Incentivizing Quantity and Quality of Output: An Experimental Investigation of the Quantity-Quality Trade-off

Jared Rubin [a]
Anya Samek [b]
Roman M. Sheremeta [c,*]

[a] Argyros School of Business and Economics, Chapman University
One University Drive, Orange, CA 92866, U.S.A.
[b] Center for Economic and Social Research, University of Southern California
635 Downey Way, Los Angeles, CA 90098, U.S.A.
[c] Weatherhead School of Management, Case Western Reserve University
11119 Bellflower Road, Cleveland, OH 44106, U.S.A.

May 10, 2016

Abstract

Firms face an optimization problem that requires a maximal quantity output given a quality constraint. How firms should incentivize quantity and quality to meet these dual goals remains an open question. We provide a theoretical model and conduct an experiment in which participants are paid for both quantity and quality of a real effort task. Consistent with the theoretical predictions, higher quality incentives encourage participants to shift their attention from quantity to quality and to decrease the error rate at the expense of lowering quantity of output. This quantity-quality trade-off is significantly impacted by the participant's ability and level of loss aversion.

JEL Classifications: D24, J24, J31, J41
Keywords: quantity, quality, experiment, incentives, real effort, loss aversion

* Corresponding author: Roman Sheremeta, rshereme@gmail.com
We thank David Clingingsmith, Catherine Eckel, Sue Helper, Jonathan Meer, Matthew Sobel, Scott Shane, Jingjing Zhang and seminar participants at Case Western Reserve University, Texas A&M, University of Southern California, and the University of Technology Sydney for helpful comments. We also thank Kevin Guo, Christa Gibbs, Kathryn Carroll and students at the Behavioral and Experimental Economics Research Group for excellent research assistance. Any remaining errors are ours.

1. Introduction

Firms face a quantity-quality output trade-off. For instance, an orchard owner wants to incentivize her workers to pick as many apples as possible, but if workers are paid only based on the number of apples they pick, they may be careless, bruising the crop or picking spoiled apples. Yet if the owner rewards workers solely based on the number of non-bruised apples picked, much of the orchard may remain unpicked. Understanding how workers respond to the incentive schemes arising from such quantity-quality trade-offs is essential for understanding the conditions under which different wage schemes are efficient.

How to incentivize workers is a question fundamental to economics, and an active literature exists on the effect of different incentive compensation schemes on worker effort.[1] Worker productivity and quantity of output have been focuses of theoretical and empirical economic research for decades (Laffont and Martimort, 2009; Syverson, 2011). Some important works also consider the quality side of the trade-off. Holmstrom and Milgrom (1991) and Baker (1992) lay out seminal principal-agent models that incorporate the multi-dimensional aspects of worker incentives, and explain why incentivizing quantity may cause agents to ignore the quality of their output. However, to the best of our knowledge, an empirical investigation of how workers respond to different quantity-quality incentives is missing from the literature.

Recent work has begun to investigate the optimal incentive contracts for workers in situations when the firm cares about multiple dimensions of worker output. A series of papers have used existing data or field experiments to investigate the relative merits of flat rate versus

[1] For instance, economists have used behavioral economics theories of gift exchange and framing to induce greater productivity of workers in a field setting – see Gneezy and List (2006) for gift exchange and Hossain and List (2012) on framing. Other notable papers include the merits of competitive or piece rate incentive schemes, including the gender gap in competitiveness (Gneezy et al., 2003), and various profit-sharing compensation schemes (Nalbantian and Schotter, 1997). While many of these papers have incorporated quality considerations into their work, none of them have evaluated quality of output directly.

piece rate incentive schemes in the workplace (Lazear, 2000; Paarsch and Shearer, 2000; Shearer, 2004; Copeland and Monnet, 2009; Helper et al., 2010; Ederer and Manso, 2013; Al-Ubaydli et al., 2015).[2] The above papers find a positive impact of piece rates on quantity of output, but the evidence is mixed for its impact on quality.[3] For instance, Al-Ubaydli et al. (2015) find increases in quality, while Johnson et al. (2015) and Ederer and Manso (2013) find decreases in quality from pay-for performance compensation. Confounds in the field may explain these disparate results – for example, Al-Ubaydli et al. (2015) note that workers may have uncertainty about the observability of the quality of output.

We overcome the confounds inherent in field data by designing a laboratory experiment aimed at investigating the trade-offs associated with incentivizing quantity and quality in worker effort. The laboratory environment gives us control over problems that exist in the field, such as difficulty in observing and measuring quality of output and the effects of selection from workers moving in and out of the firm in response to changes in their compensation package. In addition, while the existing literature examines how different contracts (e.g., pay-for-performance or fixed wage) affect quantity or quality, it does not address the quantity-quality trade-off directly.

Importantly, our results broaden our understanding of incentivizing the multiple dimensions of productivity. Consider the following example. The firm hires a worker to produce a certain product (as in our introductory example). If the worker's payment depends only on *quantity* of units produced, she may choose to put all of her effort in maximizing quantity, while ignoring the quality of the production, as predicted by the theoretical models of Holmstrom and

[2] Additional related work includes Eriksson et al. (2009) who use a real-effort experiment to examine how feedback about performance of others impacts quantity and quality under pay-for-performance and tournament payment schemes, Charness and Grieco (2014) who investigate the impact of incentives on creativity, and Bracha and Fershtman (2013) who study how competitive incentive schemes affect the combination of cognitive and labor efforts provided by workers.

[3] Helper et al. (2010) suggest that a piece rate may actually have a negative impact on quantity when the production process is complex and quality is unobservable. Similarly, Rubin and Sheremeta (2016) show that even in the gift-exchange context uncertainty about quality can significantly decrease quantity.

Milgrom (1991) and Baker (1992). On the other hand, if the worker's payment depends mainly on *quality* of units produced, she may choose to put too much effort in maintaining high quality, while producing low quantity (an example of such a setup would be workers paid mostly on commission from successful endeavors). The optimal compensation scheme should involve a balance between rewarding quantity and quality in accordance with specific quality considerations in the market for the final product.

In this study, we examine how individuals respond to different incentives in the quantity-quality context. How do quality incentives impact productivity? Does incentivizing quality increase the quality of output? Does the quantity-quality trade-off depend on the agent's ability or behavioral factors? The theoretical model we outline provides insights into the answers to these questions, while the experiment we conduct provides empirical evidence. Specifically, our model of the quantity-quality tradeoff provides baseline predictions consistent with those found in the theoretical literature (Holmstrom and Milgrom 1991; Baker 1992), even though in our model quality is perfectly observable. To test this model, we conduct an experiment in which individuals solve math problems and their output quantity (number of problems attempted) and quality (number of problems answered correctly) is measured when (i) only quantity is incentivized, (ii) some quality is incentivized, and (iii) the bulk of the incentives are on quality.

In the experiment, we find evidence consistent with the theoretical predictions: Higher quality incentives encourage participants to shift their attention from quantity to quality and to decrease the error rate (i.e., number incorrect/number attempted) at the expense of lowering quantity of output. We also find that, consistent with the theoretical predictions, higher ability participants choose to focus more on quality and have lower error rates. Based on these findings, we discuss what the potential implications are for the optimal type of labor contracts that employers should offer workers.

We also observe a behavioral component in responsiveness to the quality incentive. There is heterogeneity in the impact of treatment, with more loss-averse participants displaying greater changes to their output from a change in quality incentives. Overall, we find that loss aversion leads participants to focus more on quantity and less on quality, while increasing the number of problems answered incorrectly. In addition, we characterize participants by whether they focus on pursuing quality or quantity during the experiment, and find that higher quality incentives increase the number of participants whose primary focus is quality.

In what follows, Section 2 describes the theoretical model and predictions. Section 3 outlines the experimental design. Section 4 summarizes the results, and Section 5 provides a discussion and conclusion.

2. Theoretical Model and Predictions

2.1. Theoretical Model

In this model, we provide insight into how economic agents exert effort under different reward schemes for the quantity and quality of their output. Consider an agent who exerts two-dimensional effort $e = (e_1, e_2)$, where $e_1 \geq 0$ is effort used to produce quantity and $e_2 \geq 0$ is effort used to produce quality. The agent has one unit of effort to provide, so $e_1 + e_2 = 1$. The agent has ability $a > 0$, and agents with higher ability produce high quality output at lower cost (for a given level of effort).

The expected quantity of high-quality output produced, $E[q^H] = e_1 p(e_2)$, depends on effort e_1 used to produce quantity and effort e_2 used to increase the probability of successful production $p(e_2)$, where $p' > 0$, $p'' < 0$, $p(0) = 0$, and $p(1) = 1$. The expected low-quality output is produced with a remaining probability, i.e., $E[q^L] = e_1(1 - p(e_2))$. The cost of exerting effort to produce quality is $c(e_2, a)$, where $c_1 > 0$, $c_2 < 0$, $c_{11} > 0$, $c_{12} < 0$, and $c(0, a) = 0$. We use a simplifying assumption that the cost to produce quantity is not a function of ability and it is normalized to zero.[4] The agent receives wage $w_1 \geq 0$ for each output (payment for quantity) and wage $w_2 \geq 0$ for each high-quality output (payment for quality). We assume that quality is perfectly observable.

The expected utility of the agent is:

$$E[U] = w_1 E[q^H + q^L] + w_2 E[q^H] - c(e_2, a) = w_1 e_1 + w_2 e_1 p(e_2) - c(e_2, a) \quad (1)$$

Since e_1 does not enter the worker's cost function, we have $e_1 = 1 - e_2$ at the worker's optimum. Therefore, the agent's first order condition $\frac{\partial E[U]}{\partial e_2} = 0$ is:

$$-w_1 - w_2 p(e_2) + w_2(1 - e_2)p'(e_2) - c_1(e_2, a) = 0 \quad (2)$$

2.2. Predictions

From the first order condition in (2), we can derive comparative statics related to how optimal effort levels respond to changes in relative wages. Consider first how effort changes as the relative return from producing quality increases (i.e., w_2 increases relative to w_1). From (2), there are increasing differences in $\{e_2, w_2\}$ if and only if $e_2 \leq 1 - \frac{p(e_2)}{p\prime(e_2)}$. Note that this also means that there are increasing differences in $\{e_2, w_2\}$ if and only if $\frac{\partial E[q^H]}{\partial e_2} \geq 0$, since $\frac{\partial E[q^H]}{\partial e_2} = -p(e_2) + (1 - e_2)p'(e_2)$ and thus $\frac{\partial E[q^H]}{\partial e_2} \geq 0$ implies $e_2 \leq 1 - \frac{p(e_2)}{p\prime(e_2)}$. Intuitively, it must be true

[4] Including the cost of e_1 in the agent's utility function would change none of the comparative static results.

that $\frac{\partial E[q^H]}{\partial e_2} \geq 0$ at any level of e_2 chosen by the agent: otherwise, increasing e_2 would decrease the expected level of both high-quality output q^H and low-quality output q^L.[5] Hence, there are increasing differences in $\{e_2, w_2\}$, and e_2 is increasing in w_2. Finally, there is some e^*, which solves $e^* = 1 - \frac{p(e^*)}{p\prime(e^*)}$, which the optimal value of e_2 never exceeds. This intuition is summarized in the following proposition and represented graphically in Figure 1.

Proposition 1: As w_2 increases, the optimal value e_1^* weakly decreases and e_2^* weakly increases.

It follows directly from Proposition 1 that $\frac{\partial E[q^H+q^L]}{\partial w_2} \leq 0$ since $E[q^H + q^L] = e_1$ and $\frac{\partial e_1}{\partial w_2} \leq 0$, implying that higher quality incentives decrease the total output (the sum of high-quality and low-quality output). We state this as Prediction 1:

Prediction 1: The average quantity of output $q^H + q^L$ is weakly decreasing in w_2.

It also follows from Proposition 1 that $\frac{\partial E[q^H]}{\partial w_2} \geq 0$. Recall that $E[q^H] = e_1 p(e_2) = (1 - e_2)p(e_2)$. Therefore, $\frac{\partial E[q^H]}{\partial w_2} = \frac{\partial e_2}{\partial w_2}(-p(e_2) + (1 - e_2)p'(e_2)) \geq 0$ since $\frac{\partial e_2}{\partial w_2} \geq 0$ and the term in brackets is always non-negative in equilibrium. The intuition behind this result is that an increase in w_2 encourages the agent to spend more effort in a manner where more high-quality

[5] To see this, note that $E[q^L] = (1 - e_2)\big(1 - p(e_2)\big)$, which is clearly decreasing in e_2.

units are produced. Sometimes this means reducing the effort e_1 spent on producing quantity, as noted above. This brings us to the next prediction:

Prediction 2: The average level of high-quality output q^H is weakly increasing in w_2.

Next, we define the error rate as the fraction of low-quality output relative to total output, or $E[\frac{q^L}{q^H+q^L}]$. From Proposition 1, it follows that $\frac{\partial E[\frac{q^L}{q^H+q^L}]}{\partial w_2} \leq 0$. To show this, note that $E[\frac{q^L}{q^H+q^L}] = \frac{e_1(1-p(e_2))}{e_1} = 1 - p(e_2)$. Therefore, $\frac{\partial E[\frac{q^L}{q^H+q^L}]}{\partial w_2} = \frac{\partial(1-p(e_2))}{\partial w_2} = -\frac{\partial e_2}{\partial w_2} p'(e_2) \leq 0$ since $\frac{\partial e_2}{\partial w_2} \geq 0$ and $p_1 > 0$. This brings us to the next prediction:

Prediction 3: The average error rate $\frac{q^L}{q^H+q^L}$ is weakly decreasing in w_2.

Finally, consider how the agent's ability affects her decision to focus on quality effort e_2. Since ability only enters into the cost function, it follows directly from (2) that there are increasing differences in $\{e_2, a\}$, and hence e_2 is increasing in a. The intuition underlying this result is straight-forward: higher ability agents face a lower marginal cost from exerting effort used to produce quality, so they choose greater e_2. This intuition is summarized in the following proposition:

Proposition 2: As a increases, the optimal value e_1^* weakly decreases and e_2^* weakly increases.

Predictions 4, 5, and 6 follow the same mathematical logic as Predictions 1, 2, and 3, respectively. For the sake of brevity we do not repeat the mathematics nor the intuition, but simply note that the sign of $\frac{\partial e_1^*}{\partial w_2}$ equals the sign of $\frac{\partial e_1^*}{\partial a}$ and the sign of $\frac{\partial e_2^*}{\partial w_2}$ equals the sign of $\frac{\partial e_2^*}{\partial a}$, entailing that the above comparative statics are the same with respect to a as they are with respect to w_2.

Prediction 4: The average quantity of output $q^H + q^L$ is weakly decreasing in a.

Prediction 5: The average level of high-quality output q^H is weakly increasing in a.

Prediction 6: The average error rate $\frac{q^L}{q^H+q^L}$ is weakly decreasing in a.

3. Experimental Design and Procedures

The experiment used participants drawn from the population of undergraduate students at the University of Wisconsin-Madison. Computerized experimental sessions were run using the Zurich Toolbox for Readymade Economics Experiments (z-Tree, Fischbacher, 2007) at the Behavioral Research Insights Through Experiments (BRITE) Laboratory. A total of 287 participants participated in 21 experimental sessions. Upon arriving at the laboratory, participants were randomly assigned to a computer station. The experiment proceeded in seven parts. All participants were given written instructions (available in Appendix A) at the beginning of each part, and an experimenter also read the instructions aloud.

In part 1, participants performed a real effort task: adding up sets of five randomly generated 2-digit numbers by hand, as quickly as possible, with no assistance other than a pen and paper (no calculators), for 5 minutes. The 2-digit numbers task is commonly used in the experimental literature because it is easy to explain, does not require previous experience and performance is not associated with a particular gender, socioeconomic background, or physical conditioning (Niederle and Vesterlund, 2007; Cason et al., 2010). In each treatment, participants were provided with up to 60 problems (one at a time) they could attempt to solve during 5 minutes. Participants could see only one problem at a time and they could not skip any problems. Each time a participant arrived at a new problem, she had 5 seconds to review it before the submit button appeared. After spending at least 5 seconds, the computer allowed participants to enter their answers. The 5 second delay can be considered an opportunity cost of skipping a problem by submitting any random answer.[6]

In all treatments, as shown in Table 1, participants received w_1 = \$0.10 for each attempted problem (i.e., for quantity). Depending on the treatment, participants also received an additional bonus for each attempted problem answered correctly (i.e., for quality), varying from w_2 = \$0.00 in the T-0.00 treatment to w_2 = \$3.00 in the T-3.00 treatment.

In part 2, we elicited beliefs about output quality by asking participants to provide a guess about how many of the attempted problems they solved correctly in part 1. Participants received an additional \$3 if their guess was equal to the number of correct answers they provided part 1. Participants were not aware of part 2 until after they finished part 1 of the experiment. The main purpose of eliciting participants' beliefs about their performance was to test whether the measured quantity and quality of output from part 1 matched the participants' own beliefs about how much quality they attempted.

[6] This is an important element of our design since several studies show that in real-effort experiments participants do not respond to incentives unless opportunity costs are introduced (Corgnet et al., 2015; Araujo et al., 2016).

In order to learn whether behavioral motivations play a role in responsiveness to quality incentives, in parts 3-5, we elicited participants' preferences toward ambiguity, risk and loss. While our theoretical model did not make a clear prediction about the role of behavioral motivations, such motivations have been shown to be important in principal-agent relationships in the field (Haigh and List, 2005; Hossain and List, 2012).

In part 3, we elicited participants' preferences toward ambiguity by presenting them with a set of 20 lotteries (see Table B1 in Appendix B). In each lottery, participants were asked to state whether they prefer an ambiguous option A ($0.00 or $10.00 with unknown chance each) or a safe option B (increasing monotonically from $0.50 to $10.00).[7] Parameters were set in such a way that more ambiguity-averse participants would choose safer options (and switch earlier to a safe option) than less ambiguity-averse participants. Again, participants were not aware of this part until after they finished the preceding parts.

In part 4, we elicited participants' preferences toward risk from a set of 20 lotteries (see Table B2 in Appendix B). In each lottery, participants were asked to state whether they prefer a risky option A ($0.00 or $10.00 with 50% chance each) or a safe option B (increasing monotonically from $0.50 to $10.00). As in previous parts, participants were not aware of this part until after they finished the preceding parts.

In part 5, we elicited participants' preferences toward losses from a set of 20 lotteries (see Table B3 in Appendix B). In each lottery, participants were asked to state whether they prefer a risky option A (50% chance of losing a certain amount between -$0.50 to -$10.00) or a safe option B of $0. As in previous parts, participants were not aware of this part until after they finished the preceding parts.

[7] Our elicitation procedure is similar to Shupp et al. (2013).

Part 6 was used to obtain a measure of participants' abilities on the math task, independent of incentive concerns. In this part, participants again performed a real effort task (as in the first part of the experiment): adding up sets of five randomly generated 2-digit numbers by hand, as quickly as possible. This time, participants had only 2.5 minutes to complete the task. The computer provided participants with up to 30 math problems (one at a time) that they could attempt to solve during the allotted time. As before, participants could see only one problem at a time and they could not skip any problems. Each time a participant arrived at a new problem, she had 5 seconds to review it before the submit button appeared. Participants received \$0.50 for each problem answered correctly, regardless of the treatment. Contrary to the first part, participants made no earnings from attempted problems that were incorrect.

Finally, in part 7, participants were asked to provide a guess about how many of the attempted problems they solved correctly in part 6. Participants received an additional \$3 if their guess was equal to the number of correct answers they provided in part 6. Participants were not aware of this task until after they finished the preceding parts of the experiment. The main purpose of eliciting participants' beliefs about their performance in part 6 was to obtain a measure of confidence, which may be linked to participants' decision to put more effort into quality or quantity. This measure is comparable across treatments, since it is not affected by the quantity-quality incentives that differ across treatments (unlike the guess in part 2, which may be a function of the different quantity-quality trade-offs faced in part 1).

At the end of the experiment, each participant received earnings from parts 1, 2, 6 and 7. For parts 3-5, in order to avoid portfolio effects, only one part and one line was paid out at random. Each session lasted approximately 90 minutes. Participants' earnings ranged from

$10.50 to $119.70, with a median of $25.60. In addition to their earnings in the experiment, participants also received a $7.00 show-up fee.

4. Results

4.1. How Incentives Impact Quantity and Quality

The summary statistics of our experiment are reported in Table 2 and represented graphically in Figures 2-4. First, we examine how higher quality incentives (i.e., higher reward for solving problems correctly) impact quantity (i.e., the number of problems attempted). Prediction 1 states that the level of total output $q^H + q^L$ should decrease with higher quality incentives w_2.

We begin by noting that there is a significant difference in the number of problems attempted between treatments T-0.00 and T-0.05 (31.42 versus 23.73; Wilcoxon rank-sum test, p-value = 0.03). In the analysis that follows, we denote the T-0.00 treatment as "zero quality incentive" and the T-0.05 treatment as "low quality incentive". There are no statistically significant differences between treatments T-0.25 and T-0.50 where quality incentives are medium (17.00 versus 17.71; Wilcoxon rank-sum test, p-value = 0.65) and treatments T-1.00 and T-3.00 where quality incentives are high (13.35 versus 13.60; Wilcoxon rank-sum test, p-value = 0.65). In the analysis that follows, we report pooled data from the "medium quality incentive" treatments T-0.25 and T-0.50, and the "high quality incentive" treatments T-1.00 and T-3.00.[8]

Figure 2 suggests that there are clear differences in the number of problems attempted between treatments with zero quality incentive (i.e., T-0.00), low quality incentive (i.e., T-0.05), medium quality incentives (i.e., T-0.25 and T-0.50) and high quality incentives (i.e., T-1.00 and T-3.00). Pairwise comparisons show that the differences in distributions are statistically

[8] We find no statistically significant differences for *any* of the outcomes reported in this section when comparing T-0.25 and T-0.50 or when comparing T-1.00 and T-3.00.

significant (Wilcoxon rank-sum test, five p-values < 0.01 and one p-value = 0.03).[9] We also find significant differences when comparing all treatments jointly (Kruskal-Wallis test, p-value < 0.01). Therefore, consistent with Prediction 1, we find that higher incentives for quality decrease quantity of output.

Result 1: Higher quality incentives decrease quantity of output.

Second, we examine how higher quality incentives impact quality (i.e., the number of problems solved correctly). Recall that Prediction 2 states that the level of high-quality output q^H should increase with higher quality incentives w_2.

Figure 3 suggests that there are clear differences in the number of problems answered correctly between treatments for all sets of pooled treatments except for medium quality incentives (i.e., T-0.25 and T-0.50) versus high quality incentives (i.e., T-1.00 and T-3.00). Indeed, we find a significant difference in quality between each of the other pooled groups (Wilcoxon rank-sum test, four p-values < 0.01 and one p-value = 0.04). Meanwhile, there is no statistically significant difference in number of problems answered correctly between the medium- and high-quality incentive treatments (Wilcoxon rank-sum test, p-value = 0.31). We will attempt to provide an explanation for this result in Section 4.4. The general differences across treatments are also significant when comparing all treatments jointly (Kruskal-Wallis test, p-value < 0.01). Therefore, consistent with Prediction 2, we find that higher incentives for quality increase quality of output.

[9] These p-values are for comparison between pooled treatments. Similar results hold for comparisons for unpooled treatments. This is true of all comparisons presented in this section. Unpooled results are available upon request.

Result 2: Higher quality incentives increase quality of output.

Third, we examine how higher quality incentives impact the error rate. To calculate the error rate, we use the ratio of the number of problems solved incorrectly to the number of problems attempted. Recall that Prediction 3 states that the error rate $\frac{q^L}{q^H+q^L}$ should decrease with higher quality incentives w_2.

Figure 4 suggests that there are clear differences in the error rates between the four pooled treatments. Indeed, this is what we find (Wilcoxon rank-sum test, five p-values < 0.01 and one p-value = 0.03). The differences are also significant when comparing all treatments jointly (Kruskal-Wallis test, p-value < 0.01), suggesting that, consistent with Prediction 3, the error rate decreases with higher quality incentives.

Result 3: Higher quality incentives decrease the error rate.

Together, Results 1, 2, and 3 provide strong support for the theoretical predictions of our model: Higher quality incentives encourage participants to shift their attention from quantity to quality by increasing quality of output and decreasing the error rate at the expense of lowering quantity of output.

4.2. Individual Characteristics

Next, we explore whether individual characteristics impact the choice of quality versus quantity. To answer this question, we elicited different individual characteristics summarized in Table 3. In part 6 of the experiment, we elicited an independent measure of participants' ability by having participants perform a real effort task for 150 seconds. In part 7, we elicited participants' beliefs about their performance in part 6 (see Appendix A for details). Using these beliefs, we compute an individual measure of overconfidence, defined as the predicted number of problems solved correctly in part 6 minus the number of problems actually solved correctly. From Table 3, we see that the median participant is overconfident, overestimating his performance by 1 correct problem (the mean participant overestimates performance by 0.84 correct problems).

Knowing that preferences towards uncertainty play an important role in decision making, we also elicited preferences regarding ambiguity, risk, and losses using multiple lottery choice mechanisms (see Table B1, Table B2, and Table B3 in Appendix B). Parameters of the elicitation procedure were set in such a way that the more ambiguity-, risk- and loss-averse participants would choose 'safer' options relative to 'riskier' options (and switch earlier from a risky option to a safe option) than the less ambiguity-, risk- and loss-averse participant. For example, a participant who in Table B2 first chooses four risky options A ($0.00 or $10.00 with 50% chance) and then switches to choose sixteen safe options B ($2.50-$10.00 for sure), would be characterized as very risk averse, while a participant who first chooses sixteen risky options and then four safe options would be characterized as very risk seeking. Potentially, one could even calculate the range of risk aversion coefficients for each participant that match their decisions (Holt and Laury, 2002). However, such calculations would necessarily have to rely on a specific utility functional form and would require a much larger sample of responses per each participant in order to consistently estimate such coefficients (Wilcox, 2008). Therefore, in the analysis that follows we use the number of safe options chosen by each participant in each elicitation task as an approximation of their preferences regarding ambiguity, risk, and losses.

Although the three elicitation tasks are not directly comparable, in all three tasks, a higher number of safe options implies a higher level of aversion toward ambiguity, risk, and losses.[10]

Next, we examine whether the elicited characteristics of participants are predictive of the number of problems attempted, the number of problems solved, and the error rate. First, we examine what factors influence the number of problems attempted. Table 4 shows the estimation results of different OLS regressions in which the dependent variable is quantity of output (the number of problems attempted), and the independent variables are dummies for the various pooled treatments, a measure of ability, and various behavioral measures. Specifications (2)-(7) support the non-parametric results by showing that low, medium, and high quality incentives decrease quantity of output relative to zero quality incentives, and more generally that higher quality incentives decrease quantity of output relative to lower quality incentives (as indicated by the p-value at the bottom of Table 4). Also note that specifications (2)-(7) indicate that there is a positive and statistically significant relationship between the participant's ability and quantity of output. This is contrary to Prediction 4, which suggested that total quantity should be decreasing in ability. A potential explanation is that the relationship between ability and quantity is non-linear: participants of sufficiently high ability may have been able to exert little effort on quality while still answering problems correctly, thus achieving higher quality *and* quantity. This insight does not follow directly from the model,[11] but it does make intuitive sense. To address this possibility, we include an ability-squared term in regressions reported in Table B5 in Appendix B. The coefficient on the squared term is positive and statistically significant while the coefficient on the ability variable is negative but statistically insignificant, indicating that at

[10] A simple correlation analysis shown in Table B4 in Appendix B indicates that there is a strong correlation between ambiguity-aversion and risk-aversion ($\rho = 0.67$), and somewhat weaker correlation between loss-aversion and ambiguity-aversion ($\rho = 0.30$) and loss-aversion and risk-aversion ($\rho = 0.35$).

[11] Although a simple addition to the model, where the probability of producing high-quality output is in part a function of ability, could yield this result.

sufficiently high ability levels participants can focus on quantity without, presumably, losing much quality; see Table 5.

We next examine the impact of elicited individual characteristics on quantity of output, reported in specifications (3)-(7). We find that overconfidence, risk, and ambiguity are not predictive of quantity; see specifications (3), (4), and (5). However, we find that loss aversion is a significant predictor of quantity, with participants who are more loss-averse choosing to focus on quantity by attempting more problems; see specification (6). This finding is intuitive. By focusing on quantity, participants can always guarantee a certain amount of payment for their performance, while focusing on quality involves the possibility of not solving the problem correctly. Therefore, a loss-averse participant may choose to focus mainly on quantity in order to minimize potential losses. To check the robustness of our findings, we include additional interaction terms; see specification (7).[12] Besides confirming our previous findings, we also find that higher quality incentives affect loss-averse participants less. The most straight-forward interpretation of this result is that loss aversion is less salient when incentives for quality are high, because the loss incurred from spending more time on a problem is smaller relative to the potential gain of getting the problem correct. We report additional robustness checks with individual treatment dummies in Table B6 in Appendix B.

Result 4: Participants focus more on quantity (attempt more problems) if they have higher ability or are more loss-averse.

[12] Including interaction terms with overconfidence and ambiguity does not yield any statistically significant results, and we therefore do not report these results for the sake of brevity. The interaction terms with risk aversion do yield statistically significant results when the number attempted is the dependent variable, but not in regressions with the other dependent variables reported in this section. These results are available upon request.

Next, we examine what factors influence the choice of quality (the number of problems solved). The estimation results reported in Table 5 provide support for the non-parametric results that higher quality incentives increase quality of output. Also, we find that consistent with Prediction 5, ability is positively and significantly correlated with quality, suggesting that participants of higher ability are more likely to focus on quality of output. This is not simply a matter of participants who are better at math in part 6 being better at math in part 1: whether participants focus on quantity is a *decision* in part 1. Furthermore, we find that loss aversion is again a significant predictor of quality, with participants who are less loss-averse choosing to focus on quality by solving more problems; see specifications (6) and (7). Again, this finding is intuitive, since focusing on quality entails losing out on the sure wage, i.e., $w_1 = \$0.10$, associated with focusing on quantity. Hence, more loss averse participants are less willing to take such a loss. As was the case in the quantity regressions, loss aversion only shows up as salient in the low quality incentive treatments; see specification (7). Intuitively, in the low quality incentive treatments, the benefit of focusing on quality is low relative to the loss of the sure wage associated with focusing on quantity. As the quality incentive increase, the latter loss becomes relatively less salient. We report additional robustness checks with individual treatment dummies in Table B7 in Appendix B.

Result 5: Participants focus more on quality (solve more problems correctly) if they have higher ability and if they are less loss-averse.

Finally, we examine what factors influence the error rate. The estimation results reported in Table 6 provides support for the non-parametric results that higher quality incentives decrease the error rate. Moreover, consistent with Prediction 6, in all specifications we find that ability is negatively and significantly correlated with the error rate, suggesting that participants of higher ability have lower error rates. There is a positive and significant relationship between loss aversion and the error rate in specifications (6) and (7), confirming our previous findings relating loss aversion to quantity and quality. We report additional robustness checks with individual treatment dummies in Table B8 in Appendix B.

Result 6: Participants have higher error rates if they have lower ability or if they are more loss-averse.

To summarize, Results 4, 5, and 6 indicate that there are important individual characteristics impacting the quantity-quality trade-off. First, we find that there is heterogeneity in the impact of treatment, with more loss-averse individuals displaying greater changes to their output from a change in quality incentives. In the zero and (occasionally) low quality incentive treatments, loss aversion leads participants to focus more on quantity and less on quality, while increasing the error rates. Also, we find that, consistent with the theoretical predictions, higher ability participants choose to focus more on quality and have lower error rates.

4.3. Classification of Participants

Next, we characterize participants by response time to identify how treatment differences affected the incentives of participants to focus primarily on quantity or quality. We begin by examining how much time participants spend on average on a given problem, which we consider an indicator of how much effort participants exert on quality. We assume that participants who spend more time on a problem than the average are more likely to be focusing on quality. As suggested by column 1 of Table 7, there are significant differences in the average time spent on a problem when comparing pooled treatments (zero quality incentive, low quality incentive, medium quality incentive, and high quality incentive). Pairwise comparisons for all show that the differences in distributions are statistically significant (Wilcoxon rank-sum test, four p-values < 0.01 and two p-values < 0.03). The difference are also significant when comparing all treatment jointly (Kruskal-Wallis test, p-value < 0.01).

Table 7 also reports the fraction of problems answered 'quickly' (signifying that a participant is focusing on quantity) by treatment. Recall that each participant had to spend a minimum of 5 seconds on each problem since the 'submit' button did not appear on the screen until 5 seconds had passed. We therefore look at different cut-off points – 6, 7, and 10 seconds – to see whether participants answer more quickly when quality is not incentivized. We find that 38% of problems are answered within 6 seconds when the reward for solving problems is not incentivized, i.e., T-0.00, while only 1% answer within 6 seconds when the reward is highly incentivized, i.e., T-1.00 and T-3.00 (Kruskal-Wallis test across all six treatments, p-value < 0.01). A similar pattern is observed for participants answering within 7 seconds (Kruskal-Wallis test, p-value < 0.01) and within 10 seconds (Kruskal-Wallis test, p-value < 0.01).

Finally, Figure 6 and the last column in Table 7 shows the fraction of participants choosing to focus only on quality. We define a participant as focusing on quality on a specific question if they either answered the question correctly or they spent at least 10 seconds answering the question.[13] As expected, we find that higher quality incentives increase the number of quality types (Kruskal-Wallis test, p-value < 0.01).

[13] We calculated numerous metrics of choosing "quality" or "quantity" (also see Table 9). For instance, another metric we considered was that a participant chose quality if they spent as much time answering the problem as the

Result 7: Higher quality incentives increase the number of participants focusing on quality and decrease the number of participants focusing on quantity.

4.4. "Close but not quite": Fine-Tuning the Quality Metric

In this section, we re-visit a puzzle laid out in Section 4.1: although participants in the medium quality incentive treatments had higher error rates than those in the high quality incentive treatments, they correctly answered a similar number of questions. In other words, we found no statistically significant difference in the quality of output between participants in the two sets of treatments, although we did find a statistically significant difference in the error rate. We also reported in Section 4.1 that participants in the medium quality incentive treatments had higher quantity (i.e., number attempted) than those in the high quality incentive treatments. Combining these insights suggests that participants in the medium quality incentive treatments answered quicker – leading to a higher error rate – but not so quickly that they never answered correctly. In other words, these results indicate the possibility that participants in the medium quality treatments made quick, educated guesses at the correct answer.

To test this possibility, we fine tune our measure of quality by considering "guesstimates": answers that are within 20 of the correct answer but not correct.[14] Such answers suggest some effort – they are not merely the result of participants flying through the questions

minimum time it took them to answer a question in part 6 (where quantity was not incentivized and payouts were the same across treatments). Results are similar in all specifications, and the statistics associated with other metrics are available upon request. Moreover, in all of the definitions we do not count decisions made in the last 30 seconds or decisions made in the participant's last answer because the decision-making calculus at the end of the five minute period may be different than in the first four minutes. For instance, one who can correctly answer a problem in 10 seconds (meaning that she should focus on quality in most of the treatments) has incentive to input a quick answer if there are only 6 seconds remaining.

[14] We have calculated similar results at cutoff points at within 5 and 10 of the correct answer and results are qualitatively similar.

to pocket the $0.10 per question attempted. Figure 7 reports the mean by treatment. Not surprisingly, "guesstimating" is decreasing in the quality incentive, and the differences between treatments are statistically significant (Kruskal-Wallis test, p-value < 0.01). The logic behind this result is clear: since participants are only incentivized to get the problem *exactly* correct (and not simply close to correct), the benefit to spending more time on a problem is increasing in the amount paid for quality. This finding is also consistent with the results reported in Section 4.3, where we found that higher quality incentives led participants to spend more time on problems.

These non-parametric results are confirmed in Table 8, which reports OLS estimates where the dependent variable is our metric of guesstimates. Again, the number of guesstimates is decreasing in the quality incentive. Perhaps unsurprisingly, overconfidence is positively correlated with guesstimates; see specification (3). Those who are overconfident in their ability may suspect they can answer more correctly and with greater speed than they actually can.

Result 8: Higher quality incentives decrease the number of participants "guesstimating" the correct answer.

These results therefore suggest an answer to the puzzle noted at the beginning of the section. Participants in medium quality treatment treatments "guesstimated" about one more problem on average than those in high quality incentive treatments. In the context of our experiment, this suggests that enough of these guesstimates were correct that the higher number attempted offset the higher error rate in the medium quality incentive treatments. More broadly, these results suggest that high quality output can be achieved with modest quality incentives, so long as it does not matter to the principal that the agents occasionally err.

4.5. Optimal Choice of Quantity and Quality

A participant making a decision of whether to focus on quantity or quality should take into account her ability to perform the task. As we have already shown, such ability is indeed important in making this decision. However, another important factor is the payment the participant is rewarded for quality. For example, when the reward is w_2 = \$0.25, the participant should expect to earn \$0.35 (w_1 = \$0.10 for quantity and w_2 = \$0.25 for quality) for successfully completing a task, which comes at the cost of spending time on that task (say x seconds depending on the ability). However, the participant also has an option to focus solely on quantity, which would result in a reward of w_1 = \$0.10 at the cost of a minimum 5 seconds spent on the task. Therefore, each participant should make a choice of whether to focus on quantity or quality depending on their relative ability to complete the task in x seconds and prices w_1 and w_2. Specifically, if $(w_1 + w_2)/x > w_1/5$ then a participant should focus on quality, and otherwise they should focus on quantity. One immediate implication is that higher w_2 should lead participants to pay more attention to quality. See a more formal discussion of this argument in Section 2, Proposition 1.

We can therefore calculate how many participants *should* have chosen to focus on quality given their ability. As a proxy for ability, we use the average time a participant needs to solve one problem correctly in part 6. Table 9 summarizes the average ability of participants across treatments: the first column reports the average number of seconds participants spent on each problem in part 6 in each treatment. Not surprisingly, since participants were randomly assigned to each treatment, there is no difference in ability between treatments (Kruskal-Wallis test, p-value = 0.46). However, since the reward for quality is different across treatments, the expected earnings are different. For example, when the reward is \$0.25 per correct answer, a participant who spends 30 seconds to solve one problem correctly should expect to earn \$3.50 if she chooses to focus on quality, i.e., (\$0.25 + \$0.10) × 300/30 = \$3.50. However, if instead, such a participant chooses to focus solely on quantity, she can earn \$6 since the opportunity cost is 5 seconds of moving to the next problem, i.e., \$0.10 × 300/5 = \$6.00. Therefore, a rational decision maker who can solve only one problem during 30 seconds should choose to focus on quantity when the reward for quality is \$0.25. Similar computations can be performed for all participants in each treatment. Table 9 reports the fraction of participants who should choose quality over quantity based on their ability and quality incentives.

Obviously, when the reward for quality is \$0.00, nobody should focus on quality.[15] The same is true when the reward is only \$0.05 for all but the most mathematically gifted (none of whom took part in this treatment). When the reward is \$0.25, 15% of participants should choose to focus on quality. When the reward is \$0.50 this number increases to 79%, and further to 98% when the reward is \$1.00. Finally, when the reward is \$3.00, all participants should focus on quality. Using this information, we can calculate the portion of participants in each treatment that chose to correctly focus on quantity or quality. We first calculate their average earnings from focusing on quality, as measured by the average time they spent deriving a correct answer in part 6 (see Table 9). Using this measure, we calculate their expected earnings from focusing on quality, which equals (300 / average seconds per correct answer) × (\$0.10 + w_2), where w_2 differs by treatment. Any participant whose expected earnings from focusing on quality exceed \$6 (the amount one could earn from solely focusing on quantity) should focus on quality; otherwise they should focus on quantity. We consider it a mistake for a participant to focus on

[15] It is also possible that some participants may choose to focus on quality simply because they enjoy adding numbers. Holmstrom and Milgrom (1991) note that “we shall not suppose that all work is unpleasant. A worker on the job may take pleasure in working up to some limit.”

quality (even once) when she should focus on quantity or for a participant to focus on quantity (even once) when she should focus on quality.[16] Table 9 shows that 80% of participants make mistakes when quality is not incentivized, 96% of participants make a mistake when there is a low quality incentive, while only 15-16% make mistakes when the reward is highly incentivized (i.e., T-1.00 and T-3.00). These differences are jointly significant (Kruskal-Wallis test, p-value < 0.01).[17]

Result 9: Higher quality incentives encourage participants to make better trade-offs between quantity and quality, reducing inefficient decision making.

4.6. Optimal Wage Schemes for Employers

Finally, our experiment has implications for the optimal type of labor contracts that employers offer workers. Of course, the optimal contract will depend on the conditions faced by the employer. Employers whose product receives a high price for quality will want to incentivize quality more than employers whose product's price is less dependent on quality.

Although we do not wish to push the specific results of our experiment too far, these results have implications for the type of contracts that employers offer their workers. Specifically, assume that employers face the following profit equation:

$$\pi = p_L q^L + p_H q^H - w_1(q^H + q^L) - w_2 q^H = (p_L - w_1)q^L + (p_H - w_1 - w_2)q^H, \qquad (4)$$

[16] We do not use the term "mistake" in a pejorative manner; it is likely that money maximization is not the only aspect of participants' utility functions. So, while participants very well may have been maximizing their utility by making a "mistake", our measure allows us to see how many participants are not making the correct *money maximizing* choices. Similar to before, we do not count decisions made in the last 30 seconds or decisions made in the participant's last answer in this calculation.

[17] The same conclusion stands when we drop the first 34 seconds of experiment, which is one standard deviation above the mean time taken to answer a question in part 6 (Kruskal-Wallis test, p-value < 0.01).

where $q^H \geq 0$ and $q^L \geq 0$ are the quantities of high quality and low quality output, $p_H \geq 0$ is the price received for high quality output, and $p_L \in \mathbb{R}$ the price received for low quality output. Note that p_L may be negative (if a company has to refund low quality or is sued) or positive. $w_1 \geq 0$ is the wage paid for quantity of output and $w_2 \geq 0$ is the wage paid for quality of output. In our experiment, w_1 = \$0.10 and w_2 varies by treatment.

In this profit equation, p_L and p_H are exogenous parameters (assuming the market is competitive), and w_1 and w_2 are choice variables. The question that our experiment answers – for a *specific* set of parameters – is which wage scheme among six possible ones is optimal for a given set $\{p_L, p_H\}$. Figure 8 reports the results. Employing $w_1 = \$0.10$ and the summary statistics of each treatment to determine q^L and q^H, these figures indicate that providing zero quality incentives (i.e., w_2 = \$0.00) is an optimal strategy when the price for quantity is positive and the price for quality is not sufficiently large. When there is a penalty for low quality output, it is optimal to incentivize for quality. As long as the penalty is not too large, it is optimal to provide a medium quality incentive (i.e., w_2 = \$0.25) – participants produce enough high-quality output that it is not worthwhile to pay them more to receive marginally more high-quality output. Also note that w_2 = \$0.50 is not optimal in the price range shown in Figure 8, indicating diminishing returns to w_2 once quality is sufficiently incentivized.[18] When the costs associated with low-quality output are high (i.e., p_L is sufficiently negative), it is optimal for employers to impose greater quality incentives (i.e., w_2 = \$1.00). But the marginal benefits attained by much higher wages (i.e., w_2 = \$3.00) are only worth it if p_L is highly negative or p_H is very large (indeed, this figure would need a much larger range to show where w_2 = \$3.00 is optimal). In other words, although greater quality incentives are optimal to impose when the return on quality

[18] It is true, however, that if the range of p_L and p_H is sufficiently large, w_2 = \$0.50 and w_2 = \$3.00 are optimal wages in part of the parameter space.

is large, the return on higher wages is rapidly diminishing past a certain point. Where the breakeven point lies depends on the characteristics of the workers (e.g., their ability and loss aversion) and can be determined by the employer through trial and error.

The specifics of this exercise should obviously not be taken too far; the actual numbers apply only to the present experiment. But one upshot of the analysis is that when firms are punished for low quality – via lawsuit, return policies, or warehouse fees for unsold goods – even modest quality incentives are enough to encourage high quality effort. When punishment is minimal in our experiment, "medium" quality incentives are optimal, with the less expensive wage incentive (i.e., w_2 = \$0.25) dominating the more expensive wage incentive (i.e., w_2 = \$0.50). Likewise, when punishment for low quality is severe, high quality wage incentives are optimal, again with the less expensive wage incentive (i.e., w_2 = \$1.00) dominating the more expensive wage incentive (i.e., w_2 = \$3.00).

5. Discussion and Conclusion

Firms face an optimization problem that requires a maximal quantity output given a quality constraint. It is not trivial to incentivize economic agents to care about both the quantity and quality of their output. A large literature suggests that incentives designed to encourage certain behaviors may backfire (Bowles, 2009; Gneezy et al., 2011; Bowles and Polania-Reyes, 2012). For example, incentives that are 'too small' may crowd out intrinsic motivation to put forth effort (Gneezy and Rustichini, 2000).[19] The problem becomes even more complicated when quality is

[19] Along these lines, Mellstrom and Johannesson (2008) provide evidence that monetary incentives may decrease (instead of increasing) blood donations. Rietz et al. (2013) show that imposing restrictive rules may have a detrimental impact on a gift-exchange relationship.

incorporated into consideration. On the one hand, incentivizing quality may increase individual performance. On the other hand, incentivizing quality may decrease quantity of output.

We provide a theoretical model and conduct an experiment to examine how incentivizing quality impacts individual decisions to focus on quality versus quantity. Consistent with theoretical predictions, we find that higher quality incentives encourage participants to shift their attention from quantity to quality and decrease the error rate at the expense of lowering quantity of output. We also find that, consistent with the theoretical predictions, higher ability participants choose to focus more on quality and have lower error rates.

Our findings have direct practical relevance for managers and employers. First, we show both theoretically and experimentally that there are important quantity-quality trade-offs that should be taken into account when designing contracts. For example, a manager who is highly concerned with the quality of output may choose to incentivize high-quality output. This should lead to higher quality of output and a lower error rate. However, it will most likely decrease quantity of output. Moreover, the results of our experiment show that although greater quality incentives are optimal to impose when the return on quality is large, the return on higher wages diminishes rapidly past a certain point. Therefore, the optimal compensation scheme should involve a balance between rewarding quantity and quality. For instance, Mauboussin (2012) provides the example of the Wallace Company, a pipe and valve distributor that won the prestigious Malcolm Baldrige National Quality Award in 1990 but had to file for bankruptcy two years later. Mauboussin concludes that "both too little and too much quality can be bad for a company's financial performance." Our study provides both theoretical and empirical evidence for this insight.

Second, our findings contribute to the literature examining how behavioral components can be used to improve work outcomes (Haigh and List, 2005; Hossain and List, 2012). Hossain and List (2012), for example, show that a manager can significantly improve the performance of workers by framing contracts in terms of "losses". Imas et al. (2016) show in a laboratory experiment that loss averse workers are more likely to prefer to enter contracts framed as losses, and they also work harder under loss contracts. We also find in our experiment that more loss-averse participants display greater changes to their output from a change in quality incentives. Participants who are more loss-averse choose to focus more on quantity, increasing the error rates. Therefore, a manager who is highly concerned with the quality of output may choose to avoid framing contracts in terms of losses to reduce the tendency of loss averse workers to focus on quantity rather than quality.

Another interesting practical application of our findings relates to an ongoing discussion in health economics on how to reward physicians in order to improve medical practice and increase social welfare. One part of the debate is whether to reward physicians solely for the volume of services they order (quantity) or to incorporate certain quality measures (quality).[20] Our findings suggest that rewarding quality is indeed effective in increasing quality of output and decreasing the error rate. However, in our experiment, it is easy to define and measure quality, which is not always the case in the medical field where quality is ill-defined and may be difficult to measure (Hennig-Schmidt et al., 2011; Godager et al., 2016).[21]

[20] See the following article in the New York Times: http://www.nytimes.com/2013/01/12/nyregion/new-york-city-hospitals-to-tie-doctors-performance-pay-to-quality-measures.html?pagewanted=2&_r=1&hp

[21] Incorporating quality incentives into a payment scheme can be difficult because measures of quality are not well defined and it can be difficult to monitor the quality. Other concerns with using pay-for-performance payment schemes are that some doctors will choose to pass on hard patients, while others will choose to perform too many treatments to assure good quality.

References

Al-Ubaydli, O., Andersen, S., Gneezy, U., & List, J.A. (2015). Carrots that look like sticks: Toward an understanding of multitasking incentive schemes. Southern Economic Journal, 81, 538-561.

Araujo, F.A., Carbone, E., Conell-Price, L., Dunietz, M.W., Jaroszewicz, A., Landsman, R., Lame, D., Vesterlund, L., Wang, S.W., & Wilson, A.J. (2016). The slider task: An example of restricted inference on incentive effects. Working Paper.

Baker, G.P. (1992). Incentive Contracts and Performance Measurement. Journal of Political Economy, 100, 598-614.

Bowles, S. (2009). When economic incentives backfire. Harvard Business Review, 87, 22-23.

Bowles, S., & Polania-Reyes, S. (2012). Economic incentives and social preferences: Substitutes or complements? Journal of Economic Literature, 50, 368-425.

Bracha, A., & Fershtman, C. (2013). Competitive incentives: Working harder or working smarter? Management Science, 59, 771-781.

Cason, T.N., Masters, W.A., & Sheremeta, R.M. (2010). Entry into winner-take-all and proportional-prize contests: An experimental study. Journal of Public Economics, 94, 604-611.

Charness, G., & Grieco, D. (2014). Creativity and financial incentives. Working Paper.

Copeland, A., & Monnet, C. (2009). The welfare effects of incentive schemes. Review of Economic Studies, 76, 93-113.

Corgnet, B., Hernán-González, R., & Schniter, E. (2015). Why real leisure really matters: Incentive effects on real effort in the laboratory. Experimental Economics, 18, 284-301.

Dickinson, D., & Villeval, M.C. (2008). Does monitoring decrease work effort? The complementarity between agency and crowding-out theories. Games and Economic Behavior, 63, 56-76.

Ederer, F.P., & Manso, G. (2013). Is pay-for-performance detrimental to innovation? Management Science, 59, 1496-1513.

Eriksson, T., Poulsen, A., & Villeval, M.C. (2009). Feedback and incentives: Experimental evidence. Labour Economics, 16, 679-688.

Falk, A., & Kosfeld, M. (2006). The hidden costs of control. American Economic Review, 96, 1611-1630.

Fehr, E. and Rockenbach, B. (2003). Detrimental effects of sanctions on human altruism. Nature, 422, 137-140.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. Experimental Economics, 10, 171-178.

Frey, B.S. (1993). Does monitoring increase work effort? The rivalry with trust and loyalty. Economic Inquiry, 31, 663–670.

Gneezy, U., & List, J.A. (2006). Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments. Econometrica, 74, 1365-1384.

Gneezy, U., & Rustichini, A. (2000a). A fine is a price. Journal of Legal Studies, 29, 1-17.

Gneezy, U., & Rustichini, A. (2000b). Pay enough or don’t pay at all. Quarterly Journal of Economics, 115, 791-810.

Gneezy, U., Meier, S., &. Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. Journal of Economic Perspectives, 25, 191-209.

Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in competitive environments: Gender differences. Quarterly Journal of Economics, 118, 1049-1074.
Godager, G., Henning-Schmidt, H., & Iversen, T. (2016). Does performance disclosure influence physicians' medical decisions? An experimental analysis. Journal of Economic Behavior and Organization, forthcoming.
Griffith, T.L. (1993). Monitoring and performance: A comparison of computer and supervisor monitoring. Journal of Applied Social Psychology, 23, 549-572.
Haigh, M.S., & List, J.A. (2005). Do professional traders exhibit myopic loss aversion? An experimental analysis. Journal of Finance, 60, 523-534.
Helper, S., Kleiner, M.M., & Wang, Y. (2010). Analyzing compensation methods in manufacturing: piece rates, time rates, or gain-sharing? NBER Working Paper.
Hennig-Schmidt, H., Selten, R., & Wiesen, D. (2011). How payment systems affect physicians' provision behavior – An experimental investigation. Journal of Health Economics, 30, 637-646.
Holmstrom, B., & Milgrom, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. Journal of Law, Economics, & Organization, 7, 24-52.
Holt, C.A., & Laury, S.K. (2002). Risk aversion and incentive effects. American Economic Review, 92, 1644-1655.
Hossain, T., & List, J.A. (2012). The behavioralist visits the factory: Increasing productivity using simple framing manipulations. Management Science, 58, 2151-2167.
Johnson, R.M., Reiley, D.H., & Muñoz, J.C. (2015). "The war for the fare": How driver compensation affects bus system performance. Economic Inquiry, 53, 1401-1419.
Imas, A., Sadoff, S., & Samek, A. (2016). Do people anticipate loss aversion? Management Science, forthcoming.
Laffont, J.J., & Martimort, D. (2009). The theory of incentives: The principal-agent model. Princeton University Press.
Lazear, E.P. (2000). Performance pay and productivity. American Economic Review, 90, 1346-1361.
Mauboussin, M.J. (2012). The success equation: Untangling skill and luck in business, sports, and investing. Harvard Business Press.
Mellstrom, C., & Johannesson, M. (2008). Crowding out in blood donation: Was Titmuss right? Journal of the European Economic Association, 6, 845–863.
Nalbantian, H.R., & Schotter, A. (1997). Productivity under group incentives: An experimental study. American Economic Review, 87, 314-341.
Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? Quarterly Journal of Economics, 122, 1067-1101.
Paarsch, H.J., & Shearer, B. (2000). Piece rates, fixed wages, and incentive effects: Statistical evidence from payroll records. International Economic Review, 41, 59-92.
Rietz, T., Schniter, E., Sheremeta, R.M., & Shields, T.W. (2013). Trust, reciprocity and rules. Working Paper.
Rubin, J., & Sheremeta, R. (2016). Principal-agent settings with random shocks. Management Science, forthcoming.
Shearer, B. (2004). Piece rates, fixed wages and incentives: Evidence from a field experiment. Review of Economic Studies, 71, 513-534.
Shupp, R., Sheremeta, R.M., Schmidt, D., & Walker, J. (2013). Resource allocation contests: Experimental evidence. Journal of Economic Psychology, 39, 257-267.

Syverson, C. (2011). What determines productivity? Journal of Economic Literature, 49, 326-365.

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. Quarterly Journal of Economics, 106, 1039-1061.

Wilcox, N. (2008). Stochastic models for binary discrete choice under risk: A critical stochastic modeling primer and econometric comparison. In J.C. Cox, G.W. Harrison (Eds.), Research in Experimental Economics Vol. 12: Risk Aversion in Experiments, Emerald, Bingley, UK, pp. 197-292.

Figure 1: Optimal values of e_1^* and e_2^* over different values of w_2

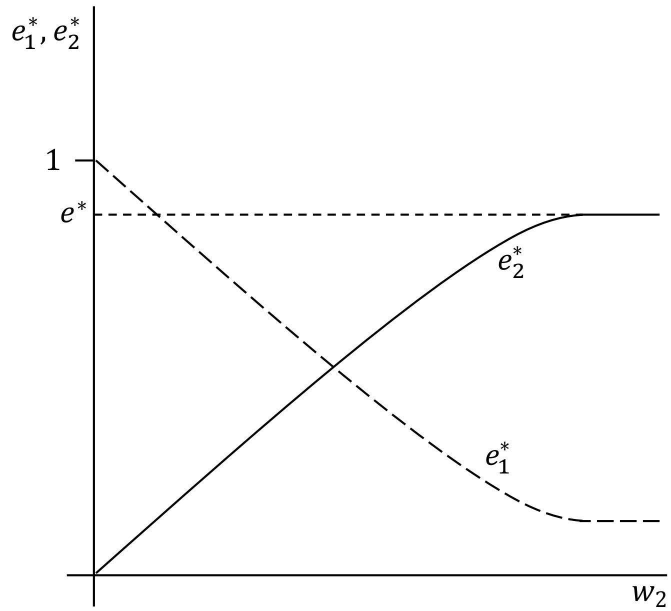

Figure 2: Measure of quantity (average problems attempted) by treatment

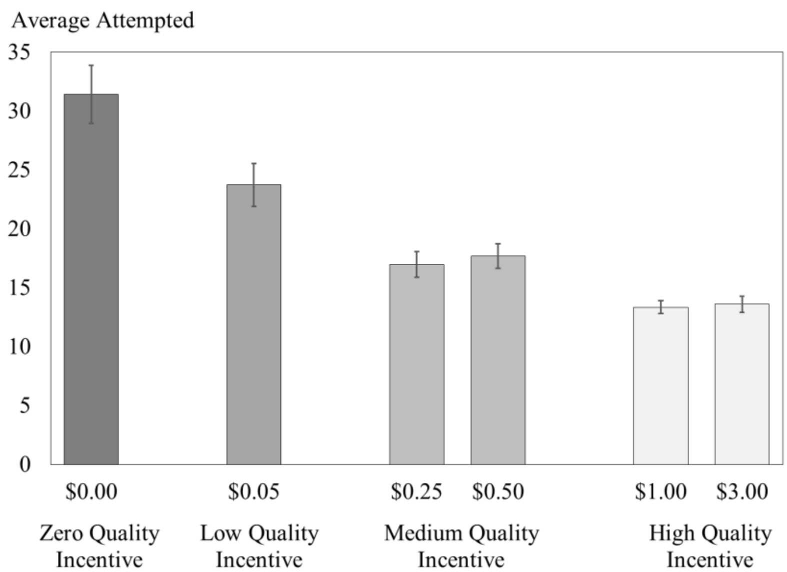

Figure 3: Measure of quality (average problems correct) by treatment

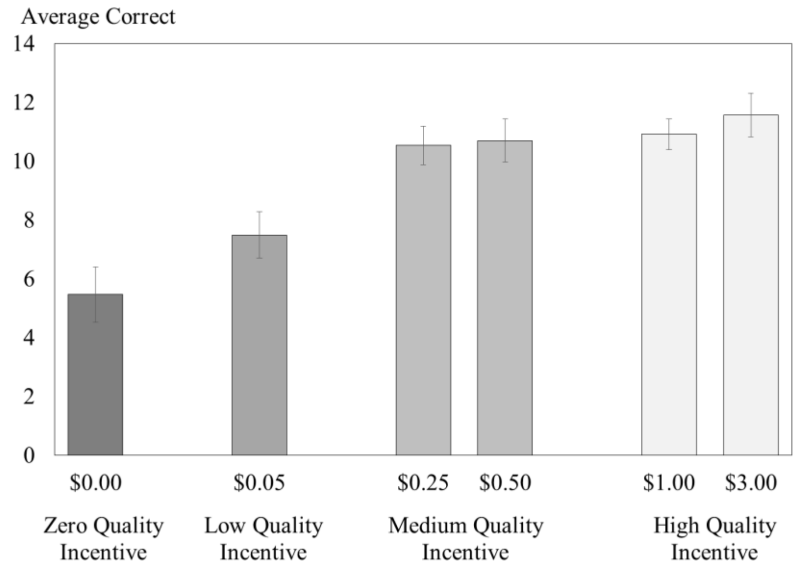

Figure 4: Error rate by treatment

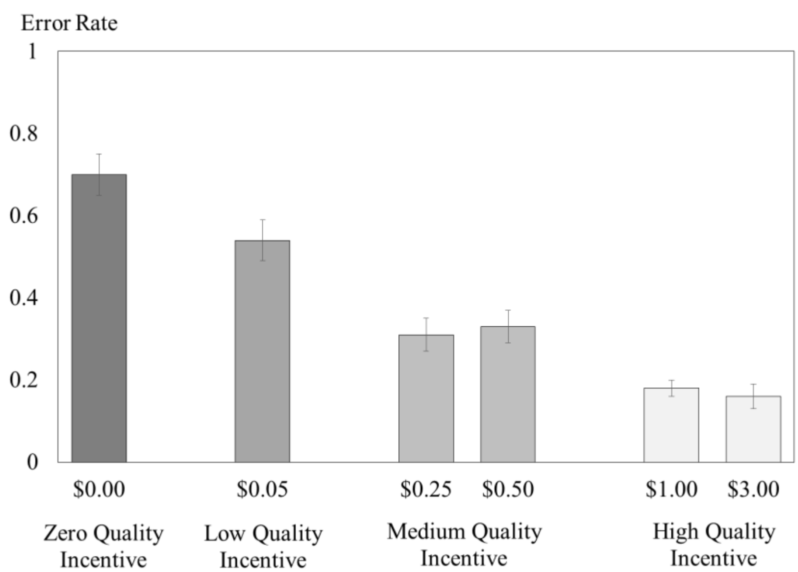

Figure 5: Fraction of problems answered in less than 10 seconds

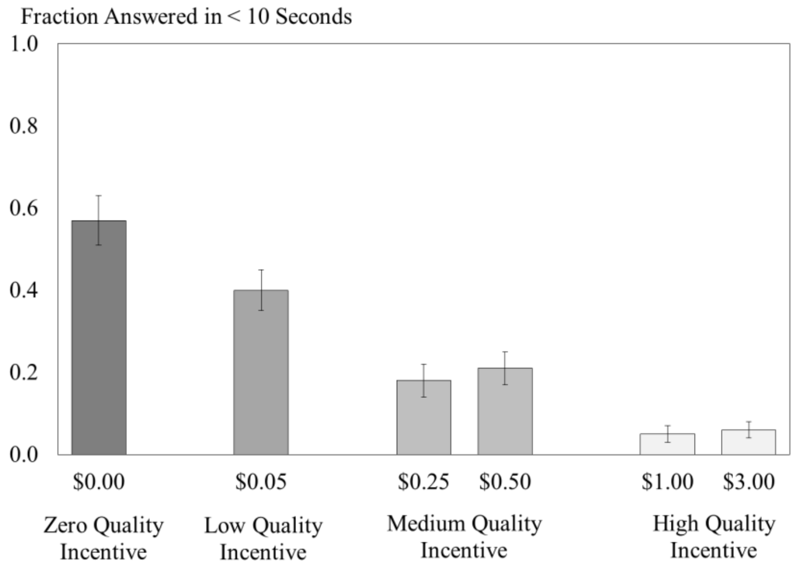

Figure 6: Fraction of participants focusing only on quality

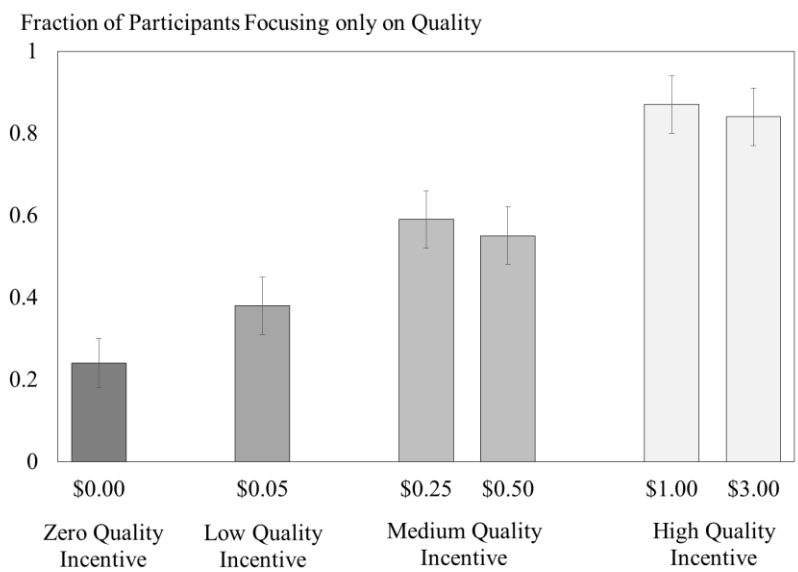

Figure 7: "Guesstimates" (answer within 20 of correct but not correct) by treatment

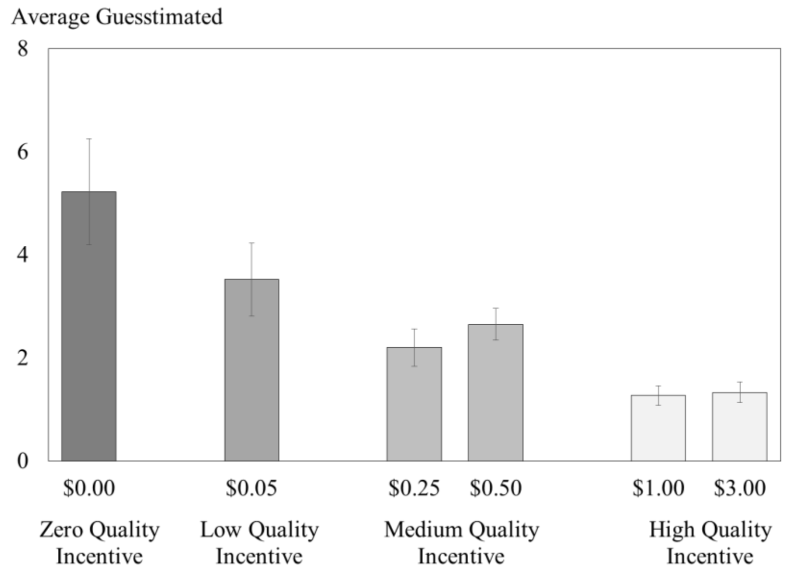

Figure 8: Optimal wage scheme for $p_H \leq 5, p_L \in [-5, 5]$

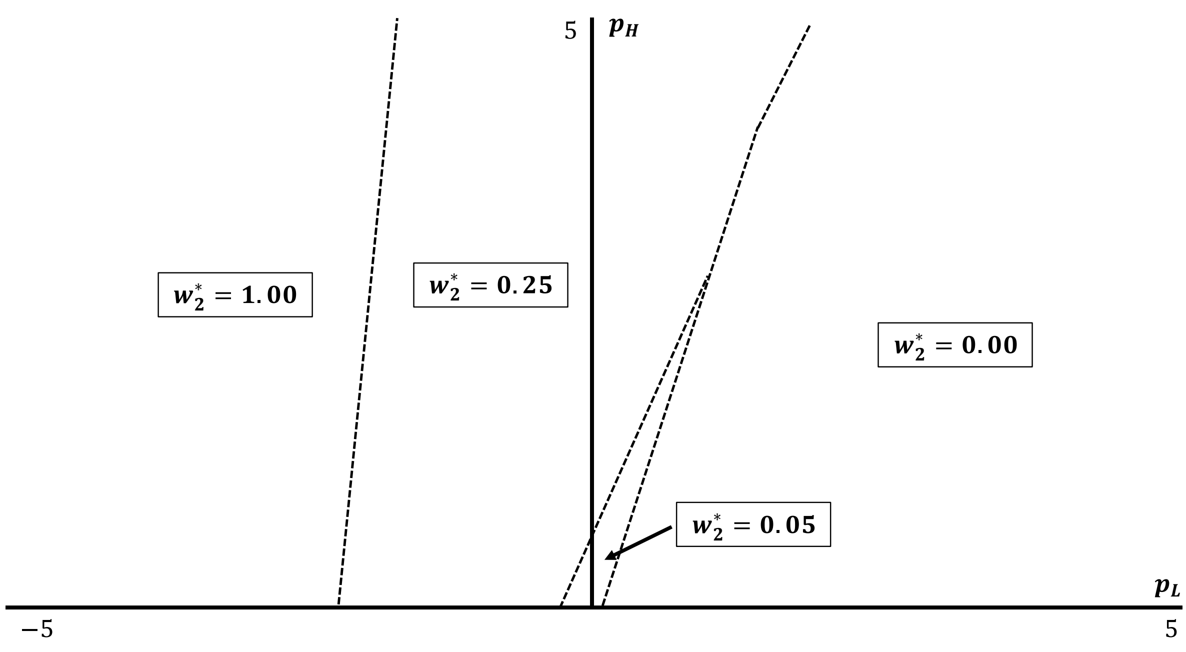

Table 1: Summary of treatments

Treatment	Payment for each problem: Attempted	Payment for each problem: Correct	N
T-0.00	$0.10	$0.00	45
T-0.05	$0.10	$0.05	48
T-0.25	$0.10	$0.25	46
T-0.50	$0.10	$0.50	51
T-1.00	$0.10	$1.00	52
T-3.00	$0.10	$3.00	45

Table 2: Summary statistics

Reward	Average attempted	Average correct	Average incorrect	Error rate = incorrect/attempted	N
$0.00	31.42	5.47	25.96	0.70	45
	(2.47)	(0.94)	(3.10)	(0.05)	
$0.05	23.73	7.50	16.23	0.54	48
	(1.83)	(0.79)	(2.42)	(0.05)	
$0.25	17.00	10.54	6.46	0.31	46
	(1.09)	(0.65)	(1.36)	(0.04)	
$0.50	17.71	10.71	7.00	0.33	51
	(1.04)	(0.73)	(1.28)	(0.04)	
$1.00	13.35	10.92	2.42	0.18	52
	(0.53)	(0.52)	(0.45)	(0.02)	
$3.00	13.60	11.58	2.02	0.16	45
	(0.69)	(0.74)	(0.34)	(0.03)	

Standard errors are in parentheses.

Table 3: Elicited characteristics

Percentile	Ability (correct in part 6)	Overconfidence (guess – correct)	Ambiguity	Risk	Loss
Min	0	-3	0	0	0
25%	5	0	10	10	14
50%	6	1	11	11	15
75%	7	1	13	12	17
Max	17	5	20	20	20

Table 4: OLS regressions of quantity

Specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Dependent variable = Quantity (number attempted)						
Ability	0.43	0.59***	0.67***	0.58**	0.59***	0.56**	0.47**
(correct in part 6)	(0.27)	(0.23)	(0.24)	(0.23)	(0.23)	(0.23)	(0.22)
Low quality incentives		-7.42***	-7.40***	-7.46***	-7.44***	-7.80***	13.73
(T-0.05)		(1.99)	(1.99)	(2.00)	(2.00)	(1.98)	(9.47)
Medium quality incentives		-14.40***	-14.39***	-14.39***	-14.42***	-14.92***	17.27*
(T-0.25 and T-0.50)		(1.74)	(1.74)	(1.74)	(1.74)	(1.73)	(8.90)
High quality incentives		-18.04***	-17.96***	-17.96***	-18.02***	-18.18***	16.65**
(T-1.00 and T-3.00)		(1.73)	(1.73)	(1.74)	(1.74)	(1.72)	(7.92)
Overconfidence			0.53				
			(0.49)				
Risk aversion				0.13			
				(0.20)			
Ambiguity aversion					0.06		
					(0.19)		
Loss aversion						0.45**	2.32***
						(0.18)	(0.47)
Loss aversion ×							-1.53**
Low quality incentives							(0.64)
Loss aversion ×							-2.23***
Medium quality incentives							(0.59)
Loss aversion ×							-2.43***
High quality incentives							(0.54)
Constant	16.74***	27.92***	26.99***	26.54***	27.26***	21.73***	-4.42
	(1.76)	(1.96)	(2.14)	(2.91)	(2.89)	(3.21)	(6.78)
Observations	287	287	287	287	287	287	287
R-squared	0.01	0.31	0.32	0.31	0.31	0.33	0.38
p-value, Low = Medium	--	0.00	0.00	0.00	0.00	0.00	--
p-value, Low = High	--	0.00	0.00	0.00	0.00	0.00	--
p-value, Medium = High	--	0.01	0.01	0.01	0.01	0.02	--

Table 5: OLS regressions of quality

Specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Dependent variable = Quality (number correct)						
Ability	1.43***	1.39***	1.48***	1.39***	1.39***	1.40***	1.41***
(correct in part 6)	(0.10)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
Low quality incentives		2.66***	2.68***	2.67***	2.67***	2.79***	-2.51
(T-0.05)		(0.76)	(0.75)	(0.76)	(0.76)	(0.76)	(3.70)
Medium quality incentives		4.35***	4.36***	4.35***	4.36***	4.53***	-4.26
(T-0.25 and T-0.50)		(0.66)	(0.65)	(0.66)	(0.66)	(0.66)	(3.48)
High quality incentives		5.56***	5.66***	5.54***	5.55***	5.61***	-1.33
(T-1.00 and T-3.00)		(0.66)	(0.65)	(0.66)	(0.66)	(0.65)	(3.10)
Overconfidence			0.60***				
			(0.18)				
Risk aversion				-0.03			
				(0.08)			
Ambiguity aversion					-0.04		
					(0.07)		
Loss aversion						-0.15**	-0.58***
						(0.07)	(0.18)
Loss aversion ×							0.37
Low quality incentives							(0.25)
Loss aversion ×							0.60**
Medium quality incentives							(0.23)
Loss aversion ×							0.49**
High quality incentives							(0.21)
Constant	0.81	-2.71***	-3.77***	-2.38**	-2.24**	-0.59	5.35**
	(0.63)	(0.75)	(0.80)	(1.11)	(1.10)	(1.22)	(2.65)
Observations	287	287	287	287	287	287	287
R-squared	0.44	0.56	0.58	0.56	0.56	0.57	0.58
p-value, Low = Medium	--	0.01	0.01	0.01	0.01	0.01	--
p-value, Low = High	--	0.00	0.00	0.00	0.00	0.00	--
p-value, Medium = High	--	0.02	0.01	0.02	0.03	0.04	--

Table 6: OLS regressions of the error rate

Specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Dependent variable = Error rate (incorrect/attempted)						
Ability	-0.04***	-0.04***	-0.04***	-0.04***	-0.04***	-0.04***	-0.04***
(correct in part 6)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Low quality incentives		-0.17***	-0.17***	-0.18***	-0.18***	-0.18***	0.09
(T-0.05)		(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.26)
Medium quality incentives		-0.35***	-0.35***	-0.35***	-0.35***	-0.36***	0.20
(T-0.25 and T-0.50)		(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.24)
High quality incentives		-0.52***	-0.52***	-0.52***	-0.52***	-0.53***	-0.04
(T-1.00 and T-3.00)		(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.22)
Overconfidence			0.01				
			(0.01)				
Risk aversion				0.01			
				(0.01)			
Ambiguity aversion					0.00		
					(0.00)		
Loss aversion						0.01**	0.04***
						(0.00)	(0.01)
Loss aversion ×							-0.02
Low quality incentives							(0.02)
Loss aversion ×							-0.04**
Medium quality incentives							(0.02)
Loss aversion ×							-0.03**
High quality incentives							(0.01)
Constant	0.62***	0.92***	0.90***	0.84***	0.87***	0.78***	0.39**
	(0.05)	(0.05)	(0.06)	(0.08)	(0.08)	(0.09)	(0.18)
Observations	287	287	287	287	287	287	287
R-squared	0.10	0.41	0.41	0.42	0.41	0.42	0.44
p-value, Low = Medium	--	0.00	0.00	0.00	0.00	0.00	--
p-value, Low = High	--	0.00	0.00	0.00	0.00	0.00	--
p-value, Medium = High	--	0.00	0.00	0.00	0.00	0.00	--

Table 7: Classification of participants by response time

Reward	Average time per problem	Fraction guessed < 6 seconds	Fraction guessed < 7 seconds	Fraction guessed < 10 seconds	Fraction only choosing quality
\$0.00	13.43	0.38	0.47	0.57	0.24
	(1.38)	(0.06)	(0.06)	(0.06)	(0.06)
\$0.05	16.26	0.24	0.33	0.40	0.38
	(1.19)	(0.04)	(0.05)	(0.05)	(0.07)
\$0.25	19.54	0.10	0.14	0.18	0.59
	(1.00)	(0.03)	(0.03)	(0.04)	(0.07)
\$0.50	18.85	0.09	0.15	0.21	0.55
	(0.88)	(0.02)	(0.03)	(0.04)	(0.07)
\$1.00	23.61	0.01	0.02	0.05	0.87
	(0.88)	(0.01)	(0.01)	(0.02)	(0.07)
\$3.00	23.11	0.01	0.02	0.06	0.84
	(0.88)	(0.01)	(0.01)	(0.02)	(0.07)

Standard errors are in parentheses.

Table 8: OLS regressions of "guesstimates"

Specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Dependent variable = "Guesstimate" (incorrect but within 20 of correct answer)						
Ability	0.00	0.03	0.08	0.03	0.03	0.02	0.01
(correct in part 6)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
Low quality incentives		-1.69**	-1.68**	-1.68**	-1.68**	-1.72**	5.04
(T-0.05)		(0.78)	(0.77)	(0.78)	(0.78)	(0.78)	(3.83)
Medium quality incentives		-2.81***	-2.80***	-2.81***	-2.80***	-2.84***	5.74
(T-0.25 and T-0.50)		(0.68)	(0.67)	(0.68)	(0.68)	(0.68)	(3.60)
High quality incentives		-3.93***	-3.87***	-3.94***	-3.93***	-3.94***	1.84
(T-1.00 and T-3.00)		(0.67)	(0.67)	(0.68)	(0.68)	(0.68)	(3.20)
Overconfidence			0.34*				
			(0.19)				
Risk aversion				-0.02			
				(0.08)			
Ambiguity aversion					-0.02		
					(0.07)		
Loss aversion						0.03	0.43**
						(0.07)	(0.19)
Loss aversion ×							-0.47*
Low quality incentives							(0.26)
Loss aversion ×							-0.59**
Medium quality incentives							(0.24)
Loss aversion ×							-0.41*
High quality incentives							(0.22)
Constant	2.68***	5.06***	4.46***	5.28***	5.33***	4.63***	-1.00
	(0.61)	(0.76)	(0.83)	(1.13)	(1.13)	(1.26)	(2.74)
Observations	287	287	287	287	287	287	287
R-squared	0.00	0.12	0.13	0.12	0.12	0.12	0.14
p-value, Low = Medium	--	0.09	0.09	0.09	0.10	0.09	--
p-value, Low = High	--	0.00	0.00	0.00	0.00	0.00	--
p-value, Medium = High	--	0.04	0.05	0.04	0.04	0.04	--

Table 9: Ability and expected earnings from quality

Reward	Average seconds per correct answer (part 6)	Expected earnings from quality	Fraction should focus on quality	Mistake
$0.00	30.47	1.28	0.00	0.80
	(3.76)	(0.08)	(0.00)	(0.06)
$0.05	33.69	1.73	0.00	0.96
	(3.80)	(0.09)	(0.00)	(0.03)
$0.25	23.91	4.84	0.15	0.93
	(1.13)	(0.23)	(0.05)	(0.04)
$0.50	28.90	8.11	0.76	0.61
	(3.07)	(0.50)	(0.06)	(0.07)
$1.00	28.10	13.75	0.98	0.15
	(2.22)	(0.65)	(0.02)	(0.05)
$3.00	27.49	40.50	1.00	0.16
	(2.46)	(2.49)	(0.00)	(0.05)

Standard errors are in parentheses.

Appendix A (For Online Publication): Instructions

GENERAL INSTRUCTIONS

This is an experiment in the economics of decision-making. Various research agencies have provided funds for this research. The instructions are simple.

The experiment will proceed in 7 parts. Each part contains decision problems that require you to make a series of choices that determine your total earnings. The currency used in all parts of the experiment is U.S. Dollars. You have already received a $7.00 participation fee. Your earnings from 7 parts of the experiment will be added to your participation fee. At the end of today's experiment, you will be paid in private and in cash.

It is very important that you remain silent and do not look at other people's work. If you have any questions, or need assistance of any kind, please raise your hand and an experimenter will come to you. If you talk, laugh, exclaim out loud, etc., you will be asked to leave and you will not be paid. We expect and appreciate your cooperation.

At this time we proceed to Part 1 of the experiment.

PART 1

In this part of the experiment, you will work on your own and have the chance to earn money by solving 2-digit math problems. At the end of the whole experiment, your entire earnings will be paid out to you immediately and in cash.

You will have 5 minutes (300 seconds) for this part. The computer will provide you with up to 60 math problems (one at a time) that you can attempt to solve during this 5 minutes. Each problem will consist of adding 5, two-digit numbers. All of the problems are about the same level of difficulty. You will see the problems one at a time and you will not be able to skip any problems. You will not be able to go back to any problems.

Your earnings for each problem depend on your responses in the following way:

- For each problem you **attempt**, you will receive **$0.10**.
- For each attempted problem you answer **correctly**, you will receive a bonus of **$0.50**.
- For each attempted problem you answer **incorrectly**, you will receive a penalty of **-$0.00**.

Answering a problem correctly means that you have provided the correct answer, for example, 2+2=4 is correct while 2+2=3 is incorrect.

The time remaining will be displayed on the overhead. When 5 minutes are up, time will be called. You will not be able to respond to any more problems after time is up because your computer will be on pause. After time is called, you will need to enter "0" to move on to the outcome screen, and the last problem you answer will not count as an attempt. An example of a problem screen is shown below.

Note that you will know which problem you are on. The 5 numbers that you should add are listed in the middle of the screen. In this example, you should be adding 28+39+48+23+25.

Each time you arrive at a new problem, you will have 5 seconds to review it before the submit button appears. After spending at least 5 seconds, the computer will allow you enter your answer. Although you will be required to spend *at least* 5 seconds on each problem, you can also spend *more than* 5 seconds.
Press “Submit” when you are ready to go on to the next problem.

You will not know if you answered any one problem correctly or incorrectly until the end of the experiment, when you will learn your total number of correct and incorrect responses.

The actual earnings for this part of the experiment will be determined at the end of the experiment, and will be independent of other parts of the experiment.

PART 2

In this part of the experiment, you will be asked to provide a guess about how many of the attempted problems in Part 1 you solved **correctly**. You will receive an additional $3 if your guess is equal to the number of correct answers that you provided us in Part 1.

Please enter your guess on your screen. Record your answer (and outcome) below. The actual earnings for this part of the experiment will be determined at the end of the experiment, and will be independent of other parts of the experiment.

Use the following table for records:

	Record your Results Here
Number of Problems Attempted	
Guess About the Number of Problems Correct	

PARTS 3-5

In PARTS 3-5 of the experiment, you will be asked to make a series of choices in decision problems. How much you receive will depend partly on **chance** and partly on the **choices** you make.

In each PART, you will see a table with 20 lines. You will state whether you prefer Option A or Option B in each line. You should think of each line as a separate decision you need to make. However, only **one line** in PARTS 4-6 will be the ‘line that counts’ and will be paid out.

- At the end of the experiment, we will draw a card from a deck of cards numbered 3, 4, 5. Depending on which card is chosen, either PART 3, PART 4, or PART 5 will “count”
- Then, we will draw a card from a deck of cards numbered 1, 2, ….20. The number on the card chosen indicates which **line** in that part will be paid out

Because each line is equally likely to be selected, and because you do not know which line will be selected when you make your choices, you should pay close attention to the choices you make in each line. **In some lines, depending on the decisions you make, you may earn up to $10.**

PART 3

For each line in the table, please state whether you prefer option A or option B. Notice that there are a total of **20 lines** in the table – you should think of each line as a separate decision you need to make.

Your earnings for the selected line depend on which option you chose: If you chose option B in that line, you will receive an amount of money specified by option B – between **$0.50** and **$10**, depending on the line. If you chose option A in that line, you will receive either **$10** or **$0**. To determine your earnings in the case you chose option A we will randomly draw a ball from a bag containing twenty balls. The balls are either **white** or **orange**, but you do not know the exact number of white and orange balls before you make your decision. Before you draw the ball you choose a color. For example, suppose that you choose white. If the drawn ball is really white, you will receive $10. If the drawn ball is orange, you will receive $0.

While you have all the information in the table, you should input all your 20 decisions into the computer. The actual drawing of the ball for this part of the experiment will be determined at the end of the experiment.

Use the following tables for records:

	Record Your Response Here
CHOOSE YOUR COLOR:	☐ WHITE ☐ ORANGE

Decision Number	Option A		Option B	Choose A or B
1	**$10.00** with unknown chance	**$0.00** with unknown chance	**$0.50** for sure	
2	**$10.00** with unknown chance	**$0.00** with unknown chance	**$1.00** for sure	
3	**$10.00** with unknown chance	**$0.00** with unknown chance	**$1.50** for sure	
4	**$10.00** with unknown chance	**$0.00** with unknown chance	**$2.00** for sure	
5	**$10.00** with unknown chance	**$0.00** with unknown chance	**$2.50** for sure	
6	**$10.00** with unknown chance	**$0.00** with unknown chance	**$3.00** for sure	
7	**$10.00** with unknown chance	**$0.00** with unknown chance	**$3.50** for sure	
8	**$10.00** with unknown chance	**$0.00** with unknown chance	**$4.00** for sure	
9	**$10.00** with unknown chance	**$0.00** with unknown chance	**$4.50** for sure	
10	**$10.00** with unknown chance	**$0.00** with unknown chance	**$5.00** for sure	
11	**$10.00** with unknown chance	**$0.00** with unknown chance	**$5.50** for sure	
12	**$10.00** with unknown chance	**$0.00** with unknown chance	**$6.00** for sure	
13	**$10.00** with unknown chance	**$0.00** with unknown chance	**$6.50** for sure	
14	**$10.00** with unknown chance	**$0.00** with unknown chance	**$7.00** for sure	
15	**$10.00** with unknown chance	**$0.00** with unknown chance	**$7.50** for sure	
16	**$10.00** with unknown chance	**$0.00** with unknown chance	**$8.00** for sure	
17	**$10.00** with unknown chance	**$0.00** with unknown chance	**$8.50** for sure	
18	**$10.00** with unknown chance	**$0.00** with unknown chance	**$9.00** for sure	
19	**$10.00** with unknown chance	**$0.00** with unknown chance	**$9.50** for sure	
20	**$10.00** with unknown chance	**$0.00** with unknown chance	**$10.00** for sure	

PART 4

For each line in the table, please state whether you prefer option A or option B. Notice that there are a total of **20 lines** in the table – you should think of each line as a separate decision you need to make.

Your earnings for the selected line depend on which option you chose: If you chose option B in that line, you will receive an amount of money specified by option B – between **$0.50** and **$10**, depending on the line. If you chose option A in that line, you will receive either **$10** or **$0**. To determine your earnings in the case you chose option A we will randomly draw a ball from a bag containing twenty balls. There are **ten orange** and **ten white** balls in the bag. That means that when we draw a ball, there is a 50% chance that it is white and a 50% chance that it is orange. Before you draw the ball you choose a color. For example, suppose that you choose white. If the drawn ball is really white, you will receive $10. If the drawn ball is orange, you will receive $0.

While you have all the information in the table, you should input all your 20 decisions into the computer. The actual drawing of the ball for this part of the experiment will be determined at the end of the experiment.

Use the following tables for records:

	Record Your Response Here
CHOOSE YOUR COLOR:	☐ WHITE ☐ ORANGE

Decision Number	Option A		Option B	Choose A or B
1	**$10.00** with 50% chance	**$0.00** with 50% chance	**$0.50** for sure	
2	**$10.00** with 50% chance	**$0.00** with 50% chance	**$1.00** for sure	
3	**$10.00** with 50% chance	**$0.00** with 50% chance	**$1.50** for sure	
4	**$10.00** with 50% chance	**$0.00** with 50% chance	**$2.00** for sure	
5	**$10.00** with 50% chance	**$0.00** with 50% chance	**$2.50** for sure	
6	**$10.00** with 50% chance	**$0.00** with 50% chance	**$3.00** for sure	
7	**$10.00** with 50% chance	**$0.00** with 50% chance	**$3.50** for sure	
8	**$10.00** with 50% chance	**$0.00** with 50% chance	**$4.00** for sure	
9	**$10.00** with 50% chance	**$0.00** with 50% chance	**$4.50** for sure	
10	**$10.00** with 50% chance	**$0.00** with 50% chance	**$5.00** for sure	
11	**$10.00** with 50% chance	**$0.00** with 50% chance	**$5.50** for sure	
12	**$10.00** with 50% chance	**$0.00** with 50% chance	**$6.00** for sure	
13	**$10.00** with 50% chance	**$0.00** with 50% chance	**$6.50** for sure	
14	**$10.00** with 50% chance	**$0.00** with 50% chance	**$7.00** for sure	
15	**$10.00** with 50% chance	**$0.00** with 50% chance	**$7.50** for sure	
16	**$10.00** with 50% chance	**$0.00** with 50% chance	**$8.00** for sure	

17	**$10.00** with 50% chance	**$0.00** with 50% chance	**$8.50** for sure	
18	**$10.00** with 50% chance	**$0.00** with 50% chance	**$9.00** for sure	
19	**$10.00** with 50% chance	**$0.00** with 50% chance	**$9.50** for sure	
20	**$10.00** with 50% chance	**$0.00** with 50% chance	**$10.00** for sure	

PART 5

For each line in the table, please state whether you prefer option A or option B. Notice that there are a total

For each line in the table, please state whether you prefer option A or option B. Notice that there are a total of **20 lines** in the table – you should think of each line as a separate decision you need to make.

Your earnings for the selected line depend on which option you chose: If you chose option B in that line, you will receive **$0**. If you chose option A in that line, you can receive either a loss between -**$1** and **-$20,** depending on the line, or a gain of **$10**. To determine your earnings in the case you chose option A we will randomly draw a ball from a bag containing twenty balls. There are **ten orange** and **ten white** balls in the bag. Before you draw the ball you choose a color. For example, suppose that you choose white. If the drawn ball is really white, you will receive -$x (the exact amount depends on the line chosen). If the drawn ball is orange, you will receive $10.

While you have all the information in the table, you should input all your 20 decisions into the computer. The actual drawing of the ball for this part of the experiment will be determined at the end of the experiment.

Use the following tables for records:

	Record Your Response Here
CHOOSE YOUR COLOR:	☐ WHITE ☐ ORANGE

Decision Number	**Option A**		**Option B**	**Choose A or B**
1	**-$1.00** with 50% chance	**$10.00** with 50% chance	**$0.00** for sure	
2	**-$2.00** with 50% chance	**$10.00** with 50% chance	**$0.00** for sure	
3	**-$3.00** with 50% chance	**$10.00** with 50% chance	**$0.00** for sure	
4	**-$4.00** with 50% chance	**$10.00** with 50% chance	**$0.00** for sure	
5	**-$5.00** with 50% chance	**$10.00** with 50% chance	**$0.00** for sure	
6	**-$6.00** with 50% chance	**$10.00** with 50% chance	**$0.00** for sure	
7	**-$7.00** with 50% chance	**$10.00** with 50% chance	**$0.00** for sure	
8	**-$8.00** with 50% chance	**$10.00** with 50% chance	**$0.00** for sure	
9	**-$9.00** with 50% chance	**$10.00** with 50% chance	**$0.00** for sure	
10	**-$10.00** with 50% chance	**$10.00** with 50% chance	**$0.00** for sure	
11	**-$11.00** with 50% chance	**$10.00** with 50% chance	**$0.00** for sure	
12	**-$12.00** with 50% chance	**$10.00** with 50% chance	**$0.00** for sure	
13	**-$13.00** with 50% chance	**$10.00** with 50% chance	**$0.00** for sure	
14	**-$14.00** with 50% chance	**$10.00** with 50% chance	**$0.00** for sure	
15	**-$15.00** with 50% chance	**$10.00** with 50% chance	**$0.00** for sure	
16	**-$16.00** with 50% chance	**$10.00** with 50% chance	**$0.00** for sure	
17	**-$17.00** with 50% chance	**$10.00** with 50% chance	**$0.00** for sure	
18	**-$18.00** with 50% chance	**$10.00** with 50% chance	**$0.00** for sure	
19	**-$19.00** with 50% chance	**$10.00** with 50% chance	**$0.00** for sure	
20	**-$20.00** with 50% chance	**$10.00** with 50% chance	**$0.00** for sure	

PART 6

In this part of the experiment, you will work on your own and have the chance to earn money by solving 2-digit math problems.

You will have 2 and a half minutes (150 seconds) for this part. The computer will provide you with up to 30 math problems (one at a time) that you can attempt to solve during this 2 and a half minutes. Each problem will consist of adding 5, two-digit numbers. All of the problems are about the same level of difficulty. You will see the problems one at a time and you will not be able to skip any problems. You will not be able to go back to any problems.

Your earnings for each problem depend on your responses in the following way:

- For each problem you answer **correctly**, you will receive **$0.50**.
- There is no penalty for incorrect problems, and no earnings from attempted problems that are not correct

Answering a problem correctly means that you have provided the correct answer, for example, 2+2=4 is correct while 2+2=3 is incorrect.

The time remaining will be displayed on the overhead. When 2 and a half minutes are up, time will be called. You will not be able to respond to any more problems after time is up because your computer will be on pause. After time is called, you will need to enter "0" to move on to the outcome screen, and the last problem you answer will not count as an attempt.

The actual earnings for this part of the experiment will be determined at the end of the experiment, and will be independent of other parts of the experiment.

PART 7

In this part of the experiment, you will be asked to provide a guess about how many of the attempted problems in Part 6 you solved **correctly**. You will receive an additional $3 if your guess is equal to the number of correct answers that you provided us in Part 6.

Please enter your guess on your screen. Record your answer (and outcome) below. The actual earnings for this part of the experiment will be determined at the end of the experiment, and will be independent of other parts of the experiment.

Use the following table for records:

	Record your Results Here
Number of Problems Attempted	
Guess About the Number of Problems Correct	

Earnings Sheet

		Result	**Your Earnings**
PART 1 – Adding Numbers	Number of Problems Attempted		
	Number of Problems Correct		
	Number of Problems Incorrect		
PART 2 – Guessing Game	Guess About the Number of Problems Correct		
	Actual Number of Problems Correct		
PARTS 3 - 5 Line Games	Which Part is Chosen? ☐ PART 3 ☐ PART 4 ☐ PART 5		
	Line that Counts (1-20)		
	Color Chosen (White or Orange)		
PART 6 – Adding Numbers	Number of Problems Attempted		
	Number of Problems Correct		
	Number of Problems Incorrect		
PART 7 - Guessing Game	Guess About the Number of Problems Correct		
	Actual Number of Problems Correct		
TOTAL:			$

Appendix B (For Online Publication): Additional Information

Table B1: Elicitation of ambiguity aversion preferences

Choice	Option A ambiguous option	Option B safe option
# 1	$0.00 or $10.00 with unknown chance	$0.50 for sure
# 2	$0.00 or $10.00 with unknown chance	$1.00 for sure
# 3	$0.00 or $10.00 with unknown chance	$1.50 for sure
# 4	$0.00 or $10.00 with unknown chance	$2.00 for sure
# 5	$0.00 or $10.00 with unknown chance	$2.50 for sure
# 6	$0.00 or $10.00 with unknown chance	$3.00 for sure
# 7	$0.00 or $10.00 with unknown chance	$3.50 for sure
# 8	$0.00 or $10.00 with unknown chance	$4.00 for sure
# 9	$0.00 or $10.00 with unknown chance	$4.50 for sure
# 10	$0.00 or $10.00 with unknown chance	$5.00 for sure
# 11	$0.00 or $10.00 with unknown chance	$5.50 for sure
# 12	$0.00 or $10.00 with unknown chance	$6.00 for sure
# 13	$0.00 or $10.00 with unknown chance	$6.50 for sure
# 14	$0.00 or $10.00 with unknown chance	$7.00 for sure
# 15	$0.00 or $10.00 with unknown chance	$7.50 for sure
# 16	$0.00 or $10.00 with unknown chance	$8.00 for sure
# 17	$0.00 or $10.00 with unknown chance	$8.50 for sure
# 18	$0.00 or $10.00 with unknown chance	$9.00 for sure
# 19	$0.00 or $10.00 with unknown chance	$9.50 for sure
# 20	$0.00 or $10.00 with unknown chance	$10.00 for sure

Participants choose between an ambiguous option A ($0.00 or $10.00 with unknown chance) or a safe option B (a certain amount for sure).

Table B2: Elicitation of risk preferences

Choice	Option A ambiguous option	Option B safe option
# 1	$0.00 or $10.00 with 50% chance	$0.50 for sure
# 2	$0.00 or $10.00 with 50% chance	$1.00 for sure
# 3	$0.00 or $10.00 with 50% chance	$1.50 for sure
# 4	$0.00 or $10.00 with 50% chance	$2.00 for sure
# 5	$0.00 or $10.00 with 50% chance	$2.50 for sure
# 6	$0.00 or $10.00 with 50% chance	$3.00 for sure
# 7	$0.00 or $10.00 with 50% chance	$3.50 for sure
# 8	$0.00 or $10.00 with 50% chance	$4.00 for sure
# 9	$0.00 or $10.00 with 50% chance	$4.50 for sure
# 10	$0.00 or $10.00 with 50% chance	$5.00 for sure
# 11	$0.00 or $10.00 with 50% chance	$5.50 for sure
# 12	$0.00 or $10.00 with 50% chance	$6.00 for sure
# 13	$0.00 or $10.00 with 50% chance	$6.50 for sure
# 14	$0.00 or $10.00 with 50% chance	$7.00 for sure
# 15	$0.00 or $10.00 with 50% chance	$7.50 for sure
# 16	$0.00 or $10.00 with 50% chance	$8.00 for sure
# 17	$0.00 or $10.00 with 50% chance	$8.50 for sure
# 18	$0.00 or $10.00 with 50% chance	$9.00 for sure
# 19	$0.00 or $10.00 with 50% chance	$9.50 for sure
# 20	$0.00 or $10.00 with 50% chance	$10.00 for sure

Participants choose between a risky option A ($0.00 or $10.00 with 50% chance) or a safe option B (a certain amount for sure).

Table B3: Elicitation of loss aversion preferences

Choice	Option A risky option	Option B safe option
# 1	-$0.50 or $5.00 with 50% chance	$0.00 for sure
# 2	-$1.00 or $5.00 with 50% chance	$0.00 for sure
# 3	-$1.50 or $5.00 with 50% chance	$0.00 for sure
# 4	-$2.00 or $5.00 with 50% chance	$0.00 for sure
# 5	-$2.50 or $5.00 with 50% chance	$0.00 for sure
# 6	-$3.00 or $5.00 with 50% chance	$0.00 for sure
# 7	-$3.50 or $5.00 with 50% chance	$0.00 for sure
# 8	-$4.00 or $5.00 with 50% chance	$0.00 for sure
# 9	-$4.50 or $5.00 with 50% chance	$0.00 for sure
# 10	-$5.00 or $5.00 with 50% chance	$0.00 for sure
# 11	-$5.50 or $5.00 with 50% chance	$0.00 for sure
# 12	-$6.00 or $5.00 with 50% chance	$0.00 for sure
# 13	-$6.50 or $5.00 with 50% chance	$0.00 for sure
# 14	-$7.00 or $5.00 with 50% chance	$0.00 for sure
# 15	-$7.50 or $5.00 with 50% chance	$0.00 for sure
# 16	-$8.00 or $5.00 with 50% chance	$0.00 for sure
# 17	-$8.50 or $5.00 with 50% chance	$0.00 for sure
# 18	-$9.00 or $5.00 with 50% chance	$0.00 for sure
# 19	-$9.50 or $5.00 with 50% chance	$0.00 for sure
# 20	-$10.00 or $5.00 with 50% chance	$0.00 for sure

Participants choose between a risky option A (which has 50% chance of losing certain amount) or a safe option B ($0.00 for sure).

Table B4: Correlation between ambiguity, risk and loss-aversion

			Correlations		
Variable	Observations	Average	Ambiguity aversion	Risk aversion	Loss aversion
Ambiguity aversion	287	11.61	1		
[# safe choices]		(3.05)			
Risk aversion	287	10.91	0.60***	1	
[# safe choices]		(2.81)			
Loss aversion	287	14.89	0.28***	0.32***	1
[# safe choices]		(3.09)			

*** significant at 1%; standard errors are in parentheses.

Table B5: OLS regressions of quantity, with ability^2

Specification	(1)	(2)	(3)	(4)	(5)
Dependent variable = Quantity (number attempted)					
Ability	-0.65	-0.85	-0.82	-0.92	-0.87
(correct in part 6)	(0.78)	(0.75)	(0.75)	(0.75)	(0.74)
Ability^2	0.09*	0.10**	0.10**	0.11**	0.10**
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Medium quality incentives	-10.53***	-10.50***	-10.53***	-10.83***	7.46
(T-0.25 and T-0.50)	(1.43)	(1.43)	(1.43)	(1.43)	(7.70)
High quality incentives	-14.08***	-14.02***	-14.10***	-14.07***	6.00
(T-1.00 and T-3.00)	(1.42)	(1.43)	(1.43)	(1.41)	(6.37)
Overconfidence	0.38				
	(0.50)				
Risk aversion		0.13			
		(0.21)			
Ambiguity aversion			0.06		
			(0.19)		
Loss aversion				0.41**	1.31***
				(0.19)	(0.32)
Loss aversion ×					-1.23**
Medium quality incentives					(0.50)
Loss aversion ×					-1.37***
High quality incentives					(0.43)
Constant	27.21***	26.83***	27.53***	22.51***	9.33*
	(2.98)	(3.42)	(3.42)	(3.71)	(5.35)
Observations	287	287	287	287	287
R-squared	0.29	0.29	0.29	0.30	0.33
p-value, Medium = High	0.01	0.01	0.01	0.02	--

Table B6: OLS regressions of quantity, with individual treatment dummies

Specification	(1)	(2)	(3)	(4)
	Dependent variable = Quantity (number attempted)			
Ability	0.43	0.60***	0.57**	0.45**
(correct in part 6)	(0.27)	(0.23)	(0.23)	(0.23)
Treatment T-$0.05		-7.42***	-7.80***	13.83
		(2.00)	(1.99)	(9.53)
Treatment T-$0.25		-14.81***	-15.40***	20.70**
		(2.02)	(2.02)	(10.47)
Treatment T-$0.50		-14.03***	-14.50***	12.66
		(1.97)	(1.97)	(11.02)
Treatment T-$1.00		-18.14***	-18.01***	17.12*
		(1.96)	(1.94)	(8.89)
Treatment T-$3.00		-17.93***	-18.39***	16.69*
		(2.03)	(2.02)	(9.14)
Loss aversion			0.45**	2.33***
			(0.19)	(0.47)
Loss aversion x				-1.54**
Treatment T-$0.05				(0.64)
Loss aversion x				-2.47***
Treatment T-$0.25				(0.69)
Loss aversion x				-1.90***
Treatment T-$0.50				(0.73)
Loss aversion x				-2.48***
Treatment T-$1.00				(0.62)
Loss aversion x				-2.42***
Treatment T-$3.00				(0.61)
Constant	16.74***	27.91***	21.63***	-4.38
	(1.76)	(1.97)	(3.23)	(6.82)
Observations	287	287	287	287
R-squared	0.01	0.31	0.33	0.38

** significant at 5%, *** significant at 1%; standard errors are in parentheses.

Table B7: OLS regressions of quality, with individual treatment dummies

Specification	(1)	(2)	(3)	(4)
	Dependent variable = Quality (number correct)			
Ability	1.43***	1.39***	1.40***	1.43***
(correct in part 6)	(0.10)	(0.09)	(0.09)	(0.09)
Treatment T-$0.05		2.66***	2.79***	-2.58
		(0.76)	(0.76)	(3.71)
Treatment T-$0.25		4.17***	4.38***	-6.57
		(0.77)	(0.77)	(4.08)
Treatment T-$0.50		4.51***	4.68***	-1.64
		(0.75)	(0.75)	(4.29)
Treatment T-$1.00		5.30***	5.26***	-0.72
		(0.75)	(0.74)	(3.47)
Treatment T-$3.00		5.86***	6.03***	-1.48
		(0.77)	(0.77)	(3.56)
Loss aversion			-0.16**	-0.58***
			(0.07)	(0.18)
Loss aversion x				0.38
Treatment T-$0.05				(0.25)
Loss aversion x				0.74***
Treatment T-$0.25				(0.27)
Loss aversion x				0.44
Treatment T-$0.50				(0.28)
Loss aversion x				0.42*
Treatment T-$1.00				(0.24)
Loss aversion x				0.52**
Treatment T-$3.00				(0.24)
Constant	0.81	-2.71***	-0.47	5.32**
	(0.63)	(0.75)	(1.23)	(2.66)
Observations	287	287	287	287
R-squared	0.44	0.56	0.57	0.58

** significant at 5%, *** significant at 1%; standard errors are in parentheses.

Table B8: OLS regressions of the error rate, with individual treatment dummies

Specification	(1)	(2)	(3)	(4)
	Dependent variable = Error rate (incorrect/attempted)			
Ability	-0.04***	-0.04***	-0.04***	-0.04***
(correct in part 6)	(0.01)	(0.01)	(0.01)	(0.01)
Treatment T-$0.05		-0.17***	-0.18***	0.10
		(0.05)	(0.05)	(0.26)
Treatment T-$0.25		-0.36***	-0.38***	0.30
		(0.05)	(0.05)	(0.29)
Treatment T-$0.50		-0.34***	-0.35***	0.07
		(0.05)	(0.05)	(0.30)
Treatment T-$1.00		-0.51***	-0.51***	-0.02
		(0.05)	(0.05)	(0.24)
Treatment T-$3.00		-0.53***	-0.54***	-0.07
		(0.05)	(0.05)	(0.25)
Loss aversion			0.01**	0.04***
			(0.00)	(0.01)
Loss aversion x				-0.02
Treatment T-$0.05				(0.02)
Loss aversion x				-0.05**
Treatment T-$0.25				(0.02)
Loss aversion x				-0.03
Treatment T-$0.50				(0.02)
Loss aversion x				-0.03**
Treatment T-$1.00				(0.02)
Loss aversion x				-0.03**
Treatment T-$3.00				(0.02)
Constant	0.62***	0.92***	0.77***	0.39**
	(0.05)	(0.05)	(0.09)	(0.19)
Observations	287	287	287	287
R-squared	0.10	0.41	0.42	0.44

** significant at 5%, *** significant at 1%; standard errors are in parentheses.

Table B9: OLS regressions of "guesstimates", with individual treatment dummies

Specification	(1)	(2)	(3)	(4)
	Dependent variable = "Guesstimate" (incorrect but within 20 of correct answer)			
Ability	0.00	0.03	0.03	0.02
(correct in part 6)	(0.09)	(0.09)	(0.09)	(0.09)
Treatment T-$0.05		-1.69**	-1.72**	4.98
		(0.78)	(0.78)	(3.85)
Treatment T-$0.25		-3.04***	-3.09***	3.93
		(0.79)	(0.79)	(4.23)
Treatment T-$0.50		-2.59***	-2.62***	7.71*
		(0.77)	(0.77)	(4.45)
Treatment T-$1.00		-3.96***	-3.95***	2.25
		(0.76)	(0.76)	(3.59)
Treatment T-$3.00		-3.89***	-3.93***	1.36
		(0.79)	(0.79)	(3.69)
Loss aversion			0.03	0.43**
			(0.07)	(0.19)
Loss aversion x				-0.47*
Treatment T-$0.05				(0.26)
Loss aversion x				-0.48*
Treatment T-$0.25				(0.28)
Loss aversion x				-0.70**
Treatment T-$0.50				(0.29)
Loss aversion x				-0.44*
Treatment T-$1.00				(0.25)
Loss aversion x				-0.37
Treatment T-$3.00				(0.25)
Constant	2.68***	5.06***	4.61***	-1.02
	(0.61)	(0.76)	(1.27)	(2.75)
Observations	287	287	287	287
R-squared	0.00	0.12	0.12	0.14

** significant at 5%, *** significant at 1%; standard errors are in parentheses.