Incentive Systems: Benefits of Sorting, Effort, and Some Potential Dangers

Marton Lukacs
Candidate for Economics M.A.
Adviser: John Earle

Department of Economics
Central European University

In partial fulfillment for the requirement of the degree of Masters of Arts…

Budapest, Hungary
2015

I, the undersigned [Marton Lukacs], candidate for the degree of Master of Economics at the Central European University Department of Economics, declare herewith that the present thesis is exclusively my own work, based on my research and only such external information as properly credited in notes and bibliography. I declare that no unidentified and illegitimate use was made of work of others, and no part the thesis infringes on any person's or institution's copyright. I also declare that no part the thesis has been submitted in this form to any other institution of higher education for an academic degree.

Budapest
June 4th, 2015

____
Signature

Incentive Systems: Benefits of Sorting, Effort, and Some Potential Dangers

By: Marton Lukacs
June 4th, 2015

Abstract:

Incentive schemes are often argued to possess both sorting and effort related benefits for firms and employees. These efficiency gains can theoretically be utilized in a Pareto improving manner. Conversely, theory also suggests there are potential dangers associated with incentive schemes, such as increased risk taking (deterioration of quality), and the ability of firms to annex employee surplus through information asymmetries. In this paper I test some of the basic theoretical concepts behind incentive schemes with the help of replicated and self-devised empirical analysis, and I show that if a number of important conditions are met, [convex] incentive schemes can be a vital tool for firms to increase their efficiencies in Pareto improving manner.

Table of Contents

List of Tables and Figures: 4

I. Introduction and Theory 5
- **I. Motivation** 5
- **II. Theoretical background** 6

II. Incentive Systems and Sorting 15
- **I. Experimental Setup** 15
- **II. Statistical Analysis** 17
- **III. Summary of the Results** 25

III. Incentive System and Effort 28
- **I. Introduction** 28
- **II. Regressions and Results** 31

IV. Conclusion, Tables, Bibliography, and Appendix 34
- **I. Conclusion** 34
- **II. Tables and Figures** 37
- **III. Bibliography** 42
- **IV. Appendix** 43

A special thanks to:
Stephen Leider and Ian Larkin for kindly allowing me to use their dataset for empirical analysis.
Kelsey Brooke and Raymond Guiteras for graciously allowing me access to their data, while their research is in progress.

List of Tables and Figures:

The following tables, figures, and charts are found in the Tables and Figures Section, or the Appendix section:

1. **Figure 2.2.1**—Average Overconfidence Score for Complete Sample
2. **Figure 2.2.2**—Average Overconfidence Score for Overconfident Sample
3. **Table 2.2.3**—Effect of Time (Period) on Overconfidence Score (OLS and RE)
4. **Table 2.2.4**—Effect of Period on Overconfidence Score for Later periods (RE)
5. **Table 2.2.5**—Change in Pay Scheme Preferences
6. **Figure 2.2.6**—Probability of Correct Incentive Choice (All Subjects)
7. **Figure 2.2.7**—Probability of Correct Incentive Choice (Overconfident)
8. **Table 2.2.8**—The Learning Curve
9. **Table 3.2.1**—Effect of Draw and Surplus on Quantity and Quality
10. **Table 3.2.2**—Fixed Effect of Incentive Surplus on Quality and Quantity
11. **Chart 1.2.1**—Dead Weight Loss From Information Asymmetries
12. **Table A.2.2.1**—Rerun of Larkin Leider Table 6, with added Pooled OLS Regression
13. **Table A.2.2.2**—Rerun of Larkin Leider Table 3
14. **Table A.2.2.3**—Rerun of Larkin Leider Table 5
15. **Figure A.2.2.4**—Frequency of Overconfidence Between Control and Treatment
16. **Figure A.2.2.5**—Average Number Performance for Groups
17. **Table A2.2.6**—Rerun of Larkin Leider Table 1
18. **Table A2.2.7**—Rerun of Larkin Leider Table 2

I. Introduction and Theory

I. Motivation

Incentive schemes take on a vital role in the modern day labor market. Despite the potential economic costs and dangers associated with them, incentive schemes can allow for numerous efficiency gains. From an employer's perspective incentives can increase the performance of employees. From an employee's perspective they can address a number of agency problems, such as eliminating the danger of being disadvantaged from free-riding colleagues. There are two main methods in which incentive systems can increase the aggregate efficiency of a firm and of the employee. One is through sorting—employees who agree to work within the frameworks of incentive schemes will generally be more productive, and as the data examined in this thesis suggests, often exhibit beneficial characteristics for firms (motivation, higher performance, and overconfidence); the consequences from sorting are especially beneficial for hard working employees. Another method is through effort—the incentivization of employees to allocate a larger portion of their time and effort towards productive behavior.

This thesis will focus on analyzing these two benefits of incentives schemes, along with some dangers. First, in this chapter a brief theoretical analysis is conducted to introduce the reader to basic economic concepts regarding these dangers and benefits. In the next chapter, using their data set, I review Ian Larkin and Stephen Leider's (2012) findings on positive sorting effects of incentive schemes, as well as conduct an empirical analysis of their data set by examining the potential for companies to take advantage through information asymmetries. In the third chapter, I use a data set provided by Kelsey Brooke and Raymond Guiteras (2014, *in progress*) to replicate a small portion of their findings regarding the encouragement and quality effects of incentive systems, as well as briefly study the key empirical question from an alternative perspective—the effect of incentive surplus on quality and quantity dimensions of labor.

II. Theoretical background

Allocation of Work and Leisure

Ideally, incentive schemes are designed to encourage employees to redistribute their allocation of work and leisure. In other words, incentives can encourage employees to work more hours, or to work with greater attentiveness. This allocation of time between leisure and work is natural, and is not a negative phenomenon. The study of this allocation, the methods of influencing it, and the consequences of doing so are especially important in labor economics. In classical labor economic theory there is an equilibrium point where the diminishing marginal utility from wages equals the increasing alternative costs of sacrificing leisure (working more hours) to obtain that wage. Today, because most employees work forty-hour workweeks, the allocation of work and leisure, as well as effort, often takes place within an eight-hour workday. This is why many firms have contracts that directly utilize performance based incentive schemes. These incentive systems have the potential to increase utility for both the employee and the employer.

Perhaps the best way to illustrate the potential efficiency gains from the effort effect is to take a concrete example. The following model is a simple supply model, but it can help illustrate how firms can take advantage of incentive schemes. Let us assume that there is a firm producing burritos from an unlimited and free resource of burrito ingredients. The price of each burrito is ten dollars[1]. The profit function of this firm would be:

[1] At $10 a burrito, these burritos better be top-notch quality.

$$\pi(b,w,h)=10b\text{-}wh$$

Where b is burritos, w is wage (incentive in our case), h is hours of labor. Let us also assume that an employee has the following utility function:

$$U(l,w)=\sqrt{wh}+\sqrt{l}$$

Where l is leisure. Hours here can be analogous to effort, and can be used to model how efficiently an employee is working. If I assume that an employee can work a maximum of 80 hours or 80 units of efficiency a week, we can easily derive that

$$U(w,h)=\sqrt{wh}+\sqrt{80\text{-}h}$$

Finally, let us assume a production function of

$$b(h)=h$$

Because of this, a firm's profit function can be simplified to:

$$\pi(b,w)=10b\text{-}wb$$

From a firm's perspective, it would set the wage as to maximize its profit. As the firm increases the wage, employees will work more, producing more burritos but the firm's costs per burrito will also increase. An employer must increase its wages until the marginal revenue from increasing its wages is equal to the marginal cost of doing so. From the utility function (derivation with respect to hours), we know that the hours worked as a function of wage is

$$b(w)=\frac{80w}{w+1}$$

By substitution, the firm's profit function with respect to wage can be simplified to

$$\pi(w)=10\frac{80w}{(w+1)}\text{-}\frac{80w^2}{(w+1)}$$

Which, if we take the first order condition, has a global maximum around \$2.32.

The first step to increasing efficiencies is to pay employees for what provides the firm with profit, not for the time worked; hence, theory suggests the firm should set an incentive contingent on burritos production, and not on hours worked. By doing so employees focus on encouraging profitable activity, and not the time worked.[2]

Setting the wage to $2.32 dollars per burrito created, however, is only efficient if the firm works within the framework of fixed wages, or a fixed incentive system per burrito. If the firm could utilize convex incentive schemes, with the help of perfect information, the firm could include a convex incentive system encouraging employees to produce more burritos. Because the marginal revenue from a burrito is constant at ten dollars, a firm could elect to provide the employee incrementally higher incentive for each burrito made, peaking at an incentive of ten dollars for the final burrito. Instead of about fifty-five burritos being created, seventy-three burritos would result; the additional production of burritos allows for increased total utility. In short, not only do properly devised incentive schemes encourage workers to focus on performance (number of burritos made over the time spent on a job), but convex schemes can potentially increase total utility too. If properly devised, because of the nature of diminishing utility of leisure and work, even basic convex incentive schemes can increase total utility in a Pareto efficient manner.[3] Clearly, perfect information is impossible to obtain for any firm, and thus designing the ideal incentive scheme to increase effort is impossible; however, through trial and error firms may approximate such an incentive scheme.

[2] In this model, there is no difference between the two incentive methods (paying for hours vs burritos) because of the fixed production function; however, in reality there generally is.

[3] As an example, if instead of paying a flat $2.32 per burrito, the above firm paid the employee $2 for the first 50 burritos, but $4 for the following burritos, 64 burritos would be made, the profit of the firm would increase from $422 to $484, while the utility of the employee would increase from just under 16.3 to about 16.5.

Incentive Schemes and Freeriding

In addition to effort, there are the obvious benefits of sorting from incentive schemes. Ideally, the agency-problem related efficiency gains are even greater than that of the effort effect. Employees and firms face numerous game theory related problems when working. Commonly in large firms a number of employees will take advantage of a lack of monitoring and generally fixed wages. This can be best illustrated with a simple model. If I assume a model with a two employee firm, and a production function of:

$$F(N)=1.5N_1+1.5N_2$$

a profit function for the employee will become

$$\pi\_X (L,N)= (F(N))/2+L\_x$$

where

$$L_x=1-N_x$$

Then the following prisoners' dilemma type setup would result for the two employees. The Nash equilibrium would be (Relax, Relax), which is not only inefficient, but also a non-Pareto efficient outcome.

Firm with two employees

Without Incentives	Work Hard	Relax
Work Hard	(1.5,1.5)	(0,1.75)
Relax	(1.75,0)	**(1,1)**

However, with the proper incentive system, the prisoners' dilemma can be resolved. Under a perfect monitoring assumption (a big assumption), N and L are observable, so it is possible to organize the wage structure in a way to punish the employee who is taking on leisure during the work, and reward the employee who is working.

$$U_X(L,N)= \frac{F(N)}{2}-L_x+L_x+L_y$$

or simplified:

$$U_X(L,N)= \frac{F(N)}{2}+L_y$$

Which would lead to a Nash equilibrium that is Pareto efficient.

Firm with two employees

With Incentives	Work Hard	Relax
Work Hard	**(1.5,1.5)**	(1.75,0)
Relax	(0,1.75)	(1,1)

Consequently, wages also become more "fair", a function of productivity per individual rather than that of the firm. As long as efficient monitoring costs are less than one, a Pareto improvement in the Nash equilibrium is possible through incentivization and monitoring in this model. Those who work more are awarded more, while those who work less are encouraged to stop freeriding. Though theoretically freeriding can be addressed through such incentivization, there are potentially major issues with incentive schemes—especially pertaining to the effect they might have on quality.

Fears Regarding Quality and Risk Taking

Unsurprisingly, one of the most commonly cited fears of incentive schemes is that they could negatively affect the quality of work (Veldman, Gaalman, 2013). Employees will rush with their work to complete as many tasks as possible, sacrificing preciseness and accuracy. Ideally monitoring, as a partial solution to the agency problem, exists within firms with incentive schemes. In a firm with monitoring, employees factor the probability of being monitored and resulting consequences of a decrease in quality into their utility function. Modern day workdays should posses enough flexibility (coffee time, break time, lack of concentration of work) for many employees to increase quantity without decreasing quality. Solutions for agency problems vary from one firm to another. In the banking environment, as a cheap monitoring mechanism sales employees are generally expected not only to acquire new customers, but also to maintain the new customer for a minimum number of months.

In higher-level and more complex positions, risk taking becomes an analogous problem to deterioration of quality. With increasingly steep and more convex incentive schemes there is a possibility that employees will act irresponsibly; some, such as Rui Figueiredo (2013) in his paper *Bed Bets, Excessive Risk Taking, Convex Incentives, and Performance* cite the risk taking stemming from incentive schemes as a contributing cause to the Great Recession. He argues that because of an effort based drive to sell more houses, shares, investments, investors and banks were driven to sell at a lower price than what the market could handle. In an oversimplification, the quantity of the sells grew, but quality of the sells dived. [4] Figueiredo argues that in managerial positions (and other positions similarly complex), incentive schemes could easily have a net negative effect for a firms due to an increase in risk taking.

> In fact, given the baseline assumptions of our empirical model the data suggest that 82% of the performance drop observed by managers, who are not very far below their threshold, is due to the performance costs of risk-taking and only 18% of the performance decline is due to effort reduction. (Figueiredo, 2013, 5)

Whatever the solution to addressing these moral hazard and agency problems, there are varying costs to doing so from one firm to the next and from one position to another. Undoubtedly in some instances, especially in complex positions, the introduction of incentive schemes can be inefficient. In many cases, however, moral hazards and agency problems result from improperly designed incentive systems. The incentivized unit should not just be a product, but rather the quality of a product. For example, in the case of sales-workers, acquisition *and* retention; in the case of waiters number of tables attended *and* the percentage tip received, etc. Admittedly devising incentive schemes properly is especially difficult in a number of professions; however, especially with simpler tasks proper incentivization with cheap monitoring should be achievable.

The data that has been provided to me by Professor Kelsey Brooke of Tufts University and Raymond Guiteras of University of Maryland allows for an empirical analysis in one real-world instance. Among others, Brooke and Guiteras (2014, *in progress*) study monitoring effects, and show empirically that monitoring can increase quality at a cost to quantity. [5] The empirical analysis included in the third chapter of this thesis utilizes their data and shows that higher

[4] Figueiredo page 1-5
[5] Brooke Guiteras page 21

incentives, at least at low-level occupations, offer an increase in quantity without a large or significant sacrifice in quality. Furthermore, with the help of a fixed effect model I show that at least in their data set monitoring is not necessary for a productivity level increase without quality sacrifice. Admittedly, bean sorting, which is the task they studied, is an ideal job for incentivization, as it possesses easy to measure quantity and quality characteristics. Nonetheless the results are surprising.

Besides the potential effect on quality and the moral hazard, there are other legitimate fears with respect to incentive schemes. For example, Uri Gneezyin in the Journal of Economic Perspectives claims that monetary incentives can block out other motivational effects; there is evidence to support this claim too (Gneezyin, 2011). Similarly, Ola KvalIoy and Anja Schöttner argue that monetary incentives can have block-out effects against non-monetary incentives. (Kvalloy, Schöttner, 2014). Another common fear is incentive schemes leading to psychological and health related damage. These are common general fears; however, there are more specific fears regarding incentivization as well. Ian Larkin and Stephen Leider introduce a highly specific but real danger. They argue that with the help of selection firms may have the opportunity to underpay employees by utilizing the differences in expected income and actual income of employees—a form of manipulation through expectations. In fact, in their experiment overconfident individuals lost 15% of their potential wage because of higher expectations. [6]Though this situation is far from universal, the presence of complex incentive systems today suggests that companies are often fine-tuning their incentives in an attempt to try to find the most cost-efficient solution to increase productivity. It is my experience working in the financial

[6] Larkin Leider page 210

environment that incentive system testing is a practiced phenomenon; firms, unaware of the reasoning, might find certain incentive systems more efficient, when in actuality they are only increasing their share of the aggregate consumer/producer surplus.

According to the empirical analysis of the data provided by Larkin and Leider, overconfident employees are drawn to incentive schemes that offer them larger rewards if they perform at higher levels (convex incentive schemes). This makes sense, as the expected income from such an employee's perspective will be higher than their realized income. Overconfident employees can thus potentially be paid under the market equilibrium. Through such manipulation not only is there a dead weight loss[7], but firms also gain some of the surplus that would have originally been allocated by the market to the employee. One major question regarding this manipulation is whether employees learn to adapt to these incentive schemes, or whether firms can take advantage of employees in the long run. With the help of the experimental data collected by Larkin and Leider, I empirically test and analyze this question in the next chapter.

Answering any universal questions about incentive schemes is beyond the scope of this thesis. My personal empirical contributions are focused on the testing of the ability of firms to manipulate employees through information asymmetries, as well as testing the effect of relative incentives on quality and quantity with and without monitoring.[8] I hope that my writing will help empirically and theoretically introduce two potential benefits of incentive schemes—sorting and effort, and that my writing will show that if a few of the introduced conditions are met, incentive schemes possess great potential for firms to achieve Pareto improvement in productivity.

[7] See chart 1.2.1 at the end of Tables and Figures section for DWL
[8] Guiteras and Brooke focus their research and work on nominal incentive effects

II. Incentive Systems and Sorting

I. Experimental Setup

Ian Larkin and Stephen Leider's data set utilizes the scientific method for experimental data collection with a control and treatment groups. This makes it possible to conduct rather conclusive analysis with basic treatment to control group comparisons, pooled OLS regressions, and panel based random effect regressions. In both, random effect and pooled OLS regressions it is relatively easy to compare the results of the control and the treatment group with the help of dummy variables; I also utilize two-dimensional graphs to examine a variety of trends between the control and treatment groups over the examined periods. Because of the experimental nature of the dataset, these types of comparisons, especially when done in unison, takes care of most endogeneities. The authors examine an increasingly important question: the sorting effects of incentive systems. For most firms having the ability to sort overconfident and high-performing employees from underconfident and underperforming employees is crucial in finding a competitive edge.

Larkin and Leider's experiment examined 255 subjects in multiple waves by providing trivia and mathematical tasks for each to complete in a set amount of time (150 seconds), and then collecting demographic information from them. 16% of the subjects were randomly assigned (control group) while 84% of the subjects were given a choice (treatment group) between a linear and convex incentive system individually personalized to their performances (a third incentive system was also introduced for the escalated group, but this is beyond the scope of this thesis). Subjects were asked to predict their multiplication scores before each round. Based on their actual results, their predictions in the first four periods, and their responses to survey questions, subjects were assigned as having either an overconfident personality, or an underconfident personality, and were given two personalized incentive systems; various analyses were conducted on the differences between the two types of personality. The effectiveness of the two incentive systems on subjects was also compared. In general, relatively simple statistical regressions were conducted, and appropriate distributions (chi, t) were used to calculate significance. Larkin and Leider utilized the Kolmogorov-Smirnov test for continuous one dimensional probability statistics, and the Mann-Whitney test for hypothesis testing between populations.[9] Their regressions were predominantly utilizing a random effect model; however, I reconfirmed their analyses with a simple pooled OLS model for increased robustness.

The following sections include a brief analysis Larkin and Leider's results. Subsequently I argue that one of the main dangers Larkin and Leider introduce, a potential for firms to underpay employees through incentive systems, is not studied extensively enough. With the help of their data, I show that Larkin and Leider overlooked the presence of a learning curve in their paper. The potential for a learning curve could negate the long-term negative effects associated with the information asymmetries between a company and an employee. In short, companies might not be able to "take advantage" of their employees through incentives in the long run.

[9] Larkin Leider page 205

II. Statistical Analysis

Convexity and Sorting

Before conducting my own analysis on their data, I attempted to rerun the regressions found in Larkin and Leider's paper[10] and concluded that Larkin and Leider's core regressions are correctly conducted, well coded, and their tables show the appropriate results.[11] Most of their regressions are conducted utilizing a random effect model with robust standard errors. Because of the nature of the data, I decided conducting simple pooled OLS regressions for added robustness to their results was appropriate. Tables containing replicated results of Larkin and Leider's regressions are found in the Appendix section (IV.IV), while tables with my regressions and results are found in the Tables and Figures section (IV.II).

One of Larkin and Leider's most important findings can be found in the Appendix on Table A.2.2.1[12]. The regression suggests that individuals who choose linear schemes achieve significantly lower results than those who do not. The random effect regressions in table A.2.2.1, which are a replication of Larkin and Leider's results, show that subjects choosing convex curves perform about two more multiplication tasks (~10%) than those choosing linear incentive curves, while the controlled OLS regressions (which were conducted by me for robustness) show similar differences (~15%) between the convex and linear curves. Larkin and Leider argue the differences are due partly to incentivization; however, I have major reservations about the conclusions the authors arrive at from their findings.

[10] With the help of consultation through email with the authors

[11] There are a few regressions where I was not able to replicate their exact results; however, the output tables were always consistent with their conclusions and findings.

[12] Larkin Leider Table 6

Leider and Larkin claim in their paper:

> 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. (Larkin, Leider, 207)

The nature of the experiment suggests that the differences are more likely due to selection bias than an incentive based effort effect. The tasks and the environment that was provided to the subjects were simply not of the nature where incentivization could make a significant difference. With very little time provided (150 seconds), and a relatively high level of incentive offered, all subjects were likely working quickly to answer as many questions as possible. I hypothesize that statistically significantly less efficient labor force chooses the linear incentive scheme, which is a form of sorting; this hypothesis is supported by the fact that within the control group the differences between those assigned to the convex scheme and those assigned to the linear scheme disappear (see table A.2.2.1). The hypothesize that the differences in results are due solely to sorting, is not only backed by Larkin and Leider's regressions, my regressions, but also by basic control-treatment group analysis.

Overconfidence and Convexity

Table A.2.2.2 and A.2.2.3 (Appendix), which summarize regressions analyzing the correlation between overconfidence and pay scheme, are similarly crucial tables for this thesis.[13] Those choosing convex schemes are not only more efficient, but are significantly more likely to be highly overconfident. Larkin and Leider find that those in the top 10% of overconfidence are 44% more likely to choose the convex incentive schemes than those who are not overconfident.

[13] Rerun of Larkin, Leider Table 3 and 5

Higher practice scores also strongly correlate with higher magnitude of overconfidence, or in other words, those who perform above average are more likely to be overconfident. This suggests that overconfident individuals are generally either more motivated or more talented (in this case, because of the nature of the experiment, more talented is the likely answer).

Further regressions suggest that those who are initially overconfident were significantly more likely to mistakenly choose the convex scheme. This random effect regression is controlled for periods, suggesting that this relationship is maintained throughout the experiment. When interaction terms are included controlling for incentive pay type (escalate vs. not) and overconfidence (A.2.2.3)[14], the magnitude of the correlation for incorrectly choosing the convex scheme increases even further, proving empirically an important hypothesis: there is a strong and significant relationship between those who are overconfident and those who choose convex pay schemes. For example, if someone was overconfident of his/her multiplication score by 5 (predicted answering 25, answered only 20), they were about 10% more likely to choose the convex pay scheme than someone who guessed their score correctly. The difference between actual outcome and predicted outcome is what allows for information asymmetries between the employees and employers. Furthermore, Larkin and Leider shows which personality and employment related traits correlate with overconfidence. Unsurprisingly, those in management, as well as those who are extraverted are generally (and significantly) more confident in their abilities than average. Larkin and Leider found that males are 58% more confident than Females

[14] Larkin, Leider Table 5

in predicting their results. Subsequently males are more likely to choose convex schemes. [15]

Based on these results convex schemes seem to be a form of sorting that companies can utilize if they wish to hire a larger proportion of eager and overconfident employees. This makes theoretical sense, as the expected wage of less-motivated and less-confident employees is significantly lower with these types of incentive schemes. Though the result is not the most groundbreaking, it is likely the most important finding of Larkin and Leider's data set from an economic perspective: their data provide an empirical and robust proof from an experimental setup that incentive schemes have the capability of filtering out and sorting certain types of employees, allowing companies to hire a higher proportion of efficient employees, which increases utility both for the firm and for efficient employees. By rerunning the regressions with their parameterizations and through controlled OLS model, and by testing the sorting hypothesis in a treatment-control comparison, I have done nothing more than strengthened the robustness of their findings, and interpreted some of their results.

Learning Curves and the Dangers of Information Asymmetries

Sorting by incentive schemes can increase efficiencies in the aggregate, but as mentioned in the theory section, the differences in an employee's expected incentives from the actual gained incentives can theoretically allow for information asymmetries between the firm and the employee. These differences can allow employers to gain some of the surplus that would normally be allocated to employees. Larkin and Leider's data provides insight on this potential

[15] Most of the tables and regressions in Larkin and Leider's paper were the same as my outputs; however, when recalculating Figure A.2.2.4 with their data, I surprisingly arrived at slightly different results. The conclusions pertaining to the figure are not altered whatsoever, but I did feel it important to include this figure into this replication, because of the slight differences in the choice group.

occurrence. Pertaining to expectation learning curves[16], Larkin and Leider do not acknowledge an existence of a statistically substantive improvement in guessing one's performance. On page 204 of their article, they discuss a significant, but small and [unimportant], learning curve through their regressions. This is because in their article the authors tend to look at overconfidence as a dummy variable. Instead of measuring how overconfident a participant is, Larkin and Leider measure whether subjects are overconfident in a binary sense (admittedly with a number of thresholds for overconfidence). Larkin and Leider argue, "After an initial learning period there is a stable and persistent level of overconfidence throughout the experiment" (Larkin, Leider, 197). Furthermore, they claim "Across all treatments, 38.5 percent of the subjects in the first treatment round (period 4) were overconfident, while 38.0 percent in the last experimental round (period 9) were overconfident" (Larkin, Leider, 197). Albeit the percentage of those overconfident does not change, the level of their overconfidence changes significantly.

Like in their findings, Figure A.2.2.5 shows that the performance for each tested group is monotonically, but asymptotically improving.[17,18] Instead of looking at overconfidence from their binary perspective, I attempted to examine the magnitude of overconfidence in Figures 2.2.1 and 2.2.2. I created a variable called the overconfidence score, which is defined by the ratio of the *Difference in Belief* to the *Number Correct* variable. Figures 2.2.1 and 2.2.2 show the overconfidence score on the (y-axis) within the whole group (Figure 2.2.1), and of those who are characterized as overconfident (Figure 2.2.2) plotted for each period. The raw data suggest that

[16] Ability of employees to improve their expectations accuracy
[17] Similar to Figure 1 of Larkin and Leider's chart; test groups are organized differently in this graph, but the results match their findings;
[18] No choice and feedback group exhibit monotonic increasing qualities, no escalate and escalate do not; however, the total group does

the power associated with employer-employee information asymmetries may not be sustainable over time, and there is a substantial learning curve present.

There is a clear asymptotic approach of accuracy in prediction for the whole group over time. Although admittedly the derivative of the asymptotic approach to perfect prediction decreases in magnitude in later periods, the groups still show a definite improvement. For example, among the Overconfident, the group utilizing the *escalate incentive system* in Leider and Larkin's observations decreases in the overconfidence index from 0.35 in the first period to a little over 0.16 in the fourth period, while further significantly decreasing to an all-period-low of 0.09 by the 9th period. Such raw results incentivized me to conduct a deeper analysis regarding learning curves.

Because the data is provided in a controlled experiment, a simple pooled OLS regression with few controls is sufficient in gauging whether there is a significant learning curve present; however, for added robustness it would be ideal to utilize a random effect model as well. I conducted a regression utilizing the overconfidence score to analyze the persistence of learning curves. In Table 2.2.3 I show that there is in fact a highly significant learning curve present for both the whole population of the experiment, and for those who are overconfident. Both the OLS and random effect models show strikingly similar results. The complete sample approaches an *overconfidence score* of 0 from the negative direction (underconfident) at a rate of about 0.1 per period, while the overconfident individuals approach an overconfidence index of 0 from the positive direction (overconfident) at a rate of about -0.2 per period. These results are persistent and highly significant whether or not they are controlled for race, gender, and age. F-tests were conducted to include a variety of available controls. Regressions with more controls were also conducted and showed similar results.

Since the authors hypothesize that learning exists only initially, in order to ensure that the relationship exists beyond the first three periods, I ran the similar regressions (random effect) for periods four through nine in table 2.2.4. Both regressions suggest although the correlation and the strength of the learning curve weaken, there is a statistically significant relationship present throughout all nine periods. In fact, while the statistical strength learning curve subsides for those who are not overconfident, it continues for the overconfident group, albeit at a significantly slower rate.

On way to assess the ability of corporations to underpay employees with the help of incentive expectations is by looking at whether subjects who were overconfident, and became less overconfident, became less likely to select the convex pay scheme. I created a *learned* variable, which I coded as one if a subject has learned between period four and period nine, and as zero if a subject has not. I only looked at those 100 subjects who had a choice in their incentive schemes and who were overconfident in period four. Interestingly, as table 2.2.5 shows, whether I controlled for gender and race or not, overconfident subjects who were more accurate in period nine than four did not show a higher tendency to choose the linear, or safer schemes over time (both random effect and OLS regressions show similar results). That is, in Larkin's experiment even though overconfident subjects are more and more accurate in predicting their scores, they are not any less likely to choose convex schemes.

Admittedly, the second portion of the table 2.2.5 show that those who are overconfident at period four (first period with a choice) are 13% less likely to learn than those who are not overconfident. Nonetheless as previous regressions and analyses prove, both the underconfident and the overconfident group show improved results in guessing their scores over periods. Even the right side of 2.2.5 shows that among the overconfident, a majority of the tested population still does improve in predicting their performance. These results beg the question: why do participants not switch to the linear and safer schemes?

There is a possibility that individuals learn and perform better from one period to the next in their multiplication tasks. Because of this, measuring the rate of selection of convex schemes in overconfident individuals is insufficient in testing whether subjects are making the choice with a higher payoff or not. Perhaps the improvement of the performance of subjects justifies the choice of the convex scheme. For this reason, I created another variable, *correct choice*, which measures the probability of correct incentive choice. The behavior of the *correct choice* variable within different groups can be seen in Figure 2.2.6 and 2.2.7.[19]

The raw data suggests there is a substantial increase in the percentage of subjects who choose the correct incentive scheme. Thus, an observed learning curve found in the treatment group can be used to conclude some form of adaptation. For this reason, as before, I conducted pooled OLS regressions and random effect regressions utilizing my new dependent *correct choice* variable

[19] After checking my data and calculation multiple times, Figure 2.4.5 continues to suggest highly suspicious results for the no choice control group. It is especially bothersome that the overconfident control group (no choice) exhibits a steep learning curve in choosing the correct incentive scheme. The results of data do seem to question the legitimacy of random assignment within this group. Nonetheless, I cannot do anything about the peculiarity, and my results do not have to rely on the control group, so I am forced to ignore the questionable control group.

(Table 2.2.8). The regressions suggests that, controlled for main demographic variables, each period leads to a significant 2.5-3.4 percentage point increase in probability that a subject will choose the correct incentive scheme. This RE regression suggests that overconfident subjects actually learn at a quicker pace (3.4 percentage points). There is no indication that this learning curve subsides in later periods, as learning seems to maintain a large magnitude.

III. Summary of the Results

Larkin and Leider's data allowed me to come to similar conclusions pertaining to the sorting effects of incentive systems as they have. Convex incentive schemes are clearly capable of achieving positive sorting for firms. Those who choose convex schemes achieved statistically higher results in multiplication, and are both motivated and overconfident, which are beneficial qualities for firms allowing many agency problems to be addressed. Based on the regressions and the analysis, I argued that contrary to what Larkin and Leider claim in their analysis, the differences between results are due solely to sorting through incentive schemes rather than attributing a part of the difference to effort based effects.

In general my OLS regressions and random effect regressions are sufficient in this environment due to the nature of a controlled experimental environment. In many instances I compared the results of the control group to the treatment group and arrived at similar conclusions as through the regressions[20]. These three methods were utilized in unison in most case for added robustness of the analysis.

Although selection bias for overconfident individuals to incorrectly pick a lower paying incentive scheme is clearly present in the data, the level of the overconfidence (measured by the overconfidence score) diminishes over periods, while the level of performance does not. This is emphasized both graphically in a two dimensional sense, and through both pooled OLS and random effect regressions. The expectation learning curve, as well as the proportion of those making the *correct choice* is present and significant for both the early (1-3) and the late stages (4-9) of the experiment.[21] The only reason linear schemes do not become much more popular in later periods, is because individuals improve on their personal results. Furthermore, in real life there are significantly more than nine periods to learn about incentive schemes, and the tasks are often considerably longer, meaning there is more time for employees to adapt and learn.

The findings found in figure 2.2.8 suggest that corporations can both succeed in positive sorting, as well as underpay employees at an early stage with the help of convex incentive schemes; however, the capability to underpay overconfident employees is likely only possible in the short run. The regressions suggests that subjects learn from one period to the next period about both their expected results and their expected payout; consequently, both, general subjects as well as overconfident subjects, show an increasing tendency to choose the correct incentive system as time progresses. This could undermine corporations' ability to extract some of the employee

[20] Exception: Control group of overconfident subjects misbehaved in learning curve (control group should exhibit no learning curve), and thus I was forced to rely purely on regression analysis.
[21] Correct choice cannot be interpreted for earlier periods, only the learned coefficient can.

surplus from overconfident employees. On the other hand, the more important (and beneficial) ability to preselect overconfident individuals through the offer of a convex incentive scheme seems to be backed both by theory and data studied.

This situation and these incentive schemes are highly specific. It would be unfair to extrapolate these results to the entire labor market without further research. There also needs to be further studies regarding learning curves conducted, as these multiplication problems are not real work scenarios in complexity, nor in time. Furthermore, I cannot say whether switching the incentive schemes provided to employees on a regular basis would diminish the shown learning curve. Nonetheless, the results show that firms do have the ability to use incentive schemes for sorting, and that employees exhibit the ability to adapt to manipulations through incentives.

III. Incentive System and Effort

I. Introduction

The empirical data and theory in the previous chapter suggests that incentive schemes have the potential for sorting. Beyond the advantages of sorting, incentives allow for firms to potentially take advantage of information asymmetries. The previous chapter has shown that individuals in the examined scenario also possess the ability to learn from their mistakes, and adapt to these information asymmetries, albeit at a slow pace. So far the benefits of convex incentive systems seem to be empirically backed, while the dangers are not. A much more frequent and universal economic fear regarding incentive schemes, however, is with respect to quality sacrifice and moral hazard. Unfortunately, the Larkin-Leider data set utilized in the previous chapter does not have relevant data to examine this issue.

Thankfully, Kelsey Brooke of Tufts University and Professor Raymond Guiteras of University of Maryland graciously provided the second dataset utilized in this thesis. Their data was collected in Rural Malawi, and unlike the data from the Larkin-Leider dataset, their observations were of real-life tasks in a full workday—sorting beans by type and quality. These experiments were conducted over many months, in the high and low agriculture season.[22] Subjects were asked to provide a minimum incentive rate at which they would work at, and then drew an incentive rate at random. If the drawn rate was higher than their minimum wage rate, then they received the contract. There were four sessions per employee, which is sufficient to allow for fixed effect

[22] Brooke Guiteras page 14-15

regressions. The race of the participants was mostly the same, as this part of Malawi is rather homogenous. The gender and the time of year were collected for each subject. This data set is ideal for measuring whether quality is sacrificed when incentives increase. Unlike in the Larkin and Leider experimental dataset, workers are working for full workdays, thus they have the ability to allocate their leisure and labor hours throughout the day; the scenario is much more similar to real life working situations. This task is relatively simple, and does not possess many of the risk taking dangers addressed by Figueiredo. Furthermore, in contrast to the Larkin Leider experimental dataset, the data is collected utilizing real life task conducted by workers on a regular basis. Lastly, bean sorting can easily be quantified (quantity), while mistakes can easily be identified (quality).

Kelsey Brooke and Raymond Guiteras are currently working on an in-depth research on the quality versus quantity issue, and have an advanced working paper. Beyond their extensive theoretical analysis, their empirical research thus far utilizes random and fixed effect regressions, which among others, look at the relationship between awarded incentives (nominal) and productivity (nominal) as well as the effects of monitoring.

I want to study the effect of higher than market incentive rates on quality and quantity of production. For this question fixed effect are ideal regressions that allow me to control for interpersonal endogeneities. By looking only at the effect of additional incentive for an individual over different periods, I can study the effects without worrying about a number of endogeneities, such as the filtering effects of lower incentives. Furthermore, there are 1338 individuals studied for 4 periods (689 unique individuals, some studied multiple times), which is a large enough sample. Though fixed effect is an appropriate model to use in this instance, I have conducted controlled pooled OLS regressions for added robustness and have found similar results; however, I must note that with pooled OLS individuals with lower incentive reservation prices will be weighted more heavily.[23]

My independent variable studied is incentive surplus. The surplus of an employee is measured by the difference between their reservation incentive rate, and the drawn incentive rate. The higher the surplus, the higher the relative incentive is for individual workers. Individuals with higher reservation prices will inherently act differently than those with lower reservation prices; among other sources of endogeneity, those with higher reservation prices likely have higher opportunity costs, suggesting that on average they might be more skilled or more efficient. From the fixed effect approach, admittedly, a regression looking at incentive surplus and just at nominal incentive should result in similar conclusions; however, I believe that looking at relative incentive allows me to address an important question from an employee's perspective. The question is not merely how much a nominal increase in incentives leads to quality-quantity changes, but rather how much an increase from a reservation incentive wage changes an employee's behavior. In short, I address the question of whether it is worth paying over the market incentive wage for a firm to achieve higher productivity and profit. Furthermore, I find striking results pertaining to monitoring at higher incentive surplus levels.

[23] Brooke Guiteras page 10

II. Regressions and Results

Initially, I conducted an OLS regression of quantity and quality on nominal incentive controlled for gender, (race is mostly the same), reservation price, and week (I included the peak labor season dummy labor in a subsequent regression, but obtained almost exactly the same coefficients and standard errors). My pooled OLS regression, showcased in Table 3.2.1, shows results similar to that conducted by Kelsey Brooke and Raymond Guiteras, and suggests that an increase in the drawn incentive rate by 5 is correlated with one more scoop of beans being sorted (draw rates ranged between 5 and 25, average scoops in workday around 7.4). The number of errors does correlate with higher draws though not as significantly. Unfortunately, with pooled OLS as the draw rate increases, the aggregate effect on productivity may be biased, as many with higher reservation prices enter the market. For them the nominally higher incentive, relatively to their reservation incentive provides no motivation. Furthermore there is a skew in the sample size for different draw rates, as there is greater weight on those with lower draw rates (there are more data points for them). [24]

Consequently, I also conducted a pooled OLS regression on incentive surplus, seen in table 3.2.1. Pooled OLS looking at incentive surplus mitigates the muting effect of those with higher incentive reservations entering the market at higher draw rates. Importantly, the interpretation of the results is different too. Instead of looking at nominal incentive rate (an approach used by Guiteras and Brooke), I study the relative incentive rate for each individual. The results suggest that there is a strong relationship between higher relative incentives and higher results: an

[24] Important to recognize that errors are measured as errors per scoop, while quantity is measured in scoops of beans

increase of incentive surplus by one leads to 0.3 more scoops of beans being sorted during the day. In this regression, there is a borderline statistically significant drop in quality of bean sorting (about 0.3-0.4 more errors per scoop) too.

Econometrically, a better approach to analyze how quantity and quality are affected by higher than reservation incentives is to utilize a fixed effect regression that examines the relationship between incentive surplus and quantity/quality. By the nature of fixed effects I can control for the endogeneities that might be present between individuals with different reservation prices. Because I am looking at the same individual over different sessions and different reservation prices, the only additional variable I need to consider controlling for is weeks and peak labor.

With this fixed effect I am only looking at how individuals perform with higher incentive surpluses and lower incentive surpluses relative to themselves. Both the fixed effect regression that I have run (3.2.2 utilizing incentive surplus) and those conducted by Raymond Guiteras and Kelsey Brooke (utilizing drawn incentive) show that incentive schemes do not sacrifice quality for quantity. As theory would suggest, relative incentive leads to workers sacrificing less of their work hours for leisure. Specifically, and similarly to my pooled OLS regressions, an increase of surplus by 5 leads to about 1 more scoop of beans being sorted in the 6-hour workday period[25]. This relation is highly significant in this case too. At the same time the regressions show that there is no significant proof that there is a sacrifice in quality of bean sorting.

[25] Compared to average scoops per day of around 7.4, and draw rates between 5-25; surplus incentive is analogous to relative incentive

One surprising result of my regressions is that pertaining to motivation. This data suggests that higher incentive rates relative to one's reservation incentive do causes higher production (as predicted), but the drop in quality is not dependent on monitoring. In fact, if anything, there is a borderline significant drop in quality when there is monitoring, but no drop when there is none if a control for quantity of beans is included. Without the quantity control, there is borderline significant increase in errors when there is no monitoring, and insignificant increase when there is. Suffice to say, with relative incentive as the dependent variable, the results are inconclusive with respect to the effects of monitoring; however, higher relative incentive rates do seem to increase quantity of beans sorted without a large (and in multiple regressions without a significant) sacrifice in quality.

IV. Conclusion, Tables, Bibliography, and Appendix

I. Conclusion

My replications, regressions, and rudimentary theoretical analysis have shown that regarding simple tasks, a well-designed incentive scheme seems to be beneficial in increasing competitiveness. The key in this statement is well designed, as incentive schemes can have potential negative effects if misused. Basic labor economics theory seem to suggest that due to diminishing returns of utility to labor and leisure, convex incentive schemes on average allow for the firm to reach higher isoprofit curves than linear schemes, and even elementary versions of convex incentive schemes can result in Pareto efficient improvements (see footnote 2).

With respect to sorting, theory suggests that individuals who are less efficient and motivated than average will theoretically seek out opportunities where wage is not dependent on performance. Vise versa, those who are highly talented and eager will theoretically seek out incentive-based pay schemes. My regressions support this theory, showing the ability of incentive schemes to sort between different types of employees.[26] This ability to sort with the help of incentive schemes reaffirms my hypothesis that firms can address many game theory and agency related problems utilizing incentive schemes.

Sorting allows employers to self-select those employees who are eager, talented, and overconfident in their abilities. Eager and talented employees are beneficial for obvious reasons,

[26] Sorting is also argued by Stephen Leider and Ian Larkin; however, they explain a portion of the differences in results to effort based effects.

while overconfident employees overestimate their incentive-pay, and thus employers can theoretically underpay them. The empirical data, however, suggests that a learning curve does exist both among overconfident and underconfident employees. Employees become more and more familiar with the incentive system they are provided, and cannot be taken advantage of in the long run. This isolated instance of manipulation has relatively strong external validity, and though more research is needed, learning can likely be extrapolated to other instances—though firms have the potential to manipulate employees through complex incentive schemes, employees seem to show an innate ability to learn from their mistakes and adapt in the future. Further research needs to be conducted in this area; however, my research seems to suggest that fears of inefficiencies through manipulation are not well grounded in the long run.

Incentive schemes obviously have an effort-based effect too. This effect is backed both by basic theory, as well as by empirics.[27] Theory suggests that convex based incentive schemes may have additional effort based effects, and could increase net utility in a Pareto improving manner. Labor economists often fear that a drop in quality, or an increase in risk taking negates such effort-based gains. Though the empirical data I have looked at does not suggest strong evidence of this (there is correlation, but it is not statistically significant unless controlled for monitoring), theory does back up this claim. The empirical study conducted and replicated in this thesis is highly specific to a low-skill low-pay labor situation where quantity and quality is easily defined. As economists suggest, the quality-based issue might be more significant for more complex tasks. Further econometric analysis needs to be conducted in this field as well; however, at least

[27] With the help of data and analysis provided by Raymond Guiteras and Kelsey Brooke

for simple tasks an increase in productivity seems to be achievable with higher relative and nominal incentive rates, without a significant decrease in quality.

Incentive schemes are not ideal in every case. Misused incentive schemes can cause more damage than good. Fears and dangers, from health related issues, to sacrifice in quality, to moral hazards, always need to be considered; however, frequently incentive based pay solves more problems than it creates, and has to potential to increase efficiencies substantially through sorting and effort. As a general rule, the easier it is to measure quantity and quality, the easier it is to implement a incentive scheme successfully. The continuous testing of these incentives schemes is vital for both the employee and the employer, but a Pareto improving impact on efficiency seems to be theoretically and empirically possible.

II. Tables and Figures

Figure 2.2.1—Average Overconfidence Score for Complete Sample

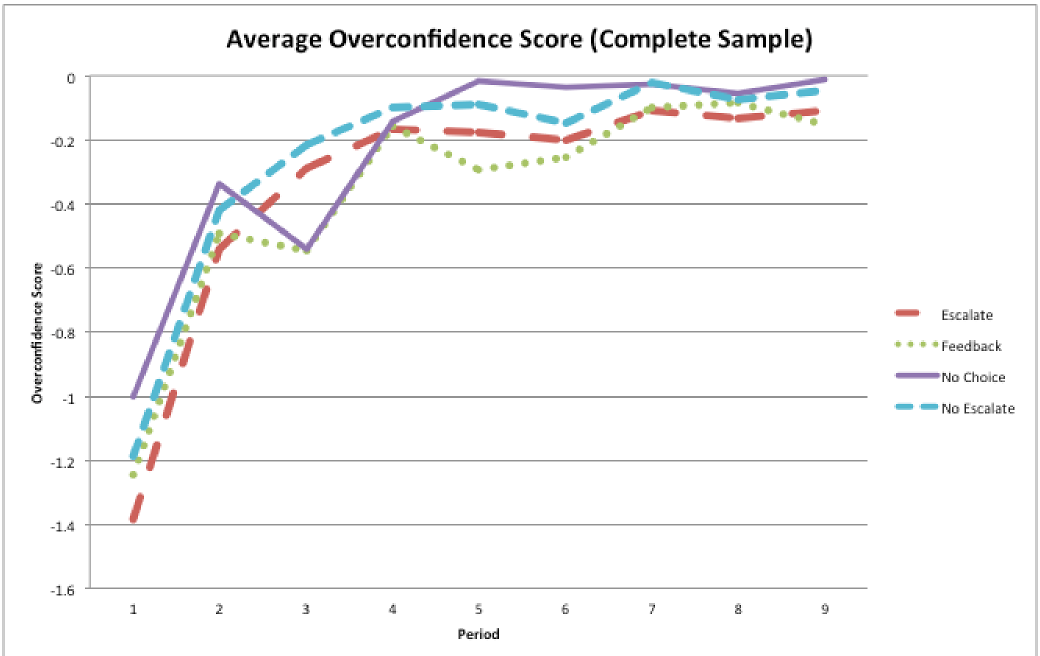


Figure 2.2.2—Average Overconfidence Score for Overconfident Sample

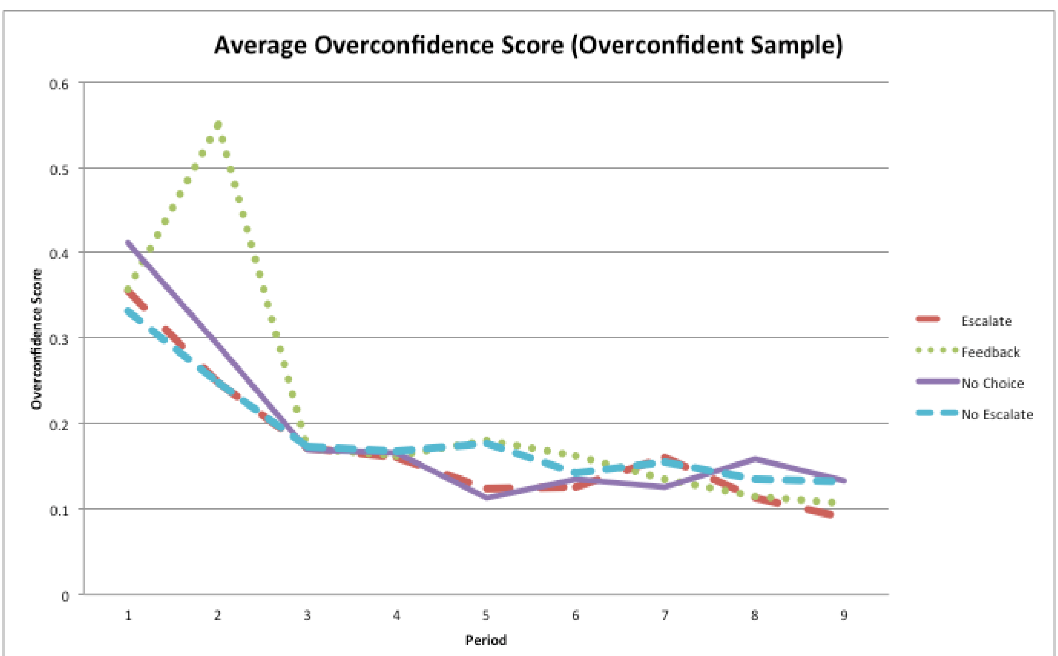


Table 2.2.3—Effect of Time (Period) on Overconfidence Score (OLS and RE)

Pooled OLS (all periods)	**Overconfidence Score**	**Overconfidence Score**
period	0.1017*** (0.0100)	-0.2238*** (0.0033)
male	0.0950* (0.0535)	-0.0031 (0.0168)
black	0.0696 (0.0854)	0.0114 (0.0276)
white	-0.0258 (0.0580)	0.0385** (0.0183)
age	0.00019 (0.0079)	-0.0013 (0.002)
constant	-0.8560*** (0.1839)	0.2776*** (0.0552)
number of observations	2214	898
Random Effect (all periods)	**Overconfidence Score**	**Overconfidence Score**
period	0.1017*** (0.0117)	-0.2238*** (0.0035)
male	0.0951 (0.0871)	-0.0033 (0.0219)
black	0.0696 (0.0787)	0.0114 (0.0207)
white	-0.0258 (0.0905)	0.0385* (0.0216)
age	0.00019 (0.00912)	-0.0013 (0.0016)
constant	-0.860*** (0.216)	0.2776*** (0.0556)
number of observations	2214	898
*** p<0.01	** p<0.05	*p<0.1

Table 2.2.4—Effect of Period on Overconfidence Score for Later Periods (RE)

(periods 4-9)	**OC score Underconfident (RE)**	**OC score Overconfident (RE)**
period	**0.0160* (0.0936)**	**-0.0091*** (0.0026)**
male	0.1303*** (0.0535)	-0.0076 (0.0168)
black	-0.0318 (0.0526)	0.0278 (0.0151)
white	-0.0853 (0.0357)	0.0303*** (0.00988)
age	0.0060 (0.0049)	-0.0001 (0.0013)
constant	-0.3678 (0.1647) *** (0.1839)	0.1580*** (0.0343)
number of observations	1476	659
*** p<0.01	** p<0.05	*p<0.1

Table 2.2.5—Change in Pay Scheme Preferences

Pooled OLS	**Payscheme**	**Learned**
learned	**0.030 (0.099)**	N/A (dependent variable)
black	0.140 (0.176)	0.0359 (0.0344)
white	-0.191 (0.115)	0.0207 (0.0234)
male	0.120 (0.100)	0.0249 (0.0216)
constant	1.599*** (0.110)	0.6800*** (0.0745)
number of unique id	100	246
confident at 4	N/A	**-0.1283 (0.0212)**
Random Effect		
learned	**-0.0227 (0.0536)**	
black	0.0465 (0.0796)	
white	-0.191 (0.115)	
male	0.111** (0.0549)	
constant	1.581*** (0.0587)	
number of unique id	179	
confident at 4	N/A	

Figure 2.2.6—Probability of Correct Incentive Choice (All Subjects)

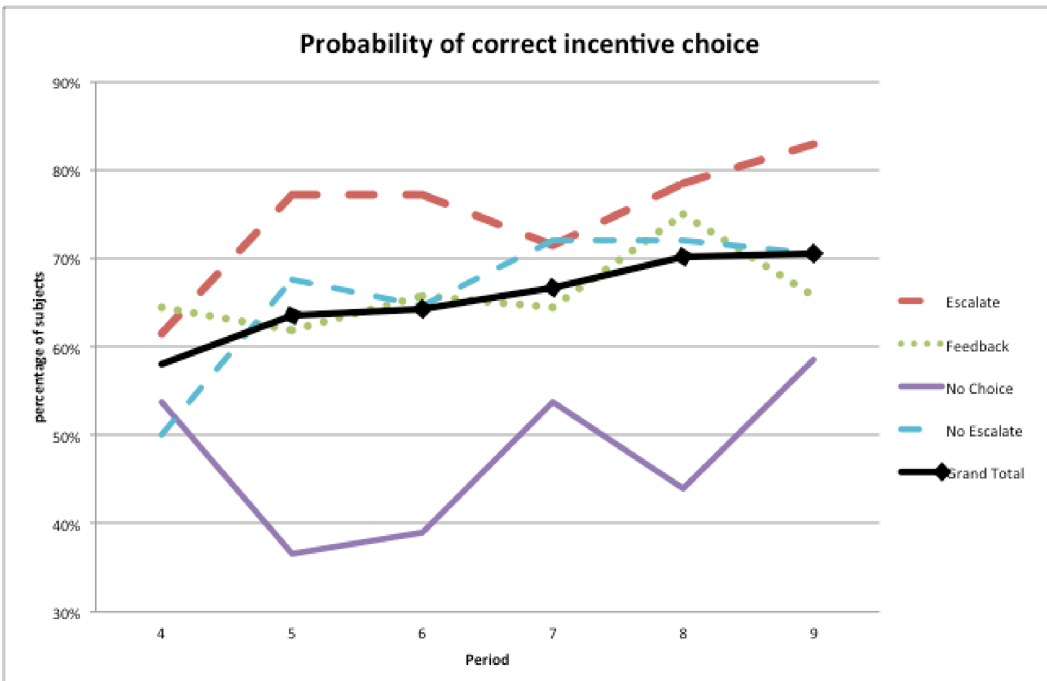


Figure 2.2.7—Probability of Correct Incentive Choice (Overconfident)

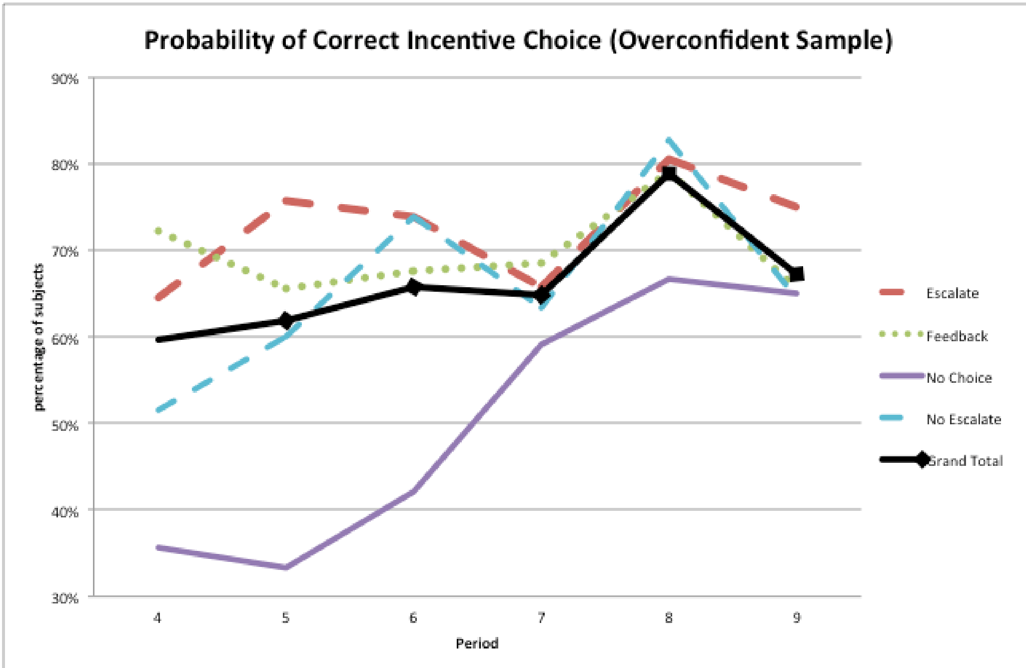


Table 2.2.8—The Learning Curve

	Correct Choice Pooled OLS	Correct Choice Random Effect	Correct Choice Random Effect if Overconfident
period	**0.0256*** (0.0071)**	**0.0256***(0.0076)**	**0.0337*** (0.0111)**
confident	0.0123 (0.0244)	0.0160 (0.0293)	N/A
choice	0.1970*** (0.0347)	0.1969*** (0.0376)	0.1952*** (0.0528)
escalate	0.0964*** (0.0287)	0.0964***(0.0287)	0.0454 (0.0479)
age	0.0022 (0.0037)	0.0022 (0.0039)	0.0074 (0.00528)
black	-0.1280*** (0.0533)	-0.1274** (0.0597)	-0.1458* (0.0887)
asian	0.0172 (0.4999)	0.0178 (0.0576)	-0.0040 (0.0755)
white	-0.0346 (0.0447)	-0.0343 (0.0509)	-0.0834 (0.0467)
other	0.0020 (0.0674)	0.0021 (0.0694)	0.0704 (0.0951)
male	0.0219 (0.0253)	0.0217 (0.0295)	-0.0030 (0.0467)
constant	0.2684*** (0.1102)	0.2668** (0.1240)	0.1780 (0.1796)
*** p<0.01	** p<0.05	*p<0.1	
n=1476			
periods 4-9			

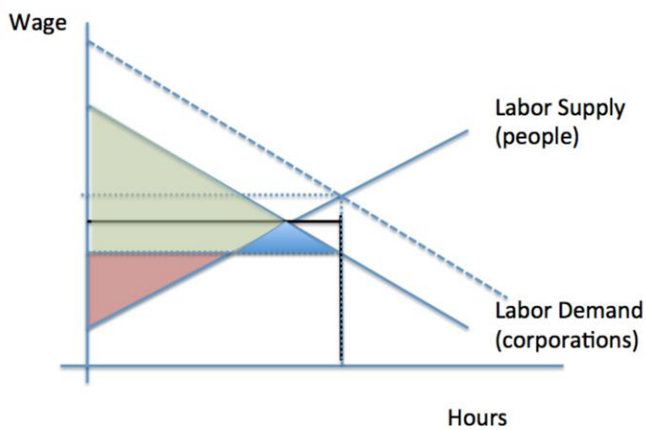
Table 3.2.1—Effect of Draw and Surplus on Quantity and Quality

Pooled OLS	Surplus	Draw
min_WTA	-0.5632 (0.0320) ***	0.4315 (0.0320)***
day	-0.1168 (0.1528)	-0.1398 (0.1533)
week	-0.2262 (0.0966)**	-0.2098 (0.0968)**
beans_quant	**0.2903 (0.0884) *** **	**0.2781 (0.0886) *** **
beans_errors	**0.3469 (0.1601) ** **	**0.3631 (0.1606)** **
female	-0.4281 (0.3534)	-0.3341 (0.3538)
constant	12.1521 (0.8821) ***	12.1662 (0.8854) ***
n=1476		

Table 3.2.2—Fixed Effect of Incentive Surplus on Quality and Quantity

Fixed Effect	Monitoring	No Monitoring	Monitoring	No Monitoring	Monitoring	No Monitoring
beans_quant	**0.3057 (0.1398) ** **	**0.2118 (0.1296)**	**0.1751 (0.1306)**	**0.1817 (0.1249)**	N/A	N/A
beans_errors	**0.7119 (0.2825)** **	**0.2224 (0.2554)**	N/A	N/A	**0.1111 (0.2464)**	**0.4864 (0.2639)* **
week	-0.1617 (0.1479)	-0.4620 (0.1640)***	-0.1394 (0.1477)	-0.4480 (0.1633)***	-0.4826 (0.1639)**	-0.1578 (0.1473)
peaklabor	0.3322 (0.4886)	-0.3354 (0.5294)	0.3485 (0.4887)	-0.3270 (0.5296)	-0.2478 (0.5281)	0.4591 (0.4829)
constant	5.8113 (1.3130) ***	8.1897 (1.4085) ***	7.7631 (1.0677) ***	8.8594 (1.1790) ***	10.0843 (0.7997)***	8.2481 (0.6948)***
n=min 712; 452 groups						

Chart 1.2.1—Dead Weight Loss From Information Asymmetries



III. Bibliography

Figueiredo, Rui J. P., Jr., Evan Rawley, and Orie Shelef. (2013) Bad Bets: Excessive Risk Taking, Convex Incentives, and Performance. STANFORD INSTITUTE FOR ECONOMIC POLICY RESEARCH 13.002 Online

Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and Why Incentives (Don't) Work to Modify Behavior. Journal of Economic Perspectives, 191-210.

Guiteras, R., & Jack, B. (2014). Incentives, Selection and Productivity in Labor Markets: Evidence from Rural Malawi. Nber Working Paper Series

Kvaloy, O., & Schöttner, A. (2014). Incentives to Motivate. CEIS Ifo, 2-35.

Larkin, I., & Leider, S. (2012). Incentive Schemes, Sorting and Behavioral Biases of Employees: Experimental Evidence. SSRN Journal SSRN Electronic Journal.

Veldman, J., & Gaalman, G. (2013) A model of strategic product quality and process improvement incentives. International Journal of Production Economics, 202-210.

IV. Appendix

Table A.2.2.1—Rerun of Larkin Leider table 6, with added Pooled OLS Regression

	Random Effect Incentive effect (number score)	OLS Incentive effect (number score)
nochoice_convex	0.148 (0.467)	0.0253 (0.538)
choice_linear	-1.402* (0.828)	-2.674*** (0.514)
choice_convex	0.0443 (0.770)	0.681 (0.462)
escalate_linear	-1.558** (0.738)	-3.061*** (0.513)
escalate_convex	-0.395 (0.692)	-0.362 (0.461)
practice_score_max	0.896*** (0.0322)	0.904 (0.0181)
period	0.494*** (0.0640)	0.429*** (0.0759)
constant	2.422*** (0.892)	2.732*** (0.730)
observations	1074	1074
Number of unique	179	179
*** p<0.01	** p<0.05	*p<0.1

Table A.2.2.2—Rerun of Larkin Leider Table 3

Random Effect

	payscheme	mistake_convex	mistake_convex_cost_pct	mistake_linear	mistake_linear_cost_pct
difference_belief	0.0184***	0.0363***	0.0124***	-0.0265***	-0.0109***
	(0.00350)	(0.00494)	(0.00185)	(0.00441)	(0.00215)
payescalate	0.0385	0.0207	0.000357	-0.0218	-0.00610
	(0.0433)	(0.0335)	(0.00952)	(0.0332)	(0.0137)
_IpayXdiffe_1	-0.00661	0.0115	0.000996	0.00768	0.000178
	(0.00529)	(0.00885)	(0.00298)	(0.00625)	(0.00511)
risk_cert_equiv	0.280**	0.130**	0.0348**	-0.164***	-0.0413*
	(0.115)	(0.0506)	(0.0154)	(0.0603)	(0.0217)
practice_score_max	-0.00592	-0.00172	-0.000130	-0.00126	-0.000273
	(0.00432)	(0.00210)	(0.000698)	(0.00215)	(0.000856)
period	0.0308***	-0.0161*	-0.00446*	-0.0136	-0.00124
	(0.00891)	(0.00835)	(0.00249)	(0.00855)	(0.00275)
Constant	0.282*	0.182**	0.0475**	0.407***	0.0929***
	(0.165)	(0.0758)	(0.0233)	(0.0984)	(0.0310)
Observations	816	816	816	816	816
Number of uniqueid	136	136	136	136	136
	*** p<0.01	** p<0.05	* p<0.1		

Table A.2.2.3—Rerun of Larkin Leider Table 5

Random Effect

	mistake_convex	mistake_convex	mistake_convex	mistake_convex	mistake_convex
init_overconf	0.00715**				
	(0.00361)				
payescalate	-0.0166	0.0320	0.0379	-0.0474	-0.0380
	(0.0384)	(0.0342)	(0.0383)	(0.0408)	(0.0363)
_IpayXinit_1	-0.00633				
	(0.00482)				
risk_cert_equiv	0.822* (0.0485)	0.765	0.0757	0.0875*	0.0858
		(0.0548)	(0.0500)	(0.0479)	(0.0616)
practice_score_max	0.00138	0.00152	0.000683	0.000300	-0.000699
	(0.00240)	(0.00253)	(0.000250)	(0.00247)	(0.000282)
period	-0.0206***	-0.0201*	-0.00208**	-0.0206**	-0.0222*
	(0.00919)	(0.00869)	(0.00929)	(0.00929)	(0.00932)
is_rel_overconf		0.216***			
		(0.0470)			
_Ipayx_is_re_1		0.0314			
		(0.0843)			
trivia_abs_diffb			0.00242		
			(0.00275)		
_IpayXtrivi_1			-0.00326		
			(0.0121)		
tc_over_5				0.0301	
				(0.0191)	
_IpayXtc_ov_1				-0.0217	
				(0.0291)	
z_overconfidence					0.0388***
					(0.0148)
_IpayXz_ove_1					-0.0569**
					(0.0241)
Constant	0.181**	0.115	0.181**	0.160**	0.203**
	(0.0797)	(0.0837)	(0.0826)	(0.0781)	(0.0869)
Observations	816	816	816	816	738
Number of uniqueid	136	136	136	136	123
	*** p<0.01	** p<0.05	* p<0.1		

Figure A.2.2.4 Frequency of Overconfidence Between Control and Treatment

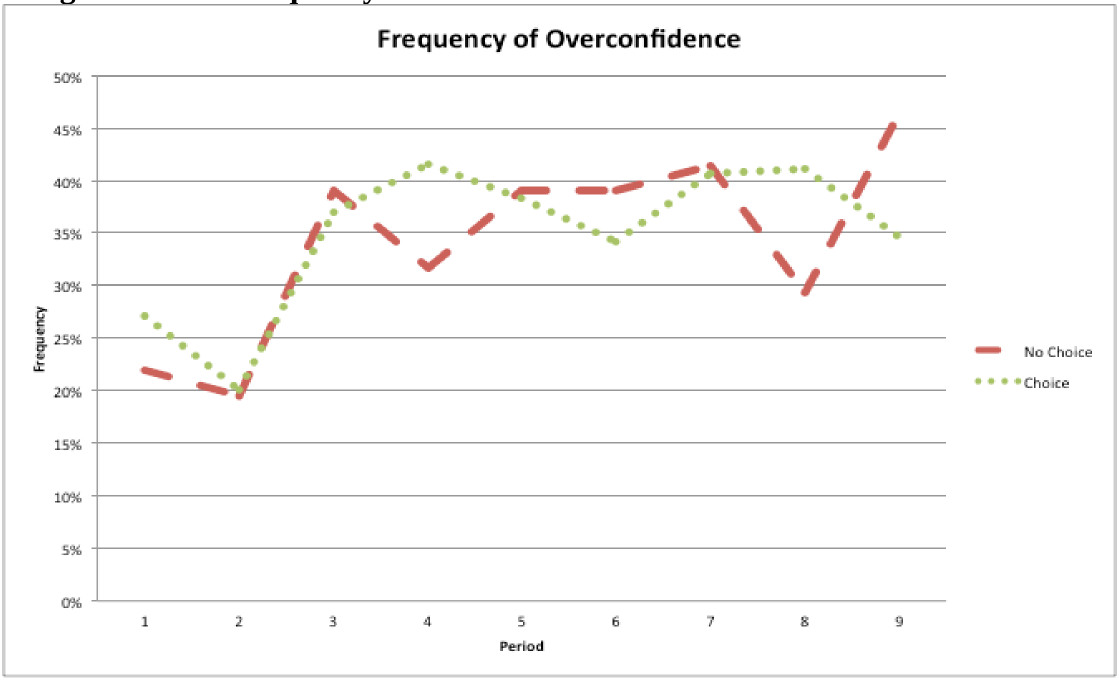


Figure A.2.2.5 Average Number Performance for Groups

Table A2.2.6—Rerun of Larkin Leider Table 1

Random Effect

Regression 1	difference_belief	difference_belief	difference_belief	difference_belief	difference_belief
init_overconf	0.119*** (0.0242)				
choice	0.0592 (0.410)	-0.0944 (0.452)	-0.176 (0.471)	-0.264 (0.440)	-0.162 (0.461)
escalate	-0.0160 (0.418)	-0.101 (0.454)	-0.0202 (0.447)	-0.0214 (0.447)	0.0271 (0.442)
practice_score_max	0.0740*** (0.0243)	0.0679*** (0.0257)	0.0581** (0.0256)	0.0520** (0.0251)	0.0426* (0.0259)
period	0.0203 (0.0788)	0.0215 (0.0744)	0.0203 (0.0788)	0.0203 (0.0788)	0.0203 (0.0788)
is_rel_overconf		3.909*** (0.308)			
trivia_abs_diffb			-0.0201 (0.0311)		
tc_over_5				0.428* (0.237)	
z_overconfidence					0.430** (0.216)
Constant	-2.217*** (0.643)	-3.079*** (0.662)	-1.988*** (0.673)	-2.035*** (0.631)	-1.712*** (0.643)
Observations	1074	1074	1074	1074	1074
Number of uniqueid	179	179	179	179	179
Robust standard errors in parentheses	*** p<0.01	** p<0.05	* p<0.1		

[28]

[28] Replication regressions and tables conducted as in original article with the help of Ian Larking through email correspondence.

Table A2.2.7—Rerun of Larkin Leider Table 2

Random Effect

	difference_belief	difference_belief
salary5		0.00143 (0.00107)
job_media		0.753 (0.485)
job_doctor		0.278 (0.529)
job_engineer		1.240 (1.033)
job_management		1.591* (0.887)
job_marketing		0.718 (0.596)
job_teacher		1.294 (0.938)
job_researcher		0.0909 (1.050)
job_whitecollar		0.0861 (0.546)
choice	-0.0956 (0.529)	-0.153 (0.428)
escalate	-0.0920 (0.420)	0.0562 (0.387)
practice_score_max	0.0424 (0.0287)	0.0460 (0.0328)
period	-0.00273 (0.0816)	0.00324 (0.0833)
male	0.580* (0.329)	
z_extraversion	0.316** (0.138)	
z_agreeableness	0.00831 (0.209)	
z_conscientiousness	-0.128 (0.190)	
z_emotionalstability	-0.0298 (0.186)	
z_openess	0.264 (0.232)	
Constant	-1.839*** (0.616)	-2.114*** (0.278)
Observations	978	948
Number of unique id	163	158
*** p<0.01	** p<0.05	* p<0.1