- Scheme A
- Scheme B

OK

ROUNDS 7-9 Escalate Treatment:

Score	Up to 26 points	26.5 to 30 points	30.5 to 34 points	34.5+ points
Scheme A	30 cents/point	30 cents/point	30 cents/point	30 cents/point
Scheme B	20 cents/point	25 cents/point	35 cents/point	75 cents/point

FEEDBACK TREATMENT (shown after Round 6 results):

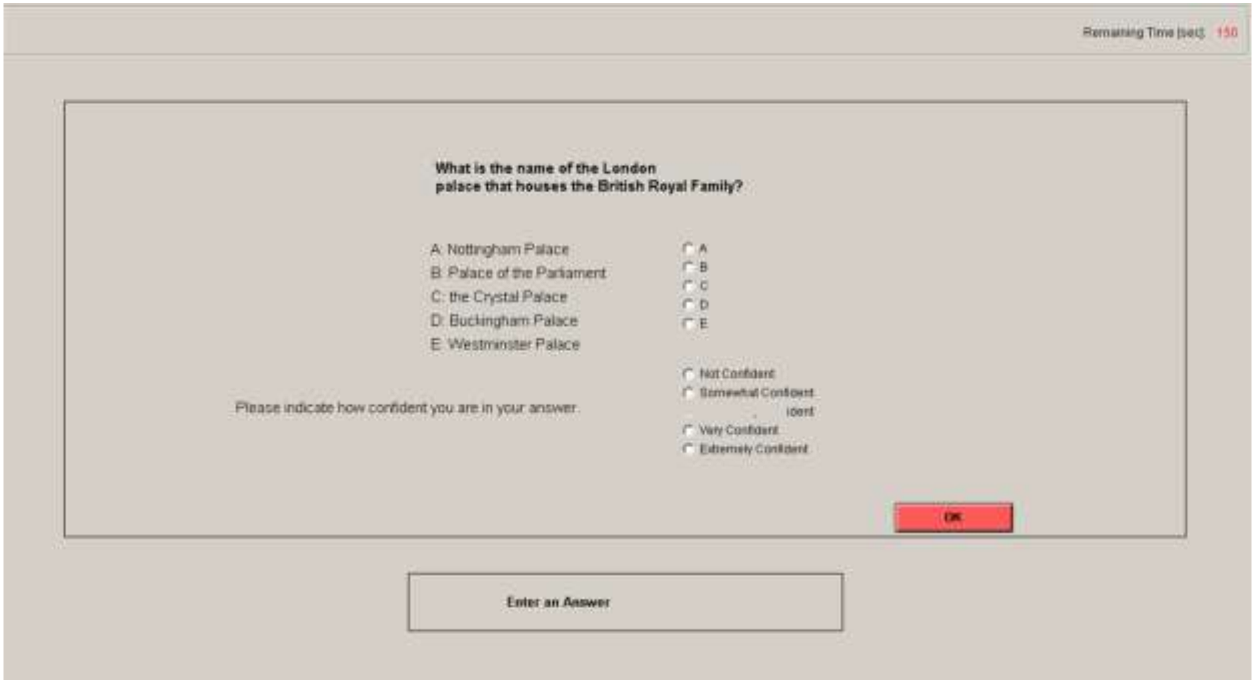


TRIVIA:


We're now going to do the Trivia Task. It will work the same as the Multiplication task, except the questions will be different.
The task is two and a half minutes long, and consists of 10 trivia questions.
Your score will be the number of correct answers you gave. There is no penalty for incorrect answers.
Your payoff for this task will equal 75 cents times your score.
Here is an example of the kind of problem you will be asked:
What is the name of the Roman Emperor who fiddled while Rome burned?
A: Augustine B: Nero C: Tiberius D: Caligula E: Claudius
OK


Remaining Time [sec]:
Please guess how many correct answers you will have (between 0 and 10).
You will earn an additional $0.50 if you guess correctly, or an additional $0.25 if your guess
is within one of the number of correct answers.
OK

(Sample Trivia Question):



RISK:

For each Situation, please select either the Fixed Payoff, or the Lottery.

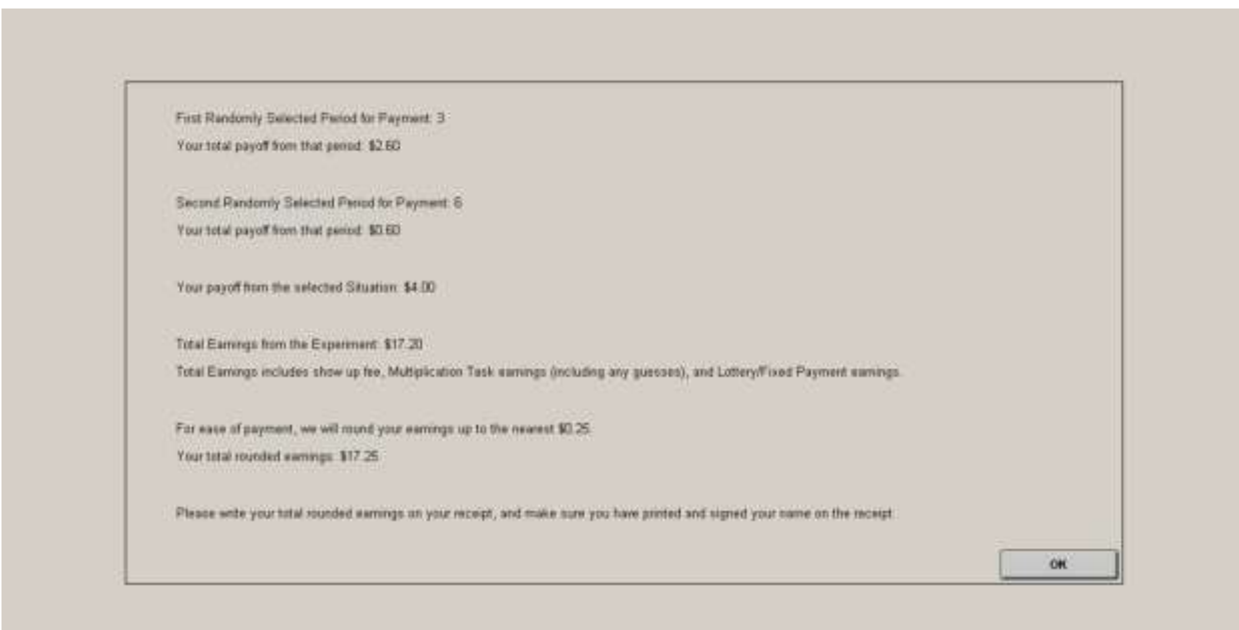
Situation	Lottery	Fixed Payoff	I choose
1	50% Chance $4.00 and 50% Chance $0.00	$0.25	Lottery ○ ○ Fixed Payoff
2	50% Chance $4.00 and 50% Chance $0.00	$0.50	Lottery ○ ○ Fixed Payoff
3	50% Chance $4.00 and 50% Chance $0.00	$0.75	Lottery ○ ○ Fixed Payoff
4	50% Chance $4.00 and 50% Chance $0.00	$1.00	Lottery ○ ○ Fixed Payoff
5	50% Chance $4.00 and 50% Chance $0.00	$1.25	Lottery ○ ○ Fixed Payoff
6	50% Chance $4.00 and 50% Chance $0.00	$1.50	Lottery ○ ○ Fixed Payoff
7	50% Chance $4.00 and 50% Chance $0.00	$1.75	Lottery ○ ○ Fixed Payoff
8	50% Chance $4.00 and 50% Chance $0.00	$2.00	Lottery ○ ○ Fixed Payoff
9	50% Chance $4.00 and 50% Chance $0.00	$2.25	Lottery ○ ○ Fixed Payoff
10	50% Chance $4.00 and 50% Chance $0.00	$2.50	Lottery ○ ○ Fixed Payoff
11	50% Chance $4.00 and 50% Chance $0.00	$2.75	Lottery ○ ○ Fixed Payoff
12	50% Chance $4.00 and 50% Chance $0.00	$3.00	Lottery ○ ○ Fixed Payoff
13	50% Chance $4.00 and 50% Chance $0.00	$3.25	Lottery ○ ○ Fixed Payoff
14	50% Chance $4.00 and 50% Chance $0.00	$3.50	Lottery ○ ○ Fixed Payoff
15	50% Chance $4.00 and 50% Chance $0.00	$3.75	Lottery ○ ○ Fixed Payoff

Select either the Fixed Payoff or the Lottery for each Situation before clicking OK.

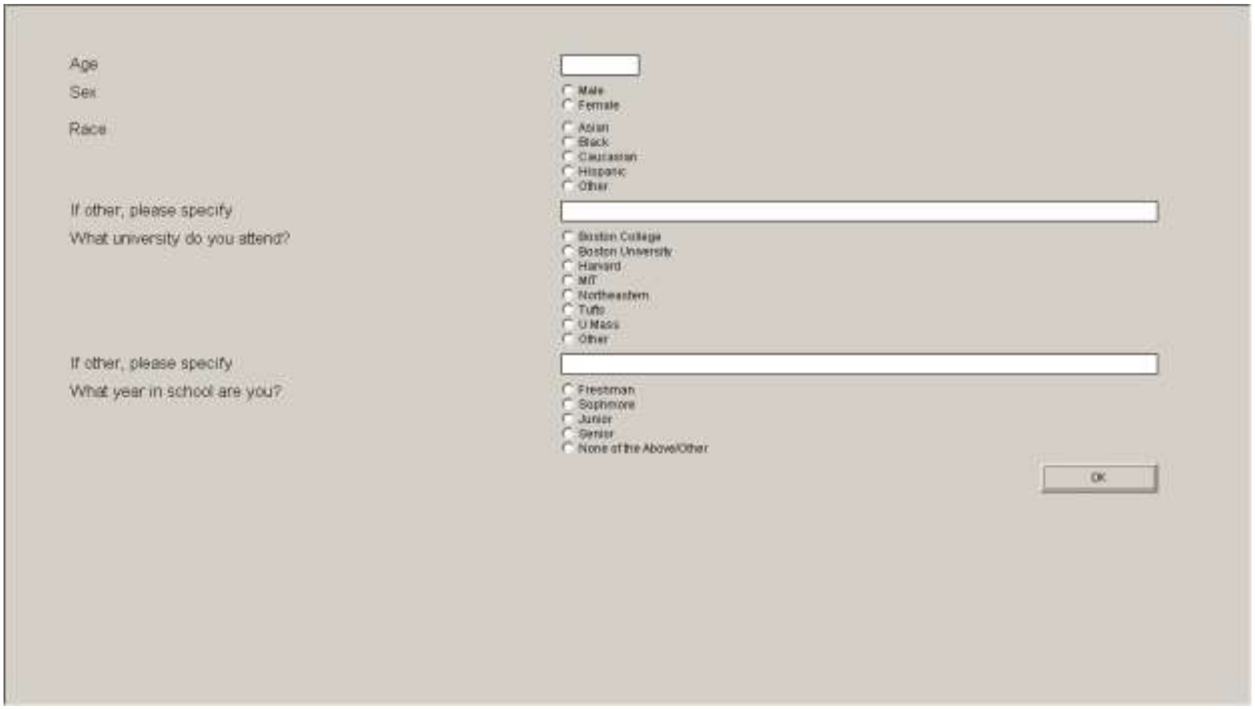
OK

(Sample Result):

TOTAL PAYOFF REPORT:



QUESTIONNAIRE:



Age

Sex
- Male
- Female

Race
- Asian
- Black
- Caucasian
- Hispanic
- Other

If other, please specify

What university do you attend?
- Boston College
- Boston University
- Harvard
- MIT
- Northeastern
- Tufts
- U Mass
- Other

If other, please specify

What year in school are you?
- Freshman
- Sophmore
- Junior
- Senior
- None of the Above/Other

OK

What kind of company or organization do you expect to work at after graduation?
- Traditional manufacturing (automotive, chemical, aerospace, etc.)
- Technology (hardware, software, technology services, etc.)
- Financial Services (banking, insurance, investment products, etc.)
- Other professional services (legal, consulting, accounting, real estate, etc.)
- Arts/entertainment/media/journalism
- Health care
- Education
- Government services (including military)
- None of the above
- Don't know

What kind of job do you expect to have after graduation?
- Engineer or computing/technical professional
- Management
- Sales
- Marketing
- White-collar (banking, consultant, lawyer)
- Doctor or other health care professional
- Professor or teacher
- Researcher
- Administrative/clerical
- Artist/entertainment/writer
- None of the above
- Don't know

What do you expect your salary to be in 5 years (to the nearest $5,000

What do you expect your salary to be in 20 years (to the nearest $5,000

Have you participated in a CLER experiment before?
- No
- Yes

OK

On the following screens, there are phrases describing an event or outcome. Please use the rating scale below to describe how how likely the event is to occur to you, compared to to other people you know of the same sex as you are, and roughly your same age? Please read each statement carefully, and then fill in the bubble that describes you best.

	Much Less Likely	Somewhat Less Likely	About As Likely	Somewhat More Likely	Much More Likely
Living past 90 years old					
Developing a drinking problem at some point in life					
Tripping and breaking a bone					
Getting a long-term job within 5 years that you like very much					
Not getting a serious cold or flu in the next 12 months					
Getting a job in the next 5 years that pays more than $75,000/year					
Having a mid-life crisis around age 40					
Getting laid off at some point in the future because the company downsizes or goes out of business					
Buying a home in the next five years that will double in value within ten years					
Getting cancer before the age of 60					

OK

	Much Less Likely	Somewhat Less Likely	About As Likely	Somewhat More Likely	Much More Likely
Being fired from a job at some point in your career					
Spontaneously going to a restaurant you know nothing about and finding it has fantastic food					
Getting spotted in a public place by a talent agent and asked to join a reality TV show					
Having your achievements written about in a major newspaper in the next 5 years					
Within the next five years, having a period of at least 12 months where you cannot find a job					
Getting hit by a car that runs a red light					
Having a very nice person move in near you in the next year					
Visiting more than 10 countries in the next 5 years					
Being selected by the IRS for a random audit of your tax return in the next 5 years					
Getting divorced (or splitting up with a life partner) after only a few years together					

OK

On the following screens, there are phrases describing people's behaviors and attitudes. Please use the rating scale below to describe how accurately each statement describes you. Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age. Please read each statement carefully, and then fill in the bubble that describes you best.

	Very Inaccurate	Moderately Inaccurate	Neither Inaccurate nor Accurate	Moderately Accurate	Very Accurate
I see myself as someone who is reserved and quiet					
I am a better driver than most people					
I am generally trusting					
I believe in Murphy's Law. If something can go wrong, it will					
I tend to be lazy					
Compared to others I know, I am worse at using computers					
I see myself as someone who is relaxed and handles stress well					
In uncertain times, I usually expect the best					
I see myself as someone who has few artistic interests					
I am better than most of my friends at card and board games					
I am outgoing and sociable					
I'm probably going to have more problems in my future life than most of my friends					
I tent to find fault with others					
I'm always optimistic about my future					
I see myself as someone who does a thorough job					
I am worse at standardized tests than most of my friends					
I tend to get nervous easily					
When I have problems, I am usually confident they will resolve themselves quickly					
I have an active imagination					
I am worse at spelling that most people I know					

OK

	Very Inaccurate	Moderately Inaccurate	Neither Inaccurate nor Accurate	Moderately Accurate	Very Accurate
I see myself as someone who is dependable and systematic					
I have better decision making skills than most people					
I see myself as someone who is complex and open to new experiences					
I can't rely on family an dfriends when things go wrong, and have to rely on myself					
I am sympathetic and warm					
Even when I'm sure about the right choice, I still ask others for advice before making an important decision					
I am disorganized and careless					
I think people who are not satisfied with their lives just need to work harder to be happy					
I see myself as someone who is conventional and uncreative					
When I'm feeling down, it's hard for me to focus on the good things in life					
I am bashful and shy					
I am better at very difficult tasks than most of my friends					
I am bold and energetic					
I shy away from hard problems because I am not good at them					
I tend to be rude and harsh					
I feel failure is a positive thing becaues I learn a lot from it					
I see myself as someone who is moody and temperamental					
When working on a team project, I want to do the most difficult task because I know I'll be able to do it well					
I tend not to be envious of others					
I find it is better not to get my hopes up, so I do not end up disappointed					

OK


Thank you for participating